



6-2014

# Mathematical Methods of Analysis for Control and Dynamic Optimization Problems on Manifolds

Robert J. Kipka

Western Michigan University, [rob.kipka@gmail.com](mailto:rob.kipka@gmail.com)

Follow this and additional works at: <https://scholarworks.wmich.edu/dissertations>



Part of the [Dynamical Systems Commons](#)

## Recommended Citation

Kipka, Robert J., "Mathematical Methods of Analysis for Control and Dynamic Optimization Problems on Manifolds" (2014). *Dissertations*. 283.

<https://scholarworks.wmich.edu/dissertations/283>

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact [maira.bundza@wmich.edu](mailto:maira.bundza@wmich.edu).



MATHEMATICAL METHODS OF ANALYSIS FOR CONTROL AND  
DYNAMIC OPTIMIZATION PROBLEMS ON MANIFOLDS

by

Robert J. Kipka

A dissertation submitted to the Graduate College  
in partial fulfillment of the requirements  
for the Degree of Doctor of Philosophy  
Mathematics  
Western Michigan University  
June 2014

Doctoral committee:

Yuri S. Ledyev, Ph.D., Chair  
Qiji Jim Zhu , Ph.D.  
Jay Treiman, Ph.D.  
Anthony Bloch, Ph.D.

MATHEMATICAL METHODS OF ANALYSIS FOR CONTROL AND  
DYNAMIC OPTIMIZATION PROBLEMS ON MANIFOLDS

Robert J. Kipka, Ph.D.

Western Michigan University, 2014

Driven by applications in fields such as robotics and satellite attitude control, as well as by a need for the theoretical development of appropriate tools for the analysis of geometric systems, problems of control of dynamical systems on manifolds have been studied intensively during the past three decades. In this dissertation we suggest new mathematical techniques for the study of control and dynamic optimization problems on manifolds. This work has several components including: an extension of the classical Chronological Calculus to control and dynamical systems which are merely measurable in time and evolve on manifolds modeled over Banach space; novel proofs of Pontryagin Maximum Principle in settings more general than those currently existing in the literature; necessary optimality conditions for dynamic optimization problems on manifolds in which the dynamics are constrained by a differential inclusion; and a generic existence and uniqueness theorem for problems of optimal control posed on manifolds. Our studies of optimal control and dynamic optimization include exact penalization and metric regularity results for problems with initially and terminally constrained states which are new even in the case  $M = \mathbb{R}^n$ . This work also includes generalizations of the classical Chow-Rashevskii theorem from geometric control theory and the Fundamental and duBois-Reymond lemmas from classical Calculus of Variations to the setting of infinite-dimensional manifolds.

© 2014 Robert Kipka

## ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to the mathematics department at Western Michigan University, who have educated me in mathematics and supported my travel to San Diego, Baltimore, and Ravello. In particular, I wish to express my thanks to Rebecca Powers, Sue Simons, and Cathie Wilson for their invaluable assistance on everything nonmathematical; to Dr. Terrell Hodge, for her academic mentorship and friendship; and to Dr. Kirsty Eisenhart, who provided me with encouragement, mathematics texts, and clothes for my kids. My thanks also to my committee members, Drs. Jim Zhu, Jay Treiman, and Anthony Bloch for sharing their time and their expertise.

I am especially grateful to Dr. Yuri Ledyev for his professional mentorship; for very patiently teaching me applied analysis; and for sharing with me many cups of espresso during a lengthy dissertation process. He has been a strong advisor and a good friend. I am forever indebted.

My sincere thanks to my parents, Bud and Kris Kipka, for their support during our time in Kalamazoo and for the summers in Marquette. We would not have made it without you. Jen and Mike DeWaele were also always there for my family while we took our car to the shop or our son to the emergency room – I am very grateful. I wish also to thank my sister Nancy for being the person she is and for sharing with me a love of education. Finally, I wish to mention my grandfathers, Homer William Kuebler and Donn George Kipka, for their inspiration and advice. I hope they would be proud.

To my wife and my boys I owe everything. Thank you for your patience during the past six years.

Robert J. Kipka

For Lissa, who did everything else.

# Table of Contents

Acknowledgements	ii
Dedication	iii
Chapter 1. Introduction	1
1.1. Overview of Main Results	1
1.2. Background	10
Chapter 2. Extension of the Chronological Calculus	19
2.1. Extension of Chronological Calculus	20
2.2. Operator Differential Equations and Their Applications	30
2.3. Calculus of Little $o$ 's	37
2.4. Commutators of Flows and Vector Fields	40
2.5. Chow-Rashevskii Theorem for Infinite-Dimensional Manifolds	43
Chapter 3. Vector Fields, Flows, and Lagrangian Charts	49
3.1. Interpretations of Operator Formulae	49
3.2. Maps Into Vector Bundles	51
3.3. Flows Induced Through Pushforward or Pullback	55
3.4. Lagrangian Charts	63
Chapter 4. The Fundamental Lemma for Calculus of Variations and Optimal Control on Manifolds	70
4.1. Pontryagin Maximum Principle in Linear Spaces	74
4.2. Integration by Parts, Fundamental and duBois-Reymond Lemmas	96
4.3. Pontryagin Maximum Principle	102
4.4. Calculus of Variations	112

## Table of Contents - Continued

Chapter 5. Optimality Conditions via Nonsmooth Analysis	117
5.1. Relaxed Controls and Sliding Variations	120
5.2. Maximum Principle for Nonsmooth Mayer Problem with Free Terminal Point	124
5.3. Metric Regularity and Penalization in $\mathbb{R}^n$	127
5.4. Maximum Principle in $\mathbb{R}^n$ – Nonsmooth Mayer Problem with Terminal Constraints	137
5.5. Metric Regularity and Penalization for Manifolds	139
5.6. Maximum Principle on Manifolds – Nonsmooth Mayer Problem with Terminal Constraints	144
5.7. Smooth Constraints	146
Chapter 6. Generic Existence and Uniqueness of Optimal Trajectories in Geometric Optimal Control	150
6.1. The Value Function	153
6.2. Existence and Uniqueness for Mayer Problems	160
6.3. Existence and Uniqueness for Bolza Problems	167
6.4. Convergence Theorems	170
Chapter 7. Differential Inclusions	175
7.1. Necessary Optimality Conditions and Exact Penalization in $\mathbb{R}^n$	180
7.2. Necessary Optimality Conditions and Exact Penalization on General Manifolds	198
Bibliography	208



## CHAPTER 1

### Introduction

In this dissertation we develop techniques of analysis for problems of dynamic optimization and control in which the state evolves on a manifold, possibly of infinite dimension. Our intent is to present a framework in which such problems can be analyzed and to develop new results within this framework. To this end we present a generalization of the *Chronological Calculus*, a powerful computational technique for the study of nonlinear control systems first introduced by Agrachev and Gamkrelidze in 1978 [2]. We also suggest a new technique of *Lagrangian charts*, which is in close analogy with idea of Lagrangian coordinates in classical fluid dynamics [8]. These techniques, along with the recently developed theory of nonsmooth analysis for smooth manifolds [66], are used to develop results in the geometric theory of dynamic optimization regarding: global controllability; existence and uniqueness of optimal controls; and necessary optimality conditions. A central focus of this dissertation is the development of necessary optimality conditions for geometric problems and we devote three chapters to this topic.

#### 1.1. Overview of Main Results

This introductory chapter provides a brief overview of our main results and a short outline of background material.

**1.1.1. Extension of the Chronological Calculus.** In 1978, Agrachev and Gamkrelidze introduced in [2] the *Chronological Calculus* as a technique for the

study of nonlinear control systems. The central idea of this technique is to study nonlinear objects of control and dynamical systems as linear operators on a particular function space. The Chronological Calculus has been further developed in [3, 4, 47] and greatly simplifies many calculations of nonlinear control theory. Applications of the existing Chronological Calculus are many and can be found in studies of abnormal extremals [5]; averaging techniques for control systems [20, 78, 88]; controllability [6, 55, 85]; higher order necessary conditions [85, 86]; motion planning [83]; series expansions such as those of Volterra or Chen-Fleiss [56, 57, 58, 70, 88]; and stability [78].

However, the existing Chronological Calculus is limited in that: it is valid only for control systems on finite-dimensional manifolds; the dynamical systems considered must be  $C^\infty$ -smooth; controls should be piecewise continuous for the full calculus; and the function space in question is a Fréchet space – a feature which complicates a number of proofs. We introduce an extension of the Chronological Calculus which is valid for  $C^k$ -smooth Banach manifolds; applies to dynamical systems which are merely  $C^k$ -smooth and controls which are merely measurable; and replaces the Fréchet-space structure with a simplified remainder-term calculus.

The Chronological Calculus has classically been very useful in applications such flows of perturbed vector fields; Volterra series expansions of flows; and derivative of flows with respect to parameter. We expand each of these applications to flows of  $C^k$ -smooth vector fields on Banach manifolds.

A 1992 paper of Mauhart and Michor [68] defines a useful notion of *bracket of flows*: for flows  $P_t$  and  $Q_t$  of autonomous vector fields  $X, Y$  one may define

$$[P_t, Q_t] := Q_{-t} \circ P_{-t} \circ Q_t \circ P_t.$$

In the same paper the authors establish a formula relating brackets of flows on manifolds of infinite dimension to Lie brackets of their vector fields. The Chronological

Calculus greatly simplifies the study of Lie brackets of vector fields and associated flows and to demonstrate the effectiveness of our remainder term calculus we provide a new proof of their formula.

Lie brackets play an important role in the study of controllability of nonlinear systems and an important classical result for finite-dimensional nonlinear control problems is the *Chow-Rashevskii* theorem [22, 75]. This result provides sufficient conditions for global controllability of finite-dimensional, affine control systems in the absence of drift. The second chapter of this dissertation concludes by combining techniques of nonsmooth analysis with the bracket formula of Mauhart and Michor to establish the following generalization of the Chow-Rashevskii theorem:

**THEOREM.** *Consider a control system*

$$\dot{q}(t) = \sum_{i=1}^{\infty} u_i(t) V_i(q(t))$$

*evolving on a manifold  $M$  modeled over a smooth Banach space  $E$ . We suppose that at each time controls  $u_i(t)$  take values in  $\{-1, 0, 1\}$  and all but finitely many are zero.*

*Define a distribution  $\mathcal{L} \subset TM$  through*

$$\mathcal{L} = \text{span} \{[V_{i_1}, [V_{i_2}, [\dots, V_{i_k}]]] : k \in \mathbb{N}\}.$$

*If for each  $q \in M$ ,  $\mathcal{L}_q$  is dense in  $T_qM$ , then the system is globally approximately controllable.*

**1.1.2. Vector Fields, Flows, and Lagrangian Charts.** In the third chapter we provide a careful exposition of properties of vector field flows, as well as flows on  $TM$  and  $T^*M$  induced by their pushforward or pullback. In addition, we provide an introduction to the method of *Lagrangian chart* and establish some elementary properties of these charts.

**1.1.3. Necessary Optimality Conditions.** In Chapters Four, Five, and Seven, we consider dynamic optimization problems in which the velocities are constrained either according to a control system:

$$(1.1.1) \quad \dot{q}(t) = f(t, q(t), u(t))$$

with measurable control  $u$  taking values in a metric space  $\mathbb{U}$  or according to a differential inclusion:

$$(1.1.2) \quad \dot{q}(t) \in F(t, q(t)),$$

where  $F : [0, T] \times M \rightrightarrows TM$  is a set-valued map satisfying  $F(t, q) \subset T_qM$  for all  $(t, q)$ . For problems of the first type we derive geometric versions of the Pontryagin Maximum Principle while for problems of the second we derive a geometric version of Clarke's Hamiltonian inclusion. We briefly describe these approaches below.

1.1.3.1. *Pontryagin Maximum Principle.* A classical condition for characterizing optimal controls is the *Pontryagin Maximum Principle* [74], a central result in the field of dynamic optimization. Since its appearance in the 1950's, the Maximum Principle has inspired considerable effort in the study of optimal control problems in  $\mathbb{R}^n$ . During the past few decades, optimal control problems on manifolds have also been studied intensively and statements of the Maximum Principle for such problems can be found in [4, 7, 12, 17, 54, 79, 84]. However, of these, only a handful (see e.g. [4, 7, 17]) offer a full proof of the Maximum Principle for problems on general manifolds and these papers are limited to special cases. Among the restrictions are assumptions that the set  $S$  constraining the terminal point  $q(T)$  must be an immersed submanifold or singleton and the terminal cost  $\ell(q(T))$  must be at least  $C^1$ -smooth. Further, none of the above references establish the Maximum Principle for problems in which the state evolves on a Banach manifold.

We demonstrate in Chapter Four that, for a broad class of problems, the first-order theory of Calculus of Variations is contained in the Maximum Principle. In this chapter we arrive at the Maximum Principle through a kind of nonsmooth calculus of variations. This approach is similar to that introduced by Clarke in [31] for the study of differential inclusions in which a penalty function of the type introduced by Filippov [44] is used to decouple the controls from the trajectories. We apply the method of Lagrangian charts to derive a similar penalty function for problems on manifolds and obtain a general statement of the Maximum Principle for Bolza problems on Banach manifolds.

This approach, in which optimal control is approached through the nonsmooth Calculus of Variations, relies on a geometric generalization of the *fundamental lemma* of Calculus of Variations. This lemma has classically played a central role in the derivation of necessary optimality conditions in Calculus of Variations and one source [35] dates the appearance of this lemma as far back as 1854. In spite of its age and importance in the linear theory, there does not appear to be an analogue for this lemma when the underlying space is a Banach manifold, even of finite dimension.

Chapter Four also provides a generalization of the classical duBois-Reymond lemma, whose linear analogue dates back to 1879 [36], and a theory of integration by parts for maps into  $TM$  and  $T^*M$ .

In the fifth chapter we study geometric control problems with terminal constraints. Here there is a nice analogy with classical optimization theory. Consider the problem of minimizing a function  $c : \mathbb{R}^n \rightarrow \mathbb{R}$  subject to smooth inequality constraints  $g_i(x) \leq 0$ ,  $1 \leq i \leq r$ . Suppose that  $\bar{x}$  is a local minimizer. Under the assumption that the gradients of the active constraints are convex independent at

$\bar{x}$  one can show that  $\bar{x}$  is an unconstrained local minimizer for the function

$$(1.1.3) \quad c(x) + K \max \{0, g_1(x), \dots, g_r(x)\},$$

with  $K$  sufficiently large. One may then employ techniques of nonsmooth analysis to arrive at the Lagrange multiplier rule: there exist  $\lambda_i \geq 0$  satisfying  $\lambda_i g_i(\bar{x}) = 0$  for  $1 \leq i \leq r$  such that

$$(1.1.4) \quad \nabla c(\bar{x}) + \sum_{i=1}^r \lambda_i \nabla g_i(\bar{x}) = 0.$$

On the other hand, if the gradients of the active constraints are not convex independent at  $\bar{x}$  then there exist  $\lambda_i \geq 0$  satisfying  $\lambda_i g_i(\bar{x}) = 0$  for  $1 \leq i \leq r$  such that

$$(1.1.5) \quad \sum_{i=1}^r \lambda_i \nabla g_i(\bar{x}) = 0.$$

Further, in this second case, the  $\lambda_i$  are not all equal to zero. Thus exact penalization, abnormality of minimizers, and necessary optimality conditions are closely related through nonsmooth analysis, even when the data  $c$  and  $g_i$  are smooth.

This approach to the theory of Lagrange multipliers is related to the *metric regularity* of the smooth inequality constraints. In the fifth chapter we approach the derivation of a geometric Maximum Principle for problems with terminal constraints by developing a pseudometric for controls and studying the metric regularity of the constraint  $q(T) \in S$  with respect to this pseudometric. Interestingly, we are able to arrive at results which are analogous to the Lagrange multiplier result described above. In particular, we prove the following:

THEOREM. Let  $u^0$  be an optimal control with trajectory  $q^0$ . Suppose that every absolutely continuous solution  $\zeta$  to the adjoint equations

$$(1.1.6) \quad \dot{\zeta}(t) = \vec{H}(t, \zeta(t), u^0(t)),$$

which satisfies both the transversality condition  $-\zeta(T) \in N_S^L(q^0(T))$  and the nondegeneracy condition  $\zeta(T) \neq 0$  fails to satisfy the maximum principle

$$(1.1.7) \quad H(t, \zeta(t), u^0(t)) = \max_{u \in \mathbb{U}} H(t, \zeta(t), u)$$

on a subset of  $[0, T]$  with nonzero measure.

Then  $u^0$  is an unconstrained local minimizer for the penalized Mayer problem whose cost is given by

$$(1.1.8) \quad q \mapsto \ell(q) + Kd(q),$$

where  $d$  is the locally defined function  $d(q) := d_{\theta(S)} \circ \theta(q)$  for a coordinate chart  $\theta$  whose domain includes  $q^0(T)$ .

The precise problem studied and surrounding assumptions are given explicitly in Chapter Five. This theorem is analogous to the Lagrange multiplier problem in that the absence of abnormality implies that the constraint can be removed through exact penalization. These techniques are used to prove the following version of the Maximum Principle:

THEOREM. Suppose that  $u^0$  is an optimal control with trajectory  $q^0$ . There exist  $\lambda^0 \in \{0, 1\}$  and  $-\zeta_T \in \lambda^0 \partial_L \ell(q^0(T)) + N_S^L(q^0(T))$  such that if  $\zeta : [0, T] \rightarrow T^*M$  is the solution to

$$(1.1.9) \quad \dot{\zeta}(t) = \vec{H}(t, \zeta(t), u^0(t)), \quad \zeta(T) = \zeta_T$$

then for almost all  $t$

$$H(t, \zeta(t), u^0(t)) = \max_{u \in \mathbb{U}} H(t, \zeta(t), u).$$

Further, either  $\lambda^0 = 1$  or  $\zeta(t) \neq 0$  for all  $t$ .

