Necessary Conditions in Dynamic Optimization

Francis Clarke [1]

Université de Lyon I et
Institut universitaire de France

July 15, 2002

[1]Written in part while the author was visiting the Università di Roma Tor Vergata. Initial version May 8, 2002.
Address: Institut Desargues, Université de Lyon 1, 69622 Villeurbanne, France.
E-mail: Clarke@desargues.univ-lyon1.fr

Key words. optimal control, calculus of variations, differential inclusion, nonsmooth analysis.
AMS subject classifications. 49K15, 49K05.

Abstract

This monograph derives necessary conditions of optimality for a general control problem formulated in terms of a differential inclusion. These conditions constitute a new state of the art, subsuming, unifying, and substantially extending the results in the literature. The Euler, Weierstrass and transversality conditions are expressed in their sharpest known forms. No assumptions of boundedness or convexity are made, no constraint qualifications imposed, and only weak pseudo-Lipschitz behavior is postulated on the underlying multifunction. The conditions also incorporate a 'stratified' feature of a novel nature, in which both the hypotheses and the conclusion are formulated relative to a given radius function. When specialized to the calculus of variations, the results yield necessary conditions and regularity theorems that go significantly beyond the previous standard. They also apply to parametrized control systems, giving rise to new and stronger maximum principles of Pontryagin type. The final chapter is devoted to a different issue, that of the Hamiltonian necessary condition. It is obtained here, for the first time, in the case of nonconvex values and in the absence of any constraint qualification; this has been a longstanding open question in the subject. Apart from the final chapter, the treatment is self-contained, and calls upon only standard results in functional and nonsmooth analysis.

To Gail, my necessary (and sufficient) condition

Contents

Chapter 1

Introduction

1.1 Classical necessary conditions

The subject of this article has its origins in the following basic problem in the calculus of variations: to minimize the integral functional

$$\int_a^b \Lambda_t(x(t), \dot{x}(t))\, dt$$

over a class of functions x on the interval $[a, b]$, where the function $\Lambda_t(x, v)$ (the *Lagrangian*) is given. The fundamental necessary condition of Euler (1744) asserts that a smooth solution x_* of this problem must satisfy the following differential equation:

$$\frac{d}{dt} D_v \Lambda_t(x_*(t), \dot{x}_*(t)) = D_x \Lambda_t(x_*(t), \dot{x}_*(t)).$$

To accommodate the need to treat nonsmooth solutions, the class of admissible x was later extended to that of piecewise smooth functions. In that setting, the Euler equation continues to hold except at finitely many points, but additional information is required in order to give uniqueness. This was provided by the first Erdmann condition: the function

$$t \mapsto D_v \Lambda_t(x_*(t), \dot{x}_*(t))$$

must have removable discontinuities. In 1879 du Bois-Raymond introduced a strengthened (integral) form of the Euler equation; it affirms the existence of an absolutely continuous function p satisfying

$$(\dot{p}(t), p(t)) = D_{x,v} \Lambda_t(x_*(t), \dot{x}_*(t)) \text{ a.e.} \tag{$\mathcal{E}$}$$

Note that both the Euler equation and the Erdmann condition are subsumed by this formulation. (In mechanics, p is the *generalized momentum*; in control theory it corresponds to the *adjoint variable*.)

Concurrently with these later developments, Weierstrass established the other fundamental necessary condition in the subject: it can be phrased as asserting for every t the global inequality

$$\Lambda_t(x_*(t), v) - \Lambda_t(x_*(t), \dot{x}_*(t)) \geq \langle p(t), v - \dot{x}_*(t) \rangle \quad \forall\, v. \tag{$\mathcal{W}$}$$

Of course, in the classical smooth setting, the $p(t)$ in this inequality can be nothing else than $D_v \Lambda_t(x_*(t), \dot{x}_*(t))$, the same $p(t)$ that occurs in $(\mathcal{E})$.

The central question of this monograph is the following: In more general settings than that of the basic problem, in what form and under what hypotheses do these necessary conditions continue to hold? As we see from the above, the question was already a central one even for the classical basic problem itself. In fact, it is relatively complex even in that setting. For example, Tonelli's pioneering work (1915) demonstrated that in order to have access to an existence theory for the basic problem, we should treat it within the class of absolutely continuous (rather than piecewise smooth) functions. But when the solution x_* lies in this more general class, the (strengthened) Euler equation is no longer a valid necessary condition (there may not exist p satisfying $(\mathcal{E})$). It *would* be possible to affirm $(\mathcal{E})$ if we knew *a priori* that x_* has essentially bounded derivative (that is, if x_* were known to be Lipschitz). This leads to the question of the *regularity* of the solution, which has grown into a central theme in the calculus of variations.

As the theory of existence and regularity was evolving, parallel developments were taking place in the formulation of more general variational problems, and their attendant necessary conditions. The initial impetus was due to Bolza (1913), who formulated a generalization of the basic problem that incorporates more general endpoint behavior as well as pointwise constraints of the form $\psi(x(t), \dot{x}(t)) = 0$ (for example). When a cost term of the form $\ell(x(a), x(b))$ (depending on the endpoints of x) is added to the integral functional of the basic problem, it was shown that the function p referred to above must satisfy the *transversality conditions*

$$(p(a), -p(b)) = D\ell(x_*(a), x_*(b)).$$

As regards the pointwise constraint, interest turned towards proving a complete *multiplier rule*. This refers to an application of the Lagrange multiplier idea to deduce that for a certain function $\lambda(t)$, the augmented Lagrangian $\Lambda + \langle \lambda, \psi \rangle$ satisfies the necessary conditions for the basic problem. Among these should figure the Euler equation, of course, but preferably the other necessary conditions as well (the Weierstrass condition turned out to be particularly difficult to establish.) A satisfying theorem of this kind became the Holy Grail of the subject for some decades.[1]

As this quest was coming to a close, another paradigm was developing that would subsume the basic variational problem in a different way, one which proved more suitable for applications. This used the model of a *controlled differential equation*

$$\dot{x}(t) = \phi_t(x(t), u(t)),$$

where the control function u is constrained to take its values in a prescribed set U. Subject to this, the goal is to minimize a cost involving an integral functional

$$\int_a^b \Lambda_t(x(t), u(t))\, dt.$$

We recover the basic variational problem as a special case of the above by taking $\phi_t(x, u) = u$ and U to be the whole space; the control is then identified with $\dot{x}$, unconstrained. The celebrated Maximum Principle of Pontryagin (1959) establishes a set of necessary conditions for the *optimal control problem* which can be viewed as (substantial) extensions of the classical results of Euler, du Bois-Raymond, Erdmann and Weierstrass, incorporating both transversality and a form of the multiplier rule.

We have said that the central question of this monograph concerns the form and the validity of the necessary conditions; we shall address this issue both for problems stemming from the calculus of variations and for problems of optimal control more general than the standard Pontryagin formulation mentioned above. It is convenient to single out three paradigms for *dynamic optimization.*

The first, commonly referred to now as the *generalized problem of Bolza*, is defined very succinctly: to minimize the integral functional

$$\ell(x(a), x(b)) + \int_a^b \Lambda_t(x(t), \dot{x}(t))\, dt$$

over the class of absolutely continuous functions $x : [a, b] \to \mathbb{R}^n$ (we refer to such functions as *arcs*). It is important to note that ℓ and Λ will in general be extended-valued (that is, permitted to assume the value $+\infty$).

[1] The Chicago school of Bliss led this project, which culminated in the work of McShane [45] and Hestenes [34].

This allows constraints to be present *implicitly*. For example, if we wish to minimize merely the integral term in the cost, subject to the endpoint constraints $(x(a), x(b)) \in S$ (where S is a given set), we can formulate this as a generalized problem of Bolza by defining ℓ to be the *indicator* of S: the function which takes the value 0 on S and $+\infty$ elsewhere.

The second paradigm we refer to as the *differential inclusion problem*. It consists of minimizing $\ell(x(a), x(b))$ over the arcs x that satisfy the boundary constraints $(x(a), x(b)) \in S$ and the differential inclusion

$$\dot{x}(t) \in F_t(x(t)) \text{ a.e.},$$

where for each t, F_t is a given mapping from R^n to its subsets. We take ℓ to be locally Lipschitz in this case, so that all the constraints on x have been rendered as explicit as possible, in total contrast to the generalized problem of Bolza, where all constraints have been made implicit.

Our third paradigm is the optimal control of parametrized families of vector fields. Given a family $\mathcal{F}$ of vector fields $f(t, x)$, we wish to find among the arcs x which satisfy $\dot{x} = f(t, x)$ for some $f \in \mathcal{F}$ one which minimizes $\ell(x(a), x(b))$, subject again to the constraint $(x(a), x(b)) \in S$. This extension of the standard Pontryagin model is sometimes referred to as a *generalized control system*.

Of course these three paradigms are not mutually exclusive; specific problems can very well fit into more than one of them. And each paradigm has certain advantages. But from the technical point of view, we believe that it is the differential inclusion problem (which is the least interesting historically) that is most convenient. It also lends itself well to taking a first look at the new results obtained in this monograph. Let us discuss it now in some detail.

1.2 Differential inclusions

We study the optimal control of a system governed by the differential inclusion

$$\dot{x}(t) \in F_t(x(t)) \text{ a.e., } t \in [a, b],$$

or more precisely, the derivation of necessary conditions for optimality. Here $[a, b]$ is a given interval in $\mathbb{R}$ and F a multifunction (for each t, F_t is a map from $\mathbb{R}^n$ to the subsets of $\mathbb{R}^n$). A *trajectory* of F is an arc[2] satisfying the differential inclusion. A typical optimization problem $\mathcal{P}$ involving trajectories

[2] Recall that an absolutely continuous function $x : [a, b] \to \mathbb{R}^n$ is referred to as an arc.

is the following:

$$\mathcal{P}: \text{ to minimize } \quad \ell(x(a), x(b)) : \dot{x}(t) \in F_t(x(t)) \text{ a.e., } (x(a), x(b)) \in S,$$

where $\ell : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz and S is a closed subset of $\mathbb{R}^n \times \mathbb{R}^n$. (When we treat this problem in Chapter 3, an additional 'stratified' constraint will be admitted.)

The interest of such differential inclusions stems in part from the fact that, at least formally, a number of different problems can be reformulated in such terms. We examine now two important instances of this fact. Consider first the basic problem in the calculus of variations defined above:

$$\text{minimize} \quad \int_a^b \Lambda_t(x(t), \dot{x}(t))\, dt$$

subject to given boundary constraints on $x(a)$ and $x(b)$. If we introduce an auxiliary variable y and define a multifunction F^Λ via

$$F_t^\Lambda(x, y) := \{(v, \Lambda_t(x, v)) : v \in \mathbb{R}^n\},$$

then the problem is equivalent to that of minimizing $y(b)$ over the trajectories (x, y) of F^Λ which satisfy $y(a) = 0$ together with the given boundary constraints on x. This has the form of the problem $\mathcal{P}$ above.

Next, consider the basic problem in optimal control . Here we consider the arcs x which satisfy a system of ordinary differential equations

$$\dot{x}(t) = \phi_t(x(t), u(t)) \text{ a.e., } t \in [a, b],$$

where the function ϕ is given, and where $u(\cdot)$ is a *control*: a measurable function from $[a, b]$ to $\mathbb{R}^n$ taking values in a given subset U. The goal is to choose the control u and the associated *state* x so as to respect the boundary constraints $(x(a), x(b)) \in S$, and so as to minimize $\ell(x(a), x(b))$. This is the Mayer form of the standard Pontryagin formulation for optimal control, which has been widely adopted in a large number of modeling applications.

If now a multifunction F^ϕ is defined as follows:

$$F_t^\phi(x) := \phi_t(x, U),$$

then under very mild assumptions on ϕ, the problem under consideration is equivalent to the resulting instance of problem $\mathcal{P}$. This assertion amounts to a measurable selection theorem; it is well-known since the pioneering work of Filippov (concerning primarily existence theory) in the 1960s.

Thus differential inclusions can potentially serve as a unified setting for considering a variety of problems in dynamic optimization; two other contexts giving rise to differential inclusions include that of differential inequalities, and cases in which control sets depend upon the state. But of course this potential is realizable only to the extent that we are able to treat the multifunctions that arise. For example, the multifunction F^Λ defined above in connection with the calculus of variations has unbounded sets as values, so if we are to use it, then unbounded multifunctions must be admissible in the theory. This is but one example of the importance of being able to treat differential inclusions in as broad a setting as possible, which is one of our goals.

It is not evident how to even express the necessary conditions of optimality for the differential inclusion problem; we address that question now as regards the Euler equation $(\mathcal{E})$. The key is to reformulate the problem to give it the appearance of a classical basic problem in the calculus of variations, and then let history (and notation) be our guide. Accordingly we define $\Lambda_t(x,v)$ to be the function which equals 0 when v belongs to $F_t(x)$ and $+\infty$ otherwise. Then the differential inclusion problem amounts to minimizing the Bolza functional

$$\ell(x(a), x(b)) + \int_a^b \Lambda_t(x(t), \dot{x}(t))\, dt$$

(subject to the endpoint constraints). Note that here Λ_t is the indicator of the *graph* of the multifunction F_t; that is, the set

$$G(t) := \operatorname{gr} F_t := \{(x,v) \in \mathbb{R}^n \times \mathbb{R}^n : v \in F_t(x)\}.$$

If Λ were smooth we would proceed to write the Euler equation $(\mathcal{E})$:

$$(\dot{p}(t), p(t)) = D_{x,v}\Lambda_t(x_*(t), \dot{x}_*(t)) \text{ a.e.}$$

Since this is not the case, we are led to considering generalizations of derivatives that could be applied to such nonsmooth functions as indicators. This is the subject of *nonsmooth analysis*, which offers several such constructs, notably in terms of (multivalued) *subdifferentials*.

Without being specific for the moment, suppose that we dispose of such a subdifferential $\partial\Lambda_t$. Then the Euler equation could perhaps be expressed as an inclusion:

$$(\dot{p}(t), p(t)) \in \partial\Lambda_t(x_*(t), \dot{x}_*(t)) \text{ a.e.}$$

It is a metatheorem in nonsmooth analysis that subdifferentials of indicator functions coincide with the set (cone) of normal vectors to the underlying set. Then, when Λ_t is taken to be the indicator of $G(t)$, the Euler equation may be expressed in the following form:

$$(\dot{p}(t), p(t)) \in N_{G(t)}(x_*(t), \dot{x}_*(t)) \text{ a.e.,}$$

where, in generic notation, $N_G(\gamma)$ denotes the normal cone to a set G at a point $\gamma \in G$. We conclude that the analogue of the Euler equation for the differential inclusion problem may be a geometrical assertion of this general type. It may seem surprising that an inclusion of such a seemingly abstract nature can yield the sharpest results in such different contexts as the calculus of variations and optimal control theory, but this will be seen to be the case.

The early attempts to develop necessary conditions for differential inclusions date from around 1970, and are characterized by strong hypotheses of smoothness or convexity on the graph of the multifunction; this is to be expected, since a calculus of normal vectors (for example) had been developed only in those contexts. Such hypotheses allow the consideration of only very special cases. The underlying methodologies (based for example on convex analysis) do not extend to more realistic situations, which are inherently nonsmooth and nonconvex.

In 1973 and subsequently, Clarke considered such problems in the absence of smoothness or convexity hypotheses, under the assumption that the multifunction is bounded-valued and satisfies locally a (true) Lipschitz condition in x of the type[3]

$$F_t(x'') \subset F_t(x') + k(t)\,|x'' - x'|)\,\overline{B}.$$

It was shown (under various sets of hypotheses, and in terms of a new nonsmooth calculus) that if the arc x_* solves $\mathcal{P}$, then there is an arc p for which we have the following necessary condition, an *Euler inclusion*:

$$(\dot{p}(t), p(t)) \in N^C_{G(t)}(x_*(t), \dot{x}_*(t)) \text{ a.e.} \qquad (\mathcal{E}_C)$$

The notation N^C refers to the normal cone used by Clarke. (We review some basic concepts of nonsmooth analysis in the next section.) This work introduced penalization and distance function methods in conjunction with Ekeland's variational principle, an approach that has since been widely adopted in the subject.

Considerable progress has been made in recent years on extending the theory in a variety of ways, by such authors as Loewen and Rockafellar, Ioffe,

[3] The notation $\overline{B}$ refers to the closed unit ball, and B to its interior; $r\overline{B}$ is also written as $\overline{B}(0, r)$.

Vinter, Mordukhovic, Kaskosz and Lojasiewicz, Zheng, Zhu, and others. We proceed now to discuss briefly the principal issues that have arisen in connection with these developments.

Refined necessary conditions. Due in large part to the work of Rockafellar, it has become clear that a more refined Euler inclusion $\mathcal{E}_L$ can be used in the theory:

$$\dot{p}(t) \in \text{co}\left\{\omega : (\omega, p(t)) \in N^L_{G(t)}(x_*(t), \dot{x}_*(t))\right\} \text{ a.e.,} \qquad (\mathcal{E}_L)$$

where N^L refers to the limiting normal (or L-normal) cone. Because the C-normal cone N^C is the closed convex hull of the limiting one, it is clear that $\mathcal{E}_L$ implies $\mathcal{E}_C$; the reverse is false in general. We may say that $\mathcal{E}_L$ involves only a 'partial' convexification. Another, earlier refinement in the same vein, due to Mordukhovich, involves the use of L-normals and subgradients in the transversality conditions.

A different question involves the possibility of affirming that the arc p appearing in the Euler inclusion simultaneously satisfies the *Weierstrass condition*:

$$\langle\, p(t), v - \dot{x}_*(t)\rangle \le 0 \quad \forall v \in F_t(x_*(t)),\ t \in [a, b] \text{ a.e.,}$$

an important point in deriving, for example, maximum principles of Pontryagin type.[4] The necessary conditions established in this report incorporate all of these refinements. They also have a novel 'stratified' nature that breaks new ground; we describe it presently.

Convexity and boundedness of the values. The analysis of the differential inclusion is considerably simplified when F is assumed to be convex-valued (this is also known as the 'relaxed' case). For example, the issue of the simultaneous Weierstrass condition does not even arise, for in the presence of convex values it is a simple consequence of the Euler inclusion. At a technical level, such techniques as penalization and localization are much easier to implement. Consideration of the nonconvex case is particularly challenging when the Lipschitz hypothesis on the multifunction is weakened, since relaxation theorems generally require the strong Lipschitz property. Similar considerations make it simpler to consider multifunctions whose values are

[4]Note that this is in fact what the classical Weierstrass condition $(\mathcal{W})$ becomes when Λ_t is the indicator of $G(t)$.

bounded sets, another hypothesis which restricts applicability. In this article we treat the nonconvex unbounded case. Convexity is thereby seen to be a superfluous hypothesis which simply gives rise to certain refinements when it happens to be present.

The Lipschitz condition. Motivated by the need to consider multifunctions F with unbounded values, where the Lipschitz hypothesis can be too strong, certain types of *pseudo-Lipschitz* behavior near the given trajectory have been introduced by Ioffe, Loewen and Rockafellar, and Vinter. Typical conditions of this nature require the following to hold for certain values of R, and for all points x', x'' in a neighborhood of $x_*(t)$:

$$F_t(x'') \cap \overline{B}(\dot{x}_*(t), R) \subset F_t(x') + k \left| x'' - x' \right| \overline{B}.$$

The results of this article are obtained under a weak pseudo-Lipschitz condition of this type.

Constraint qualifications. A *constraint qualification* is a hypothesis concerning the nature of the boundary constraints or the behavior of the problem with respect to perturbations. Absence or nondegeneracy of endpoint constraints, normality, calmness, regularity, and relaxed solvability have all figured in this connection. We impose no such requirements in this article.

Type of local minimum. The arc x_* provides a *strong local minimum* if, for some $\varepsilon > 0$, it is optimal relative to other feasible arcs x satisfying $\|x - x_*\|_\infty \leq \varepsilon$. A *weak local minimum* is obtained if the optimality is limited to those x satisfying both $\|x - x_*\|_\infty \leq \varepsilon$ and $\|\dot{x} - \dot{x}_*\|_\infty \leq \varepsilon$. These are classically familiar notions. A more modern (and in general different) type of local minimum is that obtained by considering as always $\|x - x_*\|_\infty \leq \varepsilon$, but now together with $\|\dot{x} - \dot{x}_*\|_1 \leq \varepsilon$. This has been referred to as a $W^{1,1}$ local minimum, and has featured prominently in recent work. We consider local minima of this type. At the same time, however, our results are applicable to the case of a weak local minimum as well, because of the 'stratified' constraint that we admit to the problem formulation (see below). Furthermore, we obtain the necessary conditions in the case of 'boundary trajectories', a setting more general than that of optimality.

A new stratified framework. We now explain the new 'stratified' nature of our necessary conditions. Consider first the case of a weak local minimum, so that x_* is optimal only with respect to arcs x for which $\|x - x_*\|_\infty \leq \varepsilon$ and

$$|\dot{x}(t) - \dot{x}_*(t)| \leq \varepsilon \text{ a.e.}$$

This is the same as x_* being a strong local minimum for the problem in which the multifunction F is replaced by the 'localized' multifunction F^* defined by

$$F_t^*(x) := F_t(x) \cap \overline{B}(\dot{x}_*(t), \varepsilon).$$

Ideally, we would hope to be able to write the Euler inclusion and the Weierstrass condition for F^*, the latter being given by

$$\langle\, p(t), v - \dot{x}_*(t) \rangle \leq 0 \quad \forall v \in F_t(x_*(t)) \cap \overline{B}(\dot{x}_*(t), \varepsilon),\ t \in [a, b] \text{ a.e.}$$

The Euler inclusion for F^* is the same as for F, since the two graphs in question are locally the same near $(x_*(t), \dot{x}_*(t))$. The difficulty with this ideal picture is that even if F_t is Lipschitz, F_t^* will not be well-behaved in general (not even continuous). Thus the possibility of deriving necessary conditions for F^* is problematic. Nonetheless, our results do in effect just that, and under the following type of pseudo-Lipschitz hypothesis: for x', x'' near $x_*(t)$, we have

$$F_t(x'') \cap \overline{B}(\dot{x}_*(t), \varepsilon) \subset F_t(x') + k\,|x'' - x'|\,\overline{B}.$$

Note that both the hypotheses and the conclusion are now adapted to the *radius* of the weak local minimum: we can assert the Weierstrass condition to precisely the extent that the optimality and the pseudo-Lipschitz condition hold; this degree of coherence is a new factor.

More generally, we consider the problem $\mathcal{P}$ under an *additional* constraint distinct from that giving rise to the word 'local', and distinct from the differential inclusion, a constraint of the form

$$|\dot{x}(t) - \dot{x}_*(t)| \leq R(t) \text{ a.e.,}$$

where $R : [a, b] \to (0, +\infty]$ is a given 'radius function'. As above, the case $R \equiv \varepsilon$ allows one to treat a weak local minimum (for the first time in this general context), but we also allow $R \equiv +\infty$, thereby subsuming what has been the usual case, in which the additional constraint is absent. The (possibly extended-valued) radius function is assumed only to be positive and measurable. The nature of the conclusion in the general case preserves the symmetry between hypotheses and conclusion: if F is 'pseudo-Lipschitz of radius R' near x_*, then we obtain along with the Euler inclusion the 'Weierstrass condition of radius R':

$$\langle p(t), v - \dot{x}_*(t) \rangle \leq 0 \quad \forall v \in F_t(x_*(t)) \cap \overline{B}(\dot{x}_*(t), R(t)),\ t \in [a,b] \text{ a.e.}$$

Further, if this should hold for a sequence of radius functions going to infinity, then we obtain the usual (global) Weierstrass condition . In particular, we obtain the new (and natural) result that if the multifunction is pseudo-Lipschitz of radius R (for some k_R) for every $R > 0$, then the global necessary conditions hold. The fact that the hypotheses and the conclusion can hold for different radius functions, and even in the case of a fixed radius function for radii that depend on t, motivates our use of the word 'stratified' for this new type of result.

In summary, our results subsume and substantially extend all the current ones of this nature for the standard formulation of the differential inclusion problem under consideration, as well as for the two other paradigms that we have defined above: the problem of Bolza in the calculus of variations, and optimal control systems. We believe that the theory now exhibits a completeness and unity that were missing previously. We do not study here certain of the (many) variants such as free-time or state-constrained problems, but we expect our methods to carry over.

Plan of the monograph

The next two chapters constitute the technical core of this work. Chapter 2 first proves a preparation theorem that bears upon a bounded, Lipschitz multifunction. Despite its subsidiary role, the theorem is of independent interest due to the weak nature of the weighted local minimum that it postulates, and for its new and efficient method of proof, which is based in part on a 'decoupling' technique introduced by the author in [23]. The other part of Chapter 2 is devoted to Theorem 2.3.3, which applies to boundary trajectories in a very general context.

The chapters that follow treat in turn the three paradigms we have defined above. Chapter 3 examines the differential inclusion problem. Chapter 4 studies the Bolza problem in the calculus of variations; it presents new state-of-the-art necessary conditions and regularity criteria which significantly extend those available previously (even for smooth Lagrangians). In Chapter 5 we extend the maximum principle (including its more modern and stronger versions) in several ways; we obtain in particular a new and versatile 'hybrid maximum principle'. The final chapter discusses the Hamiltonian inclusion in generalized gradient form, and establishes this necessary condition for the first time in the case of nonconvex values and in the absence of any constraint qualification . This settles a longstanding open question in the subject.

We remark that the monograph is self-contained to an extent that is unusual in this area. We use no prior results in the subject, except in the final section, which stands apart from the rest. We do require facts from proximal nonsmooth analysis in infinite-dimensional Hilbert space. Notable among these are the variational principle of Borwein and Preiss [7, 8], and the mean value inequality of Clarke and Ledyaev [26]. These theorems are proven from first principles, along with all the results of nonsmooth analysis used here, in [27], whose notation and terminology we adopt. For the convenience of the reader, however, we briefly summarize in the next section the principal constructs and results from nonsmooth analysis that we shall call upon.

Detailed references, comparisons with the literature, and other comments are provided in the Notes which conclude each chapter. The symbol ■ indicates the end of a proof, and ♦ the end of a theorem statement.

1.3 Nonsmooth analysis

It has been a lesson of the past few decades that many different topics in mathematics and its applications require to some extent a calculus for non-differentiable functions and nonsmooth sets, one that (unlike distributions) considers the *pointwise* nature of the nondifferentiability. Optimization and feedback control design (see [25]) are two examples of this, and nonlinear partial differential equations provides another with the celebrated method of viscosity solutions, which is closely related to optimal control. The present topic, necessary conditions in dynamic optimization, is another in which the best results necessitate nonsmooth analysis for their derivation, even when the underlying problem exhibits smooth data.

Although the literature of nonsmooth analysis is dauntingly large at this point, a relatively compact knowledge set suffices for many purposes. We present here some aspects of the branch of the subject known as proximal analysis, in the setting of a Hilbert space X.

Basic definitions. Given a nonempty closed subset S of X and a point x in S, we say that $\zeta \in X$ is a *proximal normal* (vector) to S at x if there exists $\sigma \geq 0$ such that

$$\langle \zeta, x' - x \rangle \leq \sigma \left\| x' - x \right\|^2 \quad \forall\, x' \in S.$$

(This is the *proximal normal inequality*.) The set (convex cone) of such ζ, which always contains 0, is denoted $N_S^P(x)$ and referred to as the proximal normal cone.

Given a lower semicontinuous function $f : X \to \mathbb{R} \cup \{+\infty\}$ and a point x in the effective domain of f, that is, the set

$$\operatorname{dom} f := \left\{x' \in X : f(x') < +\infty\right\},$$

we say that ζ is a *proximal subgradient* of f at x if there exists $\sigma \geq 0$ such that

$$f(x') - f(x) + \sigma \left\|x' - x\right\|^2 \geq \left\langle \zeta, x' - x \right\rangle$$

for all x' in a neighborhood of x. The set of such ζ, which may be empty, is denoted $\partial_P f(x)$ and referred to as the proximal subdifferential. The Proximal Density Theorem asserts that $\partial_P f(x) \neq \emptyset$ for all x in a dense subset of $\operatorname{dom} f$.

The *limiting normal cone* $N_S^L(x)$ to S at x is obtained by applying a sequential closure operation to N_S^P:

$$N_S^L(x) := \left\{\lim \zeta_i : \zeta_i \in N_S^P(x_i), x_i \to x, x_i \in S\right\}.$$

A similar procedure defines the *limiting subdifferential*:

$$\partial_L f(x) := \{\lim \zeta_i : \zeta_i \in \partial_P f(x_i), x_i \to x, f(x_i) \to f(x)\}.$$

It can be shown that ζ belongs to $\partial_P f(x)$ iff the vector $(\zeta, -1)$ belongs to $N^P_{\operatorname{epi} f}(x, f(x))$, where $\operatorname{epi} f$, the *epigraph* of f, is the set

$$\operatorname{epi} f := \{(x, r) \in X \times \mathbb{R} : f(x) \leq r\}.$$

A similar characterization holds for $\partial_L f(x)$.

Some calculus. Let us cite a few examples of calculus rules associated with the above constructs. If I_S denotes the indicator function of S, then

$$\partial_P I_S(x) = N_S^P(x), \ \partial_L I_S(x) = N_S^L(x).$$

If the set S admits a representation of the form

$$S = \left\{x' : f(x') \leq 0\right\},$$

where f is locally Lipschitz, and if x is a point such that $f(x) = 0$, $0 \notin \partial_L f(x)$, then any vector in $N_S^L(x)$ is of the form $\lambda\zeta$ for some scalar $\lambda \geq 0$ and some element $\zeta \in \partial_L f(x)$.

Proximal characterizations of many functional properties can be proven. For example, the function f satisfies a Lipschitz condition with Lipschitz constant K on an open convex set Ω iff whenever a point $x \in \Omega$ admits an element $\zeta \in \partial_P f(x)$, then $\|\zeta\| \leq K$.