1.1.3.2. *Clarke's Hamiltonian Inclusion.* In Chapter Seven we develop necessary optimality conditions for problems of dynamic optimization which include a dynamic constraint in the form of a *differential inclusion*

$$(1.1.10) \quad \dot{q}(t) \in F(t, q(t)).$$

Such differential inclusions appear naturally in a variety of settings, including geodesic problems, in which a natural formulation is

$$\dot{q}(t) \in \left\{ v \in T_{q(t)}M : \|v\|_g \leq 1 \right\}.$$

Differential inclusions also arise naturally optimal control problems which are subject to mixed constraints of the form

$$g_j(t, x(t), u(t)) \leq 0 \quad 1 \leq j \leq r.$$

Early studies of such problems in the context of optimal control include [37]. Substantial progress in the theory of necessary conditions for differential inclusions was made in the 1970s by Clarke [31], followed by Vinter and Pappas [90] and Ioffe [52] in the early 1980s.

Our approach is modelled closely on that taken in Clarke's classic text [23]. We study a Mayer problem of minimizing a locally Lipschitz function  $\ell(q(T))$  subject to terminal constraint  $q(T) \in S$  and dynamic constraint  $\dot{q}(t) \in F(t, q(t))$ . Precise assumptions on  $F$  are given in Chapter Seven. Again it is useful to consider a locally



defined penalty function  $d : M \rightarrow \mathbb{R}$  given by  $d_{\theta(S)} \circ \theta$  for a coordinate chart  $\theta$  whose domain includes  $q^0(T)$ . We provide a careful proof of the following theorem:

**THEOREM.** *Let  $q^0 : [0, T] \rightarrow M$  be a local minimizer. Suppose there are no absolutely maps  $\zeta : [0, T] \rightarrow T^*M$  which satisfy the differential inclusion*

$$\dot{\zeta}(t) \in \overrightarrow{H}(t, \zeta(t)),$$

*the transversality condition  $-\zeta(T) \in N_S^L(q^0(T))$ , and the nondegeneracy condition  $\zeta(T) \neq 0$ .*

*Then there exists a constant  $K > 0$  such that  $q^0$  provides a local minimum for the cost function*

$$(1.1.11) \quad q \mapsto \ell(q) + Kd(q),$$

*subject only to the dynamic constraint  $\dot{q}(t) \in F(t, q(t))$ .*

Here  $\overrightarrow{H} : [0, T] \times T^*M \rightrightarrows TT^*M$  is a set-valued generalization of the Hamiltonian lift of a vector field. We also establish the following:

**THEOREM.** *If  $q^0 : [0, T] \rightarrow M$  is a local minimizer then there is a  $\lambda^0 \in \{0, 1\}$  and solution  $\zeta : [0, T] \rightarrow T^*M$  to*

$$\dot{\zeta}(t) \in \overrightarrow{H}(t, \zeta(t)).$$

*which satisfies  $-\zeta(T) \in \lambda^0 \partial_L \ell(q^0(T)) + N_S^L(q^0(T))$ .*

**1.1.4. Existence and Uniqueness.** The sixth chapter of this dissertation concerns itself with the existence and uniqueness of optimal trajectories, following [62]. In this chapter we establish existence and uniqueness theorems for Mayer

and Bolza problems on manifolds which do not rely on the convexity of the associated differential inclusion  $f(t, q, \mathbb{U})$ . Again techniques of nonsmooth analysis play an important role, even though the data for the problem are assumed smooth.

## 1.2. Background

For background in smooth manifolds including the symplectic structure of  $T^*M$ , John Lee's introduction to smooth manifolds [67] provides a useful introduction to the finite-dimensional case and Lang's text [61] provides an introduction to manifolds modeled over Banach spaces. Integration in Banach space is given a clear exposition in [34], a source which includes many applications of integration to general Banach space theory. Additional properties of the Bochner integral such as Fubini's theorem are carefully established in [38] and the functional analysis surrounding the resulting Lebesgue-Bochner spaces is summarized nicely in [42].

Though certain aspects of the finite dimensional theory of differential equations can be carried easily to the case of ODE in Banach space, certain useful and delicate results on differential equations in Banach space can be found in [33]. There it is shown, for example, that continuity of the right-hand side is not enough to ensure existence of solutions to differential equations in general Banach space. The theory of Fréchet spaces is included in the classic [60] by Köthe and a complete introduction to the classical Chronological Calculus can be found in [4].

In the following subsections we collect the basic materials from the above sources that will be called for in this dissertation.

**1.2.1. Calculus in a Banach Space.** Let  $E$  and  $F$  be Banach spaces. A map  $f : E \rightarrow F$  is said to be *differentiable* at  $x_0$  if there exists a bounded linear operator  $f'(x_0) : E \rightarrow F$  such that for all  $x \in E$  we have  $f(x) = f(x_0) + f'(x_0)(x - x_0) + o(\|x - x_0\|)$ . If  $f$  is differentiable on all of  $E$ , then we have  $f' : E \rightarrow L(E, F)$ , where  $L(E, F)$  is the Banach space of bounded linear operators from  $E$  to  $F$ . When  $f'$  is

continuous, we say that  $f$  is of class  $C^1$ . As a map between Banach spaces, we may then ask if  $f'$  is differentiable and so on. If  $f$  has  $m$  continuous derivatives, then we say that  $f$  is of class  $C^m$ . The  $m^{\text{th}}$  derivative at a point  $x_0$  may be identified with an  $m$ -multilinear map  $\underbrace{E \times \cdots \times E}_{m \text{ copies}} \rightarrow F$  and the space of such maps is again a Banach space with norm

$$\|A\| = \sup \{ \|A(x_1, \dots, x_m)\| : \|x_1\| = \cdots = \|x_m\| = 1 \}.$$

Functions which take values in a Banach space can also be integrated [34]. We briefly describe here the *Bochner integral* for functions  $f : [t_0, t_1] \rightarrow E$ , where  $E$  is a Banach space. As one might expect, a function  $f : [t_0, t_1] \rightarrow E$  is said to be *simple* if it takes on only finitely many values, say  $[t_0, t_1] = \cup_{i=1}^k A_i$ , with  $A_i$  disjoint measurable sets and  $f|_{A_i} = f_i \in E$ . For simple functions one then defines

$$\int_{t_0}^{t_1} f dt = \sum_{i=1}^k f_i \mu(A_i),$$

where  $\mu$  is Lebesgue measure. If  $E$  is a Banach space, a function  $f : [t_0, t_1] \rightarrow E$  is said to be *measurable* if it is a pointwise limit of a sequence of simple functions, say  $f_n \rightarrow f$ . Measurable function  $f$  is said to be *Bochner integrable* if  $\lim_n \int_{t_0}^{t_1} \|f - f_n\| dt = 0$  for some sequence of simple functions  $f_n$ . In this case the *Bochner integral* of  $f$  is defined as

$$\int_{t_0}^{t_1} f(t) dt = \lim_n \int_{t_0}^{t_1} f_n(t) dt$$

It is worth noting that when  $E = \mathbb{R}^n$ , the Bochner integral is the same as the Lebesgue integral. Indeed the Bochner integral has many desirable properties of the Lebesgue integral. In particular, one has

$$\frac{d}{dt} \int_{t_0}^t f(\tau) d\tau = f(t)$$

for almost all  $t$  in  $[t_0, t_1]$ . A function  $F(t)$  is called *absolutely continuous* if  $F(t) = F(t_0) + \int_{t_0}^t f(\tau) d\tau$  for some integrable  $f$ . This and other properties of Bochner integral are given a clear treatment in [34, 38, 42]. For a first reading we suggest [34].

**1.2.2. Differential Equations and Flows in Banach Space.** We recall some results from the theory of differential equations in Banach spaces. In particular, we are interested in equations of the form

$$(1.2.1) \quad \dot{x} = f(t, x) \quad x(t_0) = x_0$$

where  $f : J \times E \rightarrow E$  and  $J \subseteq \mathbb{R}$  is an interval containing  $t_0$ . We introduce the following definitions for vector fields on  $E$ :

DEFINITION 1.2.1. A *nonautonomous  $C^m$  vector field* on  $E$  is a function  $f : J \times E \rightarrow E$  which is measurable in  $t$  for each fixed  $x$  and  $C^m$  in  $x$  for almost all  $t$ .

DEFINITION 1.2.2. A nonautonomous  $C^m$  vector field on  $E$  is said to be *locally integrable bounded* if for any  $x_0 \in E$ , there exists an open neighborhood  $U$  of  $x_0$  and  $k \in L^1(J, \mathbb{R})$  such that for all  $x \in U$ , for all  $0 \leq i \leq m$ , we have  $\|f^{(i)}(t, x)\| \leq k(t)$  for almost all  $t$ , where  $f^{(i)}$  denotes the  $i^{\text{th}}$  derivative of  $f$  with respect to  $x$ .

DEFINITION 1.2.3. A nonautonomous  $C^m$  vector field on  $E$  is said to be *locally bounded* if for any  $x_0 \in E$ , there exists an open neighborhood  $U$  of  $x_0$  and  $K \geq 0$  such that for all  $x \in U$ , for all  $0 \leq i \leq m$ , we have  $\|f^{(i)}(t, x)\| \leq K$  for almost all  $t$ .

Notice that any autonomous  $C^m$  vector field is locally bounded. It can be shown that if  $f : J \times E \rightarrow E$  is a nonautonomous  $C^m$  vector field that is locally integrable bounded, then for any  $(t_0, x_0)$  there exists an open interval  $J_0 \subset J$  containing  $t_0$  and depending on  $(t_0, x_0)$  as well as a unique, absolutely continuous function  $x : J_0 \rightarrow E$  which satisfies (1.2.1) for almost all  $t \in J_0$ . This type of solution is

called a *Carathéodory* solution. In addition, the dependence of this solution upon the initial condition  $x_0$  is  $C^m$ -smooth. More precisely, if  $x(t; t_0, x_0)$  denotes the solution to (1.2.1), then  $x_0 \mapsto x(t; t_0, x_0)$  is  $k$  times continuously differentiable for appropriate values of  $t$  and  $x_0$ .

We will write  $P_{t_0, t}$  for the *local flow*  $x_0 \mapsto x(t; t_0, x_0)$ . Uniqueness of solutions gives us the following properties for the flow:

$$(1.2.2) \quad P_{s, t} \circ P_{t_0, s}(x) = P_{t_0, t}(x)$$

$$(1.2.3) \quad P_{t_0, t}^{-1}(x) = P_{t, t_0}(x)$$

When the underlying vector field is autonomous, we will write  $P_t$  for  $P_{0, t}$ . One may then obtain the following local semigroup properties for the flow:

$$P_s \circ P_t(x) = P_{s+t}(x)$$

$$P_t^{-1}(x) = P_{-t}(x),$$

provided that  $t, s, t + s$ , and  $-t$  lie in  $J_0$ , an interval which in general will depend on  $x$ .

**1.2.3. Smooth Manifolds.** In defining dynamical systems, it is enough for the underlying space to have the structure of a Banach space only locally. In this section we remind the reader of some definitions and basic results from the theory of smooth manifolds. For a greater level of detail, we suggest [61].

A *Banach manifold* of class  $C^m$  over a Banach space  $E$  is a paracompact Hausdorff space  $M$  along with a collection of coordinate charts  $\{(U_\alpha, \varphi_\alpha) : \alpha \in A\}$ , where  $A$  is an indexing set. This collection of charts should be such that the collection  $\{U_\alpha\}$  is a cover for  $M$ ; each  $\varphi_\alpha$  is a bijection of  $U_\alpha$  with an open subset of  $E$ ; and the transition maps  $\varphi_\alpha \circ \varphi_\beta^{-1} : \varphi_\beta(U_\alpha \cap U_\beta) \rightarrow \varphi_\alpha(U_\alpha \cap U_\beta)$  are of class  $C^m$ .

If  $M$  and  $N$  are Banach manifolds, a function  $f : M \rightarrow N$  is said to be  $C^m$ -smooth (or  $C^m$  for brevity) if for any coordinate charts  $\varphi : U \subseteq M \rightarrow E$  and  $\psi : V \subseteq N \rightarrow F$  the map  $\psi \circ f \circ \varphi^{-1}$  is a  $C^m$ -smooth mapping of Banach spaces. Analogously, a function  $f : M \rightarrow N$  is *differentiable* at a point  $q_0$  if  $\psi \circ f \circ \varphi^{-1}$  is differentiable at  $\varphi(q_0)$ .

The *tangent space* to  $M$  at  $q$  is defined as follows. Consider the collection of differentiable curves  $\gamma : \mathbb{R} \rightarrow M$  with  $\gamma(0) = q$  and define an equivalence relation on this collection by  $\gamma_1 \sim \gamma_2$  if and only if  $(\varphi \circ \gamma_1)'(0) = (\varphi \circ \gamma_2)'(0)$  for some coordinate chart  $\varphi$ . One can check that if this relationship holds for one coordinate chart, it will hold for all coordinate charts. We write  $[\gamma]$  for the equivalence class of a curve  $\gamma$ . The collection of these equivalence classes forms the tangent space  $T_q M$  and there is a natural isomorphism  $T_q M \leftrightarrow E$ .

Every  $C^m$  map  $f : M \rightarrow N$  induces a map from  $T_q M$  to  $T_{f(q)} N$  by  $[\gamma] \mapsto [f \circ \gamma]$  and we denote this mapping by  $f_*(q)$ . The tangent bundle  $TM$  is the union of the tangent spaces with a topology given locally by the charts  $(q, v) \mapsto (\varphi(q), \varphi_*(q)v)$ , where  $\varphi$  is a coordinate chart for  $M$ . When  $f$  is a map between linear spaces  $E$  and  $F$  we will write  $f'$  for its derivative. When  $f$  is a map between Banach manifolds, we will write  $f_*$  for the corresponding map from  $TM$  to  $TN$ . We emphasize that in local coordinates,  $f_*(q) : T_q M \rightarrow T_{f(q)} N$  is the map given by  $v \mapsto f'(q)v$ . In contrast, the map  $f_* : TM \rightarrow TN$  sends a pair  $(q, v)$  to the pair  $(f(q), f_*(q)v)$ .

**1.2.4. Vector Fields and Flows on Manifolds.** Let  $\pi : TM \rightarrow M$  be the projection  $(q, v) \mapsto q$ . A *nonautonomous vector field* is a mapping  $V : \mathbb{R} \times M \rightarrow TM$  which satisfies  $\pi \circ V_t(q) = q$ . Given  $q_0 \in M$  and a coordinate chart  $(U, \varphi)$  at  $q_0$ , the function  $J \times E \rightarrow E$  given by

$$(1.2.4) \quad (\varphi_* V_t)(x) := \varphi_* (\varphi^{-1}(x)) V_t (\varphi^{-1}(x))$$

is the *local coordinate representation for  $V_t$* . Recalling definition 1.2.2 we introduce

DEFINITION 1.2.4. A nonautonomous vector field on  $M$  is said to be a *locally integrable bounded  $C^k$  vector field* if it is  $C^k$ -smooth in  $q$  for almost all  $t$ , is measurable in  $t$ , and in some neighborhood of each  $q \in M$  there is a coordinate representation (1.2.4) which is locally integrable bounded.

Similarly, recalling definition 1.2.3, we introduce

DEFINITION 1.2.5. A nonautonomous vector field on  $M$  is said to be a *locally bounded  $C^k$  vector field* if it is  $C^k$ -smooth in  $q$  for almost all  $t$ , is measurable in  $t$ , and in some neighborhood of each  $q \in M$  there is a coordinate representation (1.2.4) which is locally bounded.

If  $x(t)$  is a solution for the differential equation  $\dot{x} = (\varphi_* V_t)(x)$  on  $E$  with initial condition  $x(t_0) = \varphi(q_0)$ , then  $q(t) = \varphi^{-1} \circ x(t)$  is a solution to the differential equation on  $M$

$$(1.2.5) \quad \dot{q} = V_t(q), \quad q(t_0) = q_0.$$

For any  $\varphi \in C^m(M, E)$  we have the following integral representation

$$(1.2.6) \quad \varphi(q(t)) = \varphi(q_0) + \int_{t_0}^t \varphi_*(q(\tau)) V_\tau(q(\tau)) d\tau.$$

With each nonautonomous vector field  $V_t$  on  $M$ , we associate a local flow  $P_{t_0, t}$  given by  $q_0 \mapsto q(t; t_0, q_0)$ , the solution to (1.2.5) with initial condition  $q(t_0) = q_0$ . In the case of autonomous vector fields  $V$  we consider a local flow  $P_t : q_0 \mapsto q(t; 0, q_0)$ . These flows are  $C^m$  diffeomorphisms of  $M$  and are of central importance in the development of our extension of the Chronological Calculus.

**1.2.5. Nonsmooth Analysis.** The field of nonsmooth analysis now plays a significant role in modern optimization and control theory and has been shown to be

effective even for problems with smooth data. Textbook introductions can be found in [14, 23, 28, 80]. For the techniques used in this dissertation we recommend [28] and the paper [66], which introduces techniques of nonsmooth analysis for smooth manifolds.

Let us recall that a Banach space  $E$  is called *smooth* if there exists a non-trivial  $C^1$ -smooth bump function (that is, a function with a bounded support). For example, Banach spaces with differentiable norm are smooth Banach spaces as, in particular, Hilbert spaces are. The spaces  $L^p([0, T], \mathbb{R}^n)$  with  $1 < p < \infty$  are all smooth.

A subgradient  $\zeta \in E^*$  of a function  $f : E \rightarrow (-\infty, +\infty]$  at the point  $x$  is defined as follows: suppose there exists a  $C^1$ -smooth function  $g : E \rightarrow \mathbb{R}$  such that the function  $f - g$  attains a local minimum at  $x$ . Then a subgradient of  $f$  at  $x$  is the vector  $\zeta = g'(x)$ . The set of all subgradients at  $x$  is called a *subdifferential*  $\partial_F f(x)$ . It can be shown that for lower semicontinuous functions  $f$  subdifferentials are nonempty on a set which is dense in the domain of  $f$ . The detailed calculus of such subdifferentials can be found in the monographs [14, 28, 80]. The monograph [28] is dedicated to the calculus of proximal subgradients in Hilbert spaces.

Nonsmooth analysis for nonsmooth semicontinuous functions on smooth finite-dimensional manifolds was suggested in [66]. The concept of a subgradient of a lower semicontinuous function from [66] is easily adapted for infinite-dimensional manifolds:

DEFINITION 1.2.6. Let  $\ell : M \rightarrow \overline{\mathbb{R}}$  be lower semicontinuous. A covector  $\zeta \in T_q^*M$  is a *Fréchet subgradient* for  $\ell$  if  $\zeta = dg_q$  for some  $C^1$ -smooth function  $g : M \rightarrow \mathbb{R}$  such that  $\ell - g$  attains a local minimum at  $q$ . The set of all such covectors is denoted  $\partial_F \ell(q)$ .

The following constructs are also of considerable importance:



DEFINITION 1.2.7. Let  $\ell : M \rightarrow \overline{\mathbb{R}}$  be lower semicontinuous. A covector  $\zeta \in T_q^*M$  is a *limiting subgradient* for  $\ell$  if there is a sequence  $q_i \in M$  for which  $(q_i, \ell(q_i)) \rightarrow (q, \ell(q))$  and Fréchet subgradients  $\zeta_i \in \partial_F \ell(q_i)$  for which  $\zeta_i \rightarrow \zeta$ . The set of all such covectors is denoted  $\partial_L \ell(q)$ .

The limiting subdifferential satisfies the following sum rule:

$$(1.2.7) \quad \partial_L(\ell_1 + \ell_2)(q) \subseteq \partial_L \ell_1(q) + \partial_L \ell_2(q).$$

The normal cone to a closed set  $S$  is defined through  $N_S^L(q) = \partial_L \psi_S(q)$ , where  $\psi_S$  is the function defined through  $\psi_S(q) = 0$  for  $q \in S$  and  $\psi_S(q) = +\infty$  otherwise. One can show that if for a closed set  $S \subset \mathbb{R}^n$ , when  $q \in S$  then  $\partial_L d_S(q) \subset N_S^L(q)$  and when  $q \notin S$ , then  $\zeta \in \partial_L d_S(q)$  implies  $\|\zeta\|_{\mathbb{R}^n} = 1$ .

DEFINITION 1.2.8. Let  $\ell : M \rightarrow \mathbb{R}$  be locally Lipschitz. The *Dini subderivate* for  $\ell$  at  $q$  in a direction  $v \in T_q M$  is defined by

$$(1.2.8) \quad D\ell(q; v) = \liminf_{t \downarrow 0} \frac{\ell(c_v(t)) - \ell(q)}{t},$$

where  $c_v$  is a differentiable curve through  $q$  satisfying  $\dot{c}_v(0) = v$ .

Definitions 1.2.6, 1.2.7, and 1.2.8, developed in [66], where they are shown to be independent of local coordinates.

Nonsmooth analysis makes extensive use of mappings whose values are sets. The integral of a set-valued map  $F : [0, T] \rightarrow \mathbb{R}^n$  is defined through

$$(1.2.9) \quad \int_0^T F(t) dt = \left\{ \int_0^T v_t dt : v_t \in F(t) \text{ a.a. } t \in [0, T] \right\}.$$

We will have need for the following results regarding set-valued maps. Proofs can be found in [28].

THEOREM 1.2.9 (Aumann). *Let  $F : [0, T] \rightarrow \mathbb{R}^n$  be a measurable set-valued mapping which is bounded and has closed, nonempty values. Then*

$$(1.2.10) \quad \int_0^T F(t) dt = \int_0^T \text{co } F(t) dt.$$

LEMMA 1.2.10 (Filippov). *Let  $F : [0, T] \rightarrow \mathbb{R}^n$  be a measurable set-valued mapping with closed values and let  $g : [0, T] \times \mathbb{U} \rightarrow \mathbb{R}^n$  be Carathéodory. Suppose that for almost every  $t \in [0, T]$  there exists  $u \in \mathbb{U}$  such that  $g(t, u) \in F(t)$ . Then there exists a measurable mapping  $u(t) : [0, T] \rightarrow \mathbb{U}$  such that  $g(t, u(t)) \in F(t)$  for almost all  $t$ .*

With this we turn to our first chapter of results – an extension of the classical Chronological Calculus.

## CHAPTER 2

### Extension of the Chronological Calculus

In the 1970s, Agrachev and Gamkrelidze suggested in [2, 3] the Chronological Calculus for the analysis of  $C^\infty$ -smooth dynamical systems on finite-dimensional manifolds (for a textbook exposition see [4]). The central idea of this calculus is to consider flows of dynamical systems as linear operators on the space of  $C^\infty$ -smooth scalar functions. This “linearization” of flows on manifolds presents significant advantages from the point of view of defining derivatives of flows, developing a calculus of such derivatives, and effective computations of formal power series representing flows.

But in addition to these desirable properties, the Chronological Calculus poses some interesting problems. The space of  $C^\infty$ -smooth scalar functions is a Fréchet space with topology given by a countable family of seminorms and this complicates the proofs of the calculus rules given in [3, 4]. The approach also requires the strong assumption of  $C^\infty$ -smoothness of dynamical systems and manifolds, even if for many applications only finite sums of Volterra-like series representing flows are enough [71].

Another restriction of the classical Chronological Calculus (which is important from the point of view of applications to control systems on manifolds) is its treatment of nonautonomous vector fields which depend on  $t$  in a measurable way. In particular, there is no variant of the product rule in the classical Chronological Calculus which can be used for such flows.

In this chapter we extend the Chronological Calculus so as to require only  $C^m$ -smoothness of dynamical systems and manifolds. The result is a computationally effective version of the Chronological Calculus without reference to Fréchet spaces. Moreover, in the framework of this extension we provide a “distributional” version of the product rule which can be applied to flows of nonautonomous vector fields which are merely measurable in  $t$ . We thus provide details for a rule which are lacking in the description of the classical Chronological Calculus [2, 3, 4, 47], even for finite-dimensional manifolds.

Further, this extension allows analysis of dynamical systems on infinite-dimensional manifolds which are interesting from the point of view of applications to the theory of partial differential equations. We also develop a calculus of remainder terms (calculus of “little  $o$ ’s”) which is used for the effective calculation of representations of brackets of flows in terms of respective brackets of vector fields on infinite-dimensional manifolds and which provides an algorithm for the computation of remainder terms in such representations. Finally, we use these results for proving a generalization of Chow-Rashevskii theorem for infinite-dimensional manifolds.

## 2.1. Extension of Chronological Calculus

The main observation behind the Chronological Calculus [2, 3, 4] is that one may trade analytic objects such as diffeomorphisms or vector fields for algebraic objects such as automorphisms or derivations of the algebra  $C^\infty(M)$ , which is the collection of  $C^\infty$  mappings  $f : M \rightarrow \mathbb{R}$ . This correspondence is developed in [2, 3, 4], where  $C^\infty(M)$  is given the structure of a Fréchet space. Below we develop a streamlined version of the theory which is effective for computations with infinite-dimensional  $C^m$ -manifolds and dynamical systems. In order to include Banach spaces in the theory, we consider the vector space  $C^m(M, E)$  of  $C^m$ -smooth functions  $f : M \rightarrow E$  rather than the algebra of scalar functions  $C^\infty(M)$ .

### 2.1.1. Chronological Calculus Formalism: Flows as Linear Operators.

We begin by defining the following operators:

- (i) Given any point  $q \in M$ , let  $\hat{q} : C^m(M, E) \rightarrow E$  be the linear map given by  $\hat{q}(\varphi) := \varphi(q)$ .
- (ii) Given  $C^m$ -manifolds  $M$  and  $N$  over a Banach space  $E$  and a  $C^m$  map  $P : M \rightarrow N$ , let  $\hat{P} : C^m(N, E) \rightarrow C^r(M, E)$  ( $0 \leq r \leq m$ ) be the linear map given by  $\hat{P}(\varphi) := \varphi \circ P$ . Note that if  $P$  is a diffeomorphism of  $M$ ,  $\hat{P}$  gives us an isomorphism of  $C^m(M, E)$ .
- (iii) Given a tangent vector  $v \in T_q M$ , let  $\hat{v} : C^m(M, E) \rightarrow E$  be the linear map given by  $\hat{v}(\varphi) := \varphi_*(q)v$ .
- (iv) Given any  $C^m$  vector field  $V$  on  $M$ , we define a linear map  $\hat{V} : C^m(M, E) \rightarrow C^{m-1}(M, E)$  by  $\hat{V}(\varphi) : q \mapsto \varphi_*(q)V(q)$ .

Of course, we can consider linear combinations of such linear operators. Notice that for any smooth manifold  $M$  the identity map  $Id_M : M \rightarrow M$  defined by  $Id_M(q) = q$  induces the identity operator  $\widehat{Id}_M : C^m(M, E) \rightarrow C^r(M, E)$  through  $\widehat{Id}_M(\varphi) = \varphi$ .

We need not restrict ourselves to the space  $C^m(M, E)$ . Given any open set  $U \subseteq M$ , we may view  $U$  as a Banach manifold in its own right and therefore consider the space  $C^m(U, E)$ . For example, the local flow  $P_{t_0, t} : J_0 \times U_0 \rightarrow U$  of a vector field  $V_t$  gives rise to a family of linear mappings  $\hat{P}_{t_0, t} : C^m(U, E) \rightarrow C^m(U_0, E)$ . We also note that when  $\varphi$  is a local diffeomorphism then under an appropriate restriction of domain, the operators defined above simply give us local coordinate expressions.

For a mapping  $P : M \rightarrow N$ , the operation  $P \mapsto \hat{P}$  is contravariant. Thus for operators  $\hat{P}_{t_0, t}$  arising as flows of vector fields the semigroup property (1.2.2) becomes

$$(2.1.1) \quad \hat{P}_{t_0, s} \circ \hat{P}_{s, t} = \hat{P}_{t_0, t}.$$

An operator  $\widehat{\sigma}(t)$  will play an important role in the calculus of remainder terms. Later we will develop a more detailed definition of such operators, as well as several useful examples. For the moment we denote by  $\widehat{\sigma}(t)$  any linear operator  $C^m(M, E) \rightarrow C^r(M, E)$  ( $0 \leq r \leq m$ ) with the following property: for any  $\varphi \in C^m(N, E)$  and  $q_0 \in M$  there exists a neighbourhood  $U$  such that

$$(2.1.2) \quad \lim_{t \downarrow 0} \frac{\|\widehat{\sigma}(t)(\varphi)(q)\|}{t} = 0$$

uniformly with respect to  $q \in U$ . For an example of such an operator, consider the flow operator  $\widehat{P}_t$  for an autonomous vector field  $V$ . It follows from (1.2.6) that for the operator

$$(2.1.3) \quad \widehat{\sigma}(t) := \widehat{P}_t - \widehat{Id}_M - t\widehat{V}$$

and a function  $\varphi \in C^m(M, E)$  we have that

$$(2.1.4) \quad \widehat{\sigma}(t)(\varphi) = \int_0^t \varphi_*(P_s(q))(V(P_s(q)) - \varphi_*(q)V(q)) ds.$$

This representation implies that the operator (2.1.3) satisfies (2.1.2).

### 2.1.2. Differentiation and Integration of Operator-Valued Functions.

In this section we define integration and differentiation for operators depending on a parameter  $t \in \mathbb{R}$ .

DEFINITION 2.1.1. Consider an operator-valued function  $t \rightarrow A_t$  whose values are linear mappings  $A_t : C^m(M, E) \rightarrow C^p(M, E)$ . This function is called *integrable* if for any  $\varphi \in C^m(M, E)$  and  $q \in M$  the function  $t \rightarrow A_t(\varphi)(q)$  is integrable and the assignment

$$(2.1.5) \quad q \mapsto \int_0^T A_t(\varphi)(q) dt$$

is  $C^p$ -smooth.

The resulting linear operator  $\left(\int_{t_0}^{t_1} A_\tau d\tau\right) : C^m(M, E) \rightarrow C^p(M, E)$  is defined as follows

$$\left(\int_{t_0}^{t_1} A_\tau d\tau\right)(\varphi)(q) := \int_{t_0}^{t_1} A_\tau(\varphi)(q) d\tau$$

It follows immediately from (1.2.5) and (1.2.6) that the flow operator  $\widehat{P}_{t_0, t}$  representing the flow of diffeomorphisms for a nonautonomous vector field  $V_t$  satisfies the integral equation

$$(2.1.6) \quad \widehat{P}_{t_0, t} = \widehat{Id}_M + \int_{t_0}^t \widehat{P}_{t_0, \tau} \circ \widehat{V}_\tau d\tau.$$

Moreover, the unique operator valued solution of the integral equation (2.1.6) is the function  $t \rightarrow \widehat{P}_{t_0, t}$ . Next we introduce a concept of *differentiability* for an operator-valued function  $A_t$ .

DEFINITION 2.1.2. An operator-valued function  $A_t : C^m(M, E) \rightarrow C^r(M, E)$  is called *differentiable* at  $t$  if there exists a linear operator  $B_t : C^r(M, E) \rightarrow C^s(M, E)$

$$(2.1.7) \quad A_{t+h} = A_t + h B_t + \widehat{o}(h).$$

The operator  $\frac{dA_t}{dt} := B_t$  is the *derivative* of  $A_t$ .

This definition is well-suited for the operator  $\widehat{P}_{t_0, t}$  arising from the flow of diffeomorphisms representing differential equation (1.2.5) in the case where the nonautonomous vector field  $V_t$  is continuous in  $t$ . In particular, the semigroup property (2.1.1) implies that

$$\widehat{P}_{t_0, t+h} - \widehat{P}_{t_0, t} - h\widehat{P}_{t_0, t} \circ \widehat{V}_t = \widehat{P}_{t_0, t} \circ (\widehat{P}_{t, t+h} - \widehat{Id}_M - h\widehat{V}_t).$$

The last expression can be represented as

$$(2.1.8) \quad \widehat{P}_{t_0, t} \circ \int_t^{t+h} (\widehat{P}_{t, s} \circ \widehat{V}_s - \widehat{V}_t) ds.$$

Using a representation for (2.1.8) similar to the one from (2.1.4) and continuity  $V_t$  in  $t$ , we obtain that (2.1.8) is  $\widehat{o}(h)$ . Thus, we have derived the representation

$$(2.1.9) \quad \widehat{P}_{t_0, t+h} = \widehat{P}_{t_0, t} + h\widehat{P}_{t_0, t} \circ \widehat{V}_t + \widehat{o}(h).$$

This means that  $t \rightarrow \widehat{P}_{t_0, t}$  is differentiable and for every  $t$ ,

$$\frac{d}{dt}\widehat{P}_{t_0, t} = \widehat{P}_{t_0, t} \circ \widehat{V}_t.$$

We see that when  $V_t$  is continuous in time, the operator-valued function  $\widehat{P}_{t_0, t}$  satisfies the differential equation

$$(2.1.10) \quad \frac{d\widehat{P}_{t_0, t}}{dt} = \widehat{P}_{t_0, t} \circ \widehat{V}_t, \quad \widehat{P}_{t_0, t_0} = \widehat{Id}_M.$$

It is easy to check that  $\widehat{P}_{t_0, t}$  is the unique solution of this operator differential equation and also of the operator integral equation (2.1.6).

However, in the case when the vector field  $V_t$  is only integrable in  $t$  then a Carathéodory solution  $q(t)$  of the differential equation (1.2.5) is an absolutely continuous function and we cannot guarantee that  $\widehat{P}_{t_0, t}$  is differentiable for every  $t$ .

DEFINITION 2.1.3. An operator-valued function  $\widehat{A}_t$  is called *absolutely continuous* on  $[a, b]$  if  $\widehat{A}_t = \widehat{A}_{t_0} + \int_{t_0}^t \widehat{B}_\tau d\tau$  for any  $t \in [a, b]$  for some integrable operator-valued function  $\widehat{B}_t$ . We denote  $\widehat{B}_t$  as  $\frac{d}{dt}\widehat{A}_t$  and understand this derivative in the sense of distributions<sup>1</sup>: for any  $t_1, t_2 \in [a, b]$ , for any  $\varphi \in C^m(M, E)$  and  $q \in M$

$$\widehat{A}_{t_2}(\varphi)(q) - \widehat{A}_{t_1}(\varphi)(q) = \int_{t_1}^{t_2} \frac{d}{dt}\widehat{A}_t(\varphi)(q) dt.$$

REMARK 2.1.4. Let  $W$  be a  $C^m$  vector field and  $\widehat{A}_t$  is absolutely continuous then  $\widehat{A}_t \circ \widehat{W}$  is also absolutely continuous and  $\frac{d}{dt}(\widehat{A}_t \circ W) = \frac{d}{dt}\widehat{A}_t \circ W$ .

<sup>1</sup>We use a term *distribution* by analogy with a concept of a generalized derivative as a distribution in the theory of linear partial differential operators (see [51]).



Note that in the case when the absolutely continuous operator-valued function  $\widehat{A}_t$  is defined by a flow of diffeomorphisms  $P_t : M \rightarrow M$  then for any  $q \in M$  the derivative  $\frac{d}{dt}P_t(q)$  exists for a.a.  $t \in [a, b]$ .

As demonstrated above, for measurable in  $t$  vector fields  $V_t$  the flow operator  $\widehat{P}_{t_0,t}$  is the unique absolutely continuous solution of the integral operator equation (2.1.6). In view of Definition 2.1.3,  $\widehat{P}_t$  is also the unique solution of the operator differential equation (2.1.10) in the sense of distributions.

**2.1.3. Product Rules.** In this subsection we discuss product rules for operator-valued functions  $\widehat{P}_t$  and  $\widehat{Q}_t$ . We first establish a product rule for the case in which these functions are differentiable at  $t$  in the sense of (2.1.7), namely

$$(2.1.11) \quad \widehat{P}_{t+h} = \widehat{P}_t + h \frac{d}{dt} \widehat{P}_t + \widehat{o}_1(h), \quad \widehat{Q}_{t+h} = \widehat{Q}_t + h \frac{d}{dt} \widehat{Q}_t + \widehat{o}_2(h)$$

for some operators  $\frac{d}{dt} \widehat{P}_t$  and  $\frac{d}{dt} \widehat{Q}_t$ .