If g is a locally Lipschitz function, then we have the limiting sum rule:

$$\partial_L(f+g)(x) \subset \partial_L f(x) + \partial_L g(x),$$

with equality if g is C^2. Now let $f, g : X \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous functions having x in their effective domains, and for simplicity, take X to be finite dimensional. Suppose that ζ belongs to $\partial_P(f+g)(x)$. Then for any $\varepsilon > 0$ there exist x', x'' in $B(x, \varepsilon)$ with

$$\left|f(x') - f(x)\right| < \varepsilon, \ \left|g(x'') - g(x)\right| < \varepsilon$$

together with $\zeta' \in \partial_P f(x')$ and $\zeta'' \in \partial_P g(x'')$ such that

$$\zeta \in \zeta' + \zeta'' + \varepsilon B.$$

(This is Ioffe's *fuzzy sum rule.*)

A smooth variational principle. Now assume that the lower semicontinuous function $f : X \to \mathbb{R} \cup \{+\infty\}$ is bounded below, and that the point x_0 satisfies

$$f(x_0) < \inf_{x \in X} f(x) + \varepsilon$$

for some $\varepsilon > 0$. The Borwein-Preiss variational principle [7] asserts the existence, for any $\lambda > 0$, of points y and z in X with

$$\|z - x_0\| < \lambda, \ \|y - z\| < \lambda, \ f(y) \leq f(x_0)$$

and having the property that the function

$$x \mapsto f(x) + \frac{\varepsilon}{\lambda^2} \|x - z\|^2$$

has a unique minimum at $x = y$.

A multidirectional mean value theorem. We take X finite-dimensional for simplicity. Let Y be a compact, convex subset of X and $f : X \to \mathbb{R} \cup \{+\infty\}$ a lower semicontinuous function. Let any $x \in \operatorname{dom} f$ be given, as well as $\varepsilon > 0$. Consider any real number r no greater than the quantity $\min_Y f - f(x)$ (which may equal $+\infty$). The mean value inequality (Clarke and Ledyaev [26]) asserts the existence of a point z in the ε-neighborhood of the set

$$[x, Y] := \text{co}\,[Y \cup \{x\}]$$

together with an element $\zeta \in \partial_P f(z)$ such that

$$r < \langle \zeta, y - x \rangle + \varepsilon \quad \forall\, y \in Y, \quad f(z) < \inf_{[x,Y]} f + |r| + \varepsilon.$$

Singular subgradients. The vector ζ is said to be a singular limiting subgradient of f at $x \in \text{dom}\, f$ if $(\zeta, 0)$ belongs to $N^L_{\text{epi}\, f}(x, f(x))$. The set of such elements is the singular limiting subdifferential $\partial^0_L f(x)$. It is occasionally useful to employ the alternative notation $\partial^1_L f(x)$ for $\partial_L f(x)$.

The generalized gradient. One can develop nonsmooth calculus in an arbitrary Banach space via the theory of *generalized gradients* (see Chapter 2 of [27]). In the case of a Hilbert space X and a locally Lipschitz real-valued function f on X, the generalized gradient $\partial_C f(x)$ coincides with $\text{co}\, \partial_L f(x)$; further, the associated normal cone $N^C_S(x)$ to a set S at a point x coincides with $\text{cl}\,\text{co}\, N^L_S(x)$.

1.4 Notes

§1.1 Basic references for the calculus of variations include the classical books of Bliss [5], Morrey [48], Cesari [10], and Tonelli [57]; we also refer the reader to Buttazzo et al. [9], Giaquinta et al. [31], Goldstine [32], and Ioffe and Tikhomirov [37]. The article [21] gives a more thorough account of the existence and regularity issues than we have given here; see also [22].

The book of Hestenes [34] systematically develops the multiplier rule approach, while that of Young [61] connects the calculus of variations to optimal control. For the latter, Lee and Marcus [40], Pontryagin et al. [49], and Warga [60] are standard references. The books of Clarke [19][22] and of Clarke et al. [27] foreshadow in certain respects many of the developments in this monograph, and give examples and applications in nonsmooth analysis.

§1.2 Early approaches to developing necessary conditions for differential inclusions (under smoothness or convexity hypotheses) are due to Blagodatskih [4] , Boltianski [6], Fedorenko [29], Halkin [33] and Rockafellar [50]. Recent work is discussed in the Notes to Chapter 3.

§1.3 Presentations of nonsmooth analysis may be found in the books of Clarke and Clarke et al. cited above, and also in those of Loewen [41], Rockafellar and Wets [54], and Vinter [58]; see also the survey article by Borwein and Zhu [8]. For viscosity solutions, see Bardi and Capuzzo-Dolcetta [2].

Chapter 2

Boundary Trajectories

There are two principal results in this chapter. We begin by deriving some necessary conditions for optimality in the special context of a Lipschitz differential inclusion with bounded values (Theorem 2.1.1). This in turn is used as one of the building blocks for the main result, Theorem 2.3.3. We conclude the chapter with some variants of that theorem.

2.1 A preparation theorem

Statement of the problem. We are given a multifunction F mapping $[a,b] \times \mathbb{R}^n$ to the subsets of $\mathbb{R}^n$ (as in the Introduction, §1.2), together with two real-valued functions ℓ_0 and ℓ_1. We also have a trajectory x_* of F on the interval $[a,b]$ such that $x_*(a)$ belongs to a given set C. Finally, there are specified a positive summable function θ on $[a,b]$, and numbers $\rho \in [1,2]$ and $\varepsilon > 0$. It is assumed that x_* solves the following problem: to minimize

$$\ell_1(x(b)) + \ell_0(x(a))$$

over the trajectories x of F on $[a,b]$ which satisfy

$$x(a) \in C, \quad \int_a^b |\dot{x}(t) - \dot{x}_*(t)|^\rho \, \theta(t)\, dt \ \leq \ \varepsilon, \ \|x - x_*\|_\infty \leq \varepsilon.$$

Hypotheses on the data. The set C is assumed to be closed, and the functions ℓ_0 and ℓ_1 locally Lipschitz. Regarding F, we posit the basic hypotheses that are always in force: the map $(t,x) \to F_t(x)$ is $\mathcal{L} \times \mathcal{B}$-

measurable[1], and the graph of F_t, the set $G(t) := \operatorname{gr} F_t(\cdot)$, is closed for each t. We also require in this section that F be integrably bounded and Lipschitz near x_*: there exists a (nonnegative) summable function k such that for almost every $t \in [a,b]$ we have

$$\begin{aligned} x \in \overline{B}(x_*(t),\varepsilon),\ v \in F_t(x) \quad &\Longrightarrow \quad |v| \le k(t); \\ x, x' \in \overline{B}(x_*(t),\varepsilon) \quad &\Longrightarrow \quad F_t(x) \subset F_t(x') + k(t)\left|x' - x\right| \overline{B}. \end{aligned}$$

Finally, we postulate that $k^\rho \theta$ is summable.

2.1.1 Theorem There exists an arc p satisfying the transversality conditions

$$-p(b) \in \partial_L \ell_1(x_*(b)), \quad p(a) \in \partial_L \ell_0(x_*(a)) + N_C^L(x_*(a)),$$

the Euler inclusion

$$\dot p(t) \in \operatorname{co}\left\{\omega : (\omega, p(t)) \in N^L_{G(t)}(x_*(t),\ \dot x_*(t))\right\},\ t \in [a,b] \text{ a.e.},$$

and the Weierstrass condition

$$\langle\, p(t), v - \dot x_*(t)\rangle \le 0 \quad \forall v \in F_t(x_*(t)),\ t \in [a,b] \text{ a.e.} \qquad \blacklozenge$$

Proof of Theorem 2.1.1

Some reductions

It is clear that we can take $[a,b] = [0,1]$, $x_* \equiv 0$ and $k(t) \ge 1$ a.e. In fact, we can suppose that $k(t) = 1$ a.e. For let us induce a change of time scale via $s = \tau(t)$, where

$$\tau(t) := \int_0^t k(r)\, dr,$$

and where transformed arcs y correspond to original ones x via $y(s) = x(t) = x(\tau^{-1}(s))$. This defines a one-to-one correspondence between arcs x on $[0,1]$ and arcs y on $[0,T]$, where $T := \int_0^1 k(t)\,dt$. We have

$$\int_0^1 |\,\dot x(t)|^\rho\, \theta(t)\, dt = \int_0^T |\,\dot y(s)|^\rho\, \hat\theta(s)\, ds,$$

[1]This refers to the smallest σ-field containing the products of Lebesgue measurable subsets of $[a,b]$ and Borel measurable subsets of $\mathbb{R}^n$. See for example [27] for the basic theory of measurable multifunctions.

where

$$\hat{\theta}(s) := k(\tau^{-1}(s))^{\rho-1}\theta(\tau^{-1}(s))$$

is a function which lies in L^1 (by the change of variables formula). Further, x is an F trajectory iff y is a trajectory of the multifunction $\hat{F}$ defined by

$$\hat{F}_s(y) := \frac{1}{k(t)}F_t(y),\ t = \tau^{-1}(s).$$

It follows that the arc $y_* \equiv 0$ corresponding to $x_* \equiv 0$ is a local minimum for the transformed problem (with $\hat{F}$) in the same sense that x_* is for the original one. But the multifunction $\hat{F}$ is bounded and Lipschitz with function $\hat{k}(s) = 1$ a.e.

It is a simple exercise in nonsmooth analysis to show that if an arc q satisfies the set of necessary conditions for the transformed problem, then the arc $p(t) := q(\tau(t))$ satisfies them for the original problem. Therefore we can (and do) suppose from the outset that $k = 1$ a.e. We may also take $\theta \geq 1$.

There is no loss of generality in supposing that $F_t(x)$ is defined for all x in R^n, and that the bound on F and the Lipschitz condition hold globally on R^n, by the following device: for a suitable choice of δ, and for $|x| > \delta$, we set $F_t(x) := F_t(\pi x)$, where πx is the projection of x on $\overline{B}(0,\delta)$. Finally, we note that there is no loss of generality in assuming that ℓ_0 and ℓ_1 are globally defined and Lipschitz and bounded below, and that C is compact.

An infimal convolution approximation

Quadratic infimal convolution is a basic mollifying tool in nonsmooth analysis. We apply it here to the function ℓ_1. For any positive number i, we define

$$\ell_1^i(x) := \min_{z\in\mathbb{R}^n}\left\{\ell_1(z) + i\,|x-z|^2\right\}.$$

Then (see [27]) ℓ_1^i is Lipschitz with the same constant as ℓ_1, and there exists for each i a positive number δ_i such that

$$\ell_1(x) - \delta_i \leq \ell_1^i(x) \leq \ell_1(x)\ \ \forall\, x \in \mathbb{R}^n,\ \lim_{i\to\infty}\delta_i = 0.$$

Let z be a point where the minimum defining $\ell_1^i(x)$ is attained. Then we have

$$\zeta := 2i(x-z) \in \partial_P\ell_1(x),$$

and ζ satisfies

$$\ell_1^i(x') \leq \ell_1^i(x) + \left\langle\zeta, x'-x\right\rangle + i\left|x'-x\right|^2 \quad \forall\, x' \in \mathbb{R}^n. \tag{2.1}$$

A penalization result

We denote by L^2_θ the Hilbert space of all measurable functions $f : [0,1] \to \mathbb{R}^n$ such that

$$\int_0^1 |f(t)|^2 \, \theta(t) \, dt < +\infty,$$

with inner product and norm defined by

$$\langle f, g \rangle_\theta := \int_0^1 \langle f(t), g(t) \rangle \, \theta(t) \, dt, \;\; \|f\|_{\theta,2} := \left\{ \int_0^1 |f(t)|^2 \, \theta(t) \, dt \right\}^{1/2}.$$

We now consider the minimization of

$$\ell_1^i(x(1)) + \ell_0(x(0)) + i \int_0^1 |u(t) - x(t)|^2 \, dt,$$

where $x(t) := c + \int_0^t v(s) \, ds$, and where the minimization is taken over the points $c = x(0) \in C$ and the functions u and v in L^2_θ satisfying

$$(u(t), v(t)) \in G(t) \text{ a.e.}, \int_0^1 |v(s)|^\rho \, \theta(s) \, ds \leq \frac{\varepsilon}{2}, |c| + \int_0^1 |v(t)| \, dt \leq \frac{\varepsilon}{2}. \quad (2.2)$$

It is easy to see that these conditions define a closed subset of the Hilbert space $X := \mathbb{R}^n \times L^2_\theta \times L^2_\theta$. For given i, let I_i denote the infimum in the problem just described. A feasible triple for this problem is $c = 0$, $u = v = 0$, and the corresponding cost is $\ell_1^i(0) + \ell_0(0)$, whence $I_i \leq \ell_1^i(0) + \ell_0(0)$. We now prove that this triple is 'nearly optimal' for i large; that is, we claim that

$$\lim_{i \to \infty} \left\{ I_i - \ell_1^i(0) - \ell_0(0) \right\} = 0.$$

For suppose the contrary. Then there exist $\Delta > 0$ and sequences $c_i = x_i(0) \in C$, u_i and v_i in L^2_θ satisfying (2.2) such that

$$\ell_1^i(x_i(1)) + \ell_0(x_i(0)) + i \int_0^1 |u_i(t) - x_i(t)|^2 \, dt \leq \ell_1^i(0) + \ell_0(0) - \Delta.$$

It follows that $\|x_i\|_\infty \leq \varepsilon/2$ and that

$$\int_0^1 |u_i(t) - x_i(t)|^2 \, dt \to 0.$$

Then we calculate, using the Lipschitz condition (d stands for Euclidean distance)

$$\begin{aligned} \int_0^1 d(\dot{x}_i(t), F_t(x_i(t)))\, dt &\leq \int_0^1 \{d(\dot{x}_i(t), F_t(u_i(t))) + |u_i(t) - x_i(t)|\}\, dt \\ &\leq 0 + \left[\int_0^1 |u_i(t) - x_i(t)|^2\, dt\right]^{1/2} \to 0. \end{aligned}$$

We invoke the well-known approximation theorem of Filippov[2] to deduce the existence, for all i sufficiently large, of a trajectory y_i of F such that

$$y_i(0) = x_i(0), \ \int_0^1 |\dot{y}_i(t) - \dot{x}_i(t)|^\rho\, \theta(t)\, dt \to 0.$$

Then for i large we have

$$\|y_i\|_\infty < \varepsilon, \ \int_0^1 |\dot{y}_i(t)|^\rho\, \theta(t)\, dt < \varepsilon$$

and

$$\ell_1(y_i(1)) + \ell_0(y_i(0)) \leq \ell_1(x_i(1)) + \ell_0(x_i(0)) + K\,|y_i(1) - x_i(1)|$$

(where K is a Lipschitz constant for $\ell_0 + \ell_1$)

$$\begin{aligned} &\leq \ell_1^i(x_i(1)) + \delta_i + \ell_0(x_i(0)) + K\int_0^1 |\dot{y}_i(t) - \dot{x}_i(t)|\, dt \\ &\leq \ell_1^i(0) + \ell_0(0) - \Delta + \delta_i + K\left[\int_0^1 |\dot{y}_i(t) - \dot{x}_i(t)|^\rho\, \theta(t)\, dt\right]^{1/\rho} \end{aligned}$$

(recall that $\theta \geq 1$)

$$< \ell_1^i(0) + \ell_0(0) \leq \ell_1(0) + \ell_0(0)$$

for i sufficiently large, which contradicts the optimality of $x_* = 0$ and proves the claim.

It follows from the Borwein-Preiss variational principle (applied in X; see §1.3) that there exist a sequence ε_i tending to 0, points $c_i, c_i' \in C \cap B(0, \varepsilon_i)$

[2] Actually, a simple variant that builds directly upon the usual proof to account for ρ and θ. For example, in the context of [58], it suffices to use (2.18) in conjunction with (2.19) and the fact that $k^\rho\theta$ (here reduced to θ) is summable.

and functions u_i, u_i', v_i, v_i' in L^2_θ all of whose norm in that space is less than ε_i such that the solution to the problem of minimizing

$$\ell_1^i(x(1)) + \ell_0(x(0)) + \varepsilon_i \left|x(0) - c_i'\right|^2 + i\int_0^1 |u(t) - x(t)|^2\, dt$$
$$+\varepsilon_i \int_0^1 \left|u(t) - u_i'(t)\right|^2 \theta(t)\, dt + \varepsilon_i \int_0^1 \left|v(t) - v_i'(t)\right|^2 \theta(t)\, dt$$

over the elements $c = x(0) \in C$ and $u, v \in L^2_\theta$ satisfying (2.2) is given by $x(0) = c_i, u = u_i, v = v_i$. Since u_i goes to 0 in L^2_θ, it goes to 0 in L^1_θ (that is, $u_i\theta$ goes to 0 in L^1). Thus by taking subsequences if necessary we can assume that $u_i\theta$, and hence u_i, converges almost everywhere to 0. Analogous arguments allow us to suppose that all four function sequences u_i, v_i, u_i', v_i' converge in L^1_θ and almost everywhere to 0.

Variational analysis

Because $c_i \to 0$ and

$$\int_0^1 |v_i(t)|^2 \theta(t)\, dt < \varepsilon_i^2 \to 0,$$

the two integral constraints in (2.2) become inactive at v_i for i sufficiently large (the fact that $\rho \le 2$ is essential here). Fix such a value of i.

We shall now express the necessary conditions corresponding to the solution of this problem, in terms of the arc p_i defined by

$$\dot p_i(t) := 2i(x_i(t) - u_i(t)),\ p_i(1) = -\zeta_i,$$

(note that $\dot p_i$ does lie in L^1) where $\zeta_i = 2i(x_i(1) - z_i)$, and where z_i is a point at which the minimum defining $\ell_1^i(x_i(1))$ is attained. (Of course, x_i signifies here the arc with initial value c_i and derivative v_i.) Then we have

$$-p_i(1) \in \partial_P \ell_1(x_i(1)) \tag{2.3}$$

by construction. We note the identity

$$i\,|u - x|^2 = i\,|u_i - x_i|^2 - \langle \dot p_i, u - u_i\rangle + \langle \dot p_i, x - x_i\rangle + i\,|(x - x_i) - (u - u_i)|^2,$$

which, together with (2.1), implies that, subject to (2.2), the minimum of

$$\begin{aligned}&\langle -p_i(1), x(1)\rangle + i\,|x(1) - x_i(1)|^2 + \ell_0(x(0)) + \varepsilon_i \left|x(0) - c_i'\right|^2 \\ &+ \int_0^1 \{\langle \dot{p}_i(t), x(t)\rangle - \langle \dot{p}_i(t), u(t)\rangle\}\, dt \\ &+ 2i \int_0^1 \left\{|u(t) - u_i(t)|^2 + |x(t) - x_i(t)|^2\right\} dt \\ &+ \varepsilon_i \int_0^1 \left|u(t) - u_i'(t)\right|^2 \theta(t)\, dt + \varepsilon_i \int_0^1 \left|v(t) - v_i'(t)\right|^2 \theta(t)\, dt\end{aligned}$$

is attained at u_i, x_i.

We use integration by parts to rewrite the cost functional as

$$\begin{aligned}&\langle -p_i(0), x(0)\rangle + i\,|x(1) - x_i(1)|^2 + \ell_0(x(0)) + \varepsilon_i \left|x(0) - c_i'\right|^2 \\ &+ \int_0^1 \{\langle -p_i(t), v(t)\rangle - \langle \dot{p}_i(t), u(t)\rangle\}\, dt \\ &+ 2i \int_0^1 \left\{|u(t) - u_i(t)|^2 + |x(t) - x_i(t)|^2\right\} dt \\ &+ \varepsilon_i \int_0^1 \left|u(t) - u_i'(t)\right|^2 \theta(t)\, dt + \varepsilon_i \int_0^1 \left|v(t) - v_i'(t)\right|^2 \theta(t)\, dt,\end{aligned}$$

a functional that we shall label $\Phi(c, u, v)$, where we continue to identify c with $x(0)$.

Let us now fix the values $u = u_i, v = v_i$ in Φ. Then a local minimum of $\Phi(\cdot, u_i, v_i)$ over C is attained at $c = c_i$, whence

$$p_i(0) - 2\varepsilon_i(x_i(0) - c_i') \in \partial_P \{\ell_0 + I_C\}(x_i(0)), \tag{2.4}$$

where I_C is the indicator function of C. The transversality information is contained in (2.3) and (2.4).

We turn now to the Weierstrass condition, which will follow from the existence of a minimum of Φ with respect to the v variable. We claim that for almost each $t \in [0, 1]$, one has

$$\begin{aligned}&\langle p_i(t), v - v_i(t)\rangle + \varepsilon_i \left|v_i(t) - v_i'(t)\right|^2 \theta(t) \\ &\qquad\qquad \le \varepsilon_i \left|v - v_i'(t)\right|^2 \theta(t) \quad \forall\, v \in F_t(u_i(t)),\end{aligned}$$

which implies the approximate Weierstrass condition

$$\langle p_i(t), v - v_i(t)\rangle \le 8\varepsilon_i \theta(t) \quad \forall\, v \in F_t(u_i(t)). \tag{2.5}$$

Suppose the contrary. Then there is a measurable selection $\bar{v}$ of $F_t(u_i(t))$, a subset S of $[0,1]$ of positive measure, and a positive number η such that, for all t in S,

$$\langle p_i(t), \bar{v}(t) - v_i(t)\rangle + \varepsilon_i \left|v_i(t) - v_i'(t)\right|^2 \theta(t) \geq \eta + \varepsilon_i \left|\bar{v}(t) - v_i'(t)\right|^2 \theta(t).$$

Then for each $\lambda > 0$ sufficiently small, let S_λ be a subset of S having measure λ, and define a function $v_\lambda(t)$ to be equal to $\bar{v}(t)$ when t lies in S_λ, and equal to $v_i(t)$ otherwise. It follows that if x_λ signifies the arc corresponding to v_λ (with initial condition c_i), we have

$$|x_\lambda(t) - x_i(t)| \leq 2\lambda.$$

For λ small enough, v_λ belongs to the set relative to which $\Phi(c_i, u_i, \cdot)$ attains a minimum at v_i, whence

$$\begin{aligned} i\,|x_\lambda(1) - x_i(1)|^2 + \int_0^1 \langle -p_i(t), v_\lambda(t)\rangle \, dt \\ +2i\int_0^1 |x_\lambda(t) - x_i(t)|^2 \, dt + \varepsilon_i \int_0^1 \left|v_\lambda(t) - v_i'(t)\right|^2 \theta(t)\, dt \\ \geq \int_0^1 \langle -p_i(t), v_i(t)\rangle \, dt + \varepsilon_i \int_0^1 \left|v_i(t) - v_i'(t)\right|^2 \theta(t)\, dt. \end{aligned}$$

Together these imply $12i\lambda^2 \geq \eta\lambda$ for all $\lambda > 0$ sufficiently small, the required contradiction.

There remains the Euler inclusion to extract. To this end, fix $c = c_i$ in Φ and consider $\Phi(c_i, \cdot, \cdot)$. The terms

$$|x(1) - x_i(1)|^2, \quad \int_0^1 |x(t) - x_i(t)|^2 \, dt$$

are both majorized by

$$\int_0^1 |v(t) - v_i(t)|^2 \, dt,$$

so it follows that the minimum of the expression

$$\begin{aligned} &\int_0^1 \{\langle -p_i(t), v(t)\rangle - \langle \dot{p}_i(t), u(t)\rangle\}\, dt \\ &+ \quad 3i\int_0^1 \left\{|u(t) - u_i(t)|^2 + |v(t) - v_i(t)|^2\right\} dt \\ &+ \quad \varepsilon_i \int_0^1 \left|u(t) - u_i'(t)\right|^2 \theta(t)\, dt + \varepsilon_i \int_0^1 \left|v(t) - v_i'(t)\right|^2 \theta(t)\, dt \end{aligned}$$

over the u and v in L^2_θ such that

$$(u(t),v(t)) \in G(t) \text{ a.e.}, \int_0^1 |v(t)|^\rho\, \theta(t)\, dt \le \frac{\varepsilon}{2},\ |c_i| + \int_0^1 |v(t)|\, dt \le \frac{\varepsilon}{2}$$

is attained at $u = u_i, v = v_i$. Because the integral constraints are slack at v_i, it follows from an elementary measurable selection argument[3] that for almost each t, the integrand is minimized over $G(t)$ at $u = u_i(t), v = v_i(t)$. This implies that for almost all t,

$$\begin{aligned}(\dot p_i(t), p_i(t)) \in\ & N^P_{G(t)}(u_i(t), v_i(t)) \\ & + 2\varepsilon_i \theta(t)(u_i(t) - u_i'(t), v_i(t) - v_i'(t)).\end{aligned} \tag{2.6}$$

The relations (2.3)(2.4)(2.5)(2.6) will yield the conclusions of the theorem in the limit.

Because F_t is Lipschitz with constant 1, it follows from (2.6) that

$$|\dot p_i(t)| \le |p_i(t)| + 2\varepsilon_i \theta(t) \left\{ \left|u_i(t) - u_i'(t)\right| + \left|v_i(t) - v_i'(t)\right| \right\} \text{ a.e.}$$

Together with (2.3), this implies via Gronwall's Lemma that for a subsequence (we do not relabel) we have that p_i converges uniformly to a limiting arc p and $\dot p_i$ converges weakly to $\dot p$. It is easy to see that p satisfies the transversality and Weierstrass conditions in the limit.

There remains the Euler inclusion to verify. It follows from a fundamental stability property of the inclusion with respect to the type of convergence present here. We record for later reference a more general result than is needed at the moment, one which concludes the proof of the theorem. The following is a direct consequence of the theorems of Carathéodory and Mazur.[4]

2.1.2 Proposition Let F satisfy the basic hypotheses, and let p_i be a sequence of arcs on $[0,1]$ converging uniformly to an arc p, where $\dot p_i$ converges weakly in L^1 to $\dot p$. Suppose that for each index i, p_i satisfies the following (approximate) Euler inclusion at (u_i, v_i) on a subset Ω_i of $[0,1]$:

$$\dot p_i(t) + \alpha_i(t) \in \operatorname{co} \left\{ \omega : (\omega, p_i(t) + \beta_i(t)) \in N^L_{G(t)}(u_i(t), v_i(t)) \right\},\ t \in \Omega_i,$$

where the measurable functions (u_i, v_i) converge almost everwhere to a limit function (u_0, v_0), the measurable functions (α_i, β_i) converge in L^1 to $(0,0)$,

[3] Details of the measurable selection argument are given in [23].

[4] The argument is given in [58], pp. 250-1; alternatively, one can argue as in Theorem 3.1.7 of [19].

and where the measure of Ω_i converges to 1. Then p satisfies the Euler inclusion at (u_0, v_0):

$$\dot{p}(t) \in \mathrm{cl\,co}\left\{\omega : (\omega, p(t)) \in N^L_{G(t)}(u_0(t), v_0(t))\right\},\ t \in [0,1] \text{ a.e.} \qquad \blacklozenge$$

Note that the closure operation in the limiting Euler inclusion of Proposition 2.1.2 is superfluous when F_t is Lipschitz (or later, pseudo-Lipschitz), because in that case the set to which it is applied is already closed. This follows from the elementary fact that any limiting normal vector (α, β) to $G(t)$ must satisfy $|\alpha| \le k(t)\,|\beta|$.

The theorem now follows upon applying the proposition to the convergent sequence p_i above, in view of (2.6), and because the proximal normal cone is contained in the limiting one. ■

2.1.3 Remark (a) The restriction $\rho \le 2$ in the theorem is necessary for technical reasons in the proof, but we do not know if it reflects a fundamental limit to the validity of the Weierstrass condition.

(b) Passing to the limit in (2.4) yields a somewhat sharper transversality condition at a than the one given in the statement of the theorem:

$$p(a) \in \partial_L \{\ell_0 + I_C\}(x_*(a)).$$

2.2 An integral cost functional

Consider now the variant of the problem in which there is an explicit integral term in the cost functional: x_* minimizes locally (in the same sense as before) the cost functional

$$\ell_1(x(b)) + \ell_0(x(a)) + \int_a^b \Lambda_t(\dot{x}(t))\,dt,$$

the other data of the problem being unchanged. We assume that Λ is $\mathcal{L} \times \mathcal{B}$ measurable, and bounded and Lipschitz as follows: there exists a constant k_Λ such that, for almost every t, we have

$$\begin{aligned}
&x \in \overline{B}(x_*(t), \varepsilon),\ v \in F_t(x) \Longrightarrow |\Lambda_t(v)| \le k_\Lambda;\\
&x, x' \in \overline{B}(x_*(t), \varepsilon),\ v \in F_t(x) + B,\ v' \in F_t(x') + B\\
&\Longrightarrow \left|\Lambda_t(v') - \Lambda_t(v)\right| \le k_\Lambda \left|v' - v\right|.
\end{aligned}$$

2.2.1 Corollary There is an arc p which satisfies the transversality conditions

$$-p(b) \in \partial_L \ell_1(x_*(b)),\ p(a) \in \partial_L \ell_0(x_*(a)) + N_C^L(x_*(a)),$$

the Euler inclusion: almost everywhere, $\dot{p}(t)$ belongs to the set

$$\operatorname{co}\left\{\omega : (\omega, p(t)) \in N^L_{G(t)}(x_*(t), \dot{x}_*(t)) + \{0\} \times \partial_L \Lambda_t(\dot{x}_*(t))\right\},$$

and the Weierstrass condition

$$\Lambda_t(v) - \Lambda_t(\dot{x}_*(t)) \ \geq\ \langle\, p(t), v - \dot{x}_*(t)\rangle \quad \forall\, v \in F_t(x_*(t)), \quad \text{a.e. } t \in [a,b]. \ \blacklozenge$$

Proof. The added integral cost can be absorbed into the differential inclusion by the familiar device of state augmentation: we introduce an additional one-dimensional state variable y satisfying $y(a) = 0$ together with the following augmented dynamics:

$$(\dot{x}, \dot{y}) \in \{(v, w) : v \in F_t(x), \Lambda_t(v) \leq w \leq k_\Lambda + 1\}.$$

The problem now amounts to minimizing

$$\ell_1(x(b)) + \ell_0(x(a)) + y(b)$$

subject to this augmented inclusion. This has the form of the problem treated by the theorem, all of whose hypotheses are easily verified. Its application then leads to an arc (p, q) for which q is identically -1. The augmented Euler inclusion therefore involves a relation of the type (we suppress the t variable)

$$(\omega, p, -1) \in N^L_D(x, v, \Lambda(v)),$$

where

$$D := \left\{(x', v', r') : v' \in F(x'),\ \Lambda(v') \leq r' \leq k_\Lambda + 1\right\}.$$

This condition is equivalent to

$$(\omega, p, -1) \in N^L_E(x, v, \Lambda(v)),$$

where

$$E := \left\{(x', v', r') : v' \in F(x'),\ \Lambda(v') \leq r'\right\}.$$

But E is the intersection of the two sets $E_1 := G \times \mathbb{R}$ and $E_2 := \mathbb{R}^n \times G_\Lambda$, where G_Λ signifies the epigraph of Λ. Because Λ is Lipschitz, the proximal sum rule for normal cones (see [27], 1.8.4) implies (in the limit) that an element of $N_E^L(x,v,\Lambda(v))$ is the sum of one in $N_{E_1}^L(x,v,\Lambda(v))$ and one in $N_{E_2}^L(x,v,\Lambda(v))$. This leads to

$$(\omega,p)\in N_G^L(x,v)+\{0\}\times\partial_L\Lambda(v).$$

It follows now that the augmented Euler inclusion implies the desired one, and the other conclusions of the corollary follow immediately.