**THEOREM 2.1.5.** *Let operator-valued functions  $\widehat{P}_t$  and  $\widehat{Q}_t$  be differentiable at  $t$  and suppose the remainder terms  $\widehat{o}_1$  and  $\widehat{o}_2$  have the property*

$$(2.1.12) \quad \widehat{o}_1(h) \circ \frac{d}{dt} \widehat{Q}_t + \frac{d}{dt} \widehat{P}_t \circ \widehat{o}_2(h) + \widehat{o}_1(h) \circ \widehat{o}_2(h) = \widehat{o}(h).$$

*Then the operator-valued function  $\widehat{P}_t \circ \widehat{Q}_t$  is differentiable at  $t$  and*

$$(2.1.13) \quad \frac{d}{dt}(\widehat{P}_t \circ \widehat{Q}_t) = \frac{d}{dt} \widehat{P}_t \circ \widehat{Q}_t + \widehat{P}_t \circ \frac{d}{dt} \widehat{Q}_t$$

**PROOF.** It follows from (2.1.11) and (2.1.12) that

$$\widehat{P}_{t+h} \circ \widehat{Q}_{t+h} = \widehat{P}_t \circ \widehat{Q}_t + h \left( \frac{d}{dt} \widehat{P}_t \circ \widehat{Q}_t + \widehat{P}_t \circ \frac{d}{dt} \widehat{Q}_t \right) + \widehat{o}(h) + \widehat{o}_1(h) \circ \widehat{Q}_t + \widehat{P}_t \circ \widehat{o}_2(h)$$

The sum of last three terms is again operator  $\widehat{o}(h)$  (see (2.1.2)). This implies the differentiability of the product  $\widehat{P}_t \circ \widehat{Q}_t$  and the product rule (2.1.13).  $\square$

Thus the validity of a product rule in the form (2.1.13) can be reduced to the verification of the condition (2.1.12). We can verify directly that (2.1.12) holds for flow operators  $\widehat{P}_t$  and  $\widehat{Q}_t$  which are operator solutions of the operator equation (2.1.10) or equation

$$(2.1.14) \quad \frac{d}{dt}\widehat{Q}_t = \widehat{W}_t \circ \widehat{Q}_t$$

for continuous in  $t$  vector fields  $V_t$  and  $W_t$ .

Now we consider a product rule in the sense of distributions for absolutely continuous operator-valued functions  $\widehat{P}_t$  and  $\widehat{Q}_t$  which are represented for any  $t \in (a, b)$  as

$$(2.1.15) \quad \widehat{P}_t = \widehat{P}_{t_0} + \int_{t_0}^t \frac{d}{d\tau} \widehat{P}_\tau d\tau, \quad \widehat{Q}_t = \widehat{Q}_{t_0} + \int_{t_0}^t \frac{d}{d\tau} \widehat{Q}_\tau d\tau,$$

ASSUMPTION 2.1.6. *Let  $\widehat{P}_t, \widehat{Q}_t$  be absolutely continuous operator-valued functions such that for any  $\varphi \in C^m(M, E)$  and  $q \in M$*

(i) *The function  $t \rightarrow \widehat{P}_t \circ \widehat{Q}_t(\varphi)(q)$  is continuous on  $(a, b)$ ;*

(ii) *The functions*

$$(t, \tau) \rightarrow \frac{d}{d\tau} \widehat{P}_\tau \circ \widehat{Q}_t(\varphi)(q), \quad (t, \tau) \rightarrow \widehat{P}_t \circ \frac{d}{d\tau} \widehat{Q}_\tau(\varphi)(q)$$

*are integrable on  $(a, b) \times (a, b)$ ;*

(iii) *For any  $t, t_1, t_2 \in (a, b)$*

$$\int_{t_1}^{t_2} \frac{d}{d\tau} \widehat{P}_\tau d\tau \circ \widehat{Q}_t(\varphi)(q) = \int_{t_1}^{t_2} \frac{d}{d\tau} \widehat{P}_\tau \circ \widehat{Q}_t(\varphi)(q) d\tau,$$

$$\widehat{P}_t \circ \int_{t_1}^{t_2} \frac{d}{d\tau} \widehat{Q}_\tau d\tau(\varphi)(q) = \int_{t_1}^{t_2} \widehat{P}_t \circ \frac{d}{d\tau} \widehat{Q}_\tau(\varphi)(q) d\tau;$$

(iv) There exists an integrable function  $k_1(\tau)$  such that for all small  $h$ , all  $t \in [\tau - h, \tau]$  and a.a.  $\tau \in (a, b)$

$$(2.1.16) \quad \left\| \frac{d}{d\tau} \widehat{P}_\tau \circ \widehat{Q}_t(\varphi)(q) \right\| \leq k_1(\tau) \quad \left\| \widehat{P}_{t+h} \circ \frac{d}{d\tau} \widehat{Q}_\tau(\varphi)(q) \right\| \leq k_1(\tau).$$

REMARK 2.1.7. If  $\widehat{P}_t, \widehat{Q}_t$  are absolutely continuous solutions of (2.1.10) or (2.1.14), or they are of the type presented in Remark 2.1.4 with  $\widehat{A}_t$  being a solution of (2.1.10) or (2.1.14) then conditions (i)-(iv) are satisfied when  $V_t$  and  $W_t$  are locally integrable bounded.

THEOREM 2.1.8. Let absolutely continuous operator-valued functions  $\widehat{P}_t$  and  $\widehat{Q}_t$  satisfy Assumption 2.1.6. Then  $\widehat{P}_t \circ \widehat{Q}_t$  is absolutely continuous and for any  $t_1, t_2$  in  $(a, b)$

$$(2.1.17) \quad \int_{t_1}^{t_2} \frac{d}{dt} (\widehat{P}_t \circ \widehat{Q}_t) dt = \int_{t_1}^{t_2} \left( \frac{d}{dt} \widehat{P}_t \circ \widehat{Q}_t + \widehat{P}_t \circ \frac{d}{dt} \widehat{Q}_t \right) dt.$$

The proof of this theorem relies on the following Lemma, which can be established using the Fubini Theorem [38]:

LEMMA 2.1.9. Let  $g : (a, b) \times (a, b) \rightarrow E$  be an integrable function. Then for any  $t_1, t_2 \in (a, b)$  and sufficiently small  $h$

$$(2.1.18) \quad \begin{aligned} \int_{t_1}^{t_2} \int_t^{t+h} g(t, \tau) d\tau dt &= \int_{t_1}^{t_2} \int_{\tau-h}^{\tau} g(t, \tau) dt d\tau \\ &\quad - \int_{t_1}^{t_1+h} \int_{\tau-h}^{t_1} g(t, \tau) dt d\tau + \int_{t_2}^{t_2+h} \int_{\tau-h}^{t_2} g(t, \tau) dt d\tau \end{aligned}$$

We now turn to the Proof of Theorem 2.1.8:

PROOF. Fix  $t_1, t_2 \in (a, b)$ ,  $\varphi \in C^m(M, E)$  and  $q \in M$ . There holds

$$(2.1.19) \quad \begin{aligned} & \int_{t_1}^{t_2} \frac{1}{h} (\widehat{P}_{t+h} \circ \widehat{Q}_{t+h} - \widehat{P}_t \circ \widehat{Q}_t)(\varphi)(q) dt \\ &= \frac{1}{h} \int_{t_2}^{t_2+h} \widehat{P}_t \circ \widehat{Q}_t(\varphi)(q) dt - \frac{1}{h} \int_{t_1}^{t_1+h} \widehat{P}_t \circ \widehat{Q}_t(\varphi)(q) dt \end{aligned}$$

Due to (iii) of Assumption 2.1.6 we have

$$(2.1.20) \quad \begin{aligned} & \int_{t_1}^{t_2} \frac{1}{h} (\widehat{P}_{t+h} \circ \widehat{Q}_{t+h} - \widehat{P}_t \circ \widehat{Q}_t)(\varphi)(q) dt = \int_{t_1}^{t_2} dt \frac{1}{h} \int_t^{t+h} \frac{d}{d\tau} \widehat{P}_\tau \circ \widehat{Q}_t(\varphi)(q) d\tau \\ & \quad + \int_{t_1}^{t_2} dt \frac{1}{h} \int_t^{t+h} \widehat{P}_{t+h} \circ \frac{d}{d\tau} \widehat{Q}_\tau(\varphi)(q) d\tau \end{aligned}$$

Apply Lemma 2.1.9 to evaluate the first term in the right-hand side of (2.1.20):

$$\begin{aligned} & \int_{t_1}^{t_2} dt \frac{1}{h} \int_t^{t+h} \frac{d}{d\tau} \widehat{P}_\tau \circ \widehat{Q}_t(\varphi)(q) d\tau = \int_{t_1}^{t_2} d\tau \frac{1}{h} \int_{\tau-h}^{\tau} \frac{d}{d\tau} \widehat{P}_\tau \circ \widehat{Q}_t(\varphi)(q) dt \\ & \quad - \frac{1}{h} \int_{t_1}^{t_1+h} d\tau \int_{\tau-h}^{t_1} \frac{d}{d\tau} \widehat{P}_\tau \circ \widehat{Q}_t(\varphi)(q) d\tau + \frac{1}{h} \int_{t_2}^{t_2+h} d\tau \int_{\tau-h}^{t_2} \frac{d}{d\tau} \widehat{P}_\tau \circ \widehat{Q}_t(\varphi)(q) dt \end{aligned}$$

It follows from conditions (ii) and (iv) of Assumptions 2.1.6, from the Fubini theorem, and from the Lebesgue convergence theorem that

$$(2.1.21) \quad \lim_{h \rightarrow 0} \int_{t_1}^{t_2} dt \frac{1}{h} \int_t^{t+h} \frac{d}{d\tau} \widehat{P}_\tau \circ \widehat{Q}_t(\varphi)(q) d\tau = \int_{t_1}^{t_2} \frac{d}{d\tau} \widehat{P}_\tau \circ \widehat{Q}_\tau(\varphi)(q) d\tau.$$

By a similar argument we prove the limit

$$(2.1.22) \quad \lim_{h \rightarrow 0} \int_{t_1}^{t_2} dt \frac{1}{h} \int_t^{t+h} \widehat{P}_{t+h} \circ \frac{d}{d\tau} \widehat{Q}_\tau(\varphi)(q) d\tau = \int_{t_1}^{t_2} \widehat{P}_\tau \circ \frac{d}{d\tau} \widehat{Q}_\tau(\varphi)(q) d\tau.$$

Using these limits and the continuity of  $t \rightarrow \widehat{P}_t \circ \widehat{Q}_t(\varphi)(q)$ , we see from (2.1.19) and (2.1.20) that

$$(2.1.23) \quad (\widehat{P}_{t_2} \circ \widehat{Q}_{t_2} - \widehat{P}_{t_1} \circ \widehat{Q}_{t_1})(\varphi)(q) = \int_{t_1}^{t_2} (\widehat{P}_t \circ \frac{d}{dt} \widehat{Q}_t + \frac{d}{dt} \widehat{P}_t \circ \widehat{Q}_t) dt(\varphi)(q).$$

This implies that  $\widehat{P}_t \circ \widehat{Q}_t$  is absolutely continuous and its derivative satisfies the product rule (2.1.17) in the sense of distributions.  $\square$

**2.1.4. Operators Ad and ad.** Let  $V$  be a vector field and  $F : M \rightarrow M$  be a  $C^m$  diffeomorphism. For a solution  $q(t)$  of the equation  $\dot{q}(t) = V(q(t))$  the function  $r(t) = F(q(t))$  is also a solution of the differential equation

$$(2.1.24) \quad \dot{r}(t) = F_*(q(t))V(q(t)) = F_*V(r(t)),$$

where the vector field  $F_*V$  is defined by  $F_*V(r) := F_*(F^{-1}(r))V(F^{-1}(r))$ .

To obtain a representation for the operator  $\widehat{F_*V}$  corresponding to the vector field  $F_*V$  we consider the diffeomorphism flow  $R_t$  corresponding to the differential equation (2.1.24):

$$(2.1.25) \quad \frac{d}{dt}\widehat{R}_t = \widehat{R}_t \circ \widehat{F_*V}.$$

But  $\widehat{R}_t = \widehat{P}_t \circ \widehat{F}$  where  $P_t$  is the diffeomorphism flow corresponding to the vector-field  $V$ . Using the product rule and (2.1.25) we arrive at

$$\frac{d}{dt}\widehat{R}_t = \frac{d}{dt}\widehat{P}_t \circ \widehat{F} = \widehat{P}_t \circ \widehat{V} \circ \widehat{F} = \widehat{P}_t \circ \widehat{F} \circ \widehat{F_*V}.$$

This implies that

$$(2.1.26) \quad \widehat{F_*V} = \widehat{F^{-1}} \circ \widehat{V} \circ \widehat{F}.$$

Following [4] we define the operator  $\text{Ad } \widehat{F} : \widehat{V} \mapsto \widehat{F} \circ \widehat{V} \circ \widehat{F^{-1}}$ .

Recall that the Lie bracket  $[V, W]$  of vector fields  $V$  and  $W$  is the vector field<sup>2</sup> whose operator representation has form  $\widehat{[V, W]} = \widehat{V} \circ \widehat{W} - \widehat{W} \circ \widehat{V}$ . Let us prove that

---

<sup>2</sup>To show that this is a vector-field we can use the relation (2.4.1) for vector fields  $V, W$ .

the Lie bracket is invariant under diffeomorphism. We have

$$\begin{aligned}
F_*\widehat{[V, W]} &= \widehat{F^{-1}} \circ (\widehat{V} \circ \widehat{W} - \widehat{W} \circ \widehat{V}) \circ \widehat{F} \\
&= \widehat{F^{-1}} \circ \widehat{V} \circ \widehat{F} \circ \widehat{F^{-1}} \circ \widehat{W} \circ \widehat{F} - \widehat{F^{-1}} \circ \widehat{W} \circ \widehat{F} \circ \widehat{F^{-1}} \circ \widehat{V} \circ \widehat{F} \\
&= \widehat{F_*V} \circ \widehat{F_*W} - \widehat{F_*W} \circ \widehat{F_*V} = [\widehat{F_*V}, \widehat{F_*W}].
\end{aligned}$$

Since the assignment  $V \mapsto \widehat{V}$  is an injection, this proves the vector field equality  $F_*[V, W] = [F_*V, F_*W]$ .

It makes sense (as in [4]) to define an operator  $\text{ad } \widehat{V}_t$  by

$$(2.1.27) \quad (\text{ad } \widehat{V}_t) \circ \widehat{W}_t = [\widehat{V}_t, \widehat{W}_t].$$

We will see below that operators  $\text{ad}$  and  $\text{Ad}$  can be related through an operator-valued differential equation.

Finally, let  $v \in T_qM$  and  $F : M \rightarrow M$  a map of class  $C^m$ . Then  $F_*(q)v$  is a tangent vector in  $T_{F(q)}M$  and it is natural to ask for  $\widehat{F_*v}$  in terms of  $\widehat{v}$  and  $\widehat{F}$ . We claim that, as in [4], one obtains

$$(2.1.28) \quad \widehat{F_*(q)v} = \widehat{v} \circ \widehat{F}.$$

To see this, let  $\varphi \in C^m(M, E)$ . Then, applying the chain rule we have  $\widehat{F_*v}(\varphi) = \varphi_*(F(q))F_*(q)v = (\varphi \circ F)_*(q)v = \widehat{v}(\varphi \circ F) = \widehat{v} \circ \widehat{F}(\varphi)$ .

## 2.2. Operator Differential Equations and Their Applications

In this section we further develop our extension in the direction of applications to flows of vector fields.

**2.2.1. Differential and Integral Operator Equations.** Following [2, 3, 4]

we have introduced the operator differential equation

$$(2.2.1) \quad \frac{d}{dt} \widehat{P}_{t_0,t} = \widehat{P}_{t_0,t} \circ \widehat{V}_t, \quad \widehat{P}_{t_0,t_0} = \widehat{Id}$$

which has a unique solution  $\widehat{P}_{t_0,t}$  representing the flow of diffeomorphisms for a nonautonomous vector field  $V_t$  which is continuous in  $t$ .

In the more general case of measurable in  $t$  vector-field  $V_t$  we have that  $\widehat{P}_{t_0,t}$  satisfies the integral operator equation

$$(2.2.2) \quad \widehat{P}_{t_0,t} = \widehat{Id}_M + \int_{t_0}^t \widehat{P}_{t_0,\tau} \circ \widehat{V}_\tau d\tau$$

and it is the unique absolutely continuous solution of this equation or the solution of the differential equation (2.2.1) in sense of distributions. The justification of this fact is based on the relation of  $\widehat{P}_{t_0,t}$  to the Carathéodory solutions of the *ordinary* differential equation (1.2.5).

Now we consider the differential operator equation

$$(2.2.3) \quad \frac{d}{dt} \widehat{Q}_{t_0,t} = -\widehat{V}_t \circ \widehat{Q}_{t_0,t}, \quad \widehat{Q}_{t_0,t_0} = \widehat{Id}_M$$

Note that this operator equation, even in case  $M = \mathbb{R}^n$ , is related to some first-order linear *partial* differential equation.

The following result states that for a locally integrable bounded  $C^m$  vector field  $V_t$  there exists a solution  $\widehat{Q}_{t_0,t}$  of (2.2.3) in the sense of distributions. Moreover we have a representation of  $\widehat{Q}_{t_0,t}$  in terms of a solution of the equation of the type (2.2.2).

PROPOSITION 2.2.1. *Let  $V_t$  be a locally integrable bounded  $C^m$  vector field. Then absolutely continuous operator-valued solutions  $\widehat{P}_{t_0,t}$  and  $\widehat{Q}_{t_0,t}$  of differential equations (2.2.1) and (2.2.3) exist, are unique, and*

$$(2.2.4) \quad \widehat{Q}_{t_0,t} = (\widehat{P}_{t_0,t})^{-1}.$$

*Proof.* Let  $P_{t_0,t}$  be the flow of  $V_t$ , so that (2.2.2) holds. It is enough to prove the existence and uniqueness of (2.2.3).