2.3 A general theorem on boundary trajectories

As before, F is a multifunction from $[a,b]\times\mathbb{R}^n$ to $\mathbb{R}^n$ and C a closed subset of $\mathbb{R}^n$. The permanent basic hypotheses require that F be $\mathcal{L}\times\mathcal{B}$-measurable and that for each t, $F_t(\cdot)$ have closed graph. Let the arc x_* be a trajectory of F on $[a,b]$ having $x_*(a)\in C$.

We are given a measurable function $R:[a,b]\to(0,+\infty]$ which we refer to as the *radius function*, or simply *radius*. We stress that R is extended-valued; below, the choice $R\equiv+\infty$ is admissible.

Given $\varepsilon>0$, we consider the set $\mathcal{T}=\mathcal{T}(x_*,R,\varepsilon,C,F)$ of trajectories x of F on $[a,b]$ which satisfy the constraints

$$x(a)\in C,\ |\dot{x}(t)-\dot{x}_*(t)|\le R(t)\ \text{ a.e.}$$

and which are ε-close to x_* in the following $W^{1,1}$ sense:

$$\int_a^b|\dot{x}(t)-\dot{x}_*(t)|\,dt\le\varepsilon,\ \|x-x_*\|_\infty\le\varepsilon.$$

Let $\Phi:\mathbb{R}^n\to\mathbb{R}^m$ be a given locally Lipschitz function. We say that x_* is a *local boundary trajectory* if, for some $\varepsilon>0$, $\Phi(x_*(b))$ is a boundary point of the set

$$\Phi_{\mathcal{T}}:=\{\Phi(x(b)):x\in\mathcal{T}\}.$$

We proceed to define two properties constituting the principal hypotheses that will be made in the theorem.

2.3.1 Definition F is said to satisfy a pseudo-Lipschitz condition of radius R near x_* if there exist $\varepsilon>0$ and a summable function k such that, for almost all $t\in[a,b]$, for every x and x' in $B(x_*(t),\varepsilon)$, one has

$$F_t(x')\cap\overline{B}(\dot{x}_*(t),R(t))\subset F_t(x)+k(t)\left|x'-x\right|\overline{B}.$$

When R is identically $+\infty$, the above reduces to a (true) Lipschitz condition.

> **2.3.2 Definition** F is said to satisfy the tempered growth condition for radius R near x_* if there exist $\varepsilon > 0$, $\lambda \in (0,1)$, and a summable function r_0 such that for almost every $t \in [a,b]$ we have $0 < r_0(t) \leq \lambda R(t)$ and
>
> $$|x - x_*(t)| \leq \varepsilon \Longrightarrow F_t(x) \cap B(\dot{x}_*(t), r_0(t)) \neq \emptyset.$$

Note that by taking the minimum of the three parameters ε mentioned above, we may assume that the ε defining the local minimum is the same as that of both the pseudo-Lipschitz and the tempered growth conditions.

2.3.3 Theorem Suppose that x_* is a local boundary trajectory in the above sense where, for radius R, F satisfies the pseudo-Lipschitz and tempered growth conditions near x_*. Then there exist an arc p on $[a,b]$ and a unit vector γ in R^m such that the following transversality conditions hold

$$-p(b) \in \partial_L \langle \gamma, \Phi \rangle (x_*(b)),\ p(a) \in N_C^L(x_*(a)),$$

and such that p satisfies the Euler inclusion:

$$\dot{p}(t) \in \text{co}\left\{\omega : (\omega, p(t)) \in N_{G(t)}^L(x_*(t), \dot{x}_*(t))\right\} \ \text{a.e. } t \in [a,b],$$

as well as the Weierstrass condition of radius R: for almost every t in $[a,b]$ we have

$$\langle p(t), v \rangle \leq \langle p(t), \dot{x}_*(t) \rangle \ \forall v \in F_t(x_*(t)) \cap \overline{B}(\dot{x}_*(t), R(t)). \qquad \blacklozenge$$

2.4 Proof of the theorem

Some reductions

There is no loss of generality in taking the ε of the pseudo-Lipschitz condition, that of the tempered growth condition, and that of the local boundary trajectory all equal, and in supposing $x_* \equiv 0$, which we do henceforth. We also take $[a,b] = [0,1]$. Let us induce a change of time scale via $s = \tau(t)$, where

$$\tau(t) := \int_0^t r_0(\sigma)\, d\sigma,$$

and where transformed arcs y correspond to original ones x via $y(s) = x(t) = x(\tau^{-1}(s))$. This defines a one-to-one correspondence between arcs x on $[0,1]$ and arcs y on $[0,T]$, where $T := \tau(1)$, and x is an F trajectory iff y is a trajectory of the multifunction $\hat{F}$ defined by

$$\hat{F}_s(y) := \frac{1}{r_0(t)} F_t(y),\ t = \tau^{-1}(s).$$

It follows that the arc $y_* \equiv 0$ corresponding to $x_* \equiv 0$ is a local boundary trajectory for the transformed problem (with $\hat{F}$) in the same sense that x_* is for the original one, and for the radius function

$$\hat{R}(s) := \frac{R(t)}{r_0(t)}, \quad (t = \tau^{-1}(s)).$$

Furthermore, the multifunction $\hat{F}$ satisfies the tempered growth condition with $\hat{r}_0 := 1$ (and the same λ), and its pseudo-Lipschitz function is

$$\hat{k}(s) := \frac{k(t)}{r_0(t)} \quad (t = \tau^{-1}(s)),$$

which is summable over $[0,T]$. Since the conclusions of the theorem for the transformed data are easily seen to imply those for the original data, we conclude that without loss of generality we may take r_0 to be constant.

An auxiliary multifunction

Let us now proceed to select some parameters that will be used in the proof. First, fix any $N > r_0$ and set

$$R_N(t) := \min\,[N, R(t)].$$

We shall work below with the bounded radius function R_N. Next we select any $\lambda_0 \in (0,1)$ and $\eta \in (0,1)$ such that

$$r_0 < (\lambda_0 - 2\eta)\ \text{ess inf}\,\{R_N(t) : t \in [0,1]\}.$$

In the final stage of the proof, we let λ_0 and η converge to 1 and 0 respectively, and let N go to $+\infty$.

The proof of the theorem will employ the multifunction Γ defined as follows: for $(t,x) \in [0,1] \times \overline{B}(0,\varepsilon)$ we set

$$\Gamma_t(x) := \{(\lambda f, \lambda) : \lambda \in [0,1], f \in F_t(x), |f| \le (1 - \lambda\eta) R_N(t)\}$$

(note that there exist points λ and f as described, in view of (2.3.2)). Then Γ satisfies the basic hypotheses, and is uniformly bounded. Furthermore, Γ is integrably Lipschitz near 0:

Lemma 1 For almost all $t \in [0,1]$, for all x and x' in $\overline{B}(0,\varepsilon)$, one has

$$\Gamma_t(x) \subset \Gamma_t(x') + \frac{(2+R_N(t))k(t)}{\eta R_N(t)} \left|x'-x\right| \overline{B}.$$

To see this, fix t such that the Lipschitz hypothesis on F holds, and let $(\lambda f, \lambda)$ be any point in $\Gamma_t(x)$. We shall exhibit a point $(\lambda' f', \lambda')$ in $\Gamma_t(x')$ such that

$$\left|(\lambda' f', \lambda') - (\lambda f, \lambda)\right| \leq \frac{(2+R_N(t))k(t)}{\eta R_N(t)} \left|x'-x\right|.$$

Set $\delta := |x'-x|$. Consider first the case in which $\lambda \leq \delta k(t)/(\eta R_N(t))$. Then the choice $\lambda' = 0$ is suitable (with any $f' \in F_t(x')$ satisfying $|f'| < r_0$, say). Suppose then that $\lambda > \delta k(t)/(\eta R_N(t))$. By assumption we have

$$f \in F_t(x) \cap B(0,(1-\lambda\eta)R_N(t)) \subset F_t(x) \cap B(0,R(t)).$$

Apply the pseudo-Lipschitz property to deduce the existence of $f' \in F_t(x')$ such that $|f'-f| \leq \delta k(t)$. Note that $|f'| < R_N(t)$. If $|f'| \leq |f|$, the choice $\lambda' = \lambda$ leads to the required element of $\Gamma_t(x')$. So let us suppose $|f'| > |f|$. In this case we set

$$\lambda' := \lambda - \frac{|f'|-|f|}{\eta R_N(t)}.$$

It follows that $(\lambda' f', \lambda')$ is the required element of $\Gamma_t(x')$, and the lemma is proven.

An auxiliary problem

Let us define a function $\phi_t : [0,+\infty) \to [0,+\infty)$ as follows:

$$\phi_t(r) := \frac{\lambda_0 \left[r - (\lambda_0 - 2\eta)R_N(t)\right]_+^2}{4(1-\lambda_0+2\eta)R_N^2(t)},$$

where $[s]_+$ signifies $\max(s,0)$. This smooth nondecreasing function is picked so as to satisfy the following conditions, invoked later on:

$$r \in [0,(\lambda_0-2\eta)R_N(t)] \Longrightarrow \phi_t(r) = 0 \tag{2.7}$$

$$\phi_t((\lambda_0-\eta)R_N(t)) = \frac{\lambda_0\eta^2}{4(1-\lambda_0+2\eta)} > 0 \tag{2.8}$$

$$r \in [0,R_N(t)] \Longrightarrow \phi_t'(r) \leq \frac{\lambda_0}{2R_N(t)} \leq \frac{\lambda_0}{2r_0}, \quad \phi_t'(r)r \leq \lambda_0/2. \tag{2.9}$$

Since $\Phi(0) = \Phi(x_*(b))$ is a boundary point of the set $\Phi_{\mathcal{T}}$, there exists a sequence of points μ_i not lying in $\Phi_{\mathcal{T}}$ and converging to $\Phi(0)$.

We consider the problem of minimizing

$$|\mu_i - \Phi(x(b))| - y(1) + \int_0^1 \phi_t(|\dot{x}(t)|)\, dt$$

over the arcs (x, y) on $[0, 1]$ satisfying

$$(\dot{x}, \dot{y}) \in \Gamma_t(x) \text{ a.e., } x(0) \in C,\ y(0) = 0 \tag{2.10}$$

and

$$|x(0)| \le \varepsilon/2,\ \int_0^1 |\dot{x}|\, dt \le \varepsilon/2. \tag{2.11}$$

We view this problem as one in which the choice variables are $\dot{x}, \dot{y} \in L^2$, and $x(0) \in \mathbb{R}^n$. Any feasible (x, y) has

$$\|x\|_\infty \le \varepsilon,\ \int_0^1 |\dot{x}(t)|\, dt \le \varepsilon.$$

The infimum in the problem is no less than -1, and the arc $(0, t)$, which is feasible, provides the value $\varepsilon_i - 1$ to the cost, where $\varepsilon_i := |\mu_i - \Phi(0)| > 0$. (Note that ε_i goes to 0 as i goes to $+\infty$.) We invoke the Borwein-Preiss variational principle on $L^2 \times L^2 \times \mathbb{R}^n$ to deduce the existence of ε_i' (also going to 0), functions $v_i, v_i', \lambda_i, \lambda_i'$ in L^2 with

$$\|v_i\|_2 < \varepsilon_i',\ \|\lambda_i - 1\|_2 < \varepsilon_i',\ \left\|v_i'\right\|_2 < \varepsilon_i',\ \left\|\lambda_i' - 1\right\|_2 < \varepsilon_i',$$

as well as points z_i, z_i' in $\mathbb{R}^n$ with

$$|z_i| < \varepsilon_i',\ \left|z_i'\right| < \varepsilon_i'$$

such that the arc

$$(x_i(t), y_i(t)) := \left(z_i + \int_0^t v_i(s)\, ds,\ \int_0^t \lambda_i(s)\, ds\right)$$

solves the problem of minimizing

$$\begin{aligned} &|\mu_i - \Phi(x(b))| - y(1) + \varepsilon_i' \left|x(0) - z_i'\right|^2 \\ &\quad + \int_0^1 \phi_t(|\dot{x}(t)|)\, dt + \varepsilon_i' \int_0^1 \left|\dot{x} - v_i'\right|^2 dt + \varepsilon_i' \int_0^1 \left|\dot{y} - \lambda_i'\right|^2 dt \end{aligned}$$

over the arcs (x, y) satisfying (2.10) and (2.11). It follows that (x_i, y_i) converges uniformly to $(0, t)$, and we may suppose that $(v_i, v_i', \lambda_i, \lambda_i')$ converges almost everywhere to $(0, 0, 1, 1)$. We may write $\dot{x}_i(t)$ in the form $\lambda_i(t) f_i(t)$, where $f_i(t) \in F_t(x_i(t))$, and where f_i converges almost everywhere to 0.

Necessary conditions for the auxiliary problem

For i large enough, the constraints in (2.11) are slack at x_i, and the necessary conditions of Corollary 2.2.1 apply (we take $\theta \equiv 1, \rho = 1$). Fix such an i. If (p, q) is the arc corresponding to (x_i, y_i) in those necessary conditions, it follows that q is identically 1. Using the limiting chain rule, the transversality at $t = 1$ may then be written as follows:

$$-p(1) \in \partial_L \langle \gamma, \Phi \rangle (x_i(1)), \tag{2.12}$$

where γ is a vector of length not exceeding 1. At 0 we have

$$p(0) - 2\varepsilon_i'(x_i(0) - z_i') \in N_C^L(x_i(0)). \tag{2.13}$$

The Weierstrass condition affirms that almost everywhere, the expression

$$\langle p(t), \lambda f \rangle + \lambda - \phi_t(|\lambda f|) - \varepsilon_i' \left|\lambda f - v_i'\right|^2 - \varepsilon_i' \left|\lambda - \lambda_i'\right|^2 \tag{2.14}$$

is maximized relative to

$$f \in F_t(x_i(t)),\ \lambda \in [0,1],\ |f| \le (1 - \lambda\eta) R_N(t)$$

by the choice $\lambda = \lambda_i(t), f = f_i(t)$. Finally, the Euler inclusion affirms that for almost all t, $\dot{p}(t)$ belongs to the convex hull of the set of ω satisfying

$$(\omega, p(t) - \psi_i(t), 1 - \zeta_i(t)) \in N_{D(t)}^L(x_i(t), \lambda_i(t) f_i(t), \lambda_i(t)),$$

where $D(t)$ signifies the graph of $\Gamma_t(\cdot)$, and where

$$\psi_i(t) := 2\varepsilon_i'(\dot{x}_i(t) - v_i'(t)) + \phi_t'(|\dot{x}_i(t)|) \frac{\dot{x}_i(t)}{|\dot{x}_i(t)|},$$

$$\zeta_i(t) := 2\varepsilon_i'(\lambda_i(t) - \lambda_i'(t)).$$

Note that ψ_i and ζ_i are essentially bounded (independently of i) and go to 0 almost everywhere. The Lipschitz condition satisfied by Γ implies

$$|\dot{p}(t)| \le \frac{(2+N)k(t)}{\eta r_0} \{|p(t)| + |\psi_i(t)| + |\zeta_i(t)| + 1\} \quad \text{a.e. } t \in [0,1]. \tag{2.15}$$

The following simple fact, whose proof we omit, relates the Euler inclusion above (in the case of the proximal normal cone) to that for F.

Lemma 2 Let a point $(\bar{\omega}, \bar{p}, \bar{q})$ lie in $N_{D(t)}^P(\bar{x}, \bar{\lambda} \bar{f}, \bar{\lambda})$, where $\left|\bar{f}\right| < (1 - \bar{\lambda}\eta) R_N(t)$. Then we have

$$(\bar{\omega}, \bar{\lambda}\, \bar{p}) \in N_{G(t)}^P(\bar{x}, \bar{f}).$$

Because the limiting normal cone is generated by proximal normals, it follows that we have

$$\dot{p}(t) \in \text{co}\left\{\omega : (\omega, \lambda_i(t)(p(t) - \psi_i(t))) \in N^L_{G(t)}(x_i(t), f_i(t))\right\} \text{ a.e. } t \in \Omega_i,$$

where Ω_i is the set of points at which $|f_i(t)|$ is strictly less than $(1 - \lambda_i(t)\eta)R_N(t)$. Since f_i and λ_i converge in L^2 to 0 and 1 respectively, it follows that the measure of Ω_i goes to 1 as i goes to infinity.

Convergence

The arc p obtained above depends of course on i, so let us now denote it p_i. Similarly, we write γ_i for the vector that appears in (2.12). By taking a subsequence if necessary (we do not relabel), we may suppose that γ_i converges to a limit γ. In the presence of (2.15), and because $p_i(1)$ is bounded as a consequence of (2.12), a conventional argument in conjunction with Proposition 2.1.2 shows that, by passing again to a subsequence if necessary, it may be supposed that p_i converges uniformly to a limiting arc p which satisfies

$$-p(1) \in \partial_L \langle \gamma, \Phi \rangle (0),\ p(0) \in N^L_C(0)$$

and

$$\dot{p}(t) \in \text{co}\left\{\omega : (\omega, p(t)) \in N^L_{G(t)}(0,0)\right\} \text{ a.e. } t \in [0,1]. \tag{2.16}$$

We claim that we also obtain, for almost every t,

$$\langle p(t), f\rangle \leq 0 \quad \forall\, f \in F_t(0) \cap \overline{B}(0, (\lambda_0 - 2\eta)R_N(t)). \tag{2.17}$$

To see this, fix a value of t for which all the convergences hold, as well as the pseudo-Lipschitz condition, and for which the Weierstrass condition (2.14) holds for every index i (almost all t satisfy these conditions). Choose $f \in F_t(0) \cap B(0, (\lambda_0 - 2\eta)R_N(t))$. By the pseudo-Lipschitz property there is a point $w_i \in F_t(x_i(t))$ satisfying $|w_i - f| \leq k(t)\,|x_i(t)|$. Note that w_i converges to f. From the Weierstrass condition (the maximization of (2.14)) we deduce, for i large enough so that $w_i < (1 - \lambda_i(t)\eta)R_N(t)$, and taking $\lambda = \lambda_i(t)$,

$$\langle p_i(t), \lambda_i(t) w_i\rangle \leq \phi_t(|\lambda_i(t) w_i|) + \varepsilon_i' \left|v_i(t) - v_i'(t)\right|^2.$$

In the limit we derive $\langle p(t), f\rangle \leq 0$ as required, since $\phi_t(|f|) = 0$. Note that the theorem asserts the inequality (2.17) on a bigger set, namely $F_t(0) \cap \overline{B}(0, R(t))$; we return to this point later.

Nontriviality

The issue we address now is that of nontriviality of (p, γ); that is, we claim that γ and p are not both zero. If for an infinite number of indices i it is the case that $\mu_i \neq \Phi(x_i(1))$, then, for each such index, γ_i is a unit vector, so that the limit γ is a unit vector too, and nontriviality follows.

Let us examine the other case in which we have $\mu_i = \Phi(x_i(1))$ for all but finitely many i. It follows that for all i sufficiently large, $\lambda_i(t)$ must be strictly less than 1 on a set of positive measure, for otherwise x_i would be a trajectory for F, in fact, an element of $\mathcal{T}$, and hence μ_i would belong to $\Phi_{\mathcal{T}}$, a contradiction. We shall prove that in this case we have

$$\|p\|_\infty \geq \min\left(\frac{1-\lambda_0}{N+r_0}, \frac{\lambda_0,}{2N}, \frac{\lambda_0\eta^2}{8N(1-\lambda_0+2\eta)}\right) > 0. \tag{2.18}$$

One of the following cases must occur infinitely often on a set of positive measure (depending on i):

1. $\lambda_i(t) \leq \lambda_0$;
2. $\lambda_0 < \lambda_i(t) < 1$ and $|f_i(t)| < (1 - \lambda_i(t)\eta)R_N(t)$;
3. $\lambda_0 < \lambda_i(t) < 1$ and $|f_i(t)| = (1 - \lambda_i(t)\eta)R_N(t)$.

We shall examine the Weierstrass condition corresponding to (2.14) to reach the desired conclusion.

In Case 1, fix a suitable index i and consider the Weierstrass inequality with $\lambda = 1$ and any $f \in F_t(x_i(t)) \cap \overline{B}(0, r_0)$ (such an f exists by the tempered growth hypothesis, and the choice is admissible because $r_0 < (1-\eta)R_N(t)$). Since $\phi_t(|f|) = 0$, we derive that on a set of positive measure we have

$$\langle\, p_i(t), \lambda_i(t)f_i(t) - f\rangle \geq 1 - \lambda_0 - \varepsilon_i'(1 + 4N^2).$$

Because $|\lambda_i(t)f_i(t) - f|$ is bounded above by $R_N(t) + r_0$, this implies (2.18) (via the first term in the minimum).

In Case 2, we set $f = f_i(t)$ in the Weierstrass expression (2.14); the corresponding function of λ attains a local maximum at $\lambda_i(t)$. Expressing that the derivative is zero leads to

$$\langle p_i(t), f_i(t)\rangle \leq -1 + \phi_t'(\lambda_i(t)\,|f_i(t)|)\,|f_i(t)| + 2\varepsilon_i'(1 + 2N^2).$$

Invoking (2.9), we derive from this in turn

$$\langle p_i(t), \lambda_i(t)f_i(t)\rangle \leq -\frac{\lambda_0}{2} + 2\varepsilon_i'(1 + 2N^2).$$

This leads again to (2.18) (via the second term in the minimum).

There remains Case 3. From $|f_i(t)| = (1 - \lambda_i(t)\eta)R_N(t)$ we derive

$$|\lambda_i(t)f_i(t)| \geq (\lambda_0 - \eta)R_N(t).$$

With this in mind we put $\lambda = 1$ and $f \in F(t, x_i(t)) \cap \overline{B}(0, r_0)$ in the Weierstrass condition and deduce

$$\langle p_i(t), \lambda_i(t)f_i(t) - f\rangle + \varepsilon_i'(1 + 4N^2) \;\geq\; \phi((\lambda_0 - \eta)R_N(t)).$$

Again this leads to (2.18), in view of (2.8).

End of the proof

We now complete the proof of the theorem. We showed above that there exist p and γ (not both 0) satisfying the transversality and Euler conditions, and the Weierstrass condition of radius $(\lambda_0 - 2\eta)R_N(t)$. We may normalize (p, γ) as follows:

$$\|p\|_\infty + |\gamma| = 1. \tag{2.19}$$

We pause to observe a simple geometrical consequence[5] of the pseudo-Lipschitz property (we suppress t in the following, whose proof is omitted).

Lemma 3 Let (x_0, v_0) be a point in G, and suppose that F satisfies a pseudo-Lipschitz condition as follows: for some $R > 0$ and $k \geq 0$, for all x, x' near x_0, we have

$$F(x') \cap \overline{B}(v_0, R) \subset F(x) + k\left|x' - x\right|\overline{B}.$$

Then, for any $(q, p) \in N_G^L(x_0, v_0)$, we have $|q| \leq k\,|p|$.

It follows from this that as a consequence of the Euler inclusion (2.16), p satisfies

$$|\dot{p}(t)| \leq k(t)\,|p(t)| \quad \text{a.e. } t \in [0, 1], \tag{2.20}$$

a condition which is independent of the parameters λ_0, η and N. Now let us consider that such an arc p_j has been obtained for a sequence of parameter values λ_0^j, η^j, N^j converging to $1, 0$ and $+\infty$ respectively. Then, invoking (2.19) (2.20) and the usual convergence arguments, there is a subsequence such that the p_j converge uniformly to an arc p (and the associated γ_j to a limit γ) such that p and γ satisfy all the desired properties listed in the

[5] This fact explains why in writing the Euler inclusion we do not have to take the *closed* convex hull.

theorem, including the Weierstrass condition relative to the full radius R, except that γ may not be a unit vector.

The limiting p and γ continue to satisfy (2.19) and (2.20). We claim that γ is necessarily nonzero. To see this, suppose the contrary. Then, in view of transversality at $t = 1$ we have $p(1) = 0$. But then (2.20) implies that p is identically zero, a contradiction which proves the claim. Since γ is nonzero, we can normalize to make it a unit vector as stipulated in the statement of the theorem. This completes the proof. ■

2.4.1 Remark The proofs of Theorems 2.1.1 and 2.3.3 can be adapted to the case in which the integral constraint is replaced by one of the following more general type:

$$\int_a^b \psi_t(|\dot{x}(t) - \dot{x}_*(t)|)\,dt \le \varepsilon,$$

provided that the functional

$$v \mapsto \int_a^b \psi_t(|v(t)|)\,dt$$

is continuous at 0 relative to the L^2 norm. We shall consider such a local minimum in Chapter 4, in a variational setting that will allow us to dispense with this continuity restriction. Another generalization that can be made is to admit a weight function θ in Theorem 2.3.3, as in Theorem 2.1.1.

2.5 Two variants of the theorem

The tempered growth condition 2.3.2 amounts to a mild semicontinuity hypothesis on F. It is easy to see that when the pseudo-Lipschitz condition 2.3.1 holds, it is automatically satisfied when the function k belongs to L^∞ and R is essentially bounded away from 0. As we now see, this criterion can be generalized.

2.5.1 Proposition Let F be pseudo-Lipschitz of radius R near x_*, and suppose that in addition we have

$$\text{ess inf}\left\{\frac{R(t)}{k(t)} : t \in [a,b]\right\} > 0.$$

Then F satisfies the tempered growth condition for radius R near x_*. ♦

Proof. We may assume that k is positive-valued. Let us define $\delta > 0$ to be the essential lower bound in the statement of the proposition. Set

$$\varepsilon_0 := \min\{\varepsilon, \delta/2\},\ r_0(t) := \varepsilon_0 k(t),$$

where the ε is that of the pseudo-Lipschitz condition. Then, by taking $x' = x_*(t)$ in that condition, we deduce that for $x \in B(x_*(t), \varepsilon_0)$, we have

$$F_t(x) \cap \overline{B}(\dot{x}_*(t), r_0(t)) \neq \emptyset.$$

Since $R(t)/r_0(t) \geq 2$, the tempered growth condition is satisfied. ■

We remark that the proposition covers the standard case in which the radius function is identically $+\infty$ and F is pseudo-Lipschitz of infinite radius (that is, actually Lipschitz). Let us record the consequence of the proposition:

2.5.2 Corollary Theorem 2.3.3 remains valid if the tempered growth condition is replaced by the following (stronger) hypothesis:

$$\text{ess inf} \left\{ \frac{R(t)}{k(t)} : t \in [a, b] \right\} > 0. \qquad \blacklozenge$$

A global version of the theorem

The standard sequential compactness arguments used in the proof of the theorem lead to the following:

2.5.3 Corollary Suppose that the hypotheses of the theorem are satisfied for a sequence of radius functions R_i (and with all other parameters possibly depending on i) for which

$$\liminf_{i \to \infty} R_i(t) = +\infty \ \text{ a.e.}$$

Then the conclusions of the theorem hold for an arc p which satisfies the global Weierstrass condition:

$$\langle p(t), v\rangle \leq \langle p(t), \dot{x}_*(t)\rangle \ \forall v \in F_t(x_*(t)), \ \text{ a.e. } t \in [a, b]. \qquad \blacklozenge$$

Proof. Applying the theorem for each i yields a corresponding arc p_i and vector γ_i that can be normalized to satisfy $|p_i(1)| + |\gamma_i| = 1$. By Gronwall's Lemma and the standard sequential compactness argument, we may suppose (by passing to a subsequence) that γ_i converges to γ and that p_i converges uniformly to an arc p which satisfies the Euler inclusion (see Proposition 2.1.2) and such that (p, γ) satisfies the transversality conditions. It follows that γ is nonzero, so we can renormalize to make it a unit vector. It suffices now to verify the global Weierstrass condition for the unrenormalized p. Take any value of t such that for each i the Weierstrass condition of radius $R_i(t)$ holds for $p_i(t)$, and such that $\liminf_{i\to\infty} R_i(t) = +\infty$ (almost all t have this property). Fix any $v \in F_t(x_*(t))$. Then for all i sufficiently large we have $|v - \dot{x}_*(t)| < R_i(t)$, so that by the Weierstrass condition of radius R_i (for p_i) we have

$$\langle p_i(t), v\rangle \leq \langle p_i(t), \dot{x}_*(t)\rangle .$$

Passing to the limit, we confirm the global Weierstrass condition.

> **2.5.4 Remark** When $R \equiv +\infty$ is a suitable radius function, then in order to assert the global conclusion above, it suffices that, for some $\varepsilon > 0$, x_* be a boundary trajectory with respect to the constraints
>
> $$\dot{x} - \dot{x}_* \in L^\infty,\ x(a) \in C,\ \int_a^b |\dot{x}(t) - \dot{x}_*(t)|\, dt \leq \varepsilon,\ \|x - x_*\|_\infty \leq \varepsilon,$$
>
> for we can then apply the theorem with the sequence of radius functions $R_i := i$. Thus comparison need only be made with arcs having $W^{1,\infty}$ difference from x_*. The same conclusion obtains if there is a sequence of suitable radius functions R_i for which
>
> $$\lim_{i\to\infty} \{\text{ess inf } R_i(t) : t \in [a, b]\} = +\infty.$$

2.6 Notes

§2.1 Theorem 2.1.1 incorporates a novel type of weighted constraint and is unusual in considering a $W^{1,\rho}$ local minimum for $\rho \in (1, 2]$. But these elements aside, and given that the underlying multifunction is integrably bounded and Lipschitz, the set of necessary conditions that it obtains is essentially known, corresponding as it does to refinements by Mordukhovic (for the transversality conditions) and Ioffe, Loewen, and Rockafellar (for the Euler inclusion) of the necessary conditions initially proven by Clarke. The method of proof, however, is new and unusually direct, and allows this monograph to be self-contained.

§2.2 In contrast, Theorem 2.3.3 is fundamentally new because of its stratified nature with respect to both hypotheses and conclusions, and the level of generality of its hypotheses. The consideration of boundary trajectories rather than optimality has a long tradition in the subject of optimal control. It existed well before Clarke's early work in the 1970s, but the new approach to boundary trajectories introduced in that work (based in part on the variational principle of Ekeland) has been widely adopted since.

Boundary trajectories for differential inclusions have been studied recently by Mordukhovic [47], Zhu [64] and Kaskosz and Lojasiewicz [39], the latter in a vector field formulation (see Chapter 5). But the hypotheses are much stronger than those of Theorem 2.3.3. The recent results that most invite comparison are framed in optimality terms, so we discuss them in the notes following the next chapter.

Chapter 3

Differential Inclusions

We begin the chapter by deriving necessary optimality conditions for a control problem phrased in terms of a differential inclusion. These are a rather direct consequence of the boundary trajectory case treated in Theorem 2.3.3. In §3.2 we look more closely at the tempered growth condition; certain criteria implying it are developed, and its necessity as a hypothesis is illustrated. The following section considers the case of a weak local minimum. In §3.4 we prove that when the underlying system is autonomous, an extra assertion can be added to the set of necessary conditions. This is an analogue of the classical second Erdmann condition, which corresponds to conservation of energy (or constancy of the Hamiltonian) in mechanics .

The final section of the chapter obtains differential (or rather, proximal) criteria for the multifunction F which imply the pseudo-Lipschitz property that is the main hypothesis of our results; we refer to such criteria as *bounded slope* conditions. In particular, we exhibit a simple condition of this type, unrelated to any particular arc, that implies that the pseudo-Lipschitz condition always holds near any trajectory.

3.1 Stratified necessary conditions

It is a well-known technique in optimal control theory to derive necessary conditions for optimality from conditions that must be satisfied in the more general setting of boundary arcs (see for example [19] or [60]). The idea is to reformulate the optimality in boundary terms to be able to apply the boundary results, the details depending upon the specific way in which the boundary constraints and cost are expressed.

We are given a multifunction F as in the previous chapter. We consider the problem $\mathcal{P}$ of minimizing $\ell(x(a), x(b))$ over the trajectories x of F satisfying the boundary constraints $(x(a), x(b)) \in S$.

As before, let R be a measurable function on $[a, b]$ with values in $(0, +\infty]$ (the radius function). We are given an arc x_* feasible for $\mathcal{P}$ which is a *local $W^{1,1}$ minimum of radius R* in the following sense: for some $\varepsilon > 0$, for every other feasible arc x satisfying

$$|\dot{x}(t) - \dot{x}_*(t)| \leq R(t) \text{ a.e.}, \int_a^b |\dot{x}(t) - \dot{x}_*(t)| \, dt \leq \varepsilon, \; \|x - x_*\|_\infty \leq \varepsilon,$$

one has $\ell(x(a), x(b)) \geq \ell(x_*(a), x_*(b))$.

It is assumed that ℓ is locally Lipschitz and that S is closed. The usual basic hypotheses of measurability of F and graph-closedness of F_t are made.

3.1.1 Theorem Suppose that, for the radius R, F satisfies the pseudo-Lipschitz and tempered growth conditions near x_* (Definitions 2.3.1 and 2.3.2). Then there exist an arc p and a number λ_0 in $\{0, 1\}$ satisfying the nontriviality condition

$$(\lambda_0, p(t)) \neq 0 \quad \forall \, t \in [a, b]$$

and the transversality condition:

$$(p(a), -p(b)) \in \partial_L \lambda_0 \ell(x_*(a), x_*(b)) + N_S^L(x_*(a), x_*(b)),$$

and such that p satisfies the Euler inclusion:

$$\dot{p}(t) \in \operatorname{co} \left\{ \omega : (\omega, p(t)) \in N_{G(t)}^L(x_*(t), \dot{x}_*(t)) \right\} \text{ a.e. } t \in [a, b]$$

as well as the Weierstrass condition of radius R:

$$\langle p(t), v \rangle \leq \langle p(t), \dot{x}_*(t) \rangle \;\; \forall v \in F_t(x_*(t)) \cap \overline{B}(\dot{x}_*(t), R(t)), \text{ a.e. } t \in [a, b].$$

If the above holds for a sequence of radius functions R_i (with all parameters possibly depending on i) for which

$$\liminf_{i \to \infty} R_i(t) = +\infty \text{ a.e.}$$

then the conclusions hold for an arc p which satisfies the global Weierstrass condition:

$$\langle p(t), v \rangle \leq \langle p(t), \dot{x}_*(t) \rangle \;\; \forall v \in F_t(x_*(t)), \text{ a.e. } t \in [a, b]. \qquad \blacklozenge$$

Proof

We sketch the derivation from the corresponding boundary trajectory result (Theorem 2.3.3), following a well-known procedure. Consider first the following special case of the problem $\mathcal{P}$:

$$\mathcal{P}_1 : \text{to minimize } \ell(x(b)) : \dot{x}(t) \in F_t(x(t)) \text{ a.e., } x(a) = 0,\ x(b) \in S,$$

where ℓ is locally Lipschitz and S is closed.

To express this optimization problem in boundary terms we introduce two additional state variables $y \in \mathbb{R}^n$ and $z \in \mathbb{R}$, and we define

$$\begin{aligned} \Phi(x, y, z) &:= [\ell(x) - \ell(x_*(b)) + z, x - y] \\ C &:= \{0\} \times S \times [0, \infty) \\ \Gamma_t(x, y, z) &:= F_t(x) \times \{0\} \times \{0\}. \end{aligned}$$

If (x, y, z) is a trajectory of Γ with initial condition in C, and if $x(b) - y(b) = 0$, then x is a trajectory of F which is feasible for the optimal control problem $\mathcal{P}_1$. If x is also close to x_* in the given sense (relative to ε and R), then it follows from the optimality of x_* that the first component of $\Phi(x(b), y(b), z(b))$ must be nonnegative. It follows therefore, in the terminology of Theorem 2.3.3, that the point $[0, 0]$ (corresponding to the arc $(x_*(t), x_*(b), 0)$) lies in the boundary of $\Phi_{\mathcal{T}}$. If Γ satisfies the conditions of that theorem (or either of its corollaries), then we can deduce the existence of an extended arc (p, q, r) and a unit vector $\gamma = (\lambda_0, \lambda_1)$ satisfying certain conditions that we can then relate back to the data of the optimal control problem.

The Euler inclusion for the extended arc is easily seen to imply that p satisfies the desired one for x_*, and that q and r are constant. The transversality conditions assert

$$(-p(b), -q, -r) \in \partial_L \{\lambda_0 \ell(x) + \lambda_0 z + \langle \lambda_1, x - y \rangle\} (x_*(b), x_*(b), 0)),$$

as well as

$$q \in N_S^L(x_*(b)),\ r \in N_{[0,+\infty)}^L(0).$$

Together these give

$$-p(b) \in \partial_L \lambda_0 \ell(x_*(b)) + \lambda_1,\ q = \lambda_1 \in N_S^L(x_*(b)),\ r = -\lambda_0 \le 0,$$

which yield the appropriate transversality condition for $\mathcal{P}_1$. If $\lambda_0 > 0$, then p and q can be normalized to achieve $\lambda_0 = 1$. If $\lambda_0 = 0$, then (since γ is a unit vector), we have $\lambda_1 \neq 0$, from which it follows that $p(b) \neq 0$. But the

Euler inclusion implies $|\dot{p}(t)| \leq k(t)\,|p(t)|$, so that $p(t)$ is never 0 (otherwise it must be identically 0). This gives the required nontriviality condition. Finally, the Weierstrass condition of radius R follows from that for Γ, and the theorem is proved (in the case of $\mathcal{P}_1$, and for a given radius function).

Although $\mathcal{P}$ appears to be more general than $\mathcal{P}_1$, it is in fact possible to reduce it by means of purely notational devices[1] to the form of $\mathcal{P}_1$, and so to derive the general case of the theorem; we omit these details.

As for the limiting case of the theorem, it follows from the now familiar sequential compactness argument as in the proof of Corollary 2.5.3; it is helpful for this to first normalize the multiplier (p_i, λ_{0_i}) obtained for each radius R_i so as to satisfy

$$\|p_i\|_\infty + \lambda_{0_i} = 1.$$

After the limiting argument, one renormalizes at the end to get the limiting λ_0 equal to 0 or 1. ■

3.1.2 Remark (a) When a suitable sequence R_i of radius functions exists such that

$$\lim_{i\to\infty} \{\text{ess inf } R_i(t) : t \in [a,b]\} = +\infty.$$

(in particular when $R \equiv +\infty$ is a suitable radius function), then it suffices that, for some $\varepsilon > 0$, x_* provide a minimum with respect to the constraints

$$\dot{x} - \dot{x}_* \in L^\infty,\ \int_a^b |\dot{x}(t) - \dot{x}_*(t)|\,dt \leq \varepsilon,\ \|x - x_*\|_\infty \leq \varepsilon.$$

The restriction $\dot{x} - \dot{x}_* \in L^\infty$, which is an outgrowth of the stratified nature of the necessary conditions, may be relevant in connection with the Lavrentiev phenomenon (see Chapter 4).

(b) Another paradigm for the optimal control of differential inclusions is the following:

$$\mathcal{P}_2 : \text{to minimize} \quad \ell(x(a), x(b)) : \dot{x}(t) \in F_t(x(t)) \ \text{a.e.},$$

where now, instead of being locally Lipschitz, $\ell : \mathbb{R}^n \times \mathbb{R}^n \to (-\infty, +\infty]$ is merely lower semicontinuous and allowed to be

[1] See for example [19], pages 149 and 169.

extended-valued (so as to incorporate simultaneously both the cost and the boundary constraints). This can easily be recast in the form of $\mathcal{P}$ by introducing an auxiliary variable, which leads to a theorem for $\mathcal{P}_2$ that is identical to Theorem 3.1.1, except for the form of the transversality condition, which becomes:

$$(p(a), -p(b), -\lambda_0) \in N^L_{\operatorname{epi}\ell}(x_*(a), x_*(b), \ell(x_*(a), x_*(b))).$$

This can also be expressed in the notation of singular subdifferentials:

$$(p(a), -p(b)) \in \partial_L^{\lambda_0} \ell(x_*(a), x_*(b)).$$

When we now return to the case of the problem $\mathcal{P}$ treated by the theorem (by taking $\ell = \ell + I_S$) , we get the transversality condition

$$(p(a), -p(b)) \in \partial_L^{\lambda_0}(\ell + I_S)(x_*(a), x_*(b)).$$

This implies the one in the statement of the theorem (by the limiting sum rule). Thus, at the cost of additional notational complexity, we have deduced a potentially sharper transversality condition.

3.2 On the tempered growth condition

The tempered growth condition holds automatically (in the presence of the pseudo-Lipschitz condition) when the radius R is sufficiently large relative to the pseudo-Lipschitz function k, as pointed out in Proposition 2.5.1. This leads to:

3.2.1 Corollary Theorem 3.1.1 remains valid if the tempered growth hypothesis is replaced by the following essential infimum condition:

$$\operatorname{ess\,inf}\left\{\frac{R(t)}{k(t)} : t \in [a, b]\right\} > 0.$$

A structured pseudo-Lipschitz hypothesis

An alternative to postulating the tempered growth hypothesis of Theorem 3.1.1 is to require the existence of a pseudo-Lipschitz function k_R for each constant R, together with additional assumptions concerning the nature of its dependence on R. Here is a sample result of this type that will facilitate comparison with the literature; we continue to consider the problem $\mathcal{P}$ defined above, but we assume now that x_* is a solution of the problem relative to all feasible arcs x that satisfy

$$\int_a^b |\dot{x}(t) - \dot{x}_*(t)| \, dt \leq \varepsilon, \ \|x - x_*\|_\infty \leq \varepsilon$$

(that is, a local minimum in the $W^{1,1}$ sense). Thus no radius function R is specified: the result is not stratified.

3.2.2 Corollary Suppose that F is pseudo-Lipschitz near x_* in the following structured sense: for some $\varepsilon > 0$ and $\alpha \geq 0$, for some pair of nonnegative functions k_0 and β for which k_0 and βk_0^α are summable, for almost all $t \in [a, b]$, for all $x, x' \in \overline{B}(x_*(t), \varepsilon)$, one has, for all $R \geq 0$,

$$F_t(x) \cap \overline{B}(\dot{x}_*(t), R) \subset F_t(x') + [k_0(t) + \beta(t)R^\alpha] \left|x' - x\right| \overline{B}. \tag{3.1}$$

Then there exist an arc p and a number λ_0 satisfying the nontriviality, transversality and Euler conditions of Theorem 3.1.1 as well as the global form of the Weierstrass condition. ♦

Proof. There is no loss of generality in assuming that k_0 is positive-valued. Taking $x = x_*(t)$ and $R = 0$ in the pseudo-Lipschitz condition shows that for almost every t, for every $x' \in B(x_*(t), \varepsilon)$, we have

$$F_t(x') \cap \overline{B}(\dot{x}_*(t), k_0(t)) \neq \emptyset.$$

This verifies the tempered growth condition 2.3.2 with $r_0 := k_0$, for the radius $R_N(t) := Nk_0(t)$. Further, substituting $R = R_N$ shows that F is pseudo-Lipschitz of radius R_N near x_* with the (integrable) pseudo-Lipschitz function $k_N := k_0 + \beta k_0^\alpha N^\alpha$. Now Theorem 3.1.1 applies for each N, but also in the global case, since $\liminf_{i\to\infty} R_N(t) = +\infty$ a.e.. The result follows. ■

Example: tempered growth is needed

It is well known even in the bounded case that some type of Lipschitz property is required for the necessary conditions to hold; mere continuity, for example, does not suffice. The tempered growth condition is new, however, and one could ask whether it is essential. We present a simple example which shows that the tempered growth condition cannot be deleted from the hypotheses of Theorem 3.1.1, and which sheds some light on its role:

it guarantees that there are sufficiently many trajectory alternatives within the specified radius to make optimality meaningful.

Take $n = 1$, $[a,b] = [0,1]$, and let k be a positive-valued summable function on $[0,1]$ which is unbounded on any open interval. Set $F_t(x) := (-\infty, -k(t)\,|x|]$. Then, for any trajectory x of F and any choice of radius function R, the multifunction is pseudo-Lipschitz (in fact, Lipschitz) around x with pseudo-Lipschitz function k. Now take for R any positive-valued essentially bounded function.

Claim: The only trajectory x satisfying $|\dot{x}(t)| \leq R(t)$ a.e. is $x_* \equiv 0$.

A trajectory x necessarily satisfies $\dot{x}(t) \leq -k(t)\,|x(t)|$, whence

$$k(t)\,|x(t)| \leq |\dot{x}(t)| \leq R(t).$$

But this implies that x is identically 0, as claimed.

Consider now the minimization of $x(1)$ over the trajectories of F satisfying $x(0) = 0$ and $|\dot{x}(t)| \leq R(t)$ a.e. In view of the claim, $x_* \equiv 0$ solves this problem. If the conclusions of Theorem 3.1.1 hold, then $\lambda_0 = 1$ necessarily, for otherwise transversality gives $p(1) = 0$, violating nontriviality. It follows that $p(1) = -1$. But the Weierstrass condition implies that p is nonnegative, a contradiction.

Thus the conclusions of Theorem 3.1.1 fail to hold. We observe that the tempered growth condition fails too: if r_0 satisfies the nonemptiness property of 2.3.2, then r_0 cannot be essentially bounded, and so cannot be majorized by R.

3.3 The case of a weak local minimum

We suppose now that x_* is a classical *weak local minimum* for $\mathcal{P}$: for some $\varepsilon > 0$, x_* is optimal relative to feasible arcs satisfying

$$|\dot{x}(t) - \dot{x}_*(t)| \leq \varepsilon \text{ a.e.}, \ \|x - x_*\|_\infty \leq \varepsilon.$$

This postulates a weaker type of local minimum than in the preceding results.

3.3.1 Theorem Let x_* provide a weak local minimum for problem $\mathcal{P}$, where F is pseudo-Lipschitz near x_*. Further, let there exist $\delta > 0$ and a continuous increasing function $r : [0,\delta] \to [0,\infty)$ having $r(0) = 0$ such that for almost all $t \in [a,b]$ we have

$$|x - x_*(t)| \leq \delta \Longrightarrow \min\{|v - \dot{x}_*(t)| : v \in F_t(x)\} \leq r(|x - x_*(t)|).$$

Then there exist an arc p and a number λ_0 satisfying the nontriviality, transversality and Euler conditions of Theorem 3.1.1. If in addition F is convex-valued, then p satisfies as well the global form of the Weierstrass condition. ♦

Proof

There exist $\bar{\varepsilon} > 0$ and a summable function k such that for almost all t, for all $x, x' \in B(x_*(t), \bar{\varepsilon})$, one has

$$F_t(x') \cap \overline{B}(\dot{x}_*(t), \bar{\varepsilon}) \subset F_t(x) + k(t)\left|x' - x\right| \overline{B}.$$

Choose $\varepsilon_0 > 0$ such that $r(\varepsilon_0) < \min\{\varepsilon, \bar{\varepsilon}\} =: R$, and set

$$r_0(t) :\equiv \frac{r(\varepsilon_0) + R}{2}.$$

Then the tempered growth condition holds for this choice of data. Applying Theorem 3.1.1 yields the Euler inclusion, transversality, nontriviality, and a 'small' Weierstrass condition of radius R. When in addition F is convex-valued, then it is well-known (and elementary to show) that the Euler inclusion implies the global Weierstrass condition. ■

3.3.2 Remark (a) Note that a local minimum in the $W^{1,\rho}$ sense, that is, relative to

$$\int_a^b |\dot{x}(t) - \dot{x}_*(t)|^\rho \, dt \leq \varepsilon, \|x - x_*\|_\infty \leq \varepsilon$$

for *any* $\rho \geq 1$, always provides a weak local minimum, so that the Euler condition holds in these cases too.

(b) The existence of the function r postulated in the theorem follows automatically if the pseudo-Lipschitz function k is essentially bounded. In the general case, rather than postulate the existence of the function r, it suffices instead to require the existence of positive constants δ and $r_0 < \min\{\varepsilon, \bar{\varepsilon}\}$ such that almost everywhere

$$|x - x_*(t)| \leq \delta \Longrightarrow F_t(x) \cap B(\dot{x}_*(t), r_0) \neq \emptyset.$$

3.4 The second Erdmann condition

We continue to consider the optimal control problem $\mathcal{P}$:

$$\mathcal{P} : \text{to minimize} \quad \ell(x(a), x(b))$$

over the arcs x satisfying the differential inclusion and boundary constraints

$$\dot{x}(t) \in F(x(t)) \;\text{ a.e., } (x(a), x(b)) \in S,$$

but note that now the multifunction F is *autonomous*; that is, has no dependence on t.

We continue to take ℓ locally Lipschitz and S closed; the usual basic hypotheses of measurability and closedness are made on F. As before, let R be a measurable function on $[a, b]$ with values in $(0, +\infty]$ (the radius function). We are given an arc x_* feasible for $\mathcal{P}$ which is a *strong local minimum* for the problem: for some ε_*, for every other feasible arc x satisfying

$$\|x - x_*\|_\infty \le \varepsilon_*,$$

one has $\ell(x(a), x(b)) \ge \ell(x_*(a), x_*(b))$.

3.4.1 Theorem Suppose that F is autonomous and satisfies, for some radius function R, the pseudo-Lipschitz and tempered growth conditions near x_*. Then there exist an arc p and a number λ_0 in $\{0, 1\}$ satisfying the same conditions as in Theorem 3.1.1 together with an additional conclusion: there is a constant h such that

$$\langle p(t), \dot{x}_*(t)\rangle = h \;\text{ a.e.}$$

If the hypotheses hold for a sequence of radius functions R_i (with all parameters possibly depending on i) for which

$$\liminf_{i\to\infty} R_i(t) = +\infty \;\text{ a.e.},$$

then all the conclusions above hold for an arc p which satisfies the *global* Weierstrass condition:

$$\langle p(t), v\rangle \le \langle p(t), \dot{x}_*(t)\rangle \;\forall v \in F(x_*(t)), \;\text{ a.e. } t \in [a, b]. \qquad \blacklozenge$$

Proof. We shall extend the problem to one involving both arcs y with values in $\mathbb{R}^n$ as well as real-valued arcs z. We begin by choosing $\delta_0 \in (0,1)$ so that any scalar arc z satisfying $z(0) = 0$ and $|\dot{z}(t) - 1| \leq \delta_0$ a.e. must also satisfy

$$\left|x_*(t) - x_*(z^{-1}(t))\right| < \frac{\varepsilon_*}{2},\ t \in [a,b].$$

Next we choose, for almost each t, a number $\delta_t \in (0, \delta_0)$, and set

$$\tilde{R}(t) := (1 - \delta_t)R(t) - \delta_t \left|\dot{x}_*(t)\right|.$$

We choose δ_t small enough so that

$$\text{ess inf}\left\{\frac{\tilde{R}(t)}{r_0(t)} : t \in [a,b]\right\} > 1.$$

We may assume that the mapping $t \mapsto \delta_t$ is measurable. The multifunction $\tilde{F}$ of the extended problem is given as follows:

$$\tilde{F}(y,z) := \{(\lambda v, \lambda) : v \in F(y), |\lambda - 1| \leq \delta_t\}.$$

The extended problem consists of minimizing $\ell(y(a), y(b))$ over the trajectories (y,z) of $\tilde{F}$ on $[a,b]$ satisfying

$$(y(a), y(b)) \in S, z(a) = a, z(b) = b, \|y - x_*\|_\infty \leq \varepsilon_*/2.$$

We claim that $(x_*(t), t)$ (which is evidently feasible for the problem) is a solution.

Let us prove this claim by contradiction. Suppose that the feasible arc (y,z) is better than $(x_*(t), t)$. We proceed to define an arc x via

$$x(t) := y(z^{-1}(t)).$$

It follows readily that x is a trajectory for F. Furthermore, we have

$$\begin{aligned}|x(t) - x_*(t)| &= \left|y(z^{-1}(t)) - x_*(t)\right| \\ &\leq \left|y(z^{-1}(t)) - x_*(z^{-1}(t))\right| + \left|x_*(z^{-1}(t)) - x_*(t)\right| \\ &< \varepsilon_*/2 + \varepsilon_*/2 = \varepsilon_*,\end{aligned}$$

(by hypothesis for the first term, by choice of δ_t for the second) whence x lies within the uniform ε_*-neighborhood of x_*. But then x is feasible for the original problem and strictly better than x_*, a contradiction which proves the claim.

We wish to apply Theorem 3.1.1 to the solution $(x_*(t), t)$ of the extended problem, with radius function $\tilde{R}$. We pause to verify the required pseudo-Lipschitz condition. Thus let y_1, y_2 be points appropriately near $x_*(t)$ and let $(\lambda_1 v_1, \lambda_1)$ be an element of

$$\tilde{F}(y_1, z_1) \cap \overline{B}((\dot{x}_*(t), 1), \tilde{R}(t))$$

Then we have

$$|\dot{x}_*(t) - \lambda_1 v_1| \leq \tilde{R}(t),$$

which implies $|\dot{x}_*(t) - v_1| \leq R(t)$. This allows us to apply the pseudo-Lipschitz hypothesis for F; we deduce the existence of $v_2 \in F(y_2)$ such that

$$|(\lambda_1 v_2, \lambda_1) - (\lambda_1, \lambda_1)| \leq \lambda_1 k(t)\,|y_1 - y_2| \leq (1 + \delta_t) k(t)\,|y_1 - y_2|\,.$$

The required pseudo-Lipschitz condition for $\tilde{F}$ now ensues, with

$$\tilde{k}(t) := (1 + \delta_t) k(t).$$

We see that the hypotheses of Theorem 3.1.1 hold for the new problem data. Applying the theorem leads to an arc p and a constant (arc) q such that, almost everywhere, $\dot{p}(t)$ belongs to the set

$$\mathrm{co}\left\{\omega : (\omega, p(t), q) \in N^L_{D(t)}(x_*(t), \dot{x}_*(t), 1)\right\},$$

where

$$D(t) := \{(y, \lambda v, \lambda) : (y, v) \in G(t), |\lambda - 1| \leq \delta_t\}\,.$$

A straightforward analysis of this inclusion (via the proximal normal inequality) shows that p satisfies the desired Euler inclusion, and that we have

$$\langle p(t), \dot{x}_*(t)\rangle + q = 0 \text{ a.e.}$$

This completes the proof for a given radius.

The proof of the limiting case follows the standard line: one applies Gronwall's Lemma and the sequential compactness argument after normalizing via

$$\lambda_{0_i} + |p_i(1)| = 1,$$

observing that the associated constants h_i are necessarily bounded. ■

3.5 Proximal criteria for pseudo-Lipschitzness

Thus far in the chapter we have directly postulated pseudo-Lipschitz behavior of the multifunction F in a neighborhood of the local solution x_*. It is of interest to be able to verify this property *a priori* by pointwise conditions *restricted to the graph of* F, and if possible independently of the particular trajectory x_*. A most relevant analogy here is the fact that a differentiable function is (locally) Lipschitz iff its derivative is (locally) bounded, or the analogous fact for lower semicontinuous functions and proximal subgradients (see [27]). In this section we develop such criteria in geometric terms via normal vectors; in the following chapter we examine their functional counterparts.

We phrase the results in generic terms for their intrinsic interest. In this section, Γ is a multifunction from $\mathbb{R}^m$ to $\mathbb{R}^n$ having closed graph G. The pseudo-Lipschitz property is linked to the following type of condition bearing upon the proximal normals to the graph.

> **3.5.1 Definition** Let v_* be a point in $\Gamma(x_*)$. The multifunction Γ is said to satisfy the **bounded slope condition** near (x_*, v_*) (with parameters ε, R and k) if
>
> $$x \in B(x_*, \varepsilon), v \in B(v_*, R), (\alpha, \beta) \in N_G^P(x, v) \Longrightarrow |\alpha| \leq k\,|\beta|\,.$$

We remark that an equivalent definition is obtained if one replaces the proximal normal cone by the limiting one. We have mentioned earlier that if Γ satisfies a pseudo-Lipschitz condition near (x_*, v_*) with pseudo-Lipschitz constant k, then the inequality $|\alpha| \leq k\,|\beta|$ holds for all elements (α, β) of $N_G^P(x, v)$, for all (x, v) near (x_*, v_*). Our interest in this section lies in the considerably more delicate converse.

3.5.2 Theorem Let Γ satisfy the bounded slope condition 3.5.1 near the point (x_*, v_*) in G. Then for any $\eta \in (0, 1)$, Γ is pseudo-Lipschitz on $B(x_*, \varepsilon_\eta)$, with pseudo-Lipschitz constant k and radius $(1 - \eta)R$, where

$$\varepsilon_\eta := \min\{\varepsilon, \eta R/(3k)\}\,. \qquad \blacklozenge$$

More explicitly, the conclusion is that for $x_1, x_2 \in B(x_*, \varepsilon_\eta)$, we have

$$\Gamma(x_1) \cap \overline{B}(v_*, (1-\eta)R) \subset \Gamma(x_2) + k\,|x_2 - x_1|\,\overline{B}.$$

Proof. Let x_1 and x_2 in $B(x_*, \varepsilon_\eta)$ be given, together with $v_1 \in \Gamma(x_1) \cap \overline{B}(v_*, (1-\eta)R)$. We must prove the existence of $v_2 \in \Gamma(x_2)$ such that $|v_2 - v_1| \leq k\,|x_2 - x_1|$. To do so, we shall apply the Mean Value Inequality (see §1.3) with the following data:

$$f(x,v) := I_G(x,v), \quad Y := \{x_2\} \times \overline{B}(v_1, k\,|x_2 - x_1|).$$

We take the base point (x_1, v_1), and a tolerance $\tilde{\varepsilon}$ small enough in a sense to be specified presently. If $G \cap Y$ is nonempty, the existence of the required point v_2 follows. Therefore we suppose that $G \cap Y$ is empty, and we proceed to get a contradiction.