Denote by  $Q_{t_0,t}$  the flow of diffeomorphisms  $P_{t,t_0}$ . The operator-valued function  $t \rightarrow \widehat{Q}_{t_0,t}$  is absolutely continuous and together with  $\widehat{P}_{t_0,t}$  satisfies Assumption 2.1.6 for the product rule, Theorem 2.1.8.

Fix  $\varphi \in C^m(M, E)$  and  $q_0 \in M$ . There exists an interval  $(a, b)$  such that  $\widehat{P}_{t_0,t}(\varphi)(q_0)$  exists for any  $t_0, t$  in  $(a, b)$ . By the product rule we have for any  $t \in (a, b)$

$$\int_{t_0}^t \left( \frac{d}{d\tau} \widehat{P}_{t_0,\tau} \circ \widehat{Q}_{t_0,\tau} + \widehat{P}_{t_0,\tau} \circ \frac{d}{d\tau} \widehat{Q}_{t_0,\tau} \right) dt(\varphi)(q_0) = 0$$

This implies that for almost all  $t \in (a, b)$

$$\left( \widehat{P}_{t_0,t} \circ \widehat{V}_t \circ \widehat{Q}_{t_0,t} + \widehat{P}_{t_0,t} \circ \frac{d}{d\tau} \widehat{Q}_{t_0,t} \right) (\varphi)(q_0) = 0$$

and  $\widehat{Q}_{t_0,t}$  satisfies (2.2.3) in the sense of distributions.

To prove uniqueness of the solution  $\widehat{Q}_{t_0,t}$  we use the product rule (2.1.17)

$$\begin{aligned} \widehat{P}_{t_0,t} \circ \widehat{Q}_{t_0,t} - \widehat{Id}_M &= \int_{t_0}^t \frac{d}{d\tau} \left( \widehat{P}_{t_0,\tau} \circ \widehat{Q}_{t_0,\tau} \right) d\tau \\ &= \int_{t_0}^t \left( \widehat{P}_{t_0,\tau} \circ \widehat{V}_\tau \circ \widehat{Q}_{t_0,\tau} - \widehat{P}_{t_0,\tau} \circ \widehat{V}_\tau \circ \widehat{Q}_{t_0,\tau} \right) d\tau = 0 \end{aligned}$$

As a consequence, we have  $\widehat{P}_{t_0,t} \circ \widehat{Q}_{t_0,t} = \widehat{Id}_M$  for all  $t$ , hence  $\widehat{Q}_{t_0,t} = \widehat{P}_{t,t_0}$  which also proves (2.2.4).  $\square$



PROPOSITION 2.2.2. *Let  $V_t$  be locally integrable bounded  $C^m$  smooth vector field and  $\widehat{P}_{t_0,t}$  be an absolutely continuous solution of (2.2.2). Then for any  $C^m$  smooth vector field  $W$  the operator-valued function  $t \rightarrow \text{Ad } \widehat{P}_{t_0,t} \circ \widehat{W}$  is absolutely continuous and satisfies the following equation in the sense of distributions:*

$$(2.2.5) \quad \frac{d}{dt} \text{Ad } \widehat{P}_{t_0,t} \circ \widehat{W} = \text{Ad } \widehat{P}_{t_0,t} \circ \text{ad } \widehat{V}_t \circ \widehat{W}.$$

*Proof.* Note that  $\widehat{P}_{t_0,t}^{-1}$  exists and due to the assertion of Proposition 2.2.1 satisfies the differential equation (2.2.3). For any smooth vector field  $W$

$$(2.2.6) \quad \begin{aligned} \text{Ad } \widehat{P}_{t_0,t} \circ \widehat{W} &= \widehat{W} + \int_{t_0}^t \frac{d}{d\tau} \left( \widehat{P}_{t_0,\tau} \circ \widehat{W} \circ \left( \widehat{P}_{t_0,\tau} \right)^{-1} \right) d\tau \\ &= \widehat{W} + \int_{t_0}^t \left( \widehat{P}_{t_0,\tau} \circ \widehat{V}_\tau \circ \widehat{W} \circ \left( \widehat{P}_{t_0,\tau} \right)^{-1} - \widehat{P}_{t_0,\tau} \circ \widehat{W} \circ \widehat{V}_\tau \circ \widehat{P}_{t_0,\tau}^{-1} \right) d\tau \\ &= \widehat{W} + \int_{t_0}^t \left( \text{Ad } \widehat{P}_t \right) \circ [\widehat{V}_t, \widehat{W}] d\tau. \end{aligned}$$

Recall the definition (2.1.27) of the operator  $\text{ad } \widehat{V}_t$  to conclude the proof.  $\square$

The method of variation of parameters can also be easily applied to the operator differential equation

$$(2.2.7) \quad \frac{d}{dt} \widehat{S}_{t_0,t} = \widehat{S}_{t_0,t} \circ (\widehat{V}_t + \widehat{W}_t), \quad \widehat{S}_{t_0,t_0} = \widehat{Id}$$

In particular we have the following Proposition:

PROPOSITION 2.2.3. *Let  $V_t, W_t$  be locally integrable bounded  $C^m$  smooth vector fields. Then a solution of (2.2.7) can be represented in the form*

$$(2.2.8) \quad \widehat{S}_{t_0,t} = \widehat{C}_{t_0,t} \circ \widehat{P}_{t_0,t},$$

where  $\widehat{P}_{t_0,t}$  is the solution of the differential equation (2.2.1) and  $\widehat{C}_{t_0,t}$  is a solution of the differential equation

$$(2.2.9) \quad \frac{d}{dt} \widehat{C}_{t_0,t} = \widehat{C}_{t_0,t} \circ \text{Ad } \widehat{P}_{t_0,t} \circ \widehat{W}_t, \quad \widehat{C}_{t_0,t_0} = \widehat{Id}_M$$

PROOF. It follows from (2.2.8) and Proposition 2.2.1 that  $\widehat{C}_{t_0,t}$  is absolutely continuous and by the product rule

$$\begin{aligned} \widehat{C}_{t_0,t} - \widehat{Id}_M &= \int_{t_0}^t \left( \frac{d}{d\tau} \widehat{S}_{t_0,\tau} \circ \widehat{P}_{t_0,\tau}^{-1} + \widehat{S}_{t_0,\tau} \circ \frac{d}{d\tau} \widehat{P}_{t_0,\tau}^{-1} \right) d\tau \\ &= \int_{t_0}^t \left( \widehat{C}_{t_0,\tau} \circ \widehat{P}_{t_0,\tau} \circ (\widehat{V}_\tau + W_\tau) \circ \widehat{P}_{t_0,\tau}^{-1} - \widehat{C}_{t_0,\tau} \circ \widehat{P}_{t_0,\tau} \circ \widehat{V}_\tau \circ \widehat{P}_{t_0,\tau}^{-1} \right) d\tau \\ &= \int_{t_0}^t \widehat{C}_{t_0,\tau} \circ \widehat{P}_{t_0,\tau} \circ W_\tau \circ \widehat{P}_{t_0,\tau}^{-1} d\tau = \int_{t_0}^t \widehat{C}_{t_0,\tau} \circ \text{Ad } \widehat{P}_{t_0,\tau} \circ \widehat{W}_\tau d\tau \end{aligned}$$

which proves (2.2.9).  $\square$

**2.2.2. Derivatives of Flows with Respect to a Parameter.** Consider a family of nonautonomous  $C^m$  vector field  $V_t^\alpha$  which depends upon scalar parameter  $\alpha$  and corresponding flow  $P_{t_0,t}^\alpha$ . Let us assume that  $V_t^\alpha$  is differentiable in  $\alpha$  in the following sense:

$$V_t^\alpha = V_t + \alpha W_t + o_t(\alpha),$$

where  $V_t, W_t$  are nonautonomous  $C^m$  vector fields which are locally integrable bounded.

Let  $\widehat{P}_{t_0,t}$  and  $\widehat{Q}_{t_0,t}$  be absolutely continuous solutions of (2.2.1) and (2.2.3). We assume that the operator  $\widehat{o}_t(\alpha)$  satisfies (2.1.2) along with

$$\widehat{P}_{t_0,t} \circ \widehat{o}_t(\alpha) \circ \widehat{Q}_{t_0,t} = \widehat{o}(\alpha) \quad \widehat{V}_t \circ \widehat{o}_t(\alpha) \circ \widehat{W}_t = \widehat{o}(\alpha),$$

uniformly with respect to  $t \in [t_0, t_1]$ .

We use these assumptions and (2.2.9) from Proposition 2.2.3 to obtain

$$(2.2.10) \quad \widehat{P}_{t_0,t}^\alpha = \widehat{C}_{t_0,t}^\alpha \circ \widehat{P}_{t_0,t},$$

where

$$(2.2.11) \quad \widehat{C}_{t_0,t}^\alpha = \widehat{Id}_M + \alpha \int_{t_0}^t \text{Ad } \widehat{P}_{t_0,\tau} \circ \widehat{W}_\tau d\tau + \widehat{o}(\alpha).$$

Hence

$$(2.2.12) \quad \widehat{P}_{t_0,t}^\alpha = \widehat{P}_{t_0,t} + \alpha \int_{t_0}^t \text{Ad } \widehat{P}_{t_0,\tau} \circ \widehat{W}_\tau d\tau \circ \widehat{P}_{t_0,t} + \widehat{o}(\alpha).$$

This implies that  $\widehat{P}_{t_0,t}^\alpha$  is differentiable with respect to  $\alpha$  at  $\alpha = 0$  and that

$$(2.2.13) \quad \frac{\partial}{\partial \alpha} \widehat{P}_{t_0,t}^\alpha = \int_{t_0}^t \text{Ad } \widehat{P}_{t_0,\tau} \circ \widehat{W}_\tau d\tau \circ \widehat{P}_{t_0,t}.$$

This same formula is given in the context of the classical Chronological Calculus [4]. We invite the reader to check that the second representation found in [4], given below

$$(2.2.14) \quad \frac{\partial}{\partial \alpha} \widehat{P}_{t_0,t}^\alpha = \widehat{P}_{t_0,t} \circ \int_{t_0}^t \text{Ad } \widehat{P}_{t,\tau} \circ \widehat{W}_\tau d\tau$$

is easily obtained from the first.

The operator formulas (2.2.13) and (2.2.14) can be used in order to obtain the following representations for derivative of the flow  $P_{t_0,t}^\alpha$  at  $\alpha = 0$

$$(2.2.15) \quad \frac{\partial}{\partial \alpha} P_{t_0,t}^\alpha(q) = P_{t_0,t*}(q) \int_{t_0}^t P_{\tau,t*}(P_{t_0,\tau}(q)) W_\tau(P_{t_0,\tau}(q)) d\tau,$$

$$(2.2.16) \quad \frac{\partial}{\partial \alpha} P_{t_0,t}^\alpha(q) = \int_{t_0}^t P_{\tau,t*}(P_{t_0,\tau}(q)) W_\tau(P_{t_0,\tau}(q)) d\tau.$$

Here we prove (2.2.16). The proof of (2.2.15) is similar. Using (2.2.14) and (2.1.28), we obtain for any  $\varphi \in C^m(M, E)$ , the following relations:

$$\begin{aligned}
(2.2.17) \quad \varphi_*(P_{t_0,t}(q)) \frac{\partial}{\partial \alpha} P_{t_0,t}^\alpha(q) &= \widehat{q} \circ \frac{\partial}{\partial \alpha} \widehat{P}_{t_0,t}^\alpha(\varphi) = \int_{t_0}^t \widehat{q} \circ \widehat{P}_{t_0,t} \circ \text{Ad} \widehat{P}_{t,\tau} \circ \widehat{W}_\tau d\tau(\varphi) \\
&= \int_{t_0}^t \widehat{P}_{t_0,t}(q) \circ \widehat{P}_{\tau,t*} \widehat{W}_\tau(\varphi) d\tau \\
&= \int_{t_0}^t \varphi_*(P_{t_0,t}(q)) P_{\tau,t*}(P_{t_0,\tau}(q)) W_\tau(P_{t_0,\tau}(q)) d\tau
\end{aligned}$$

These relations imply (2.2.16).

### 2.2.3. Finite Sums of Volterra Series and a Remainder Term Estimate.

Let  $V_t$  be a nonautonomous  $C^m$  vector field on  $M$  and let  $P_{t_0,t} : J_0 \times U_0 \rightarrow U$  be the local flow of this field. Consider the operator integral equation (2.2.2). Replacing  $\widehat{P}_{t_0,\tau}$  in (2.2.2) with its integral form, we obtain

$$\widehat{P}_{t_0,t} = \widehat{I}d_M + \int_{t_0}^t \widehat{V}_\tau d\tau + \int_{t_0}^t \int_{t_0}^\tau \widehat{P}_{t_0,\sigma} \circ \widehat{V}_\sigma \circ \widehat{V}_\tau d\sigma d\tau.$$

Provided that  $k \leq m$ , we may continue to obtain

$$(2.2.18) \quad \widehat{P}_{t_0,t} = \widehat{I}d_M + \sum_{i=1}^{k-1} \int_{\Delta_i(t)} \widehat{V}_{\tau_i} \circ \cdots \circ \widehat{V}_{\tau_1} d\tau_i \dots d\tau_1 + \widehat{R}_k(t)$$

where

$$(2.2.19) \quad \widehat{R}_k(t) := \int_{\Delta_k(t)} \widehat{P}_{t_0,\tau_k} \circ \widehat{V}_{\tau_k} \circ \cdots \circ \widehat{V}_{\tau_1} d\tau_k \dots d\tau_1,$$

is the remainder term and  $\Delta_k(t)$  is the simplex  $\{t_0 \leq \tau_k \leq \tau_{k-1} \leq \cdots \leq \tau_1 \leq t\}$ .

Suppose that for any  $\varphi \in C^m(M, E)$  and  $q_0 \in M$  there is a neighborhood  $U$  of  $q_0$ ,  $\delta > 0$  and a constant  $C$  such that for all  $q \in U$ , for any  $t_0 \leq \tau_k \leq \cdots \leq \tau_1 \leq t \leq t_0 + \delta$ ,

we have

$$(2.2.20) \quad \left\| \widehat{V}_{\tau_k} \circ \cdots \circ \widehat{V}_{\tau_1}(\varphi)(q) \right\|_E \leq C$$

This is true, for example, when  $V_t$  is locally bounded or autonomous. Then

$$(2.2.21) \quad \begin{aligned} \left\| \widehat{R}_k(t)(\varphi)(q) \right\|_E &\leq \int_{\Delta_k(t)} \left\| \widehat{V}_{\tau_k} \circ \cdots \circ \widehat{V}_{\tau_1}(\varphi)(P_{t_0, \tau_k}(q)) \right\|_E d\tau_k \cdots d\tau_1 \\ &\leq \int_{\Delta_k(t)} C d\tau_k \cdots d\tau_1 = \frac{Ct^k}{k!}. \end{aligned}$$

It follows for any  $\varphi \in C^m(M, E)$  and any  $q_0 \in M$ , there is a neighborhood  $U$  of  $q_0$  on which the function  $\frac{1}{t^{k-1}} R_k(t)(\varphi)$  converges uniformly to zero. The family  $\widehat{R}_k(t)$  is an important example of a  $\widehat{o}(t^{k-1})$  family of operators. In the following section, we rigorously define such families, establish their properties, and prove that  $\widehat{R}_k(t) = \widehat{o}(t^{k-1})$ .

### 2.3. Calculus of Little $o$ 's

We develop in this section the theory of operators of type  $\widehat{o}(t^\ell)$  for  $C^m$ -smooth manifolds  $M$ . This theory is developed in detail for the case in which vector fields and flows are merely  $C^k$ -smooth. We then provide details for the case of  $C^\infty$ -vector fields and flows.

**2.3.1. Calculus of Little  $o$ 's:  $C^k$ -smooth Case.** We need the following definitions:

DEFINITION 2.3.1. A set  $\mathcal{F} \subset C^m(M, E)$  is called *locally bounded* at  $q_0 \in M$  if there exists a coordinate chart  $(\mathcal{O}, \psi)$  with  $q_0 \in \mathcal{O}$  and a constant  $C$  such that for any  $i = 0, \dots, m$

$$(2.3.1) \quad \sup_{x \in \psi(\mathcal{O})} \left\| (\varphi \circ \psi^{-1})^{(i)}(x) \right\| \leq C$$

DEFINITION 2.3.2. We say that a family of operators  $\widehat{A}_t : C^m(M, E) \rightarrow C^{m'}(M, E)$ ,  $t \in (-\delta, \delta)$  has defect  $k_1 := \text{def } \widehat{A}_t$  if for any  $n$ ,  $k_1 \leq n \leq m$  and any  $\varphi \in C^n(M, E)$  we have  $\widehat{A}_t \varphi \in C^{n-k_1}(M, E)$ . A smooth vector field  $V_t$  gives an example of the operator  $\widehat{V}_t$  which has defect 1.

DEFINITION 2.3.3. A family of operators  $\widehat{A}_t : C^m(M, E) \rightarrow C^{m'}(M, E)$ ,  $0 < |t| < \delta$ , with defect  $k_1$  is called  $\widehat{o}(t^k)$  if for any  $q_0 \in M$  and locally bounded at  $q_0$  set  $\mathcal{F} \subset C^m(M, E)$  there exists a coordinate chart  $(\mathcal{O}, \psi)$  with  $q_0 \in \mathcal{O}$  such that for any  $i = 0, \dots, m - k_1$

$$(2.3.2) \quad \lim_{t \rightarrow 0} \frac{1}{t^k} \left\| (\widehat{A}_t(\varphi) \circ \psi^{-1}(x))^{(i)} \right\| = 0$$

uniformly with respect to all  $\varphi \in \mathcal{F}$ ,  $x \in \psi(\mathcal{O})$ .

The following proposition gives an important example of  $\widehat{o}(t^k)$  operator.