Since (under the emptiness assumption) $\min_Y f = +\infty$, the Mean Value Inequality asserts that for any $r > 0$ we may write

$$r < \langle \zeta, x_2 - x_1 \rangle + \langle \psi, v - v_1 \rangle + \tilde{\varepsilon}$$

for all $v \in \overline{B}(v_1, k\,|x_2 - x_1|)$, where (ζ, ψ) belongs to $\partial_P f(z,w)$, and where (z,w) is a point within distance $\tilde{\varepsilon}$ of the set co $[Y \cup \{(x_1, v_1)\}]$. For $\tilde{\varepsilon}$ small enough, the point z certainly lies in $B(x_*, \varepsilon)$. As for w, we have

$$\begin{aligned} |w - v_*| &\leq |w - v_1| + |v_1 - v_*| \\ &\leq k\,|x_2 - x_1| + \tilde{\varepsilon} + (1-\eta)R \\ &< 2k\varepsilon_\eta + \tilde{\varepsilon} + (1-\eta)R \\ &\leq 2\eta R/3 + \tilde{\varepsilon} + (1-\eta)R < R, \end{aligned}$$

for $\tilde{\varepsilon}$ sufficiently small. Thus the point (z,w) lies in the open set where the bounded slope condition is valid. Also, we know that (ζ, ψ) belongs to $N_G^L(z,w)$. Since the bounded slope condition continues to hold for the limiting normal cone, this observation together with the preceding inequality gives

$$\begin{aligned} r &< |\zeta|\,|x_2 - x_1| - |\psi|\,k\,|x_2 - x_1| + \tilde{\varepsilon} \\ &\leq (|\zeta| - k\,|\psi|)\,|x_2 - x_1| + \tilde{\varepsilon} \leq \tilde{\varepsilon}. \end{aligned}$$

Since we arrive at this for any $r > 0$, we have the desired contradiction. ■

The reader will have guessed our intention of applying the criterion proven above to optimal control problems such as $\mathcal{P}$, for which we require pseudo-Lipschitz behavior of $F_t(\cdot)$ for almost every t, near $(x_*(t), \dot{x}_*(t))$. In this setting, we obtain from the above a pseudo-Lipschitz condition for x within a distance of $x_*(t)$ which is proportional to $R(t)/k(t)$. Since a uniform neighborhood of x_* is required in order to invoke Theorem 2.3.3, this suggests that we postulate the essential lower limit condition of Corollary 3.2.1 (a different approach will be given later). The immediate consequence is the following:

3.5.3 Corollary Suppose that in Theorem 3.1.1 the pseudo-Lipschitz hypothesis is replaced by the following bounded slope condition: for almost every t, for every $(x,v) \in G(t)$ with $x \in B(x_*(t), \varepsilon)$ and $v \in B(\dot{x}_*(t), R(t))$, for all $(\alpha, \beta) \in N^P_{G(t)}(x,v)$, one has $|\alpha| \leq k(t)\,|\beta|$. Suppose also that the tempered growth condition is strengthened to

$$\text{ess inf}\left\{\frac{R(t)}{k(t)} : t \in [a,b]\right\} > 0.$$

Then the necessary conditions of the theorem hold as stated, for the same radius R. ♦

Proof. For each $\eta \in (0,1)$, Theorem 3.5.2 implies that F satisfies a pseudo-Lipschitz condition of radius $(1-\eta)R$ in a uniform neighborhood of x_* of radius

$$\bar{\varepsilon}_\eta := \text{ess inf}\left\{\min\left[\varepsilon, \frac{\eta R(t)}{3k(t)}\right] : t \in [a,b]\right\} > 0.$$

This permits the application of Corollary 3.2.1, the result of which is an arc p_η satisfying the required conclusions, except that the Weierstrass condition is of radius $(1-\eta)R$ rather than R. We let η decrease to 0 and apply the now familiar sequential compactness argument to the associated arcs p_η to arrive at the stated result. ■

A global criterion

The solution x_* to $\mathcal{P}$ is not generally known in advance; in particular we may not know *a priori* whether or not its derivative is essentially bounded. As is well known in the setting of the calculus of variations, such a property is central in being able to apply necessary conditions. For this purpose it is useful to have a structural criterion assuring that F is pseudo-Lipschitz of arbitrarily large radius around any arc x_*. The following gives a bounded slope condition having this effect; we return to the generic notation.

3.5.4 Theorem Let X be an open subset of $\mathbb{R}^m$. Suppose that there exist positive constants c, d such that

$$x \in X, v \in \Gamma(x), (\alpha,\beta) \in N^P_G(x,v) \Longrightarrow |\alpha| \leq \{c\,|v| + d\}\,|\beta|\,.$$

Let any $x_* \in X$ and $v_* \in \Gamma(x_*)$ be given, together with $\varepsilon > 0$ such that $B(x_*, \varepsilon) \subset X$. Set

$$\delta_* := \min\left\{\varepsilon,\ \frac{1}{3(1+2c)}\right\}.$$

Then for any $N \geq 1$, the multifunction Γ is pseudo-Lipschitz of radius $R_N := (c\,|v_*| + d)N$ near (x_*, v_*) as follows: for any $x, x' \in B(x_*, \delta_*)$ we have

$$\Gamma(x') \cap \overline{B}(v_*, R_N) \subset \Gamma(x) + k_N\,|x' - x|\,\overline{B},$$

where $k_N := (c\,|v_*| + d)(1 + 2cN)$. ♦

Proof. Fix $N \geq 1$, and consider any point $(x, v) \in G$ with $x \in B(x_*, \varepsilon)$ and $v \in B(v_*, 2R_N)$, and for such a point any element $(\alpha, \beta) \in N_G^P(x, v)$. Then by hypothesis we have

$$|\alpha| \leq \{c\,|v| + d\}\,|\beta| \leq \{c\,|v_*| + 2cR_N + d\}\,|\beta| = k_N\,|\beta|.$$

Applying Theorem 3.5.2 with $\eta = 1/2$, we deduce that Γ is pseudo-Lipschitz with constant k_N and of radius R_N for

$$|x - x_*| < \min\left\{\varepsilon,\ \frac{R_N}{3k_N}\right\} = \min\left\{\varepsilon,\ \frac{N}{3(1+2cN)}\right\}.$$

Since $N \geq 1$, this last expression is no less than δ_*, which completes the proof. ■

The application of the above to (unstratified) necessary conditions is notable for the absence of tempered growth or pseudo-Lipschitz hypotheses explicitly related to the solution:

3.5.5 Corollary Suppose that for every bounded set X in $\mathbb{R}^n$ there exist a constant c and a summable function d such that for almost every t, for all $(x, v) \in G(t)$ with $x \in X$, for all $(\alpha, \beta) \in N_{G(t)}^P(x, v)$, one has $|\alpha| \leq (c\,|v| + d(t))\,|\beta|$. Then for any local $W^{1,1}$ minimum x_* for $\mathcal{P}$, the global conclusions of Theorem 3.1.1 hold. ♦

Proof. We may choose an open bounded set X and $\varepsilon > 0$ such that $B(x_*(t), \varepsilon) \subset X$ for each t. Let c and d be the corresponding bounded slope parameters for X. Fix $N \geq 1$ and define

$$R_N(t) := (c\,|\dot{x}_*(t)| + d(t))N,\ k_N(t) := (c\,|\dot{x}_*(t)| + d(t))(1 + 2cN).$$

In view of Theorem 3.5.4, there is a constant $\delta_* > 0$ such that, for almost each t, the multifunction F_t is pseudo-Lipschitz of radius $R_N(t)$ with constant $k_N(t)$ for $x \in B(x_*(t), \delta_*)$. Note that k_N is summable, and that we have

$$\text{ess inf}\left\{\frac{R_N(t)}{k_N(t)} : t \in [a,b]\right\} = \frac{1}{1+2cN} > 0.$$

We may therefore invoke Corollary 3.2.1 in the global case (we may suppose $d(t) \geq 1$, which implies $\liminf_{N\to\infty} R_N(t) = +\infty$ a.e. as required), which completes the proof. ■

A refined criterion for pseudo-Lipschitzness

We develop now another proximal criterion for pseudo-Lipschitz behavior. The distinction with Theorem 3.5.2 is that we avoid shrinking the neighborhood around x_* to a possibly troublesome extent by invoking a tempered growth condition and making fuller use of the Mean Value Inequality.

3.5.6 Theorem Let v_* be a point in $\Gamma(x_*)$, and let Γ satisfy the following tempered growth condition near (x_*, v_*):

$$\|x - x_*\| < \varepsilon_0 \Longrightarrow \Gamma(x) \cap \overline{B}(v_*, r_0) \neq \emptyset$$

for positive parameters ε_0 and r_0. Suppose that for some $R > r_0$, Γ satisfies the bounded slope condition 3.5.1. Set $\varepsilon' := \min\{\varepsilon, \varepsilon_0\}$ and choose any $R' \in (0, (R - r_0)/2)$. Then Γ is pseudo-Lipschitz of radius R' on $B(x_*, \varepsilon')$ with pseudo-Lipschitz constant k. ♦

Proof. Let x_1 and x_2 in $B(x_*, \varepsilon')$ be given, together with $v_1 \in \Gamma(x_1) \cap \overline{B}(v_*, R')$. We must prove the existence of $v_2 \in \Gamma(x_2)$ such that $|v_2 - v_1| \leq k\,|x_2 - x_1|$. To do so, we shall apply the Mean Value Inequality with the following data:

$$f(x,v) := I_G(x,v) + |v - v_1|\,,\ Y := \{x_2\} \times \overline{B}(v_1, \max[r_0 + R', k\,|x_2 - x_1|]).$$

We take the base point (x_1, v_1), and a tolerance $\tilde{\varepsilon}$ small enough in a sense to be specified presently.

In view of the tempered growth condition, there is a point $v_0 \in \Gamma(x_2) \cap B(v_*, r_0)$. This implies $|v_0 - v_1| \leq r_0 + R'$. As a consequence we deduce that $r := \min_Y f \leq r_0 + R'$. The Mean Value Inequality (with this choice of r) implies that we may write

$$r < \langle \zeta, x_2 - x_1 \rangle + \langle \psi, v - v_1 \rangle + \tilde{\varepsilon}, \quad \forall\, v \in \overline{B}(v_1, k\,|x_2 - x_1|), \tag{3.2}$$

where (ζ, ψ) belongs to $\partial_P f(z, w)$, and where (z, w) is a point within distance $\tilde{\varepsilon}$ of the set co $[Y \cup \{(x_1, v_1)\}]$. For $\tilde{\varepsilon}$ small enough, the point z certainly lies in $B(x_*, \varepsilon)$. We now claim that we also have $|w - v_*| < R$ (for $\tilde{\varepsilon}$ small enough).

The conclusion of the Mean Value Inequality includes an upper bound implying

$$|w - v_1| = f(z, w) < f(x_1, v_1) + r + \tilde{\varepsilon} = r + \tilde{\varepsilon}.$$

This yields

$$|w - v_*| \leq r + \tilde{\varepsilon} + R' \leq r_0 + 2R' + \tilde{\varepsilon} < R,$$

proving the claim.

Thus the point (z, w) lies in the open set where the bounded slope condition is valid. Returning now to 3.2, we have (from the limiting sum rule)

$$(\zeta, \psi) = (\alpha, \beta) + (0, \theta),$$

where (α, β) belongs to $N_G^L(z, w)$ and $|\theta| \leq 1$. Since the bounded slope condition continues to hold for the limiting normal cone, we derive from 3.2 the following estimates:

$$\begin{aligned} r &< |\alpha|\,|x_2 - x_1| - |\beta + \theta|\,k\,|x_2 - x_1| + \tilde{\varepsilon} \\ &\leq (|\alpha| - k\,|\beta|)\,|x_2 - x_1| + |\theta|\,k\,|x_2 - x_1| + \tilde{\varepsilon} \\ &\leq k\,|x_2 - x_1| + \tilde{\varepsilon}. \end{aligned}$$

Since $\tilde{\varepsilon}$ is as small as desired, this gives

$$r := \min_Y f \leq k\,|x_2 - x_1|,$$

which implies the existence of the point v_2 that we seek. ■

We obtain as a consequence the following conditions under which the Euler inclusion holds.

3.5.7 Corollary Let x_* be a $W^{1,1}$ local minimum for $\mathcal{P}$, and let there exist a positive constant ε, a positive-valued summable function r_0, and a summable function k such that for almost every t, the following two conditions are satisfied:

1. $x \in B(x_*(t), \varepsilon) \Longrightarrow F_t(x) \cap B(\dot{x}_*(t), r_0(t)) \neq \emptyset$;
2. for all $(x, v) \in G(t)$ with $x \in B(x_*(t), \varepsilon)$ and $v \in B(\dot{x}_*(t), 4r_0(t))$, for all $(\alpha, \beta) \in N^P_{G(t)}(x, v)$, one has $|\alpha| \leq k(t)\,|\beta|$.

Then there exists an arc p and a scalar $\lambda_0 \in \{0,1\}$ satisfying the nontriviality and transversality conditions as well as the Euler inclusion. ♦

We remark that (as usual) if F is convex-valued, then the global Weierstrass condition also holds.

Proof. We apply Theorem 3.5.6 with $R(t) := 4r_0(t)$ to deduce that F_t is pseudo-Lipschitz of radius $R'(t) := 5r_0(t)/4$ near $(x_*(t), \dot{x}_*(t))$ for $x \in B(x_*(t), \varepsilon')$. The result then follows from an appeal to Theorem 3.1.1. ■

Note the distinction between this result and Corollary 3.5.3: now we require that the pseudo-Lipschitz radius $(4r_0)$ be larger than the radius r_0 of tempered growth, in contrast to Corollary 3.5.3, where the pseudo-Lipschitz radius R had to be large relative to the pseudo-Lipschitz function k.

3.6 Notes

§3.1-3.2 The results of these sections subsume and extend in a variety of ways (see the Introduction) the necessary conditions for $\mathcal{P}$ in the literature, notably those obtained by Clarke [11, 17], Loewen and Rockafellar [43], Ioffe [35], Mordukhovich [47], Smirnov [55], and Vinter [58]; they also answer in the affirmative some questions raised by Ioffe in [35]. We proceed to make some detailed comments about the most closely-related literature.

Mordukhovich [47] obtains the Euler inclusion for optimal arcs (as well as for a boundary trajectory in a less general sense), but does not assert any Weierstrass condition. Although the underlying multifunction is not assumed convex-valued, it is uniformly bounded and Lipschitz, and certain extra hypotheses restrict the t-dependence. The underlying methodology rests on discrete approximations, a subject of independent interest.

The unbounded and nonconvex-valued case is treated by Ioffe (Theorem 1 of [35]) and Vinter (Theorem 7.4.1 of [58]), who obtain the Euler inclusion and the global Weierstrass condition in a standard unstratified context. They postulate pseudo-Lipschitzness with respect to *every* constant radius R and linear growth of the corresponding pseudo-Lipschitz function k as a function of R:

$$F_t(x) \cap \overline{B}(\dot{x}_*(t), R) \subset F_t(x') + [k_0(t) + \beta R] \left| x' - x \right| \overline{B}.$$

We recover this result with Corollary 3.2.2 by taking $\alpha = 1, \beta =$ constant.

In the convex-valued case, both Ioffe (Theorem 2 of [35]) and Vinter (Theorem 7.5.1 of [58]) reduce the pseudo-Lipschitz hypothesis to a local one, and postulate additional growth conditions that are easily seen to imply the tempered growth condition. Since in the convex-valued case the global Weierstrass condition is equivalent to its local version, these results are subsumed by Theorem 3.1.1.

Loewen and Rockafellar (Theorem 4.3 of [43]) consider the (unstratified) differential inclusion problem, in the convex-valued case. They posit a pseudo-Lipschitz condition in which $k(t)/R(t)$ is essentially bounded. We recover this result with Corollary 3.2.1. They also extend their result to the case in which unilateral state constraints are imposed, a situation not discussed in the present work.

§3.3 This appears to be the first result that obtains the Euler necessary condition as a consequence of a weak local minimum, even in the case of convex values. Corollary 7.4 of [47] claims to do this, but the proof is based on the false premise that the multifunction $F_t(x) \cap \overline{B}(\dot{x}_*(t), \varepsilon)$ is Lipschitz when F_t is Lipschitz.

§3.4 Note that in deriving the second Erdmann condition, x_* has been assumed to be a *strong* local minimum, in contrast to Theorem 3.1.1. However, a more complicated version of the underlying transformation device used in the proof can be used to obtain the same conclusion in the presence of a local minimum of $W^{1,1}$ type. It is also possible to extend the theorem to more general kinds of time dependence. In every case one transforms the new problem to the type for which the previously-obtained necessary conditions can be applied. These reformulations are described in Chapter 8 of Vinter's book [58].

§3.5 In an article that we shall cite again in connection with Chapter 4, Loewen and Rockafellar (Theorem 4.3 of [44]) develop bounded slope criteria for the pseudo-Lipschitz property, based on the results of [52]. These are subsumed and improved by Theorem 3.5.2; in particular, our sharper estimates allow us to obtain the full radius R in the (very natural) conclusions of Corollary 3.5.3. We remark that differential characterizations of pseudo-Lipschitzness have been used in connection with optimality or 'solvability' issues; see for example Section 3.3 of [27] and Mordukhovic [46].

The global bounded slope criterion of Corollary 3.5.5 appears to be new.

Chapter 4

The Calculus of Variations

In this chapter we study problems that lie within the framework of the calculus of variations. While these problems resemble notationally the classical basic problem, the Lagrangians that we admit are much less regular, in fact generally extended-valued.

The first section deals with a stratified context in which local pseudo-Lipschitz behavior near the given arc is either an explicit hypothesis (Theorem 4.1.1) or furnished by a local bounded slope condition (Theorem 4.1.3). The second section illustrates the use of such results in deriving multiplier rules of classical type for variational problems of Lagrange. In §4.3 we define a new general class of Lagrangians for which pseudo-Lipschitz behavior of arbitrarily large radius will automatically hold around any locally minimizing arc x_*; this allows us to dispense with hypotheses explicitly related to a particular arc x_*, and leads to the global (that is, infinite radius) form of the necessary conditions. The fourth section of the chapter obtains certain refinements in the case in which the Lagrangian is assumed to be locally finite-valued near the solution, and obtains a new state of the art for that context. The fifth and final section is devoted to deriving new criteria implying the Lipschitz regularity of solutions.

4.1 Stratified necessary conditions

The problem $\mathcal{P}_\Lambda$ and the basic hypotheses. We consider the following problem $\mathcal{P}_\Lambda$ in the calculus of variations: to minimize the functional

$$J(x) := \ell(x(a), x(b)) + \int_a^b \Lambda_t(x(t), \dot{x}(t))\, dt$$

subject to the boundary conditions

$$(x(a), x(b)) \in S,$$

where ℓ is locally Lipschitz and S is closed. The other basic hypotheses in force throughout include the $\mathcal{L} \times \mathcal{B}$ measurability of the Lagrangian Λ with respect to t and (x, v), and the lower semicontinuity of the function $(x, v) \mapsto \Lambda_t(x, v)$ for each t. An arc x is said to be *feasible* for the problem if it satisfies the endpoint constraints, and if the integral over $[a, b]$ of the function $t \mapsto \Lambda_t(x(t), \dot{x}(t))$ is well-defined and finite.

Let R be a measurable function on $[a, b]$ with values in $(0, +\infty]$ (the radius function). We are given an arc x_* feasible for $\mathcal{P}_\Lambda$ which is a *local $W^{1,1}$ minimum of radius R* in the following sense: for some $\varepsilon > 0$, for every other feasible arc x satisfying

$$|\dot{x}(t) - \dot{x}_*(t)| \leq R(t) \text{ a.e.},$$

and which is $W^{1,1}$ close to x_* as follows:

$$\int_a^b |\dot{x}(t) - \dot{x}_*(t)| \, dt \leq \varepsilon, \; \|x - x_*\|_\infty \leq \varepsilon,$$

one has $J(x) \geq J(x_*)$.

If we proceed to define a multifunction via

$$F_t(x, y) = F_t(x) := \operatorname{epi} \Lambda_t(x, \cdot) = \{(v, \Lambda_t(x, v) + \delta) : v \in \mathbb{R}^n, \delta \geq 0\}$$

(where y is scalar valued) and define

$$\tilde{\ell}(x, y, x', y') := \ell(x, x') + y', \; \tilde{S} := \left\{(x, y, x', y') : (x, x') \in S, y = 0\right\},$$

and if we set

$$y_*(t) := \int_a^t \Lambda_s(x_*(s), \dot{x}_*(s)) \, ds,$$

then it follows that the arc (x_*, y_*) is a local minimum for the resulting differential inclusion problem $\mathcal{P}$ in the sense of Theorem 3.1.1. The data also satisfy the basic hypotheses. (If we had defined $F_t(x)$ to be the graph of $\Lambda_t(x, \cdot)$ rather than the epigraph, then F_t would not necessarily have closed graph as required by the basic hypotheses.)

The normal and abnormal cases

Theorem 4.1.1 below is an immediate consequence of interpreting Theorem 3.1.1 for the data we have just defined. We refer to the case in which $\lambda_0 = 1$ as the *normal* case. When $\lambda_0 = 1$, the Euler inclusion has the form

$$\dot{p}(t) \in \text{co}\, \{\omega : (\omega, p(t)) \in \partial_L \Lambda_t(x_*(t), \dot{x}_*(t))\} \quad \text{a.e. } t \in [a, b].$$

The abnormal case $\lambda_0 = 0$ (also termed 'singular' or 'degenerate') can arise even in problems stemming from classical smooth settings, when differential constraints and endpoint constraints combine to be 'overly tight'. The normal case necessarily holds in the theorem if for all t in a set of positive measure, the function $\Lambda_t(x_*(t), \cdot)$ is continuous at $\dot{x}_*(t)$ (this follows easily from the Weierstrass condition). Another case in which λ_0 must equal 1 is that in which $x(b)$ is free (that is, S imposes locally no restriction on $x(b)$).

In the statement below, recall (§1.3) that $\partial_L^0 \Lambda$ refers to the *singular limiting subdifferential*, while ∂_L^1 is identified with the usual limiting subdifferential ∂_L. Recall also that dom f is the set of points x for which $f(x) < +\infty$. The notation $\Lambda_t(*)$ is shorthand for $\Lambda_t(x_*(t), \dot{x}_*(t))$.

4.1.1 Theorem Suppose that Λ is pseudo-Lipschitz of radius R near x_* in the following sense: there exists a summable function k such that, for almost all $t \in [a, b]$, for every x and x' in $B(x_*(t), \varepsilon)$, for every $v \in B(\dot{x}_*(t), R(t))$ satisfying

$$\Lambda_t(x', v) \le \Lambda_t(*) + R(t),$$

there exists w such that

$$|w - v| \le k(t) \left|x' - x\right|, \; \Lambda_t(x, w) \le \Lambda_t(x', v) + k(t) \left|x' - x\right|.$$

In addition, let Λ satisfy the following tempered growth condition of radius R near x_*: For some positive-valued summable function r_0 and $\lambda \in (0, 1)$ with $0 < r_0(t) \le \lambda R(t)$ a.e. we have, for almost every t, for all $x \in B(x_*(t), \varepsilon)$, the existence of $v \in \overline{B}(\dot{x}_*(t), r_0(t))$ such that

$$\Lambda_t(x, v) \le \Lambda_t(*) + r_0(t).$$

Then there exist an arc p and a number λ_0 in $\{0, 1\}$ satisfying the nontriviality condition

$$(\lambda_0, p(t)) \ne 0 \quad \forall\, t \in [a, b]$$

and the transversality condition:

$$(p(a), -p(b)) \in \partial_L \lambda_0 \ell(x_*(a), x_*(b)) + N_S^L(x_*(a), x_*(b)),$$

and such that p satisfies the (possibly abnormal) Euler inclusion:

$$\dot{p}(t) \in \operatorname{co}\left\{\omega : (\omega, p(t)) \in \partial_L^{\lambda_0}\Lambda_t(x_*(t), \dot{x}_*(t))\right\} \text{ a.e. } t \in [a,b]$$

as well as the Weierstrass condition of radius R: for almost every t we have

$$\langle p(t), v - \dot{x}_*(t)\rangle \le \lambda_0\Lambda_t(x_*(t), v) - \lambda_0\Lambda_t(*) \quad \forall\, v \in \operatorname{dom}\Lambda_t(x_*(t), \cdot) \cap V_R(t),$$

where

$$V_R(t) := \left\{v \in \overline{B}(\dot{x}_*(t), R(t)) : \Lambda_t(x_*(t), v) \le \Lambda_t(*) + R(t)\right\}.$$

If the above holds for a sequence of radius functions R_i (with possibly different pseudo-Lipschitz and tempered growth parameters depending on i) for which

$$\liminf_{i\to\infty} R_i(t) = +\infty \text{ a.e.}$$

then the conclusions hold for an arc p which satisfies the global Weierstrass condition: for almost each t,

$$\langle p(t), v - \dot{x}_*(t)\rangle \le \lambda_0\Lambda_t(x_*(t), v) - \lambda_0\Lambda_t(*) \quad \forall\, v \in \operatorname{dom}\Lambda_t(x_*(t), \cdot). \qquad \blacklozenge$$

4.1.2 Remark (a) As in the previous chapter, the tempered growth condition is automatically satisfied if the ratio $R(t)/k(t)$ is essentially bounded away from 0.

(b) To obtain the conclusions of the theorem, it suffices that x_* be a local minimum in a weaker sense than stated above; that is, relative to not only the stated constraints but also

$$\Lambda_t(x(t), \dot{x}(t)) \le \Lambda_t(*) + R(t) \text{ a.e.}$$

and

$$\int_a^b [\Lambda_t(x(t), \dot{x}(t))) - \Lambda_t(*)]_+ \, dt \le \varepsilon.$$

While the theorem's hypotheses may seem difficult to verify in practice, it encompasses a number of special cases in which verifiable criteria can be formulated. We proceed to illustrate this in various contexts, now and in the following sections, in which the advantages of both stratified necessary conditions and proximal criteria for pseudo-Lipschitz behavior become apparent.

4.1.3 Theorem Suppose that in Theorem 4.1.1 we replace the pseudo-Lipschitz and tempered growth conditions by the following bounded slope postulate: there exists a summable function k with

$$\text{ess inf}\left\{\frac{R(t)}{k(t)} : t \in [a,b]\right\} > 0$$

such that for almost all $t \in [a,b]$, for every x in $B(x_*(t),\varepsilon)$ and v in $B(\dot{x}_*(t),R(t))$, for every $(\zeta,\psi) \in \partial_P \Lambda_t(x,v)$, one has

$$|\zeta| \leq k(t)\left\{1+|\psi|\right\}.$$

Then the conclusions of Theorem 4.1.1 are valid. ♦

Proof. We wish to invoke Corollary 3.5.3, whose bounded slope condition we proceed now to verify. Accordingly, let $(\alpha,\beta,-\gamma)$ be an element of $N^P_{G(t)}(x,v,\Lambda_t(x,v)+\delta)$, where $x \in B(x_*(t),\varepsilon)$, $v \in B(\dot{x}_*(t),R(t))$.

Consider first the case $\gamma > 0$. It follows then (see 1.2.1 of [27]) that $\delta = 0$, whence

$$(\alpha/\gamma,\beta/\gamma) \in \partial_P \Lambda_t(x,v),$$

so that by hypothesis

$$|\alpha/\gamma| \leq k(t)(1+|\psi/\gamma|).$$

This gives rise to

$$|\alpha| \leq 2k(t)\left|(\psi,\gamma)\right|,$$

which confirms the bounded slope condition.

Consider now the case $\gamma = 0 = \delta$. Then (by an approximation result of Rockafellar; see 1.11.23 of [27]) there exist $\gamma' > 0, \alpha', \beta', x', v'$ arbitrarily close to $0, \alpha, \beta, x, v$ respectively, with $\Lambda_t(x',v')$ arbitrarily close to $\Lambda_t(x,v)$, such that

$$(\alpha',\beta',-\gamma') \in N^P_{G(t)}(x',v',\Lambda(x',v')).$$

Then we get the same bound as in the previous case for the perturbed data, and hence for the original in the limit.

There remains the case $\gamma = 0, \delta > 0$. In this case it follows (see 1.2.1 of [27]) that we have

$$(\alpha,\beta,0) \in N^P_{G(t)}(x,v,\Lambda(x,v)).$$

Applying the previous case, we derive

$$|\alpha| \leq 2k_R(t)\left|(\beta,0)\right|.$$

Thus we have verified the bounded slope condition of Corollary 3.5.5 in all cases. Applying that result, we obtain the desired conclusions. ■

4.2 A classical multiplier rule

We consider now the following problem: to minimize the functional $J(x)$ subject to, as in the previous section, the endpoint constraints $(x(a), x(b)) \in S$, but subject as well to an additional *pointwise* constraint

$$\psi_t(x(t), \dot{x}(t)) = 0, \; t \in [a, b],$$

where $\psi : [a, b] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ is a given function $(m \leq n)$. This is referred to classically as a problem of Lagrange, and in this section we shall consider it only under classical hypotheses. We suppose that the Lagrangian Λ and the function ψ are continuously differentiable, and that a weak local minimum x_* exists which is also of class C^1. Our purpose is to illustrate how naturally the stratified necessary conditions yield the classical *multiplier rules* of the calculus of variations.

4.2.1 Theorem Suppose that $D_v \psi_t(x_*(t), \dot{x}_*(t))$ has maximal rank for each t. Then there exist a scalar $\lambda_0 \in \{0, 1\}$, an arc p and a continuous function $\lambda : [a, b] \to \mathbb{R}^m$ with $(\lambda_0, p(t)) \neq 0$ for each t such that the transversality condition holds:

$$(p(a), -p(b)) \in \partial_L \lambda_0 \ell(x_*(a), x_*(b)) + N_S^L(x_*(a), x_*(b)),$$

as well as the Euler equation:

$$(\dot{p}(t), p(t)) = D_{x,v} \left[\lambda_0 \Lambda_t + \langle \lambda(t), \psi_t \rangle\right] (x_*(t), \dot{x}_*(t)), \; t \in [a, b]. \qquad \blacklozenge$$

Proof. The hypotheses permit us to affirm the existence of positive constants ε and k such that for every t, for any $x \in B(x_*(t), \varepsilon)$ and $v \in B(\dot{x}_*(t), \varepsilon)$, the matrix $D_v \psi_t(x, v)$ is of maximal rank and the inequality

$$|D_x \langle \lambda, \psi_t(x, v) \rangle| \leq k \, |D_v \langle \lambda, \psi_t(x, v) \rangle|$$

holds for any vector λ in $\mathbb{R}^m$. We may pick ε small enough so that x_* is a global minimum relative to the uniform neighborhood of radius ε around $(x_*, \dot{x}_*)$.

Consider now the following multifunction:

$$F_t(x, y) := \{(v, \Lambda_t(x, v) + \delta) : \delta \geq 0, v \in \mathbb{R}^n, \psi_t(x, v) = 0\},$$

which is easily seen to satisfy the basic hypotheses. As in the previous section, we consider the minimization of $\ell(x(a), x(b)) + y(b)$ over the trajectories

(x, y) of F satisfying $y(a) = 0$ and $(x(a), x(b)) \in S$. A weak local minimum is furnished by the arc (x_*, y_*), where

$$y_*(t) := \int_a^t \Lambda_s(x_*(s), \dot{x}_*(s))\, ds.$$

With an eye to applying Corollary 3.5.3, we take $R(t) \equiv \varepsilon$ and $k(t) \equiv k$. The following fact will be needed. (We suppress the t dependence to ease the notation; G signifies the graph of F_t.)