PROPOSITION 2.3.4. *Let  $C^m$ -smooth vector field  $V_t$  be locally bounded. Then the remainder term operator  $\widehat{R}_k(t)$  given by (2.2.19) is a  $\widehat{o}(t^{k-1})$  operator with defect at most  $k$ .*

*Proof.* Fix  $q_0 \in M$  and a locally bounded at  $q_0$  family of functions  $\mathcal{F} \subset C^m(M, E)$ . Then there exists a constant  $C$  such that (2.2.20) holds for any  $\varphi \in \mathcal{F}$ . This implies that (2.2.21) holds for any  $q \in \mathcal{O}$ , where  $\mathcal{O}$  is some neighbourhood of  $q_0$ . This proves uniform convergence (2.3.2) for  $i = 0$ . A similar argument will establish the uniform convergence (2.3.2) for any  $i = 0, \dots, m - k$ .  $\square$

Later we'll use the following properties of operators  $\widehat{o}(t^k)$ .

PROPOSITION 2.3.5. *Let  $\widehat{o}(t^k)$  and  $\widehat{o}(t^\ell)$  be families of operators with defects  $k_1$  and  $\ell_1$  respectively. Then*

$$(i) \quad \widehat{o}(t^k) + \widehat{o}(t^\ell) = \widehat{o}(t^{\min(k, \ell)}) \text{ if } \max\{k_1, \ell_1\} \leq m;$$

- (ii)  $\widehat{\partial}(t^k) \circ \widehat{\partial}(t^\ell) = \widehat{\partial}(t^{k+\ell})$  if  $k_1 + \ell_1 \leq m$ ;
- (iii)  $\widehat{X}_t \circ \widehat{\partial}(t^k) \circ \widehat{Y}_t = \widehat{\partial}(t^k)$  with defect at most  $k_1 + 2$  if  $k_1 \leq m - 2$  and  $C^m$  smooth vector fields  $X_t, Y_t$  are locally bounded;
- (iv)  $\widehat{P}_{0,t} \circ \widehat{\partial}(t^k) \circ \widehat{Q}_{0,t} = \widehat{\partial}(t^k)$  with defect at most  $k_1$  if  $P_{0,t}$  and  $Q_{0,t}$  are flows of locally bounded  $C^m$  smooth vector fields;
- (v)  $\widehat{P}_{0,t} = \widehat{Id}_E + \widehat{\partial}(1)$  with defect 0 if  $\widehat{P}_{0,t}$  is the flow of a locally bounded  $C^m$  smooth vector field.

*Proof.* Let us fix  $q_0 \in M$ , a locally bounded at  $q_0$  family  $\mathcal{F} \subset C^m(M, E)$ , and a coordinate chart  $(\mathcal{O}, \psi)$ . Observe that for any  $\varphi \in \mathcal{F}$ ,  $x \in \psi(\mathcal{O})$  and  $i = 0, \dots, m - \max\{k_1, \ell_1\}$

$$\left\| ((\widehat{\partial}(t^k) + \widehat{\partial}(t^\ell))(\varphi) \circ \psi^{-1}(x))^{(i)} \right\| \leq \left\| (\widehat{\partial}(t^k)(\varphi) \circ \psi^{-1}(x))^{(i)} \right\| + \left\| (\widehat{\partial}(t^\ell)(\varphi) \circ \psi^{-1}(x))^{(i)} \right\|$$

This inequality implies (i).

Note that the family of functions  $\mathcal{B} := \left\{ \frac{1}{t^\ell} \widehat{\partial}(t^\ell)(\varphi) : \varphi \in \mathcal{F}, 0 < |t| < \delta \right\}$  is locally bounded at  $q_0$ . Then (ii) follows immediately from Definition 2.3.3.

To prove (iii) we note that  $\widehat{\partial}(t^k) \circ \widehat{Y}_t$  is  $\widehat{\partial}(t^k)$  with defect  $k_1 + 1$ . Then it is easy to see that  $\widehat{X}_t \circ \widehat{\partial}(t^k) \circ \widehat{Y}_t$  is  $\widehat{\partial}(t^k)$  with defect  $k_1 + 2$ . The assertion (iv) follows from the observation that  $\widehat{\partial}(t^k) \circ \widehat{Q}_{0,t} = \widehat{\partial}(t^k)$  with defect  $k_1$  and from the uniform convergence in (2.3.2).

The last assertion (v) follows from the integral representation (2.2.2) for  $\widehat{P}_{0,t}$  and boundedness assumptions.  $\square$

**2.3.2. Calculus of Little  $o$ 's:  $C^\infty$ -smooth Case.** In the case of a  $C^\infty$ -smooth manifold  $M$  we won't require the concept of defect of operators which map  $\varphi \in C^\infty(M, E)$  into  $C^\infty(M, E)$ . We make the following modifications to the above definitions in this case:

DEFINITION 2.3.6. For  $C^\infty$  manifold  $M$  a set  $\mathcal{F} \subset C^\infty(M, E)$  is called *locally bounded* at  $q_0 \in M$  if for any natural number  $m$  there exists a coordinate chart  $(\mathcal{O}, \psi)$  with  $q_0 \in \mathcal{O}$  and a constant  $C$  such that (2.3.1) holds for any  $i = 0, \dots, m$ .

DEFINITION 2.3.7. For a  $C^\infty$  manifold  $M$ , a family of operators  $\widehat{A}_t : C^\infty(M, E) \rightarrow C^\infty(M, E)$ ,  $t \in (-\delta, \delta)$ , is called  $\widehat{o}(t^k)$  if for any  $q_0 \in M$ , for any locally bounded at  $q_0$  set  $\mathcal{F} \subset C^\infty(M, E)$ , and for any natural number  $m$  there exists a coordinate chart  $(\mathcal{O}, \psi)$  with  $q_0 \in \mathcal{O}$  such that for any  $i = 0, \dots, m$  the limit (2.3.2) takes place uniformly with respect to all  $\varphi \in \mathcal{F}$ ,  $x \in \psi(\mathcal{O})$ .

Then we have

PROPOSITION 2.3.8. For  $C^\infty$  manifold  $M$  and locally bounded  $C^\infty$  vector fields  $X_t$  and  $Y_t$  assertions (i)-(v) of Proposition 2.3.5 hold with  $k_1 = 0$ ,  $\ell_1 = 0$  and  $m = \infty$ .

## 2.4. Commutators of Flows and Vector Fields

Let  $P_t$  and  $Q_t$  be flows on a  $C^m$  manifold  $M$ , generated by  $C^m$  vector fields  $X$  and  $Y$ , so that  $\widehat{P}_0 = \widehat{Id}$ ,  $\widehat{Q}_0 = \widehat{Id}$ ,

$$\frac{d\widehat{P}_t}{dt} = \widehat{P}_t \circ \widehat{X}, \quad \frac{d\widehat{Q}_t}{dt} = \widehat{Q}_t \circ \widehat{Y}.$$

Following [68], we define a *bracket of flows*  $[P_t, Q_t] = Q_t^{-1} \circ P_t^{-1} \circ Q_t \circ P_t$  and we note that

$$[\widehat{P}_t, \widehat{Q}_t] = \widehat{P}_t \circ \widehat{Q}_t \circ \widehat{P}_t^{-1} \circ \widehat{Q}_t^{-1}.$$

In the case of finite dimensional manifolds, it follows from the classical result that

$$(2.4.1) \quad [\widehat{P}_t, \widehat{Q}_t] = \widehat{Id} + t^2 [\widehat{X}, \widehat{Y}] + \widehat{o}(t^2).$$

In the case of infinite dimensional manifolds and flows  $P_t^i$ ,  $i = 1, \dots, k$ , generated by vector fields  $X_i$ , the general formula for an arbitrary bracket expression



$B(P_t^1, \dots, P_t^k)$  was proved by Mauhart and Michor in [68]. In operator notation, the general formula is

$$(2.4.2) \quad B(\widehat{P_t^1, \dots, P_t^k}) = \widehat{Id} + t^k B(\widehat{X}_1, \dots, \widehat{X}_k) + \widehat{o}(t^k).$$

Here we use the Chronological Calculus to prove this formula. In particular, we will establish the following:

**THEOREM 2.4.1** (Mauhart and Michor). *Let  $M$  be an  $C^m$ -smooth Banach manifold and  $X_1, \dots, X_k$ ,  $k \leq m$ , be  $C^m$ -smooth vector fields. Then for any bracket expression  $B(P_t^1, \dots, P_t^k)$  we have the presentation (2.4.2) where  $\widehat{o}(t^k)$  has defect at most  $k - 1$ .*

The advantage of our approach follows from the fact that the main part of the proof is reduced to algebraic computations. Moreover, an algorithm for deriving a representation for remainder term in (2.4.2) is given.

We will need the following results for families of local diffeomorphisms of the form

$$(2.4.3) \quad \widehat{P}_t = \widehat{Id} + t^m \widehat{X} + \widehat{o}(t^m) \quad \widehat{Q}_t = \widehat{Id} + t^n \widehat{Y} + \widehat{o}(t^n).$$

**PROPOSITION 2.4.2.** *Let  $X$  be a  $C^m$  vector field. Then*

$$(2.4.4) \quad \widehat{P}_t^{-1} = \widehat{Id} - t^m \widehat{X} + \widehat{o}(t^m).$$

**PROOF.** Consider flows  $S_t$  and  $T_t$  defined by

$$(2.4.5) \quad \frac{d\widehat{S}_t}{dt} = \widehat{S}_t \circ \widehat{X}, \quad \widehat{S}_0 = \widehat{Id} \quad \frac{d\widehat{T}_t}{dt} = -\widehat{X} \circ \widehat{T}_t, \quad \widehat{T}_0 = \widehat{Id}.$$

Then  $\widehat{T}_t = \widehat{S}_t^{-1}$ ,  $\widehat{S}_t = \widehat{Id} + t\widehat{X} + \widehat{o}(t)$ , and  $\widehat{T}_t = \widehat{Id} - t\widehat{X} + \widehat{o}(t)$ . In particular,  $\widehat{P}_t = \widehat{S}_t^m + \widehat{o}(t^m)$ .

By applying  $\widehat{P}_t^{-1}$  from the right, we get  $\widehat{Id} = \widehat{S}_{t^m} \circ \widehat{P}_t^{-1} + \widehat{o}(t^m)$ . Then by applying  $\widehat{T}_{t^m}$  from the left, we get  $\widehat{T}_{t^m} = \widehat{P}_t^{-1} + \widehat{o}(t^m)$ , and so  $\widehat{P}_t^{-1} = \widehat{Id} - t^m \widehat{X} + \widehat{o}(t^m)$ .  $\square$

PROPOSITION 2.4.3. *Let families of local diffeomorphisms  $P_t$  and  $Q_t$  satisfy (2.4.3). Then*

$$(2.4.6) \quad [\widehat{P}_t, \widehat{Q}_t] = \widehat{Id} + t^{m+n} [\widehat{X}, \widehat{Y}] + \widehat{o}(t^{m+n}).$$

PROOF. Recall that  $[\widehat{P}_t, \widehat{Q}_t] := \widehat{P}_t \circ \widehat{Q}_t \circ \widehat{P}_t^{-1} \circ \widehat{Q}_t^{-1}$ . Write

$$\widehat{P}_t^{-1} = \widehat{Id}_M - \widehat{V}_1, \quad \widehat{P}_t = \widehat{Id}_M + \widehat{V}_2, \quad \widehat{Q}_t^{-1} = \widehat{Id}_M - \widehat{W}_1, \quad \widehat{Q}_t = \widehat{Id}_M + \widehat{W}_2.$$

Then

$$(2.4.7) \quad [\widehat{P}_t, \widehat{Q}_t] = (\widehat{Id}_M + \widehat{V}_2) \circ \widehat{Q}_t \circ (\widehat{Id}_M - \widehat{V}_1) \circ \widehat{Q}_t^{-1}.$$

Now,  $\widehat{Q}_t \circ (\widehat{Id}_M - \widehat{V}_1) \circ \widehat{Q}_t^{-1} = \widehat{Id}_M - \widehat{Q}_t \circ \widehat{V}_1 \circ \widehat{Q}_t^{-1} = \widehat{Id}_M - (\widehat{Id}_M + \widehat{W}_2) \circ \widehat{V}_1 \circ (\widehat{Id}_M - \widehat{W}_1) = \widehat{Id}_M - \widehat{V}_1 - \widehat{W}_2 \circ \widehat{V}_1 + \widehat{V}_1 \circ \widehat{W}_1 + \widehat{W}_2 \circ \widehat{V}_1 \circ \widehat{W}_1$ .

Substituting this expression into (2.4.7) gives

$$\begin{aligned} [\widehat{P}_t, \widehat{Q}_t] &= (\widehat{Id}_M + \widehat{V}_2) \circ (\widehat{Id}_M - \widehat{V}_1 - \widehat{W}_2 \circ \widehat{V}_1 + \widehat{V}_1 \circ \widehat{W}_1 + \widehat{W}_2 \circ \widehat{V}_1 \circ \widehat{W}_1) \\ &= \widehat{Id}_M - \widehat{V}_1 - \widehat{W}_2 \circ \widehat{V}_1 + \widehat{V}_1 \circ \widehat{W}_1 + \widehat{W}_2 \circ \widehat{V}_1 \circ \widehat{W}_1 + \widehat{V}_2 - \widehat{V}_2 \circ \widehat{V}_1 \\ &\quad - \widehat{V}_2 \circ \widehat{W}_2 \circ \widehat{V}_1 + \widehat{V}_2 \circ \widehat{V}_1 \circ \widehat{W}_1 + \widehat{V}_2 \circ \widehat{W}_2 \circ \widehat{V}_1 \circ \widehat{W}_1 \end{aligned}$$

But

$$-\widehat{V}_1 + \widehat{V}_2 - \widehat{V}_2 \circ \widehat{V}_1 = \widehat{P}_t^{-1} - \widehat{Id}_M + \widehat{P}_t - \widehat{Id}_M - (\widehat{P}_t - \widehat{Id}_M) \circ (\widehat{Id}_M - \widehat{P}_t^{-1}) = 0.$$

Therefore,

$$[\widehat{P}_t, \widehat{Q}_t] = \widehat{Id}_M - \widehat{W}_2 \circ \widehat{V}_1 + \widehat{V}_1 \circ \widehat{W}_1 + \widehat{R}$$

where

$$(2.4.8) \quad \widehat{R} := \widehat{W}_2 \circ \widehat{V}_1 \circ \widehat{W}_1 - \widehat{V}_2 \circ \widehat{W}_2 \circ \widehat{V}_1 + \widehat{V}_2 \circ \widehat{V}_1 \circ \widehat{W}_1 + \widehat{V}_2 \circ \widehat{W}_2 \circ \widehat{V}_1 \circ \widehat{W}_1$$

By (2.4.3) and (2.4.4) we have  $\widehat{V}_1 = t^m \widehat{X} + \widehat{o}(t^m)$ ,  $\widehat{V}_2 = t^m \widehat{X} + \widehat{o}(t^m)$ ,  $\widehat{W}_1 = t^n \widehat{Y} + \widehat{o}(t^n)$ ,  $\widehat{W}_2 = t^n \widehat{Y} + \widehat{o}(t^n)$ .

By using Proposition 2.4.3 we obtain from previous relations and (2.4.8) that  $\widehat{R} = \widehat{o}(t^{m+n})$  and

$$\widehat{V}_1 \circ \widehat{W}_1 - \widehat{W}_2 \circ \widehat{V}_1 = t^{m+n} [\widehat{X}, \widehat{Y}] + \widehat{o}(t^{m+n}).$$

This proves (2.4.6). □

Applying 2.4.3 inductively, we obtain Theorem 2.4.1. Note that in the process of proving the theorem, we have obtained an expression for the remainder term.

## 2.5. Chow-Rashevskii Theorem for Infinite-Dimensional Manifolds

Consider an  $n$ -dimensional manifold  $M$  with a sub-riemannian *distribution*  $\mathcal{H} \subset TM$ , which, by definition, is a vector sub-bundle of the tangent bundle  $TM$  of the manifold with an inner product on its fiber space [72]. An absolutely continuous curve  $q : [0, T] \rightarrow M$  is called *horizontal*, if its derivative belongs to  $\mathcal{H}$  for almost all  $t$ .

The classical Chow-Rashevskii theorem [22, 75] provides conditions in terms of basis vector fields  $\{V_i\}_{i=1, \dots, m}$  of the distribution  $\mathcal{H}$  and their iterated Lie brackets for connectivity of arbitrary two points of the sub-riemannian manifold by a horizontal curve.

More precisely, consider a distribution  $\mathcal{L}$  which is defined pointwise as the linear span of the set generated by iterated Lie brackets of basis vector fields  $\{V_i\}_{i=1, \dots, m}$

as follows:

$$(2.5.1) \quad \mathcal{L}[V_1, \dots, V_m](q) := \text{span} \{B(V_{i_1}, V_{i_2}, \dots, V_{i_{k-1}}, V_{i_k})(q) : k = 1, 2, \dots\}.$$

The classical Chow-Rashevskii theorem states that the condition

$$(2.5.2) \quad \mathcal{L}[V_1, \dots, V_m](q) = TM(q) \quad \forall q \in M$$

implies the connectivity of any two points on the manifold  $M$  by a horizontal curve.

Historically this theorem has played a fundamental role in nonlinear control theory [12, 32, 53, 54] by demonstrating that the condition (2.5.2) is a sufficient for the global controllability of the following affine-control system:

$$(2.5.3) \quad \dot{q} = \sum_{i=1}^m u_i(t) V_i(q).$$

Here we are interested in generalizing these sufficient conditions for global controllability for the case of infinite-dimensional manifold  $M$ . Consider an affine control system

$$(2.5.4) \quad \dot{q} = \sum_{i=1}^{\infty} u_i(t) V_i(q),$$

where  $V_i$  are smooth vector-fields on  $M$ , and  $u(t) := (u_1(t), u_2(t), \dots)$  is a control.

Let  $M$  be an infinite-dimensional  $C^\infty$  smooth connected manifold [61] with underlying smooth Banach space  $E$ .

A control  $u(t)$  will be called *admissible* if it is piecewise constant and at each  $t$  only a finite number of its components  $u_i(t)$  are different from zero and take values  $+1$  or  $-1$ . The set of all admissible controls will be denoted  $\mathcal{U}$ .