Lemma Let $(\alpha, \beta, -\gamma) \in N_G^L(x, v, \Lambda(x, v) + \delta)$, where $D\psi(x, v)$ has maximal rank. Then $\gamma \geq 0$ and there exists λ such that

$$(\alpha, \beta) = D_{x,v}\left[\gamma \Lambda(x, v) + \langle \lambda, \psi \rangle\right](x, v).$$

The proof of the lemma follows familiar lines in proximal analysis (see [27]): for the proximal normal cone, an inequality characterizes the normal vector, and a Lagrange multiplier rule gives the stated formula; then the result for limiting normals follows.

We now proceed to verify the bounded slope condition of Corollary 3.5.3. Accordingly, let $(\alpha, \beta, -\gamma)$ belong to $N_G^P(x, v, \Lambda(x, v) + \delta)$, where x and v lie within ε of $x_*(t)$ and $\dot{x}_*(t)$ respectively. According to the lemma we have

$$(\alpha - \gamma D_x \Lambda(x, v), \beta - \gamma D_v \Lambda(x, v)) = D_{x,v} \langle \lambda, \psi \rangle (x, v).$$

The choice of ε and k then implies

$$|\alpha - \gamma D_x \Lambda(x, v)| \leq k\, |\beta - \gamma D_v \Lambda(x, v)|.$$

If M is a uniform bound on the operator norm of $D_{x,v}\Lambda$ over the relevant (bounded) set of possible values of (x, v), we derive from this

$$|\alpha| \leq k\, |\beta| + M(k + 1)\gamma,$$

which yields the required bounded slope condition $|\alpha| \leq K\, |(\beta, \gamma)|$ for a certain constant K.

We now invoke Corollary 3.5.3 to deduce the existence of an arc having the form $(p, -\lambda_0)$ satisfying the nontriviality and transversality conditions together with the inclusion

$$\dot{p} \in \mathrm{co}\, \left\{\omega : (\omega, p, -\lambda_0) \in N_G^L(x_*, \dot{x}_*, \Lambda_t(*))\right\}.$$

Applying the lemma, we deduce the existence of λ (depending on t) such that

$$(\dot{p}, p) = D_{x,v}\left[\lambda_0 \Lambda + \langle \lambda, \psi \rangle\right](x_*, \dot{x}_*).$$

By focusing on the second component of this equation, and bearing in mind that $D_v \psi_t(x_*(t), \dot{x}_*(t))$ has maximal rank, one deduces the continuity of $\lambda(t)$. ■

4.2.2 Remark The proof is easily adaptable to the case of pointwise *inequality* constraints, and solutions x_* which are 'piecewise smooth'.

4.3 Generalized Tonelli-Morrey integrands

It is well-known that even in a classical smooth setting, a (non Lipschitz) solution x_* of the basic problem may not satisfy the Euler equation. This is closely linked to a different pathology called the *Lavrentiev phenomenon* (see for example Cesari [10]): the infimum of $J(x)$ over smooth (or Lipschitz) arcs may differ from that taken over all arcs.

In view of these facts, it is of evident interest to identify classes of Lagrangians for which *a priori* these two phenomena do not arise. In particular, the conclusion that a solution x_* is necessarily *Lipschitz* (or equivalently, has essentially bounded derivative) is most desirable, for it precludes the Lavrentiev phenomenon, very often allows one to assert the necessary conditions, and is generally the key to obtaining further regularity of x_* (smoothness). We discuss regularity later in §4.5. In this section, we define a new class of Lagrangians for which the necessary conditions can be asserted to hold.

4.3.1 Definition The Lagrangian Λ satisfies the **generalized Tonelli-Morrey growth condition** if for every bounded subset X of $\mathbb{R}^n$ there exist a constant c and a summable function d such that for almost every t, for every $(x, v) \in X \times \mathbb{R}^n$, for every $(\zeta, \psi) \in \partial_P \Lambda_t(x, v)$, one has

$$\frac{|\zeta|}{1 + |\psi|} \le c\,\{\Lambda_t(x, v) + |v|\} + d(t).$$

4.3.2 Theorem Let x_* be a $W^{1,1}$ local minimum for $\mathcal{P}_\Lambda$, where Λ satisfies the generalized Tonelli-Morrey growth condition. Then there exist an arc p and a scalar $\lambda_0 \in \{0, 1\}$ with

$$(p(t), \lambda_0) \neq (0, 0)\ \forall t \in [a, b]$$

satisfying the transversality condition

$$(p(a), -p(b)) \in \lambda_0 \partial_L \ell(x_*(a), x_*(b)) + N_S^L(x_*(a), x_*(b)),$$

the (possibly abnormal) Euler inclusion

$$\dot{p}(t) \in \operatorname{co}\left\{\omega : (\omega, p(t)) \in \partial_L^{\lambda_0} \Lambda_t(x_*(t), \dot{x}_*(t))\right\} \text{ a.e.,}$$

and the global Weierstrass condition: for almost all $t \in [a,b]$ and for all $v \in \operatorname{dom} \Lambda_t(x_*(t), \cdot)$, we have

$$\lambda_0 \Lambda_t(x_*(t), v) - \lambda_0 \Lambda_t(x_*(t), \dot{x}_*(t)) \geq \langle p(t), v - \dot{x}_*(t) \rangle . \qquad \blacklozenge$$

Proof. We define the usual multifunction:

$$F_t(x,y) := \{(v, \Lambda_t(x,v) + \delta) : \delta \geq 0, v \in \mathbb{R}^n\},$$

and consider the minimization of $\ell(x(a), x(b)) + y(b)$ over the trajectories (x,y) of F satisfying $y(a) = 0$ and $(x(a), x(b)) \in S$. A local minimum in the $W^{1,1}$ sense is furnished by the arc (x_*, y_*), where

$$y_*(t) := \int_a^t \Lambda_s(x_*(s), \dot{x}_*(s))\, ds.$$

We wish to invoke Corollary 3.5.5, whose bounded slope condition we proceed now to verify. Accordingly, for $\varepsilon > 0$, let X be an open bounded set containing all the balls $B(x_*(t), \varepsilon), t \in [a,b]$, and let c and d be given as in Definition 4.3.1. Now let $(\alpha, \beta, -\gamma)$ be an element of $N^P_{G(t)}(x, v, \Lambda_t(x,v) + \delta)$, where $x \in B(x_*(t), \varepsilon)$.

Consider first the case $\gamma > 0$. It follows then (see 1.2.1 of [27]) that $\delta = 0$, whence

$$(\alpha/\gamma, \beta/\gamma) \in \partial_P \Lambda_t(x,v),$$

so that (by 4.3.1)

$$\frac{|\alpha|}{\gamma + |\beta|} \leq c\,\{\Lambda_t(x,v) + |v|\} + d(t).$$

This gives rise to

$$|\alpha| \leq 2\,[c\,\{\Lambda_t(x,v) + |v|\} + d(t)]\,|(\beta,\gamma)|,$$

which confirms the bounded slope condition for this case. But the cases $\gamma = 0 = \delta$ and $\gamma = 0, \delta > 0$ are reduced to the one just treated exactly as in the proof of Theorem 4.1.3. This confirms the bounded slope condition of Corollary 3.5.5. Applying that result, we obtain the desired conclusions. ∎

4.3.3 Remark (a) The abnormal case $\lambda_0 = 0$ is excluded in the theorem if for all t in a set of positive measure the function $\Lambda_t(x_*(t), \cdot)$ is locally finite near $\dot{x}_*(t)$.

(b) A weaker form of the generalized Tonelli-Morrey growth condition is one in which the term $\Lambda_t(x, v)$ is replaced by its absolute value. This weaker form suffices for the theorem if there exist a summable function θ and a vector ψ such that

$$\Lambda_t(x, v) \geq \theta(t) + \langle \psi, v \rangle \quad \forall\, t \in [a, b], x \in B(x_*(t), \varepsilon), v \in \mathbb{R}^n,$$

for then we can reduce to the case of a nonnegative Lagrangian by considering instead

$$\tilde{\Lambda}_t(x, v) := \Lambda_t(x, v) - \theta(t) - \langle \psi, v \rangle$$

and appropriately modifying ℓ.

An example

We discuss now a simple problem of dimension $n = 1$ which is beyond the scope of previous necessary conditions in the literature. It involves the minimization of

$$\int_0^1 \left\{ \sqrt{|x(t) - \dot{x}(t)|} + \dot{x}(t) \right\} dt$$

subject to $x(0) = 0$. Observe that the Lagrangian

$$\Lambda(x, v) := \sqrt{|x - v|} + v$$

is neither locally Lipschitz in x nor convex in v. This is a special case of problem $\mathcal{P}_\Lambda$ with $\ell \equiv 0$ and $S := \{0\} \times \mathbb{R}$. The issue we address is whether the arc $x_* \equiv 0$ is a local minimum of some type.

Claim 1. Λ satisfies the Tonelli-Morrey growth condition 4.3.1. This is clear, with $c = 0, d \equiv 1$, at a point (x, v) for which $x \neq v$, since then any (ζ, ψ) in $\partial_P \Lambda(x, v)$ satisfies $\psi + \zeta = 1$. But the same relationship between ζ and ψ must hold when $x = v$, as follows from the defining inequality for a proximal subgradient.

Claim 2. x_* fails to be a local minimum in the $W^{1,1}$ sense.
If x_* is such a minimum, then there is an arc p satisfying the conclusions of Theorem 4.3.2. The normal case $\lambda_0 = 1$ must hold. The Euler inclusion for x_* gives $\dot{p}+p = 1$, and the transversality condition provides $p(1) = 0$; hence $p(t) = 1 - e^{1-t}$. The (global) Weierstrass condition reads

$$\sqrt{|v|} + v \geq \langle p(t), v\rangle \quad \forall\, v \in \mathbb{R},$$

which cannot be.

Claim 3. If x_* is a local solution of radius R, then

$$R(t) \leq e^{2(t-1)} \ \text{ a.e.}$$

If x_* is a solution of radius R, then by Theorem 4.1.3 we have, for the same $p(t)$ as above, for almost every t, the inequality

$$\sqrt{|v|} + v \geq \langle p(t), v\rangle \quad \forall\, v \in [-R(t), R(t)].$$

It is easily seen that this inequality forces $R(t) \leq (1 - p(t))^{-1} = e^{2(t-1)}$.

Claim 4. If $R(t) \leq e^{2(t-1)}$ a.e., then x_* is a local solution of radius R; thus x_* is a weak local minimum.
We prove this by an *ad hoc* argument.[1] First, observe that it suffices to consider arcs x for which $x(t) \leq 0$ and $\dot{x}(t) \leq 0$. For any such admissible arc x within radius R of x_*, we have $|\dot{x}(t)| \leq e^{2(t-1)}$, which implies $|x(t)| \leq e^{2(t-1)}$. It follows that $|x(t) - \dot{x}(t)| \leq e^{2(t-1)}$. In turn this gives

$$\begin{aligned}\sqrt{|x(t) - \dot{x}(t)|} + \dot{x}(t) &\geq |x(t) - \dot{x}(t)|\, e^{1-t} + \dot{x}(t) \\ &\geq (x(t) - \dot{x}(t))e^{1-t} + \dot{x}(t) = \frac{d}{dt}\left\{x(t)(1 - e^{(1-t)})\right\}.\end{aligned}$$

Integrating both sides from 0 to 1, we get the desired conclusion.

4.4 Finite Lagrangians

We consider the following version of the basic problem in the calculus of variations:

$$\text{minimize } J(x) := \ell(x(a), x(b)) + \int_a^b \Lambda_t(x(t), \dot{x}(t))\, dt$$

subject to

$$\dot{x}(t) \in V_t \ \text{ a.e.}$$

[1] Whose inspiration stems from the *verification function* method; see [22] and [27].

Hypotheses. We retain the basic hypotheses as regards the Lagrangian Λ, but now require it to be *finite-valued*. We no longer require ℓ to be locally Lipschitz; instead, ℓ is merely assumed to be lower semicontinuous, possibly extended-valued. (This is why it is no longer necessary to make explicit any endpoint constraint.) As for the multifunction V, it is taken to be measurable and closed-valued. This setting allows certain refinements to be made in weakening the hypotheses under which the necessary conditions can be obtained, and will be useful later in proving regularity theorems.

We introduce a function $\psi : [a,b] \times [0,\infty) \to [0,\infty)$ which is $\mathcal{L} \times \mathcal{B}$-measurable, and such that for almost every t, the mapping $r \to \psi_t(r)$ is locally Lipschitz, nondecreasing, and 0 at 0. This function will play a role in defining the generalized nature of the local minimum below, extending the case $\psi_t(r) = |r|$ that we have most often considered so far.

Let x_* be an arc feasible for this problem such that the integral of Λ along x_* is well-defined and finite and such that $\ell(x_*(a), x_*(b))$ is finite, and let $R : [a,b] \to (0,\infty]$ be a given measurable function. We posit the existence of a positive number ε and a summable function c such that, for almost each $t \in [a,b]$,

$$x \in \overline{B}(x_*(t), \varepsilon),\ v \in V_t,\ |v - \dot{x}_*(t)| \le R(t) \Longrightarrow \Lambda_t(x,v) \ge c(t).$$

We assume that x_* is a local minimum of radius R for the problem in the following general sense: for any arc x which satisfies the constraints

$$\dot{x}(t) \in V_t \text{ a.e.},\ |\dot{x}(t) - \dot{x}_*(t)| \le R(t) \text{ a.e.} \tag{4.1}$$

and which is ε-close to x_* as follows:

$$\|x - x_*\|_\infty \le \varepsilon,\ \int_a^b \psi_t(|\dot{x}(t) - \dot{x}_*(t)|)\, dt \le \varepsilon, \tag{4.2}$$

we have $J(x) \ge J(x_*)$. Note that in view of the hypotheses, the integral of Λ along x is well-defined for any such x. In the following, the Euler inclusion involves the *indicator function* I_{V_t} of the set V_t, the function which is equal to 0 on the set and $+\infty$ elsewhere.

4.4.1 Theorem Let there exist a summable function k such that, for almost all t, for all $v \in V_t$ with $|v - \dot{x}_*(t)| \le R(t)$ and for all $x, x' \in \overline{B}(x_*(t), \varepsilon)$ we have

$$\left|\Lambda_t(x', v) - \Lambda_t(x,v)\right| \le k(t) \left|x' - x\right|.$$

Suppose as well that the set

$$\Omega := \{t : \dot{x}_*(t) \in \operatorname{int} V_t\}$$

has positive measure. Then there exists an arc p which satisfies the transversality condition

$$(p(a), -p(b)) \in \partial_L \ell(x_*(a), x_*(b)),$$

the Euler inclusion

$$\dot{p}(t) \in \text{co}\, \{\omega : (\omega, p(t)) \in \partial_L \left[\Lambda_t + I_{V_t}\right](x_*(t), \dot{x}_*(t))\} \text{ a.e.,}$$

and the Weierstrass condition of radius R: for almost each $t \in [a, b]$,

$$\Lambda_t(x_*(t), v) - \Lambda_t(x_*(t), \dot{x}_*(t)) \geq \langle p(t), v - \dot{x}_*(t)\rangle$$

$$\forall\, v \in V_t \cap \overline{B}(\dot{x}_*(t), R(t)).$$

Now suppose that x_* satisfies the above hypotheses for a sequence of radius functions R_i (with possibly different c, ψ, ε, k depending on i) such that

$$\lim_{i\to\infty} \text{ess inf } \{R_i(t) : t \in [a, b]\} = +\infty \text{ a.e.}$$

and suppose that the following set has positive measure:

$$\Omega' := \{t \in \Omega : \exists\, r > 0, \sigma > 0 \ni \Lambda_t(x_*(t), v) < \sigma \quad \forall v \in B(\dot{x}_*(t), r)\}.$$

Then there is an arc p satisfying all of the above as well as the global Weierstrass condition: for almost every t, we have

$$\Lambda_t(x_*(t), v) - \Lambda_t(x_*(t), \dot{x}_*(t)) \geq \langle p(t), v - \dot{x}_*(t)\rangle \quad \forall\, v \in V_t. \qquad \blacklozenge$$

Proof of the theorem. Without loss of generality we take $x_* \equiv 0$, $[a, b] = [0, 1]$, and $\Lambda_t(0, 0) = 0$ a.e. Fix any radius function $R(\cdot)$ for which the hypotheses of the theorem hold. For t in $[0, 1]$ and for $(y, z, \alpha, \beta)) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, we define $F_t(y, z, \alpha, \beta) = F_t(y)$ to be the set of all points $(w, r, 0, \theta)$ in the same space such that

$$w \in V_t \cap \overline{B}(0, R(t)),\ \Lambda_t(y, w) \leq r,\ \theta = \psi_t(|w|).$$

The multifunction F satisfies the basic hypotheses. We consider the problem of minimizing

$$z(1) + \alpha(1)$$

over the trajectories (y, z, α, β) of F on $[0, 1]$ satisfying

$$\|y\|_\infty < \varepsilon$$

and the boundary conditions

$$(y(0), y(1), \alpha(1)) \in \operatorname{epi} \ell,\ z(0) = 0,\ \beta(0) = 0,\ \beta(1) \le \varepsilon.$$

It is a simple exercise to verify that the arc $(0,0,0,0)$ solves this problem.

We claim that F_t is Lipschitz on the set $|y| < \varepsilon$ with Lipschitz constant $k(t)$. For let y, y' be points in that set, and let $(w, r, 0, \theta)$ be any point in $F_t(y)$. Then w lies in $V_t \cap \overline{B}(0, R(t))$, θ equals $\psi_t(|w|)$, and r is of the form

$$\Lambda_t(y, w) + \delta$$

for some $\delta \ge 0$. Set

$$r' := \Lambda_t(y', w) + \delta.$$

Then the point $(w, r', 0, \theta)$ belongs to $F_t(y')$ and satisfies

$$\big|(w, r', 0, \theta) - (w, r, 0, \theta)\big| = \big|r' - r\big| = \big|\Lambda_t(y', w) - \Lambda_t(y, w)\big| \le k(t),$$

in light of the Lipschitz condition satisfied by Λ. This proves the claim.

We now apply Theorem 3.1.1 with radius function identically $+\infty$. The set S of that theorem is given by

$$\{(y_0, z_0, \alpha_0, \beta_0, y_1, z_1, \alpha_1, \beta_1) : (y_0, y_1, \alpha_1) \in \operatorname{epi} \ell,\ z_0 = \beta_0 = 0,\ \beta_1 \le \varepsilon\}.$$

It follows readily that we obtain an arc $(p, -\lambda_0, 0, 0)$ and a number λ_0 equal to 0 or 1 such that $(p(t), \lambda_0)$ is nonvanishing,

$$(p(0), -p(1), -\lambda_0) \in N^L_{\operatorname{epi} \ell}(0, 0, \ell(0,0)), \tag{4.3}$$

and such that the Weierstrass condition for F holds, which gives:

$$\lambda_0 \Lambda_t(0, w) \ge \langle p(t), w\rangle \quad \forall\, w \in V_t \cap \overline{B}(0, R(t)), \ \text{ a.e.}$$

Because the set Ω has nonempty interior on a set of positive measure, the abnormal case is excluded by this conclusion, for if $\lambda_0 = 0$ then $p(t)$ must vanish for some values of t, which contradicts the nonvanishing of $(p(t), \lambda_0)$. Thus $\lambda_0 = 1$. Then (4.3) gives

$$(p(0), -p(1)) \in \partial_L \ell(0, 0).$$

The Euler equation asserts that for almost every t, $\dot p(t)$ belongs to the convex hull of the set of points ω such that

$$(\omega, p(t), -1, 0) \in N^L_{D_t}(0, 0, 0, 0),$$

where D_t is the set

$$\left\{(y, w, \Lambda_t(y, w) + \delta, \psi_t(|w|)) : \delta \geq 0, w \in V_t \cap \overline{B}(0, R(t))\right\}.$$

Proximal analysis shows that such an ω satisfies

$$(\omega, p(t)) \in \partial_L \left[\Lambda_t + I_{V_t}\right](0, 0)$$

(the Lipschitz condition on ψ_t figures here in applying the proximal sum rule). Then p satisfies all the assertions of the theorem, which is therefore proven for the case of a given radius function $R(\cdot)$.

There remains the final assertion of the theorem in the case in which the hypotheses hold for the given sequence R_i of radius functions. Each arc p_i obtained from applying the fixed radius case satisfies

$$|\dot{p}_i(t)| \leq k_1(t) \text{ a.e.}$$

as a consequence of the Euler inclusion, where k_1 is the Lipschitz function for radius R_1. To apply the usual convergence arguments that will lead to the required limiting arc that satisfies the Weierstrass condition globally, it suffices to exhibit a uniform bound on

$$\min\left\{|p_i(t)| : t \in [0, 1]\right\}.$$

To obtain such a bound, we use the fact that the set Ω' has positive measure to deduce the existence of a subset Σ of Ω, also of positive measure, together with positive numbers r and σ such that

$$\Lambda_t(0, v) \leq \sigma \quad \forall\, v \in B(0, r) \subset V_t, \quad \forall\, t \in \Sigma.$$

But then, for any index i sufficiently large so that

$$\text{ess inf}\left\{R_i(t) : t \in [a, b]\right\} > r,$$

the Weierstrass condition implies that on a set of positive measure we have

$$|p_i(t)| \leq \sigma / r,$$

which completes the proof of the theorem. ∎

4.4.2 Remark (a) The set Ω' will have positive measure as required in the limiting case of the theorem if the hypothesis $\Lambda_t(x, v) \geq c(t)$ is strengthened to $|\Lambda_t(x, v)| \leq c(t)$. Alternatively, we can dispense with any hypothesis concerning Ω' if ℓ is locally Lipschitz in one of its arguments.

(b) As in Remark 3.1.2(a), it suffices in the limiting case that, for some $\varepsilon > 0$, x_* provide a minimum with respect to the constraints

$$\dot{x} - \dot{x}_* \in L^\infty, \int_a^b \psi_t(|\dot{x}(t) - \dot{x}_*(t)|)\, dt \leq \varepsilon, \ \|x - x_*\|_\infty \leq \varepsilon.$$

4.5 Three regularity theorems

There is a close link between necessary conditions and the regularity of the solution: If we know the solution to be regular, then we can usually assert the necessary conditions; conversely, if we can write the necessary conditions, then we may be able to deduce regularity from them. In this section we use the new necessary conditions to derive regularity consequences for solutions of the problem $\mathcal{P}_\Lambda$ when the Lagrangian is finite-valued. The following property will play a role.

4.5.1 Definition We say that the Lagrangian Λ is **coercive** if for any bounded subset X of $\mathbb{R}^n$ there exists a function $\theta : [0, \infty) \to \mathbb{R}$ satisfying

$$\lim_{r \to \infty} \frac{\theta(r)}{r} = +\infty$$

and such that

$$\Lambda_t(x, v) \geq \theta(|v|) \ \ \forall\, (t, x, v) \in [a, b] \times X \times \mathbb{R}^n.$$

We remark that coercivity is a familiar ingredient in the theory of existence of solutions (see for example Cesari [10]). The symbiosis between necessary conditions and regularity is well illustrated by the following new result.

4.5.2 Theorem Let x_* be a local minimum of $W^{1,1}$ type for $\mathcal{P}_\Lambda$, where Λ is finite-valued and satisfies the generalized Tonelli-Morrey growth condition 4.3.1. Suppose in addition that Λ is coercive, and bounded above on bounded sets. Then x_* is Lipschitz. ♦

Proof. In view of Theorem 4.3.2, we know that an arc p exists which satisfies the (normal) Weierstrass condition. Let M be an upper bound on

$$\Lambda_t\left(x_*(t), \frac{\dot{x}_*(t)}{1+|\dot{x}_*(t)|}\right)$$

for $t \in [a,b]$, and let θ be a coercivity function for Λ when x is restricted to the bounded set consisting of the values of x_* on $[a,b]$. Then, taking $v := \dot{x}_*(t)/(1+|\dot{x}_*(t)|)$ in the Weierstrass inequality leads to (almost everywhere)

$$\theta(|\dot{x}_*(t)|) \leq \Lambda_t(x_*(t), \dot{x}_*(t)) \leq M + |p(t)|\,|\dot{x}_*(t)|\,.$$

Since $|p|$ is bounded and $\lim_{r\to\infty}\theta(r)/r = +\infty$, it follows from this that $\dot{x}_*$ is essentially bounded. ■

> **4.5.3 Remark** The reasoning given in Remark 4.3.3 shows that (in the presence of coercivity), the (apparently) weaker 'absolute value form' of the Tonelli-Morrey growth condition 4.3.1 suffices.

Another regularity theorem

It is possible to formulate a purely local hypothesis (instead of global Tonelli-Morrey growth) and still obtain conclusions in regard to necessary conditions or regularity. We illustrate this now in a classical setting. We consider a Lagrangian $\Lambda_t(x,v)$ which is continuously differentiable in (x,v) for each t and satisfies: for some $\varepsilon > 0$ and summable function k we have

$$|D_x\Lambda_t(x, \dot{x}_*(t))| \leq k(t),\ x \in B(x_*(t),\varepsilon),\ t \in [a,b] \text{ a.e.} \tag{4.4}$$

Note that if x_* satisfies the Euler equation

$$(\dot{p}(t), p(t)) = D_{x,v}\Lambda_t(x_*(t), \dot{x}_*(t)) \text{ a.e.,}$$

then $|D_x\Lambda_t(x_*(t), \dot{x}_*(t))|$ must be summable. In this light, the condition (4.4) may be thought of as being 'close' to necessary for the Euler equation to hold. It turns out to be sufficient.

4.5.4 Theorem Let x_* be a weak local minimum for $\mathcal{P}_\Lambda$ which satisfies condition (4.4). Then there exists an arc p satisfying (with $\lambda_0 = 1$) the transversality condition and Euler equation. If in addition Λ is convex in v, then the global Weierstrass condition holds. And if in addition to that the Lagrangian is coercive and bounded above on bounded sets, then x_* is Lipschitz. ♦

Proof. For each t such that (4.4) holds, for every $x \in B(x_*(t), \varepsilon)$, we have

$$\Lambda_t(x, \dot{x}_*(t)) \geq \Lambda_t(*) - \varepsilon k(t).$$

For every such t there exists $R(t) > 0$ such that

$$|D_x \Lambda_t(x, v)| \leq k(t) + 1, \; x \in B(x_*(t), \varepsilon), v \in B(\dot{x}_*(t), R(t))$$

and

$$\Lambda_t(x, v) \geq \Lambda_t(*) - \varepsilon k(t) - 1, \; x \in B(x_*(t), \varepsilon), v \in B(\dot{x}_*(t), R(t)).$$

We may suppose $R(\cdot)$ measurable, and that x_* is a solution relative to the radius R. It follows that Λ possesses the properties that permit us to invoke Theorem 4.4.1 (with $V_t = \mathbb{R}^n$). This gives rise to the full set of necessary conditions, with the Weierstrass condition being of radius R. The remaining conclusions of the theorem follow now as in the proof of Theorem 4.5.2. ■

Autonomous integrands

One of several principal threads running through the theory of Lipschitz regularity involves conditions which limit the nature of the t dependence of the Lagrangian. It was proved by Clarke and Vinter [28] (see also Section 11.4 of [58]) in particular that when Λ is *autonomous* (has no explicit dependence on the t variable) and locally Lipschitz in (x, v), as well as convex in v and coercive, then any strong local minimum for the basic problem has essentially bounded derivative. We now show that it is possible to dispense with the Lipschitzness and convexity hypotheses in that result.

4.5.5 Theorem Let Λ be autonomous and coercive, and bounded above on bounded sets. If x_* provides a strong local minimum for $\mathcal{P}_\Lambda$, then x_* is Lipschitz. ♦

Proof. There is no loss of generality in assuming that Λ is nonnegative in a strong neighborhood of x_* (in view of coercivity, and by adding an appropriate constant if necessary). We define a new Lagrangian $L : [a, b] \times \mathbb{R} \to [0, +\infty]$ as follows:

$$L(t, w) := \Lambda\left(x_*(t), \frac{\dot{x}_*(t)}{w}\right) w + I_{[1/2, +\infty)}(w).$$

We consider the minimization of

$$\int_a^b L(t, \dot{y}(t)) \, dt$$

over the scalar arcs y on $[a,b]$ satisfying $y(a)=a,\ y(b)=b$, as well as

$$\|y(t)-t\|_\infty \le \delta.$$

Here, δ is a positive number picked so that for any $t \in [a,b]$, the inequality $|y-t| \le \delta$ implies $|x_*(y)-x_*(t)| < \varepsilon_*$, where ε_* is the radius of the uniform neighborhood relative to which x_* is optimal.

We claim that the arc $y_*(t) := t$ solves this problem (for which it is evidently feasible). For suppose there were a feasible arc y for the problem strictly better than y_*. Then y is strictly increasing, and the arc $x(t) := x_*(y^{-1}(t))$ is feasible for the original problem, and has the same boundary values as x_*. It also lies in the uniform ε_*-neighborhood about x_*. Furthermore we find (by the change of variables formula[2])

$$\begin{aligned}
\int_a^b \Lambda(x(\tau),\dot{x}(\tau))\,d\tau &= \int_a^b \Lambda\left(x_*(t),\frac{\dot{x}_*(t)}{\dot{y}(t)}\right)\dot{y}(t)\,dt \\
&= \int_a^b L(t,\dot{y}(t))\,dt < \int_a^b L(t,\dot{y}_*(t))\,dt \\
&= \int_a^b \Lambda(x_*(t),\dot{x}_*(t))\,dt,
\end{aligned}$$

which contradicts the optimality of x_* for the original problem, proving the claim.

Since L is independent of y, it evidently satisfies the generalized Tonelli-Morrey condition 4.3.1. We may therefore invoke the necessary conditions of Theorem 4.3.2 for the problem solved by y_*, in the case of a strong local minimum, with ℓ equal to the indicator of the singleton set $(0,0)$, and with $\lambda_0 = 1$. The Euler and Weierstrass conditions assert that for some scalar p, for almost all t, we have

$$\Lambda\left(x_*(t),\frac{\dot{x}_*(t)}{w}\right)w - \Lambda\left(x_*(t),\dot{x}_*(t)\right) \ge p(w-1) \quad \forall\, w \in [1/2,+\infty).$$

Let M be an upper bound on

$$\Lambda\left(x_*(t),\frac{\dot{x}_*(t)}{1+|\dot{x}_*(t)|}\right)$$

for $t \in [a,b]$, and let θ be a coercivity function for Λ when x is restricted to the bounded set consisting of the values of x_* on $[a,b]$. Then, taking $w := 1+|\dot{x}_*(t)|$ in the Weierstrass inequality leads to (almost everywhere)

$$\theta(|\dot{x}_*(t)|) \le \Lambda(x_*(t),\dot{x}_*(t)) \le M(1+|\dot{x}_*(t)|) + |p|\,|\dot{x}_*(t)|\,.$$

[2]See for example *Real Analysis* (Second Ed.) by H.L. Royden.

Since $\lim_{r\to\infty}\theta(r)/r = +\infty$, this implies that $\dot{x}_*$ is essentially bounded. ■

4.5.6 Remark Note that no necessary condition of Euler type for the solution of the original problem is asserted here (unsurprisingly, since no hypothesis is made with respect to the x dependence, except for measurability). However, the proof yields the existence of a scalar p satisfying

$$p \in \partial_P L(t,1) \text{ a.e.},$$

a condition which implies, when Λ is differentiable in v, the classically familiar *second Erdmann condition*:

$$\Lambda(x_*(t),\dot{x}_*(t)) - \langle \dot{x}_*(t), D_v\Lambda(x_*(t),\dot{x}_*(t))\rangle = \text{constant} \quad \text{a.e.}$$

4.6 Notes

§4.1 Theorem 4.1.1 subsumes and greatly extends the early results on the generalized problem of Bolza, such as those of Clarke [13][15]. Loewen and Rockafellar [44], in an article already mentioned in the notes for Chapter 3, prove necessary conditions for the (unstratified) problem of Bolza under the convexity hypothesis; their result is subsumed by Theorem 4.1.3.

Another line of research in connection with the problem of Bolza seeks to apply the Hamiltonian (rather than the Lagrangian) formulation of both hypotheses and conclusions; see the notes for Chapter 6.

§4.2 Nowadays, differentially constrained problems are generally treated in the context of optimal control, under much weaker hypotheses, as we do in the next chapter. The books of Bliss [5] and Hestenes [34] are principal references for the multiplier rule.

§4.3 The class of Lagrangians satisfying what we have termed the generalized Tonelli-Morrey growth condition appears here for the first time. For smooth coercive Lagrangians, the condition may be written in the equivalent form

$$|D_x\Lambda| \le c_1\{|\Lambda| + |D_v\Lambda|\} + d_1(t) + \{c_2|\Lambda| + d_2(t)\}|D_v\Lambda|.$$

By taking $c_2 = d_2 = 0$, one obtains a class of Lagrangians that has been considered before, notably by Clarke and Vinter [28] in connection with regularity (see below).

The example given in this section is of interest partly because it fails to be Lipschitz in the x variable. But Theorem 4.3.2 also extends the class of *smooth* Lagrangians for which the necessary conditions can be asserted. A simple example is provided (for $n = 1$) by

$$\Lambda_t(x, v) = \exp\left\{(1 + x^2 + t^2)v^2\right\}.$$

This Lagrangian satisfies the generalized Tonelli-Morrey growth conditions, as well as the hypotheses of the classical Tonelli existence theorem. Thus Theorem 4.3.2 (necessary conditions) and Theorem 4.5.2 (regularity) can be applied to it. However, as for the previous example, it fails to be encompassed by any class of Lagrangians previously considered in the literature.

§4.4 The necessary conditions of Theorem 4.4.1 appear to be the most general on record for the well-studied case of a finite-valued Lagrangian, one which is neither convex nor continuous in the velocity variable. In particular, the results of Clarke [13] and Ioffe and Rockafellar [36] (see also Vinter [58]) are subsumed by the theorem and extended in several ways. A weaker notion of local minimum is used, and weaker boundedness and reduced Lipschitz hypotheses imposed. In further contrast, the necessary conditions are obtained in stratified form: for each radius R for which the hypotheses are valid, and not just in the case where they hold for all R. Finally, the presence of the velocity constraint set V_t is admitted.

There exist other necessary conditions that are not subsumed by the above, however. They are obtained by invoking a generalized version of the maximum principle for optimal control problems (see Clarke [16]), and yield (together with the other necessary conditions) a *separated* form of the Euler inclusion:

$$\dot{p}(t) \in \partial_C \left\{\Lambda_t(\cdot, \dot{x}_*(t))\right\}(x_*(t)),\ p(t) \in \partial_P \left\{\Lambda_t(x_*(t), \cdot)\right\}(\dot{x}_*(t)) \text{ a.e.}$$

This separated Euler inclusion is different in general from the one given here, and the resulting necessary condition is sometimes less, and sometimes more, informative; examples appear in [36]. The maximum principles given in the next chapter similarly yield such separated conclusions.

§4.5 The article [28] of Clarke and Vinter is a principal reference for the issue of identifying structural hypotheses on the Lagrangian which give rise to Lipschitz regularity of the solution; see also the book of Vinter [58]. In regularity theory, it is the Lipschitz property that is the watershed: Once the solution is known to be Lipschitz, it is relatively simple to deduce enhanced regularity (for example, by assuming strict convexity of the Lagrangian: see [28]). The fact that the solution can fail to be Lipschitz even when the Lagrangian is a polynomial satisfying the hypotheses of Tonelli's existence theorem was proven by Ball and Mizel [1], who adapted a classical example of Maniá for this purpose.

The approach of Clarke and Vinter is based upon a generalization of Tonelli's theorem on regularity 'in the small', and gives rise to certain results of a type not obtained here. Theorem 4.5.2 subsumes their Tonelli-Morrey type result, however, which applies to Lagrangians satisfying a more restrictive growth condition (defined above in connection with §4.3) and which are locally Lipschitz in (x, v) and convex in v. In a similar vein, Theorem 4.5.5 extends their corresponding result by dispensing with these last two hypotheses. In connection with an 'indirect method' of existence and regularity, Clarke [24] had previously dispensed with Lipschitz behavior in x.

Theorem 4.5.4 appears to be new, in particular because it applies to a weak local minimum. An interesting open question is whether its hypothesis (4.4) could be replaced by the lesser requirement that $|D_x\Lambda_t(x_*(t), \dot{x}_*(t))|$ be integrable; if so, this would constitute a necessary and sufficient condition for the Euler equation to hold.

Chapter 5

Optimal Control of Vector Fields

The subject of necessary conditions in optimal control is dominated by the maximum principle of Pontryagin. In this chapter we shall prove three types of variants of this celebrated theorem. The first type (§5.1) exhibits a new stratified nature; here, we directly postulate the required pseudo-Lipschitz behavior of the data. Classical hypotheses of smoothness of the data (and boundedness of the control) provide special cases in which this behavior is guaranteed to be present, but in general it is of interest to have weaker structural conditions guaranteeing *a priori* that it will occur near the (possibly as yet unidentified) solution x_*. The two other types of maximum principles proven here have that feature. In §5.2 we give differential growth conditions having the desired effect. The third and final section is devoted to a maximum principle (Theorem 5.3.1) having a novel hybrid nature that admits unusually general cost integrands in the presence of fully nonlinear (but smooth) dynamics.

5.1 A stratified maximum principle

Let there be given a parametrized family

$$\mathcal{F} = \{f_u(t, x) : u \in \mathcal{U}\}$$

of nonautonomous vector fields. An arc x is said to be a trajectory of $\mathcal{F}$ if, for some $u \in \mathcal{U}$, one has

$$\dot{x}(t) = f_u(t, x(t)) \quad \text{a.e. } t \in [a, b].$$

We study in this section an optimal control problem defined over the trajectories of $\mathcal{F}$, deferring for the moment the discussion of the more familiar case of a standard control system.

The problem and basic hypotheses. We assume that each $f \in \mathcal{F}$ is a Carathéodory function on $[a,b] \times \mathbb{R}^n$, and that the family is *decomposable* (or closed under switching): if $f_1 \in \mathcal{F}$ and $f_2 \in \mathcal{F}$, and if Ω is a measurable subset of $[a,b]$, then the field f coinciding with f_1 for $t \in \Omega$ and with f_2 for t in the complement of Ω also belongs to $\mathcal{F}$.

The problem $\mathcal{P}_C$ consists of minimizing $\ell(x(a), x(b))$ over the trajectories of $\mathcal{F}$ that satisfy the boundary constraint $(x(a), x(b)) \in S$. We assume that ℓ is locally Lipschitz and S is closed.

Now let there be given a measurable radius function R on $[a,b]$ with values in $(0, +\infty]$, as well as an arc x_*, a trajectory of $\mathcal{F}$ corresponding to the vector field $f_* \in \mathcal{F}$. We suppose that x_* is a local $W^{1,1}$ minimum of radius R in the following sense: for some $\varepsilon_* > 0$, for every trajectory x of $\mathcal{F}$ that satisfies the boundary constraint and

$$|\dot{x}(t) - \dot{x}_*(t)| \le R(t) \text{ a.e., } \int_a^b |\dot{x}(t) - \dot{x}_*(t)| \, dt \le \varepsilon_*, \, \|x - x_*\|_\infty \le \varepsilon_*,$$

we have $\ell(x(a), x(b)) \ge \ell(x_*(a), x_*(b))$.

5.1.1 Theorem Suppose that each $f \in \mathcal{F}$ admits a number $\varepsilon > 0$ and a measurable function k (both possibly depending on f) such that, for almost every t,

$$\begin{aligned} &x, x' \in B(x_*(t), \varepsilon), \ |f(t,x) - \dot{x}_*(t)| \le R(t) \\ &\qquad\qquad \Longrightarrow \left| f(t,x') - f(t,x) \right| \le k(t) \left| x' - x \right|. \end{aligned}$$

Suppose also that the function $k_*(\cdot)$ corresponding to f_* is summable, and that we have

$$\text{ess inf } \left\{ \frac{R(t)}{k_*(t)} : t \in [a,b] \right\} > 0.$$

Then there exist an arc p on $[a,b]$ and a number $\lambda_0 \in \{0,1\}$ with

$$(p(t), \lambda_0) \ne 0 \quad \forall\, t \in [a,b]$$

such that the following transversality condition holds

$$(p(a), -p(b)) \in \lambda_0 \partial_L \ell(x_*(a), x_*(b)) + N_S^L(x_*(a), x_*(b)),$$

and such that p satisfies the adjoint inclusion:

$$-\dot{p}(t) \in \partial_C \langle p(t), f_*(t, \cdot)\rangle (x_*(t)) \text{ a.e.},$$

as well as the maximum condition of radius R: for every $f \in \mathcal{F}$, for almost every t,

$$|f(t, x_*(t)) - \dot{x}_*(t)| \le R(t) \Longrightarrow \langle p(t), f_*(t, x_*(t))\rangle \ge \langle p(t), f(t, x_*(t))\rangle .$$

Further, if x_* satisfies the above for a sequence of radius functions R_i such that

$$\liminf_{i\to\infty} R_i(t) = +\infty \text{ a.e.}$$

(with all associated parameters possibly depending on i), then there is an arc p satisfying all the conditions above as well as the *global* maximum condition:

$$\langle p(t), f_*(t, x_*(t))\rangle \ge \langle p(t), f(t, x_*(t))\rangle \text{ a.e. } \forall f \in \mathcal{F}. \qquad \blacklozenge$$

We remark that the generalized gradient ∂_C that appears in the adjoint inclusion above was defined in §1.3.

It is useful to have differential criteria for the pseudo-Lipschitz hypothesis of the theorem. The following is a straightforward consequence of Theorem 3.5.2. Recall that by Rademacher's theorem, a locally Lipschitz mapping from $\mathbb{R}^n$ to itself is differentiable almost everywhere.

5.1.2 Corollary Suppose that each $f \in \mathcal{F}$ admits a number $\varepsilon > 0$ and a measurable function k (both possibly depending on f) such that, for almost every t, the function $f(t, \cdot)$ is locally Lipschitz on the (open) set

$$f_t^{-1}\left[B(\dot{x}_*(t), R(t)\right] := \{x : |\dot{x}_*(t) - f(t, x)| < R(t)\}$$

and satisfies

$$\|D_x f(t, x)\| \le k(t) \text{ a.e. } x \in f_t^{-1}\left[B(\dot{x}_*(t), R(t)\right].$$

Suppose also that the function $k_*(\cdot)$ corresponding to f_* is summable, and that we have

$$\text{ess inf} \left\{ \frac{R(t)}{k_*(t)} : t \in [a, b] \right\} > 0.$$

Then the conclusions of Theorem 5.1.1 hold.

Standard control systems

Before proving the theorem, we pause to discuss the classical Pontryagin formulation of the optimal control problem, which involves the system of controlled differential equations

$$\dot{x}(t) = \phi_t(x(t), u(t)),\ u \in U(t),$$

where $\phi_t : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, and where the control set multifunction U is prescribed as well. Given a class $\mathcal{C}$ of functions mapping $[a,b]$ to $\mathbb{R}^m$, an *admissible control* $u(\cdot)$ is defined to be an element of $\mathcal{C}$ satisfying $u(t) \in U(t)$ a.e. We refer to the triple $(\phi, U, \mathcal{C})$ as a *standard control system.*

Note that we may define a vector field f_u for each admissible control u as follows:

$$f_u(t,x) := \phi_t(x, u(t)).$$

Taking as parameter set $\mathcal{U}$ the set of admissible controls, we obtain a parametrized family $\mathcal{F}$ of vector fields. The most common choice for $\mathcal{C}$ is the class of all measurable functions. Then, under the appropriate mild assumptions on ϕ and U, $\mathcal{F}$ will satisfy the basic hypotheses of the theorem (including decomposability). Other suitable choices for $\mathcal{C}$ include the class of bounded (or integrable) measurable functions.

Denoting by u_* the optimal control giving rise to x_*, the adjoint equation becomes:

$$-\dot{p}(t) \in \partial_C \langle p(t), \phi_t(\cdot, u_*(t)) \rangle (x_*(t)) \text{ a.e.}$$

and the Weierstrass condition (for a given radius R) implies (for almost every t):

$$u \in U(t),\ |\phi_t(x_*(t), u) - \dot{x}_*(t)| \le R(t) \Longrightarrow \\ \langle p(t), \phi_t(x_*(t), u_*(t)) \rangle \ge \langle p(t), \phi_t(x_*(t), u) \rangle .$$

Proof of Theorem 5.1.1. There is no loss of generality in taking $[a,b] = [0,1]$ and $x_* \equiv 0$. The proof follows that of Ioffe [35] in general outline, but relies upon Theorem 3.1.1. We fix a finite subfamily Σ of fields $f_1, \ldots, f_k$ in $\mathcal{F}$ and some $\delta > 0$, and define $F_t(x)$ as the union of $f_*(t,x)$ and those $f_i(t,x)$ in Σ for which

$$|f_i(t, x_*(t))| > \delta,\ k_i(t) < \frac{1}{\delta},$$

where k_i is the pseudo-Lipschitz function corresponding to f_i. Then F satisfies the basic hypotheses of Chapter 3.

Let R be the radius function for which the hypotheses of the theorem hold. It is easy to see (in light of the decomposability of $\mathcal{F}$) that x_* is a local solution (in the same sense) to the version of the problem $\mathcal{P}$ studied in Chapter 3 that corresponds to F (with ℓ and S unchanged). Furthermore, it follows readily that F is pseudo-Lipschitz of radius R near x_*. We apply Theorem 3.1.1 to deduce the existence of p, λ_0 as given there, normalizing to get

$$\lambda_0 + \|p\|_\infty = 1,$$

so that λ_0 is no longer equal to 0 or 1 necessarily. Because $\dot{x}_*(t) = 0$ is an isolated point of $F_t(x_*(t))$, the Euler inclusion implies the adjoint equation (this follows from the proximal normal inequality, and uses the fact that the generalized gradient is convex). The Weierstrass condition of radius R holds as follows: for almost every t, for any $f_i \in \Sigma$ with index i such that $k_i(t) < 1/\delta$, we have

$$\langle p(t), f_i(t, x_*(t)) \rangle \leq 0 \text{ if } \delta < |f_i(t, x_*(t))| \leq R(t).$$

Now the adjoint equation implies $|\dot{p}(t)| \leq k_*(t)\,|p(t)|$. A standard argument involving Gronwall's Lemma shows that the set Q of (normalized) (p, λ_0) satisfying this relationship is compact. If we denote by $M(\delta, \Sigma)$ the set of (p, λ_0) satisfying the transversality condition, the normalization condition, the adjoint equation, and the preceding Weierstrass condition for δ and Σ, then the sets $M(\delta, \Sigma)$ are closed subsets of Q having the finite intersection property. Compactness implies the existence of a single (p, λ_0) satisfying the transversality condition, the normalization condition, the adjoint equation, and the preceding Weierstrass condition for *every* $\delta > 0$ and finite subfamily of $\mathcal{F}$. This implies the Weierstrass condition as affirmed by the theorem. If $\lambda_0 = 0$, then (in view of the adjoint equation and Gronwall's Lemma) p is necessarily nonvanishing, whence nontriviality.

This completes the proof in the case of a fixed radius function R. If the hypotheses hold for a sequence R_i as described, the usual arguments allow us to extract a suitably convergent subsequence from the ensuing (normalized) (p_i, λ_{0_i}) whose limit gives the required conclusion. ∎

Boundary trajectories

Let us note for the record the more general version of Theorem 5.1.1 that corresponds to boundary trajectories rather than optimal ones. For this purpose we let C be a closed subset of $\mathbb{R}^n$ and x_* a trajectory of $\mathcal{F}$ on $[a, b]$ having $x_*(a) \in C$.

We consider (for a given radius function R as above) the set $\mathcal{T} = \mathcal{T}(x_*, R, \varepsilon_*, C, \mathcal{F})$ of trajectories x of $\mathcal{F}$ on $[a,b]$ which satisfy the constraints

$$x(a) \in C,\ |\dot{x}(t) - \dot{x}_*(t)| \leq R(t) \text{ a.e.}$$

and which are ε_*-close to x_* in the following $W^{1,1}$ sense:

$$\int_a^b |\dot{x}(t) - \dot{x}_*(t)|\, dt \leq \varepsilon_*,\ \|x - x_*\|_\infty \leq \varepsilon_*.$$

Let $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ be a given locally Lipschitz function. We assume below that x_* is a *local boundary trajectory* in the sense that $\Phi(x_*(b))$ is a boundary point of the set

$$\Phi_{\mathcal{T}} := \{\Phi(x(b)) : x \in \mathcal{T}\}.$$

5.1.3 Theorem Suppose that each $f \in \mathcal{F}$ admits a number $\varepsilon > 0$ and a measurable function k (both possibly depending on f) such that, for almost every t,

$$\begin{aligned} &x, x' \in B(x_*(t), \varepsilon),\ |f(t,x) - \dot{x}_*(t)| \leq R(t) \\ &\qquad\qquad \Longrightarrow \left|f(t,x') - f(t,x)\right| \leq k(t)\left|x' - x\right|. \end{aligned}$$

Suppose also that the function $k_*(\cdot)$ corresponding to f_* is summable, and that we have

$$\operatorname{ess\,inf}\left\{\frac{R(t)}{k_*(t)} : t \in [a,b]\right\} > 0.$$

Then there exist an arc p on $[a,b]$ and a unit vector γ in R^m such that the following transversality conditions hold

$$-p(b) \in \partial_L \langle \gamma, \Phi\rangle (x_*(b)),\ p(a) \in N_C^L(x_*(a)),$$

and such that p satisfies the adjoint equation:

$$-\dot{p}(t) \in \partial_C \langle p(t), f_*(t,\cdot)\rangle (x_*(t)) \text{ a.e.,}$$

as well as the maximum condition of radius R: for every $f \in \mathcal{F}$, for almost every t,

$$|f(t, x_*(t)) - \dot{x}_*(t)| \leq R(t) \Longrightarrow \langle p(t), f_*(t, x_*(t))\rangle \geq \langle p(t), f(t, x_*(t))\rangle.$$

Further, if x_* satisfies the above for a sequence of radius functions R_i such that

$$\liminf_{i\to\infty} R_i(t) = +\infty \text{ a.e.}$$

(with all associated parameters possibly depending on i), then there is an arc p satisfying all the conditions above as well as the *global* maximum condition:

$$\langle p(t), f_*(t, x_*(t))\rangle \geq \langle p(t), f(t, x_*(t))\rangle \quad \text{a.e. } \forall f \in \mathcal{F}. \qquad \blacklozenge$$

We omit the proof, which is analogous to that of Theorem 5.1.1 but based upon Theorem 2.3.3.

5.2 Indirect criteria for pseudo-Lipschitzness

In the previous section the pseudo-Lipschitz behavior of the system near the optimal arc was a direct hypothesis. Now we wish to consider *a priori* conditions on the system which, with minimal reference to a given arc, guarantee that such pseudo-Lipschitz behavior will be present, and that the full set of necessary conditions will hold in global form.

The following gives a verifiable criterion for this.

5.2.1 Theorem Let x_* be a $W^{1,1}$ local minimum for $\mathcal{P}_C$. Suppose that each vector field $f \in \mathcal{F}$ admits positive constants ε and c and a summable function d (all possibly depending on f) such that, for almost every t, $f(t, \cdot)$ is locally Lipschitz on $B(x_*(t), \varepsilon)$ and satisfies

$$\|D_x f(t,x)\| \leq c\,|f(t,x)| + d(t), \quad \text{a.e. } x \in B(x_*(t), \varepsilon).$$

Then the global conclusions of Theorem 5.1.1 hold. $\blacklozenge$

Proof. There is no loss of generality in taking $[a,b] = [0,1]$ and $x_* \equiv 0$. Let us relabel f_0 the vector field corresponding to x_*. We fix a finite subfamily of fields $f_1, \ldots, f_k$ in $\mathcal{F}$ and some $\delta > 0$, and define the multifunction $F_t(x)$ as the union of $f_0(t,x)$ and those $f_i(t,x)$ for which

$$|f_i(t, x_*(t))| > \delta.$$

Then F satisfies the basic hypotheses. It is clear, as in the proof of Theorem 5.1.1, that in order to prove the present theorem, it suffices to prove the existence of a (global) multiplier for F in the sense of Theorem 3.1.1. We shall achieve this by an appeal to Corollary 3.5.3.

Let ε_i, c_i and d_i be the parameters corresponding to f_i $(i = 0, \ldots, k)$, and set

$$\bar{\varepsilon} := \min_{0 \leq i \leq k} \varepsilon_i, \;\; \bar{c} := \max_{0 \leq i \leq k} c_i, \;\; \bar{d}(t) := \max_{0 \leq i \leq k} d_i(t).$$

We define, for any positive integer N, the radius function

$$R_N(t) := \left\{ |f_0(t, x_*(t))| + \bar{d} \right\} N.$$

Fix $t \in [a, b]$ and let $x \in B(x_*(t), \bar{\varepsilon})$ be given, together with a point $v \in F_t(x) \cap B(\dot{x}_*(t), R_N(t))$ and any element (α, β) of $N^P_{G(t)}(x, v)$. Then v is of the form $f_i(t, x)$, and the proximal normal inequality implies that the expression

$$-\langle \alpha, x' \rangle - \langle \beta, f_i(t, x') \rangle + \sigma \left\{ |x' - x|^2 + |f_i(t, x') - f_i(t, x)|^2 \right\}$$

attains a local minimum at $x' = x$. This gives

$$-\alpha \in \partial_P \langle \beta, f_i(t, \cdot) \rangle (x),$$

which, combined with the growth condition on f_i leads to

$$\begin{aligned} |\alpha| &\leq \left\{ \bar{c}\, |f_i(t, x)| + \bar{d}(t) \right\} |\beta| \\ &\leq \left\{ \bar{c}\, [|f_0(t, x_*(t))| + R_N(t)] + \bar{d}(t) \right\} |\beta| \\ &= k_N(t)\, |\beta|, \end{aligned}$$

where

$$k_N(t) := \bar{c} \left\{ |f_0(t, x_*(t))| (1 + N) + \bar{d}(t) N \right\} + \bar{d}(t).$$

Since k_N is summable, this confirms the bounded slope condition of Corollary 3.5.3; note also that the ratio $R_N(t)/k_N(t)$ is bounded away from zero. We obtain therefore the conclusions of Theorem 3.1.1 for each N. Finally, note that if we assume (as we may) that $\bar{d}(t) \geq 1$, then R_N goes to infinity with N in the required sense, giving the required global conclusion. ■

5.2.2 Remark When the family $\mathcal{F}$ of vector fields arises from a standard control system $(\phi, U, \mathcal{C})$, diverse sets of conditions on ϕ, U and $\mathcal{C}$ can be formulated to give rise to the growth hypothesis postulated in the theorem. The simplest case is to assume that ϕ is smooth and U is uniformly bounded. To give but one more example, the following will also do: take $\mathcal{C}$ to be the class of essentially bounded functions, and let ϕ be locally Lipschitz in x and satisfy

$$\|D_x \phi_t(x, u)\| \leq c\, |\phi_t(x, u)| + \theta(t, x, u) \quad \text{a.e. } x,$$

for some constant c and continuous function θ.

5.3 A hybrid maximum principle

The two preceding sections have both considered the Mayer form of the optimal control problem in which the cost is a function of the endpoint values of the state x. Many applications of standard control systems, however, involve an integral cost functional $\Lambda_t(x, u)$. When Λ satisfies the same hypotheses as the dynamics, it is a simple matter (as we have seen in Chapter 2) to absorb the integral cost into the dynamics by introducing an extra state variable. In this section we study an optimal control problem in which the cost integrand is allowed to be much less regular than the dynamics; in fact, we allow it the fullest generality treated in the previous chapter. On the other hand, we postulate smooth dynamics, in contrast to the previous sections. The new hybrid result so obtained encapsulates in a single setting the necessary conditions of the classical calculus of variations, the classical maximum principle of Pontryagin, Theorem 4.3.2 for the generalized problem of Bolza, and even multiplier rules for problems incorporating mixed state/control constraints. It is notable for not requiring Lipschitz hypotheses that refer explicitly to the optimal arc.

The problem and basic hypotheses. We consider the minimization of the functional

$$\ell(x(a), x(b)) + \int_a^b \Lambda_t(x(t), u(t))\, dt$$

subject to the boundary conditions $(x(a), x(b)) \in S$ and the standard control dynamics

$$\dot{x}(t) = \phi_t(x(t), u(t)) \text{ a.e.}$$

The minimization takes place with respect to arcs x and measurable functions $u : [a, b] \to \mathbb{R}^m$. Note that no explicit constraints are placed upon $u(t)$; if such constraints exist, they are accounted for by assigning to the extended-valued integrand Λ the value $+\infty$ whenever the constraints are violated. We assume as in Chapter 4 that Λ is $\mathcal{L} \times \mathcal{B}$ measurable and lower semicontinuous in (x, u). As usual, ℓ is taken to be locally Lipschitz and S closed.

The growth conditions. We assume that the function ϕ is Lebesgue measurable in t, continuously differentiable in (x, u), and that ϕ and Λ satisfy the following: for every bounded subset X of $\mathbb{R}^n$, there exist a constant c and a summable function d such that, for almost every t, for every $(x, u) \in$

$\operatorname{dom} \Lambda_t(\cdot, \cdot)$ with $x \in X$, we have

$$\|D_x \phi_t(x,u)\| \leq c\,\{|\phi_t(x,u)| + \Lambda_t(x,u)\} + d(t),$$

and for all (ζ, ψ) in $\partial_P \Lambda_t(x,u)$ (if any) we have

$$\frac{|\zeta|\,(1 + \|D_u \phi_t(x,u)\|)}{1 + |\psi|} \leq c\,\{|\phi_t(x,u)| + \Lambda_t(x,u)\} + d(t).$$

We also assume that for almost every t, the function Λ_t is bounded below on $X \times \mathbb{R}^m$.

In the following, $(*)$ denotes evaluation at $(x_*(t), u_*(t))$.