Note that for any initial point  $q_0$  for any admissible control  $u(t)$  there exists (at least locally) a unique solution  $q(t; q_0, u)$  of the control system (2.5.4). This solution

we call a *trajectory*. The *reachability set* for the initial point  $q_0$

$$(2.5.5) \quad \mathcal{R}(q_0) := \{q(t; u, q_0) : \forall t \geq 0, \forall u \in \mathcal{U}\}$$

consists of all points of all trajectories of (2.5.4) corresponding to all admissible controls  $u \in \mathcal{U}$ . Thus, the set  $\mathcal{R}(q_0)$  consists of all points to which the control system can be driven from the point  $q_0$  using admissible controls.

Here we provide infinitesimal conditions in terms of the vector fields  $\{V_i\}_{i=1,2,\dots}$  and their iterated Lie brackets similar to (2.5.2) which imply global approximate controllability of the system (2.5.4).

DEFINITION 2.5.1. Control system (2.5.4) is called global approximate controllable if for any  $q_0 \in M$

$$(2.5.6) \quad \overline{\mathcal{R}(q_0)} = M.$$

Thus, global approximate controllability of system (2.5.4) means that for arbitrary points  $q_0, q_1 \in M$  and any open neighbourhood  $\mathcal{O}$  of the point  $q_1$  there exists an admissible control  $u \in \mathcal{U}$  such that at some moment  $T$  the trajectory  $x(T; u, x_0)$  enters the neighbourhood  $\mathcal{O}$ .

For the family of smooth vector fields  $V_i, i = 1, 2, \dots$  define the following distribution:

$$(2.5.7) \quad \mathcal{L}[V_1, V_2, \dots](q) := \text{span} \{B(V_{i_1}, V_{i_2}, \dots, V_{i_{k-1}}, V_{i_k})(q) : k = 1, 2, \dots\}$$

THEOREM 2.5.2. Let  $M$  be an infinite-dimensional smooth manifold associated with the smooth Banach space  $E$  and a smooth affine-control system (2.5.4) satisfies

$$(2.5.8) \quad \overline{\mathcal{L}[V_1, V_2, \dots]}(q) = T_q M \quad \forall q \in M$$

Then system (2.5.4) is globally approximate controllable.

The proof of this variant of Chow-Rashevskii theorem for infinite-dimensional manifolds is based on the use of nonsmooth analysis [28, 80] and a characterization of the property of strong invariance [28] of sets with respect to solutions of the control system (2.5.4) and is similar to the proof of the analogous result for the case in which  $E$  is a Hilbert space [64].

**2.5.1. Nonsmooth Analysis on Smooth Manifolds and Strong Invariance of Sets.** The concepts of *strong* and *weak invariance* play important role in control theory. See [28] for finite-dimensional results and [27] for related results on approximate invariance in Hilbert spaces. A set  $S \subset M$  is called *strongly invariant* with respect to trajectories of a control system (2.5.4) if for any  $q_0 \in S$  and any admissible control  $u \in \mathcal{U}$  the trajectory  $q(t; q_0, u)$  stays in  $S$  for all  $t > 0$  sufficiently small. Note that the fact that the reachability set  $\mathcal{R}(q_0)$  (2.5.5) is strongly invariant follows immediately from its definition. Here we provide infinitesimal conditions for strong invariance of a closed set  $S$  in terms of normal vectors to  $S$  and iterated Lie brackets of vector fields  $V_i, i = 1, 2, \dots$

PROPOSITION 2.5.3. *Let  $q' \in S$  be a boundary point of the closed set  $S$ . Then any neighbourhood  $\mathcal{O}$  of  $q'$  contains a point  $q \in S$  such that there exists a normal vector  $\zeta \neq 0, \zeta \in N_q S$ .*

The proof of this proposition follows from the multidirectional mean-value inequality and is left to a reader.

THEOREM 2.5.4. *If the closed set  $S \subset M$  is strongly invariant with respect to solutions of the control system (2.5.4) then*

$$(2.5.9) \quad \langle \zeta, B(V_{i_1}, V_{i_2}, \dots, V_{i_{k-1}}, V_{i_k})(q) \rangle = 0$$

for any iterated Lie bracket of vector fields  $V_i$ , any normal vector  $\zeta \in N_q S$  and any  $q \in S$ .

*Proof.* Let us assume that the set  $S$  is strongly invariant,  $q \in S$  and  $\zeta = dg(q) \in N_q S$ . This implies that

$$(2.5.10) \quad \chi_S(q') - \chi_S(q) \geq g(q') - g(q)$$

for some smooth function  $g$  and all  $q'$  near  $q$ .

Let us fix some iterated Lie bracket  $B(V_{i_1}, V_{i_2}, \dots, V_{i_{k-1}}, V_{i_k})(q)$  as in (2.5.9), denote it  $v$  and relate to it an appropriate iterated flow bracket as in Section 4. For arbitrarily small  $t > 0$  we can find an admissible control  $u \in \mathcal{U}$  associated with this iterated flow bracket such that we have in accordance with Theorem (2.4.1)

$$g(q(t; q, u)) = g(q) + t^k \langle dg(q), v \rangle + o(t^k)$$

Then we obtain from (2.5.10) that

$$t^k \langle dg(q), v \rangle + o(t^k) \leq 0$$

Of course, we can easily derive (2.5.9) from this inequality.

□

### 2.5.2. Proof of an Infinite-Dimensional Variant of the Chow-Rashevskii

**Theorem.** Consider the reachability set  $\mathcal{R}(q_0)$  (2.5.5) and recall that this set is strongly invariant. Note that for a fixed admissible control  $u$ , the function  $q \rightarrow q(t; q, u)$  is continuous. This implies that the closure of the reachability set  $\overline{\mathcal{R}(q_0)}$  is also strongly invariant.

Now let us assume that Theorem 2.5.2 is not true and  $\overline{\mathcal{R}(q_0)} \neq M$  for some  $q_0 \in M$ . This implies the existence of some border point  $q'$  of  $\overline{\mathcal{R}(q_0)}$ . Due to Proposition 2.5.3 there exists a point  $q \in \overline{\mathcal{R}(q_0)}$  and a nonzero normal vector  $\zeta$  at

$q$  to it. But due to the strong invariance of  $\overline{\mathcal{R}(q_0)}$  we have for the normal vector  $\zeta$  that (2.5.9) holds for any iterated Lie bracket. In view of condition (2.5.8) it implies that  $\zeta = 0$  and this contradiction proves Theorem 2.5.2.



## CHAPTER 3

### Vector Fields, Flows, and Lagrangian Charts

Our study of dynamic optimization on manifolds will rely to a great extent on the theory of vector fields and flows on smooth manifolds. These flows induce flows on  $TM$  and  $T^*M$  through the evolution of the pushforward and pullback maps  $P_{s,t*}$  and  $P_{s,t}^*$  respectively, and such flows play an important role throughout the remaining chapters. In this chapter we provide a careful exposition of properties of vector fields, flows, and mappings into the vector bundles  $TM$  and  $T^*M$ .

We also present the details of a technique we refer to as the method of *Lagrangian charts*, which will prove useful in our derivation of necessary optimality conditions. These charts are in close analogy with the concept of Lagrangian coordinates in fluid dynamics, see e.g. [8] or [9].

This chapter is organized as follows. In the first section, we translate several operator formulae from the Chronological Calculus into the language of classical differential geometry. We then outline an elementary theory of maps into vector bundles. In the third section, we study flows on  $TM$  and  $T^*M$  induced by pushforward or pullback of a flow on  $M$ . These last two sections are essential for a careful theory of the adjoint equations of geometric optimal control. In the final section we introduce Lagrangian coordinates and derive some elementary properties.

#### 3.1. Interpretations of Operator Formulae

Our first proposition characterizes the derivative  $\frac{d}{ds}P_{s,t}(q)$ .

PROPOSITION 3.1.1. *Suppose that  $V_t$  is a locally  $L^1$ -bounded  $C^1$ -smooth vector field on a Banach manifold  $M$  with flow  $P_{s,t}$ . Then  $P_{s,t}$  is absolutely continuous in  $s$  and for a fixed  $q \in M$  satisfies, for almost all  $s$ ,*

$$(3.1.1) \quad \frac{d}{ds} P_{s,t}(q) = -P_{s,t*}(q)V_s(q).$$

PROOF. Fix  $q \in M$  and recall that  $P_{s,t} = P_{t,s}^{-1}$ . Through Proposition 2.2.1 we have, for almost all  $s$

$$(3.1.2) \quad \frac{d}{ds} \widehat{q} \circ \widehat{P}_{t,s}^{-1} = -\widehat{q} \circ \widehat{V}_s \circ \widehat{P}_{t,s}^{-1} = -\widehat{q} \circ \widehat{P}_{t,s}^{-1} \circ \text{Ad } \widehat{P}_{t,s} \widehat{V}_s.$$

Thus

$$(3.1.3) \quad \frac{d}{ds} \widehat{q} \circ \widehat{P}_{t,s}^{-1} = -\widehat{P_{s,t}(q)} \circ \widehat{P_{s,t*} V_s}$$

and this yields (3.1.1). □

We also recall the following results, which are useful for the study of flows of perturbed vector fields.

PROPOSITION 3.1.2. *Suppose that  $V_t$  and  $W_t$  are locally  $L^1$ -bounded  $C^1$ -smooth vector fields on a Banach manifold  $M$ . Let  $P_{s,t}$  be the flow of  $V_t$  and write  $P_{s,t}^\lambda$  for the flow of  $V_t + \lambda W_t$ . There holds*

$$(3.1.4) \quad \left. \frac{\partial}{\partial \lambda} \right|_{\lambda=0} P_{s,t}^\lambda(q) = \int_s^t P_{\tau,t*}(P_{s,\tau}(q)) W_\tau(P_{s,\tau}(q)) d\tau.$$

PROOF. See (2.2.16). □

PROPOSITION 3.1.3. *Suppose that  $V_t$  and  $W_t$  are locally  $L^1$ -bounded  $C^1$ -smooth vector fields on a Banach manifold  $M$ . Let  $P_{s,t}$  be the flow of  $V_t$  and  $S_{s,t}$  the flow of  $V_t + W_t$ . If  $C_{s,t}$  is the flow of the vector field  $P_{0,t}^{-1} W_t$  then we have*

$$(3.1.5) \quad S_{0,t} = P_{0,t} \circ C_{0,t}.$$

PROOF. This is a restatement of Proposition 2.2.3. □

### 3.2. Maps Into Vector Bundles

Mappings into vector bundles appear naturally in the geometric theory of dynamic optimization, appearing for example: as lifts  $\dot{q} : [0, T] \rightarrow TM$  of absolutely continuous maps  $q : [0, T] \rightarrow M$ ; as fiber derivatives of Lagrangians in the Calculus of Variations [12]; and as adjoint arcs in the Maximum Principle [4, 7, 12, 17, 54, 79, 84]. In this section we provide basic definitions for such mappings.

**3.2.1. Definitions.** We state the following definitions for the analysis of mappings  $v : [0, T] \rightarrow TM$ , though entirely analogous definitions hold for maps into  $T^*M$  and indeed into any Banach-space valued vector bundle over  $M$ . We remind the reader that we use the terms *measurable* and *integrable* in the strong Bochner sense [34].

DEFINITION 3.2.1. Let  $q : [0, T] \rightarrow M$  be a continuous mapping. We say that  $v : [0, T] \rightarrow TM$  is a *mapping along  $q$*  if for all  $t \in [0, T]$  there holds  $\pi(v(t)) = q(t)$ .

REMARK 3.2.2. Vector bundles come with natural “fibrewise” operations. For example, if  $v, w \in TM$  and  $\pi(v) = \pi(w)$  then we can add  $v$  and  $w$  to get  $v+w \in TM$ . In local coordinates, if  $v = (x, V)$  and  $w = (x, W)$  then  $v+w = (x, V+W)$ . Likewise one may define scalar multiplication, for  $\alpha \in \mathbb{R}$ , by

$$(3.2.1) \quad \alpha v = \alpha(x, v) := (x, \alpha v).$$

Thus if  $v, w : [0, T] \rightarrow TM$  are mappings along  $q : [0, T] \rightarrow TM$  then we obtain in a natural way mappings  $v+w : [0, T] \rightarrow TM$  and  $\alpha v : [0, T] \rightarrow TM$  through fibrewise operations.

DEFINITION 3.2.3. Let  $q : [0, T] \rightarrow M$  be continuous and let  $v : [0, T] \rightarrow TM$  be a mapping along  $q$ . We say that  $v$  is *measurable* if for any  $t_0 \in [0, T]$  there exist  $\delta > 0$  and a coordinate chart  $(\mathcal{O}, \varphi)$  such that  $q : (t_0 - \delta, t_0 + \delta) \rightarrow \mathcal{O}$  and such that  $v(t)$  is measurable in the local coordinates  $\varphi$  for  $t \in (t_0 - \delta, t_0 + \delta)$ .

DEFINITION 3.2.4. Let  $q : [0, T] \rightarrow M$  be continuous and let  $v : [0, T] \rightarrow TM$  be a mapping along  $q$ . We say that  $v$  is *locally  $L^p$ -bounded* if for any  $t_0 \in [0, T]$  there exist  $\delta > 0$  and a coordinate chart  $(\mathcal{O}, \varphi)$  such that  $q : (t_0 - \delta, t_0 + \delta) \rightarrow \mathcal{O}$  and such that the local coordinate representation for  $v(t)$  in coordinates  $\varphi$  is both measurable for  $t \in (t_0 - \delta, t_0 + \delta)$  and bounded in norm on this interval by an  $L^p$  function.

DEFINITION 3.2.5. Let  $q : [0, T] \rightarrow M$  be continuous and let  $v : [0, T] \rightarrow TM$  be a mapping along  $q$  which is locally  $L^1$ -bounded. A time  $t_0 \in [0, T]$  is said to be a *Lebesgue point* for  $v$  if there exists a coordinate chart  $\varphi$  such that  $t_0$  is a Lebesgue point for the local coordinate representation of  $v(t)$ .

DEFINITION 3.2.6. If  $v : [0, T] \rightarrow TM$  takes values entirely in  $T_q M$  for a fixed  $q$ , then we define  $\int_0^T v(t) dt$  to be by choosing a coordinate map  $\varphi$  and, setting  $x = \varphi(q)$ , define

$$(3.2.2) \quad \int_0^T v(t) dt := \varphi_*^{-1}(x) \circ \int_0^T \varphi_*(q)v(t) dt.$$

It can be checked that each of these definitions does not depend on choice of coordinates  $\varphi$  and that if  $v$  is locally  $L^p$ -bounded then almost every  $t \in [0, T]$  is a Lebesgue point.

It will often be useful to extend a mapping  $v : [0, T] \rightarrow TM$  which is defined along a map  $q$  to a nonautonomous vector field. We turn now to this task.

**3.2.2. Extension of Maps to Vector Fields.** Recall that a Banach space  $E$  is *smooth* if its norm is differentiable away from zero. Smooth Banach spaces admit

$C^1$ -smooth *bump functions*, that is  $C^1$ -smooth functions  $b : E \rightarrow \mathbb{R}$  which are not identically zero and whose support

$$(3.2.3) \quad \text{supp } b := \overline{\{x \in E : b(x) \neq 0\}}$$

is bounded. We will work with Banach spaces whose dual  $E^*$  is separable and it can be shown [45] that for such spaces, smoothness of the norm is equivalent to the existence of a  $C^1$ -smooth bump function  $b : E \rightarrow \mathbb{R}$ . Using standard bump functions on  $\mathbb{R}$  one can prove the following lemma:

LEMMA 3.2.7. *If  $M$  is a manifold modeled over a Banach space  $E$  whose norm is  $C^k$ -smooth away from zero then for any  $q \in M$  there is a coordinate chart  $\varphi : \mathcal{O} \rightarrow E$  along with an open set  $\mathcal{O}_0 \subset \overline{\mathcal{O}_0} \subset \mathcal{O}$  with  $q \in \mathcal{O}_0$  and a  $C^k$ -smooth bump function  $b : M \rightarrow [0, 1]$  which is identically equal to one on  $\mathcal{O}_0$  and whose support is contained in  $\mathcal{O}$ .*

We use this to prove the next lemma, which shows that one may switch between  $L^p$ -bounded mappings  $v : [0, T] \rightarrow TM$  and locally  $L^p$ -bounded  $C^k$ -smooth vector fields with little difficulty, provided the underlying Banach space is sufficiently smooth. We stress that for the Banach spaces under consideration in the chapters which follow we may always take  $k = 1$ .

LEMMA 3.2.8 (Extension Lemma). *Let  $q : [0, T] \rightarrow M$  be continuous and let  $v : [0, T] \rightarrow TM$  be an  $L^p$ -bounded mapping along  $q$ . If  $M$  is a  $C^k$ -smooth manifold which admits  $C^k$ -smooth bump functions, then there exists a locally  $L^p$ -bounded  $C^k$ -smooth vector field  $V_t$  which extends  $v$  in the sense that  $V_t(q(t)) = v(t)$  for almost all  $t$ . When  $M$  is of finite dimension we may assume that  $V_t$  is compactly supported.*

PROOF. By Lemma 3.2.7, for any fixed  $t \in [0, T]$  there exists an open neighborhood  $\mathcal{O}_t$  of  $q(t)$  and a  $C^k$ -smooth bump function  $b_t : M \rightarrow [0, 1]$  such that  $b_t \equiv 1$

on  $\mathcal{O}_t$  and  $\text{supp } b_t$  is contained in the domain of a coordinate chart  $\varphi_t$ . By compactness of  $[0, T]$  we may choose finitely many such charts and a finite partition  $0 = t_0 < t_1 < \dots < t_r = T$  such for each  $i$  we have  $q : [t_i, t_{i+1}] \rightarrow \mathcal{O}_t$  for a fixed value of  $t$ . Write  $\mathcal{O}_i$  for the set containing the image of  $[t_i, t_{i+1}]$ , let  $b_i$  be the corresponding bump function and  $\varphi_i$  the corresponding coordinate map.