5.3.1 Theorem Let the control function u_* give rise to an arc x_* which is a $W^{1,1}$ local minimum for the problem above. Then there exist an arc p on $[a,b]$ and a number $\lambda_0 \in \{0,1\}$ with

$$(p(t), \lambda_0) \neq 0 \quad \forall\, t \in [a,b]$$

such that the following transversality condition holds

$$(p(a), -p(b)) \in \lambda_0 \partial_L \ell(x_*(a), x_*(b)) + N_S^L(x_*(a), x_*(b)),$$

and such that p satisfies the hybrid adjoint inclusion: $-\dot{p}(t)$ belongs almost everywhere to the set

$$\operatorname{co}\left\{\omega : (\omega + D_x^* \phi_t(*) p(t), D_u^* \phi_t(*) p(t)) \in \partial_L^{\lambda_0} \Lambda_t(*)\right\},$$

as well as the maximum condition: for almost every t, for every u in $\operatorname{dom} \Lambda_t(x_*(t), \cdot)$, one has

$$\langle p(t), \phi_t(x_*(t), u) - \phi_t(*)\rangle \leq \lambda_0 \Lambda_t(x_*(t), u) - \lambda_0 \Lambda_t(*). \qquad \blacklozenge$$

Proof. Our strategy is to appeal to Corollary 3.5.5, much as was done in the proof of Theorem 4.3.2. Fix $\theta \in (0,1)$, and take F to be the multifunction

$$F_t(x,y) := \{(\phi_t(x,u), \Lambda_t(x,u) + \theta\,|u - u_*(t)| + \delta) : \delta \geq 0, u \in \mathbb{R}^m\},$$

which satisfies the basic hypotheses. We consider the minimization of

$$\ell(x(a), x(b)) + y(b)$$

over the trajectories (x, y) of F satisfying $y(a) = 0$ and $(x(a), x(b)) \in S$. A local minimum in the $W^{1,1}$ sense is furnished by the arc (x_*, y_*), where

$$y_*(t) := \int_a^t \Lambda_s(x_*(s), u_*(s))\, ds.$$

We proceed now to verify the bounded slope condition of Corollary 3.5.5. Accordingly, let X be an open bounded set containing the points $x_*(t), t \in [a, b]$, and let $(\alpha, \beta, -\gamma)$ be an element of

$$N^P_{G(t)}(x, v, \Lambda_t(x, u) + \theta\,|u - u_*(t)| + \delta),$$

where $x \in X$. Just as in the proof of Theorem 4.3.2, only the case $\gamma > 0, \delta = 0$ need be considered. The proximal subgradient inequality then asserts that for some $\sigma > 0$, the following function of (x', u') has a local minimum at (x, u):

$$\gamma\Lambda_t(x', u') + \gamma\theta\,\left|u' - u_*(t)\right| - \left\langle \beta, \phi_t(x', u')\right\rangle - \left\langle \alpha, x'\right\rangle + \\ \sigma\left\{\left|x' - x\right|^2 + \left|\phi_t(x', u') - \phi_t(x, u)\right|^2\right\}$$

The first-order necessary conditions for this minimum yield

$$(\alpha + D_x^*\phi_t(x, u)\beta, \gamma\theta\mu + D_u^*\phi_t(x, u)\beta) \in \partial_L\gamma\Lambda_t(x, u),$$

for some μ in the unit ball. In view of the second growth condition this gives rise to

$$|\alpha + D_x^*\phi_t(x, u)\beta| \leq \\ \frac{c\,[\{|\phi_t(x, u)| + \Lambda_t(x, u)\} + d(t)]\,[\gamma + |\gamma\theta\mu + D_u^*\phi_t(x, u)\beta|]}{1 + \|D_u\phi_t(x, u)\|}.$$

When combined with the first growth condition, this leads to

$$|\alpha| \leq 4\,[c\,\{|\phi_t(x, u| + \Lambda_t(x, u) + \theta\,|u - u_*(t)|\} + d(t)]\,|(\beta, \gamma)|,$$

which confirms the bounded slope condition of Corollary 3.5.5. Applying that result, we obtain an arc of the form $(p, -\lambda_0)$ which satisfies (along with the other conditions) the Euler inclusion almost everywhere at the point

$$(x_*(t), \dot{x}_*(t), \Lambda_t(x_*(t), u(t)) + \theta\,|u(t) - u_*(t)|),$$

for some $u(t)$ with $\dot{x}_*(t) = \phi_t(x_*(t), u(t))$ a.e. and

$$\Lambda_t(x_*(t), u(t)) + \theta\,|u(t) - u_*(t)| = \Lambda_t(x_*(t), u_*(t)) \text{ a.e.}$$

But then $u(t) = u_*(t)$ a.e., for otherwise the control u would generate x_* while giving a strictly lower cost. Proximal analysis of the ensuing Euler inclusion along the lines of the argument given above yields the following: $-\dot{p}(t)$ belongs almost everywhere to the set

$$\text{co}\left\{\omega : (\omega + D_x^*\phi_t(*)p(t), \lambda_0\theta\mu(t) + D_u^*\phi_t(*)p(t)) \in \partial_L^{\lambda_0}\Lambda_t(*)\right\}.$$

Note the presence of an unwanted term that would vanish if θ were zero. The maximum condition is similarly perturbed:

$$\langle p(t), \phi_t(x_*(t), u) - \phi_t(*)\rangle \leq \lambda_0\Lambda_t(x_*(t), u) - \lambda_0\Lambda_t(*) + \lambda_0\theta\,|u - u_*(t)|\,.$$

We consider the above for a sequence θ_i decreasing to 0 and apply the usual sequential compactness arguments to the resulting (p_i, λ_{0_i}) to arrive at the required conclusions for $\theta = 0$. ■

Some special cases

(a) A special case of the theorem is that in which for each t, Λ_t is the indicator of a closed set $U(t)$; this amounts to imposing the condition $u(t) \in U(t)$ on the admissible controls u. Note that in this case the second growth condition is trivially satisfied (since $\zeta = 0$), and the first growth condition is active only on $U(t)$. The hybrid adjoint inclusion implies

$$-\dot{p}(t) = D_x^*\phi_t(x_*(t), u_*(t))p(t),$$

and we recover the conclusions of the classical maximum principle.

(b) When we take $\phi_t(x, u) = u$ the first growth condition is trivially satisfied and the second coincides with the generalized Tonelli-Morrey growth condition 4.3.1; thus we recover Theorem 4.3.2 for the generalized problem of Bolza.

(c) Consider the optimal control problem in Mayer form in the presence of mixed state/control pointwise constraints $(x(t), u(t)) \in \Omega$ a.e. for a given closed set Ω. We encompass this by taking $\Lambda_t = I_\Omega$. Then the second growth condition reduces to the following geometric assumption: for every $(x, u) \in \Omega$, for every $(\zeta, \psi) \in N_\Omega^P(x, u)$, one has

$$\frac{|\zeta|\,(1 + \|D_u\phi_t(x, u)\|)}{|\psi|} \leq c\,|\phi_t(x, u)| + d(t).$$

By taking a suitable representation for Ω in terms of functional equalities and/or inequalities, sufficient conditions in terms of rank can be adduced which imply this property, leading to explicit multiplier rules (see §4.2). With an appropriate 'transversal intersection' condition, we can also treat the case in which *both* the constraints $(x(t), u(t)) \in \Omega$ and $u(t) \in U(t)$ are present.

5.4 Notes

§5.1 A general nonsmooth maximum principle for standard control systems was first proved by Clarke [14, 16]. Theorem 5.1.1 subsumes that result as well as the recent ones of Ioffe [35] and Vinter [58] (Theorem 6.2.1). For the more general context of parametrized families of vector fields, we recover the results of Kaskosz and Lojasiewicz [38, 39], who are responsible for some of the early work on generalized control systems. All of the foregoing are also extended to the new pseudo-Lipschitz and stratified setting. Note that, as a corollary of the theorem, we obtain the adjoint equation (and a 'small' maximum condition) when x_* provides a classical weak local minimum for the problem; this appears to be a new result at this level of generality (see also [63]).

There is a voluminous literature associated to the maximum principle, which has also been obtained in 'axiomatic' fashion, in which the existence of certain approximations is postulated (Dubovitskii and Milyutin, Halkin, Neustadt, Sussmann, Warga) . These general approaches may be of use in treating other types of nonstandard problems not discussed here, for example: hybrid control, systems whose states lie in a Banach space, or problems incorporating state constraints. We refer the reader to the discussion in Vinter [58] (pp. 228-231) for further details.

§§5.2-5.3 We are not aware of precedents for the results of these sections, and in particular for the hybrid maximum principle (Theorem 5.3.1). It appears useful to us to have versatile and verifiable criteria such as these for the maximum principle to hold.

Chapter 6

The Hamiltonian inclusion

Let us return again to the classical setting of the basic problem in the calculus of variations. The *Legendre transform* associates to the Lagrangian $\Lambda_t(x,v)$ a new conjugate function $H_t(x,p)$ called the *Hamiltonian*, by means of the following formula: we define

$$H_t(x,p) := \langle p, v\rangle - \Lambda_t(x,v),$$

where $v = v(t,x,p)$ is the solution of the equation

$$p = D_v\Lambda_t(x,v).$$

Of course it is necessary to impose certain hypotheses for this transfomation to make sense. For example, one may require that Λ be of class C^2, and that for some $\varepsilon > 0$, the matrix $D_{vv}\Lambda_t(x,v) - \varepsilon I$ be positive definite for all (x,v). This special class covers many applications in mechanics.

There is a rich classical theory of Hamiltonian methods, in connection with necessary conditions among other things. Of central importance in the theory is the Hamiltonian system of differential equations:

$$(-\dot p(t), \dot x(t)) = D_{x,p}H_t(x(t), p(t))$$

which, for the special class alluded to above, is equivalent to the Euler equation

$$(\dot p(t), p(t)) = D_{x,v}\Lambda_t(x(t), \dot x(t)).$$

Instead of defining the Hamiltonian as was done above, one can apply to the Lagrangian the *Fenchel-Moreau transform* . It yields a Hamiltonian defined by

$$H_t(x,p) := \sup_{v\in\mathbb{R}^n}\left\{\langle p, v\rangle - \Lambda_t(x,v)\right\}.$$

It is easy to check that for the special class defined above, the two Hamiltonians coincide.

The advantage of the new definition is that it can be applied to a much larger class of functions, including nonsmooth and extended-valued ones. For example, consider the case in which Λ_t is the indicator function of $G(t)$, the graph of the multifunction F_t. Then the Fenchel-Moreau formula leads to the Hamiltonian naturally associated to the multifunction F:

$$H_t(x,p) := \sup \{\langle p, v\rangle : v \in F_t(x)\}.$$

For the purposes of deriving necessary conditions associated to the differential inclusion problem, Clarke [11, 19] introduced an analogue of the classical Hamiltonian system, the following Hamiltonian inclusion $\mathcal{H}_C$:

$$(-\dot{p}(t), \dot{x}(t)) \in \partial_C H_t(x(t), p(t)) \text{ a.e.},$$

where, as we have seen, ∂_C refers to the generalized gradient of H (with respect to (x,p)). It was proved that, under certain conditions, the existence of an arc p satisfying $\mathcal{H}_C$ is a necessary condition for optimality.

In parallel to the study of potentially more refined versions of the Euler inclusion, there have been proposals to refine $\mathcal{H}_C$. Rockafellar and others [44, 53, 58] have studied the following relation $\mathcal{H}_L$:

$$-\dot{p}(t) \in \text{co}\,\{\omega : (\omega, \dot{x}(t)) \in \partial_L H_t(x(t), p(t))\} \text{ a.e.}$$

It is easy to see that $\mathcal{H}_L$ implies $\mathcal{H}_C$; the opposite is false. In a setting more general than that of a differential inclusion, but under a convexity hypothesis, Rockafellar proved a significant 'dualization' result: the Euler inclusion (in the sense of the previous chapters) implies the Hamiltonian inclusion $\mathcal{H}_L$. (See also Bessis, Ledyaev and Vinter [3] or Vinter [58] for a proof of this based upon standard tools of proximal analysis.)

In the present context, the required convexity hypothesis is simply that F be convex valued. It follows that generally speaking, whenever we can assert the Euler inclusion (as in Theorem 3.1.1, say), and if F is convex-valued, then the Hamiltonian inclusion $\mathcal{H}_L$ holds as well.

In the general case, when F may fail to have convex values, the situation is not completely clear. It is easily shown that $\mathcal{H}_L$ (or $\mathcal{H}_C$ for that matter) implies the global Weierstrass condition, which we know under rather general circumstances to be a necessary condition in tandem with the Euler inclusion. But the necessity of the Hamiltonian inclusion itself, for example in connection with the optimal control problem $\mathcal{P}$ of Chapter 3, has been established only in the presence of certain constraint qualifications bearing upon the nature of the boundary constraints or the behavior of the problem with respect to perturbations, and this only for multifunctions F which are bounded and Lipschitz. The question of whether the Hamiltonian inclusion $\mathcal{H}_C$ is a necessary condition of optimality for a strong local minimum, when F is bounded and Lipschitz but not convex-valued, and in the absence of any constraint qualification, is a longstanding open one in the subject.

We settle this question affirmatively below. That of the necessity of $\mathcal{H}_L$ (in this same basic setting, and all the more for unbounded multifunctions, and for weaker types of local minima) remains open, however.

6.1 Boundary trajectories

Let F be a multifunction from $[a,b] \times \mathbb{R}^n$ to $\mathbb{R}^n$ and C a closed subset of $\mathbb{R}^n$. We suppose as usual that F is $\mathcal{L} \times \mathcal{B}$-measurable and that for each t, $F_t(\cdot)$ has closed graph.

Now let the arc x_* be a trajectory of F on $[a,b]$ having $x_*(a) \in C$, and, for a given $\varepsilon > 0$, consider the set $\mathcal{T}$ of trajectories x of F on $[a,b]$ which satisfy $x(a) \in C$ and which are uniformly ε-close to x_*: $\|x - x_*\|_\infty \leq \varepsilon$. Let $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ be a given locally Lipschitz function. We shall assume that x_* is a local boundary trajectory in the sense that $\Phi(x_*(b))$ is a boundary point of the set

$$\Phi_{\mathcal{T}} := \{\Phi(x(b)) : x \in \mathcal{T}\}.$$

6.1.1 Theorem Let F be integrably Lipschitz and bounded near x_* as follows: there exists a summable function k such that for almost every t in $[a,b]$, we have

$$\begin{aligned} x \in \overline{B}(x_*(t),\varepsilon),\ v \in F_t(x) \quad &\Longrightarrow \quad |v| \leq k(t); \\ x, x' \in \overline{B}(x_*(t),\varepsilon) \quad &\Longrightarrow \quad F_t(x) \subset F_t(x') + k(t)\left|x' - x\right| \overline{B}. \end{aligned}$$

Then, if x_* is a local boundary trajectory in the above sense, there exist an arc p on $[a,b]$ and a unit vector γ in R^m such that the following transversality conditions hold

$$-p(b) \in \partial_L \langle \gamma, \Phi\rangle (x_*(b)),\ p(a) \in N_C^L(x_*(a)),$$

and such that p satisfies the Hamiltonian inclusion

$$(-\dot{p}(t), \dot{x}_*(t)) \in \partial_C H_t(x_*(t), p(t)) \ \text{ a.e. } t \in [a,b]$$

and the global Weierstrass condition. ♦

Necessary conditions for optimality

As in Chapter 3 the theorem leads to necessary conditions for the problem $\mathcal{P}$ of minimizing $\ell(x(a), x(b))$ over the trajectories of F satisfying $(x(a), x(b)) \in S$. We suppose as before that ℓ is locally Lipschitz and S is closed. Let x_* be a strong local minimum for the problem, and assume that F has the same Lipschitz and boundedness properties as in the theorem.

6.1.2 Corollary There exist an arc p and a number λ_0 in $\{0,1\}$ satisfying the nontriviality condition

$$(\lambda_0, p(t)) \neq 0 \quad \forall\, t \in [a,b]$$

and the transversality condition:

$$(p(a), -p(b)) \in \partial_L \lambda_0 \ell(x_*(a), x_*(b)) + N_S^L(x_*(a), x_*(b)),$$

as well as the global Weierstrass condition and the Hamiltonian inclusion

$$(-\dot{p}(t), \dot{x}_*(t)) \in \partial_C H_t(x_*(t), p(t)) \ \text{ a.e. } t \in [a,b]. \qquad \blacklozenge$$

Proof of Theorem 6.1.1

We sketch the proof of the theorem, which uses the necessary conditions of Chapter 2, but also a relaxation theorem, as well as the dualization theorem of Rockafellar cited above.

By the same transformation device used at the beginning of the proof of Theorem 2.1.1, we can suppose that k is identically 1. Again we take $x_* \equiv 0$ and $[a,b] = [0,1]$, and we suppose that $F_t(x)$ is defined for all x in R^n, that the bound on F and the Lipschitz condition hold globally on R^n, and that C is compact.

Since $\Phi(0)$ is a boundary point of the set $\Phi_{\mathcal{T}}$, there exists a sequence of points μ_i not lying in $\Phi_{\mathcal{T}}$ and converging to $\Phi(0)$. We set

$$\varepsilon_i := |\mu_i - \Phi(0)| > 0.$$

Consider now the problem of minimizing

$$|\mu_i - \Phi(x(b))| + \int_0^1 |x(t)|^2 \, dt$$

over the arcs x on $[0,1]$ satisfying

$$\dot{x} \in F_t(x) \ \text{ a.e.}, \ x(0) \in C. \tag{6.1}$$

We view this problem as one in which the choice variables are $\dot{x} \in L^2$ and $x(0) \in \mathbb{R}^n$. The set of feasible $\dot{x}, x(0)$ defined by (6.1) is closed and bounded in $L^2 \times \mathbb{R}^n$, and we may apply Stegall's variational principle (see [27]) to deduce the existence of elements $\alpha_i \in L^2$ and $\beta \in \mathbb{R}^n$ with

$$\|\alpha_i\|_2 < \varepsilon_i, \ |\beta_i| < \varepsilon_i$$

such that the problem of minimizing

$$|\mu_i - \Phi(x(b))| + \langle \beta_i, x(0) \rangle + \int_0^1 |x(t)|^2 \, dt + \int_0^1 \langle \alpha_i(t), \dot{x}(t) \rangle \, dt \tag{6.2}$$

subject to (6.1) admits a solution x_i. Since $x_* = 0$ assigns the value ε_i to the cost functional of this problem, we deduce

$$|\mu_i - \Phi(x_i(b))| + \langle \beta_i, x_i(0) \rangle + \int_0^1 |x_i(t)|^2 \, dt + \int_0^1 \langle \alpha_i(t), \dot{x}_i(t) \rangle \, dt \leq \varepsilon_i.$$

It follows from this that $\int_0^1 |x_i(t)|^2 \, dt$ converges to 0, so that for i sufficiently large we have $\|x_i\|_\infty < \varepsilon$. In consequence, we have

$$|\mu_i - \Phi(x_i(b))| \neq 0$$

for all large i (since μ_i does not lie in $\Phi_{\mathcal{T}}$). By passing to a subsequence if necessary, we can assume that x_i converges uniformly (by Arzela-Ascoli), $\dot{x}_i$ converges weakly to a limit v_0 and $x_i(0)$ converges to x_0. If y is the arc defined by $\dot{y} = v_0$, $y(0) = x_0$, it follows that x_i converges uniformly to y. But the fact that $\int_0^1 |x_i(t)|^2 \, dt$ converges to 0 forces $y = 0$, so in fact we conclude that x_i converges uniformly to 0 and $\dot{x}_i$ weakly to 0. We may also suppose that α_i converges almost everywhere to 0.

Because F is uniformly bounded and Lipschitz, and because the only term involving $\dot{x}$ in (6.2) is linear, and since there is no constraint on $x(1)$, we can apply a standard relaxation theorem[1] to deduce that x_i continues to solve the problem if in (6.1) we replace $F_t(x)$ by its convex hull $\hat{F}_t(x) := \operatorname{co} F_t(x)$.

The next step is to write necessary conditions of optimality for x_i as a solution of the relaxed problem. By the standard device of absorbing the integral terms in the inclusion (state augmentation: see for example the proof of Corollary 2.2.1), we obtain a problem to which Theorem 2.1.1 applies. The new multifunction is

$$\Gamma_t(x, y) := \left\{ (v, \langle \alpha_i(t), v \rangle + |x|^2) : v \in \hat{F}_t(x) \right\}.$$

[1] See for example Theorem 2 of [12], or [19] p.117.

Since the problem is relaxed, the dualization theorem cited above allows us to express the necessary conditions in Hamiltonian terms; the Hamiltonian G corresponding to Γ is given by

$$G_t(x, y, p, q) := H_t(x, p + q\alpha_i(t)) + q\,|x|^2\,,$$

and we may write the Hamiltonian inclusion with $\partial_C G$. Routine analysis of the inclusion and the transversality conditions leads (for i sufficiently large) to an arc p_i satisfying

$$(-\dot{p}_i(t) + 2x_i(t), \dot{x}_i(t)) \in \partial_C H_t(x_i(t), p_i(t) - \alpha_i(t)) \text{ a.e.}$$

and

$$p_i(0) - \beta_i \in N_C^L(x_i(0)),\ -p_i(1) \in \partial_L \langle \gamma_i, \Phi\rangle (x_i(1))$$

where γ_i is a unit vector. We derive from this

$$|\dot{p}_i(t)| \leq |p_i(t) - \alpha_i(t)|\,,$$

which allows us to use standard limiting arguments to pass to the limit in the two preceding relations. The theorem follows, since the global Weierstrass condition is an easy consequence of the Hamiltonian inclusion. ■

6.2 Notes

§6.1 The results of this chapter fit into a longstanding and ongoing attempt to develop a fully Hamiltonian theory of dynamic optimization, to parallel the classical theory for the basic problem in the calculus of variations. (There is no classical Hamiltonian theory for constrained variational problems.) The project can be viewed as originating with Rockafellar's work on the existence question [51] for the problem of Bolza. Another step was the formulation of the Hamiltonian inclusion $\mathcal{H}_C$ and the attendant necessary conditions by Clarke [11][18][19][20]; there followed sufficient conditions by Zeidan [62]. Further progress was made by Loewen and Rockafellar [42], who derived under certain circumstances the existence of an arc p satisfying simultaneously the Euler and Hamiltonian inclusions, an issue later clarified by the dualization theory cited earlier. We also refer the reader to the article [44] of Loewen and Rockafellar for comments on the Hamiltonian formulation of the pseudo-Lipschitz hypothesis.

The conditions under which the Hamiltonian inclusion were previously known to be a necessary condition are discussed in [19]; see also Vinter [58] (pp. 286-7), and Ioffe [35].

Bibliography

[1] J. M. Ball and V. J. Mizel. One-dimensional variational problems whose minimizers do not satisfy the Euler-Lagrange equation. *Arch. Ration. Mech. Anal.*, 90:325–388, 1985.

[2] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations.* Birkhäuser, Boston, 1997.

[3] D. Bessis, Y. S. Ledyaev, and R. B. Vinter. Dualization of the Euler and Hamiltonian inclusions. *Nonlinear Anal.*, 43:861–882, 2001.

[4] V. I. Blagodatskih. Maximum principle for differential inclusions. *Trudy Mat. Inst. Steklov*, 166:23–43, 1984. In Russian; English translation in Proc. Steklov Inst. Math. 166 (1986).

[5] G. A. Bliss. *Lectures on the Calculus of Variations.* University of Chicago Press, 1946.

[6] V. G. Boltianski. Local section method in the theory of optimal processes. *Differentsial'nye Uravneniya*, 4:2166–2183, 1968. In Russian; English translation in Differential Equations 4 (1968) .

[7] J. M. Borwein and D. Preiss. A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions. *Trans. Amer. Math. Soc.*, 303:517–527, 1987.

[8] J. M. Borwein and Q. J. Zhu. A survey of subdifferential calculus with applications. *Nonlinear Anal.*, 38:687–773, 1999.

[9] G. Buttazzo, M. Giaquinta, and S. Hildebrandt. *One-dimensional Variational Problems*, volume 15 of *Oxford Lecture Series in Mathematics and its Applications.* Clarendon Press, 1998.

[10] L. Cesari. *Optimization–Theory and Applications*, volume 17 of *Applications of Mathematics.* Springer-Verlag, 1983.

[11] F. H. Clarke. *Necessary Conditions for Nonsmooth Problems in Optimal Control and the Calculus of Variations.* Doctoral thesis, University of Washington, 1973. (Thesis director: R. T. Rockafellar).

[12] F. H. Clarke. Admissible relaxation in variational and control problems. *J. Math. Anal. Appl.*, 51:557–576, 1975.

[13] F. H. Clarke. The Euler-Lagrange differential inclusion. *J. Differential Equations*, 19:80–90, 1975.

[14] F. H. Clarke. Maximum principles without differentiability. *Bulletin Amer. Math. Soc.*, 81:219–222, 1975.

[15] F. H. Clarke. The generalized problem of Bolza. *SIAM J. Control Optim.*, 14:682–699, 1976.

[16] F. H. Clarke. The maximum principle under minimal hypotheses. *SIAM J. Control Optim.*, 14:1078–1091, 1976.

[17] F. H. Clarke. Optimal solutions to differential inclusions. *J. Optim. Th. Appl.*, 19:469–478, 1976.

[18] F. H. Clarke. Extremal arcs and extended Hamiltonian systems. *Trans. Amer. Math. Soc.*, 231:349–367, 1977. Russian translation: Differentsialnye Uravnenia 13 (1977) 427-442.

[19] F. H. Clarke. *Optimization and Nonsmooth Analysis.* Wiley Interscience, New York, 1983. Republished as vol.5 of *Classics in Applied Mathematics*, SIAM, 1990.

[20] F. H. Clarke. Hamiltonian analysis of the generalized problem of Bolza. *Trans. Amer. Math. Soc.*, 301:385–400, 1987.

[21] F. H. Clarke. Regularity, existence and necessary conditions for the basic problem in the calculus of variations. In L. Cesari, editor, *Contributions to the modern calculus of variations*, London, 1987. Longman.

[22] F. H. Clarke. *Methods of Dynamic and Nonsmooth Optimization.* S.I.A.M., Philadelphia, 1989. Regional conference Series in Applied Mathematics vol. 57.

[23] F. H. Clarke. A decoupling principle in the calculus of variations. *J. Math. Anal. Appl.*, 172:92–105, 1993.

[24] F. H. Clarke. An indirect method in the calculus of variations. *Trans. Amer. Math. Soc.*, 336:655–673, 1993.

[25] F. H. Clarke. Nonsmooth analysis in control theory: a survey. *European J. Control*, 7:63–78, 2001.

[26] F. H. Clarke and Yu. S. Ledyaev. Mean value inequalities in Hilbert space. *Trans. Amer. Math. Soc.*, 344:307–324, 1994.

[27] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth Analysis and Control Theory.* Graduate Texts in Mathematics, vol. 178. Springer-Verlag, New York, 1998.

[28] F. H. Clarke and R. B. Vinter. Regularity properties of solutions to the basic problem in the calculus of variations. *Trans. Amer. Math. Soc.*, 289:73–98, 1985.

[29] R. P. Fedorenko. Maximum principle for differential inclusions (necessity). *Zhurnal Vyc. Matematiki i Matem. Fiziki*, 11:885–893, 1971.

[30] H. Frankowska. The maximum principle for an optimal solution to a differential inclusion with endpoint constraints. *SIAM J. Control Optim.*, 25:145–157, 1987.

[31] M. Giaquinta and S. Hildebrandt. *Calculus of Variations I.* Springer-Verlag, Berlin, 1996.

[32] H. H. Goldstine. *A History of the Calculus of Variations.* Springer-Verlag, New York, 1980.

[33] H. Halkin. Extremal properties of biconvex contingent equations. In *Ordinary Differential Equations (NRL-MRC Conf.).* Academic Press, New York, 1972.

[34] M. R. Hestenes. *Calculus of Variations and Optimal Control Theory.* Wiley, New York, 1966.

[35] A. D. Ioffe. Euler-Lagrange and Hamiltonian formalisms in dynamic optimization. *Trans. Amer. Math. Soc.*, 349:2871–2900, 1997.

[36] A. D. Ioffe and R. T. Rockafellar. The Euler and Weierstrass conditions for nonsmooth variational problems. *Calc. Var. Partial Differential Eq.*, 4:59–87, 1996.

[37] A. D. Ioffe and V. Tikhomirov. *Theory of Extremal Problems.* Nauka, Moscow, 1974. English translation, North-Holland (1979).

[38] B. Kaskosz and S. Lojasiewicz. A maximum principle for generalized control systems. *Nonlinear Anal.*, 9:109–130, 1985.

[39] B. Kaskosz and S. Lojasiewicz. Lagrange-type extremal trajectories in differential inclusions. *Syst. Control Letters*, 19:241–247, 1992.

[40] E. B. Lee and L. Markus. *Foundations of Optimal Control Theory.* Wiley, New York, 1967.

[41] P. D. Loewen. *Optimal Control via Nonsmooth Analysis*, volume 2 of *CRM Proc. & Lecture Notes.* American Mathematical Society, Providence, R.I., 1993.

[42] P. D. Loewen and R. T. Rockafellar. The adjoint arc in nonsmooth optimization. *Trans. Amer. Math. Soc.*, 325:39–72, 1991.

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

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

[45] E. McShane. On multipliers for Lagrange problems. *American J. Math.*, 61:809–819, 1939.

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

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

[48] C. B. Morrey. *Multiple Integrals in the Calculus of Variations.* Graduate Texts in Mathematics, vol. 178. Springer-Verlag, New York, 1966.

[49] L. S. Pontryagin, R. V. Boltyanskii, R. V. Gamkrelidze, and E. F. Mischenko. *The Mathematical Theory of Optimal Processes.* Wiley-Interscience, New York, 1962.

[50] R. T. Rockafellar. Generalized Hamiltonian equations for convex problems of Lagrange. *Pacific J. Math.*, 33:411–428, 1970.

[51] R. T. Rockafellar. Existence theorems for general control problems of Bolza and Lagrange. *Advances in Math.*, 15:312–333, 1975.

[52] R. T. Rockafellar. Extensions of subgradient calculus with applications to optimization. *Nonlinear Anal.*, 9:665–698, 1985.

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

[54] R. T. Rockafellar and R. Wets. *Variational Analysis.* Springer-Verlag, New York, 1998.

[55] G. V. Smirnov. Discrete approximations and optimal solutions to differential inclusions. *Kibernetika*, 10:76–79, 1991. In Russian; English translation in Cybernetics, 27 (1991), pp. 101-107.

[56] H. J. Sussmann. Geometry and optimal control. In J. Baillieul and J. C. Willems, editors, *Mathematical Control Theory*, pages 140–194, New York, 1999. Springer-Verlag.

[57] L. Tonelli. *Fondamenti del calcolo delle variazioni I,II.* Zanichelli, Bologna, 1921,1923.

[58] R. B. Vinter. *Optimal Control.* Birkhäuser, Boston, 2000.

[59] R. B. Vinter and H. Zheng. The extended Euler-Lagrange condition for nonconvex variational problems with state constraints. *Trans. Amer. Math. Soc.*, 350:1181–1204, 1998.

[60] J. Warga. *Optimal Control of Differential and Functional Equations.* Academic Press, New York, 1972.

[61] L. C. Young. *Lectures on the Calculus of Variations and Optimal Control Theory.* Saunders, Philadelphia, 1969.

[62] V. Zeidan. Sufficient conditions for the generalized problem of Bolza. *Trans. Amer. Math. Soc.*, 275:561–586, 1983.

[63] V. Zeidan. New second-order optimality conditions for variational problems with C^2-Hamiltonians. *SIAM J. Control Optim.*, 40:577–609, 2001.

[64] Q. J. Zhu. Necessary optimality conditions for nonconvex differential inclusions with endpoint constraints. *J. Differential Eq.*, 124:186–204, 1996.

Index