For  $t \in [t_i, t_{i+1}]$  define a vector field  $V_t$  on  $M$  through

$$V_t(q) = b_i(q)\varphi_{i*}^{-1}(\varphi_i(q))\varphi_{i*}(q(t))v(t),$$

extended smoothly to zero outside of the domain for the coordinates  $\varphi_i$ . Note that  $V_t$  is locally  $L^p$ -bounded and  $C^k$ -smooth. Further, since  $q(t) \in \mathcal{O}_i$  and  $b_i|_{\mathcal{O}_i} \equiv 1$ , we have

$$(3.2.4) \quad V_t(q(t)) = b_i(q(t))\varphi_{i*}^{-1}(\varphi_i(q(t)))\varphi_{i*}(q(t))v(t) = b_i(q(t))v(t) = v(t).$$

Thus  $V_t$  extends  $v(t)$ . The compactness statement for finite-dimensional manifolds is apparent from the construction – simply take each  $\mathcal{O}_t$  to be such that the closure  $\overline{\mathcal{O}_t}$  is compact.  $\square$

REMARK 3.2.9. If  $q : [0, T] \rightarrow M$  is absolutely continuous with locally  $L^p$ -bounded derivative and  $M$  is a  $C^k$ -smooth manifold which admits  $C^k$ -smooth bump functions, then Lemma 3.2.8 implies the existence of a locally  $L^p$ -bounded  $C^k$ -smooth vector field  $V_t$  for which  $\dot{q}(t) = V_t(q(t))$ . As a consequence, if  $P_{s,t}$  is the flow of  $V_t$ , then  $q(t) = P_{0,t}(q_0)$ . When  $M$  is finite-dimensional we may assume that  $V_t$  is compactly supported and hence that  $P_{s,t}$  is complete.

Lemma 3.2.8 raises an interesting question: how smooth can we expect a bump function to be in a smooth Banach space? For finite-dimensional  $C^\infty$ -smooth manifolds there always exists a  $C^\infty$  bump function and so  $V_t$  may be assumed  $C^\infty$ -smooth

in  $q$ . Generally speaking, however, a smooth Banach space may admit only a  $C^1$ -smooth bump function. A survey of the theory of bump functions in Banach spaces can be found in [45]. We would like to point out two interesting results. The first is due to [13]:

PROPOSITION 3.2.10. *Consider the space  $L^p(\mathbb{R}^n, \mathbb{R})$  with norm  $\|\cdot\|_{L^p}$ . We have*

- (i) *If  $p$  is an even integer then  $\|\cdot\|_{L^p}$  is  $C^\infty$ -smooth away from zero;*
- (ii) *If  $p$  is an odd integer then  $\|\cdot\|_{L^p}$  is  $C^{p-1}$ -smooth away from zero with a locally Lipschitz  $(p-1)^{\text{st}}$  derivative;*
- (iii) *If  $p$  is not an integer then  $\|\cdot\|_{L^p}$  is  $C^{[p]}$ -smooth away from zero with a  $p - [p]$ -Hölder continuous  $(p-1)^{\text{st}}$  derivative.*

We also note that any Hilbert space has a  $C^\infty$ -smooth norm. The second result we wish to mention, established in [40], is the surprising fact that if the norms on  $E$  and  $E^*$  are sufficiently smooth then  $E$  is a Hilbert space:

PROPOSITION 3.2.11. *If  $E$  and  $E^*$  admit bump functions with locally Lipschitz derivatives then  $E$  is a Hilbert space.*

### 3.3. Flows Induced Through Pushforward or Pullback

We turn now to the study of flows on  $TM$  and  $T^*M$  induced by pushforward and pullback of flows on  $M$ . A central goal of this section is to understand the mappings  $\zeta : [0, T] \rightarrow T^*M$  defined through  $P_{t,T}^* \zeta_T$ , as we will find such mappings to be ubiquitous in our study of necessary optimality conditions.

**3.3.1. Pushforward Dynamics.** Suppose  $P_{s,t}$  is the flow of a locally integrable bounded,  $C^1$ -smooth vector field  $V_t$ . Fix  $q_0 \in M$  and write  $q(t)$  for  $P_{0,t}(q_0)$ . Given  $v_0 \in T_{q_0}M$ , define a mapping  $v : [0, T] \rightarrow TM$  along  $q$  through  $v(t) = P_{0,t*}v_0$ .

In this subsection we prove that  $v(t)$  is absolutely continuous and derive an expression for the infinitesimal generator of such pushforward flow. A proof of this result can also be found in [7]. We provide a short proof here for completeness.

Recall that the bundle  $TTM$  is equipped with a *canonical involution* [59] given locally by  $\kappa(x, v; w, V) = (x, w; v, V)$ . Fix a time  $t_0 \in [0, T]$ , coordinate chart  $(\mathcal{O}, \varphi)$ , and  $\delta > 0$  such that  $q(t) \in \mathcal{O}$  for all  $t \in (t_0 - \delta, t_0 + \delta)$ . Define the following maps, which take values in  $E$ :

$$x(t) := \varphi(q(t)) \quad w(t) := \varphi_*(q(t))v(t).$$

PROPOSITION 3.3.1. *The map  $w : [0, T] \rightarrow E$  is absolutely continuous and for almost all  $t \in (t_0 - \delta, t_0 + \delta)$  we have*

$$(3.3.1) \quad \frac{d}{dt}(x(t), w(t)) = ((\varphi_* V_t)(x(t)), (\varphi_* V_t)'(x(t))w(t)).$$

and  $\dot{v}(t) = \kappa \circ V_{t*}(q(t))v(t)$ .

PROOF. It is clear that  $\dot{x}(t) = (\varphi_* V_t)(x(t))$  and so we prove the formula for  $\dot{w}(t)$ . Fix a time  $t \in (t_0 - \delta, t_0 + \delta)$ . By the semigroup property for flows,  $v(t + \varepsilon) = P_{t, t+\varepsilon*}(q(t))v(t)$ . Since  $v(t) = \varphi_*^{-1}(x(t))w(t)$ , we find

$$(3.3.2) \quad \begin{aligned} w(t + \varepsilon) &= \varphi_*(q(t + \varepsilon))P_{t, t+\varepsilon*}(q(t))v(t) = (\varphi \circ P_{t, t+\varepsilon})_* (\varphi^{-1}(x(t)))\varphi_*^{-1}(x(t))w(t) \\ &= (\varphi \circ P_{t, t+\varepsilon} \circ \varphi^{-1})'_*(x(t))w(t). \end{aligned}$$

But  $\varphi \circ P_{t, t+\varepsilon} \circ \varphi^{-1}$  is a map from  $E$  to itself and so we may write

$$(3.3.3) \quad w(t + \varepsilon) = (\varphi \circ P_{t, t+\varepsilon} \circ \varphi^{-1})'(x(t))w(t).$$

The map  $\varphi \circ P_{t, t+\varepsilon} \circ \varphi^{-1}$  is the local flow for the  $C^1$ -smooth vector field  $\varphi_* V_t$ . By the theory of flows in Banach space the map  $(\varphi \circ P_{t, t+\varepsilon} \circ \varphi^{-1})'(x(t))$  is absolutely



continuous and for almost all  $t \in (t_0 - \delta, t_0 + \delta)$  satisfies

$$(3.3.4) \quad \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} (\varphi \circ P_{t,t+\varepsilon} \circ \varphi^{-1})'(x(t)) = (\varphi_* V_t)'(x(t)).$$

Consequently  $w$  is absolutely continuous and for almost all  $t$  satisfies

$$(3.3.5) \quad \dot{w}(t) = \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} w(t+\varepsilon) = (\varphi_* V_t)'(x(t))w(t).$$

To complete the proof recall that in local coordinates,  $V_t$  is the map  $x \mapsto (x, (\varphi_* V_t)(x))$ . Therefore, in the same local coordinates,  $V_{t*} : TM \rightarrow TTM$  acts on a pair  $(x, w) \in E \times E$  through

$$(3.3.6) \quad (x, w) \mapsto (x, (\varphi_* V_t)(x), w, (\varphi_* V_t)'(x)w).$$

Composing this map with  $\kappa$  we find that  $\dot{v}(t) = \kappa \circ V_{t*}(q(t))v(t)$ .  $\square$

**3.3.2. Pullback Dynamics.** In this subsection we derive local coordinate expressions for the infinitesimal generators of pullback-type flows such as  $\zeta(t) = P_{t,T}^* \zeta_T$  for fixed  $\zeta_T \in T^*M$ . In the following subsection, we provide details of a well-known link to coordinate-free *Hamiltonian vector fields* on  $T^*M$ .

Throughout this subsection we let  $P_{s,t}$  be denote flow of a locally  $L^1$ -bounded,  $C^1$ -smooth vector field  $V_t$ . We fix  $q_0 \in M$  and set  $q(t) = P_{0,t}(q_0)$ . Finally, we fix  $\zeta_T \in T_{q(T)}^*M$ .

We first consider mappings  $\zeta : [0, T] \rightarrow T^*M$  along  $q$  of the form  $\zeta(t) = P_{t,T}^* \zeta_T$ . Fix a time  $t_0 \in [0, T]$ ,  $\delta > 0$ , and coordinate chart  $(\mathcal{O}, \varphi)$ , with  $q(t) \in \mathcal{O}$  for all  $t \in (t_0 - \delta, t_0 + \delta)$ . Let  $(x(t), p(t)) \in E \times E^*$  be the local coordinate representation for  $\zeta(t)$ .

**PROPOSITION 3.3.2.** *For almost all  $t \in (t_0 - \delta, t_0 + \delta)$  we have*

$$(3.3.7) \quad \dot{x}(t) = (\varphi_* V_t)(x(t)) \quad \dot{p}(t) = -(\varphi_* V_t)'(x(t))^* p(t),$$

where  $(\varphi_* V_t)'(x(t))^* : E^* \rightarrow E^*$  is the adjoint of the bounded linear operator  $(\varphi_* V_t)'(x(t)) : E \rightarrow E$ .

PROOF. Only the expression for  $\dot{p}(t)$  requires proof. First note that because

$$(3.3.8) \quad \zeta(t + \varepsilon) = P_{t+\varepsilon,t}^* \zeta(t),$$

one has  $\varphi^* p(t + \varepsilon) = P_{t+\varepsilon,t}^* \circ \varphi^* p(t)$ . Rearranging, one finds

$$(3.3.9) \quad p(t + \varepsilon) = (\varphi \circ P_{t+\varepsilon,t} \circ \varphi^{-1})^* p(t).$$

Let  $v \in E$  be arbitrary and set  $v(\varepsilon) = (\varphi \circ P_{t,t+\varepsilon} \circ \varphi^{-1})_* v$ . We have

$$(3.3.10) \quad \begin{aligned} \langle p(t + \varepsilon), v(\varepsilon) \rangle &= \left\langle (\varphi \circ P_{t,t+\varepsilon} \circ \varphi^{-1})^* p(t), (\varphi \circ P_{t+\varepsilon,t} \circ \varphi^{-1})_* v(\varepsilon) \right\rangle \\ &= \langle p(t), v \rangle. \end{aligned}$$

Thus,  $\frac{d}{d\varepsilon} \Big|_{\varepsilon=0} \langle p(t + \varepsilon), v(\varepsilon) \rangle = 0$ . Equation (3.3.1) now gives us

$$(3.3.11) \quad \langle \dot{p}(t), v \rangle = - \langle p(t), (\varphi_* V_t)'(x(t))v \rangle,$$

provided  $t$  is a point of differentiability both for  $p$  and for  $(\varphi \circ P_{t,t+\varepsilon} \circ \varphi^{-1})'$ . Almost all times  $t \in [0, T]$  are such points and, since  $v$  was arbitrary, we obtain (3.3.7)  $\square$

In applications such as nonholonomic mechanics, optimal control, and the Calculus of Variations, one also comes across covector curves defined in the following way: let  $\beta : [0, T] \rightarrow T^*M$  be a locally  $L^1$ -bounded mapping along  $q$  and define

$$(3.3.12) \quad \zeta(t) = P_{t,T}^* \zeta_T + \int_t^T P_{t,\tau}^* \beta(\tau) d\tau.$$

As before we fix a time  $t_0 \in [0, T]$ , coordinate chart  $(\mathcal{O}, \varphi)$ , and  $\delta > 0$  such that  $q(t) \in \mathcal{O}$  for all  $t \in (t_0 - \delta, t_0 + \delta)$ . Let  $(x(t), p(t)) \in E \times E^*$  be the local coordinate representation for  $\zeta(t)$ .

PROPOSITION 3.3.3. *For almost all  $t \in (t_0 - \delta, t_0 + \delta)$  we have*

$$(3.3.13) \quad \dot{p}(t) = -(\varphi_* V_t)'(x(t))^* p(t) - \varphi^{-1*} \beta(t),$$

where  $(\varphi_* V_t)'(x(t))^* : E^* \rightarrow E^*$  is the adjoint of the bounded linear operator  $(\varphi_* V_t)'(x(t))$ .

PROOF. If  $\alpha, \omega : [0, T] \rightarrow T^*M$  are mappings along  $q$  then for each  $t \in [0, T]$  we have  $\alpha(t), \omega(t) \in T_{q(t)}^*M$  and we may take advantage of the underlying linear structure to define a mapping  $\alpha + \omega$  along  $q$  through

$$(3.3.14) \quad (\alpha + \omega)(t) := \alpha(t) + \omega(t).$$

Set  $\alpha(t) = P_{t,T}^* \zeta_T$  and  $\omega(t) = \int_t^T P_{t,\tau}^* \beta(\tau) d\tau$ , so that

$$(3.3.15) \quad \zeta(t) = \alpha(t) + \omega(t).$$

We are interested in the local coordinate representation of  $\dot{\omega}(t)$ . Let  $r(t)$  be the local coordinate representation for  $\omega(t)$  and check that

$$(3.3.16) \quad r(t + \varepsilon) = \varphi^{-1*} \circ P_{t+\varepsilon, T}^* \circ \varphi^* \int_{t+\varepsilon}^T \varphi^{-1*} P_{T,\tau}^* \beta(\tau) d\tau.$$

By the product rule we can see that, for almost all  $t \in (t_0 - \delta, t_0 + \delta)$  there holds

$$(3.3.17) \quad \begin{aligned} \dot{r}(t) &= -(\varphi_* V_t)'(x(t))^* \int_t^T \varphi^{-1*} P_{T,\tau}^* \beta(\tau) d\tau \\ &\quad - \varphi^{-1*} \circ P_{t,T}^* \circ \varphi^* \circ \varphi^{-1*} P_{T,t}^* \beta(t) \\ &= -(\varphi_* V_t)'(x(t))^* \int_t^T \varphi^{-1*} P_{T,\tau}^* \beta(\tau) d\tau - \varphi^{-1*} \beta(t). \end{aligned}$$

Now by Proposition 3.3.2 we have

$$\begin{aligned}
(3.3.18) \quad \dot{p}(t) &= -(\varphi_* V_t)'(x(t))^* \varphi^{-1*} P_{t,T}^* \zeta_T \\
&\quad - (\varphi_* V_t)'(x(t))^* \int_t^T \varphi^{-1*} P_{T,\tau}^* \beta(\tau) d\tau - \varphi^{-1*} \beta(t) \\
&= -(\varphi_* V_t)'(x(t))^* p(t) - \varphi^{-1*} \beta(t).
\end{aligned}$$

□

Finally, we will have need for a particular case of Proposition 3.3.3 in which  $\beta$  arises as the exterior derivative of a cost function. Suppose that  $L : [0, T] \times M \rightarrow \mathbb{R}$  is a Carathéodory function which is  $C^1$ -smooth in  $q$ . Let

$$(3.3.19) \quad \zeta(t) = P_{t,T}^* \zeta_T + \int_t^T P_{t,\tau}^* dL(\tau, q(\tau)) d\tau,$$

where  $dL(t, q)$  denotes the exterior derivative of  $L$  with respect only to  $q$ . Fix a time  $t_0 \in [0, T]$ , coordinate chart  $(\mathcal{O}, \varphi)$ , and  $\delta > 0$  such that  $q(t) \in \mathcal{O}$  for all  $t \in (t_0 - \delta, t_0 + \delta)$ . Let  $(x(t), p(t)) \in E \times E^*$  be the local coordinate representation for  $\zeta(t)$ .

PROPOSITION 3.3.4. *For almost all  $t \in (t_0 - \delta, t_0 + \delta)$  we have*

$$(3.3.20) \quad \dot{p}(t) = -(\varphi_* V_t)'(x(t))^* p(t) - (\varphi_* L)'(t, x(t)),$$

where  $\varphi_* L : [0, T] \times \varphi(\mathcal{O}) \rightarrow \mathbb{R}$  is defined by  $(\varphi_* L)(t, x) = L(t, \varphi^{-1}(x))$ .

PROOF. In Proposition 3.3.3 take  $\beta(t) = dL(t, q(t))$  and recall the formulae

$$(3.3.21) \quad \varphi^{-1*} dL = d(L \circ \varphi^{-1}) = (\varphi_* L)'.$$

□

The expressions given in equations (3.3.7), (3.3.13), and (3.3.20) are complicated by the explicit presence of the coordinate chart  $\varphi$ . Suppressing the coordinate chart

yields the following familiar expressions

$$(3.3.22) \quad \frac{d}{dt} P_{t,T}^* \zeta_T = -V'_t(q_t)^* \zeta(t)$$

$$(3.3.23) \quad \frac{d}{dt} \left( P_{t,T}^* \zeta_T + \int_t^T P_{t,\tau}^* \beta(\tau) d\tau \right) = -V'_t(q(t))^* \zeta(t) - \beta(t)$$

$$(3.3.24) \quad \frac{d}{dt} \left( P_{t,T}^* \zeta_T - \int_t^T P_{t,\tau}^* dL(\tau, q(\tau)) d\tau \right) = -V'_t(q(t))^* \zeta(t) + L'(t, q(t)).$$

**3.3.3. Hamiltonian Vector Fields and Pullback.** An important observation in the theory of geometric optimal control is that the adjoint equations evolve on the cotangent bundle  $T^*M$ . Indeed, this is a central part of every statement of the geometric Maximum Principle that we have seen [4, 7, 12, 17, 54, 79, 84]. We provide here a brief introduction to the natural symplectic structure on  $T^*M$ . The complete details of this construction for manifolds modeled over Banach spaces, including the theory of exterior derivatives of  $k$ -forms (we will need only  $k = 1$ ), can be found in Lang's text [61].

For any smooth manifold  $M$ , the cotangent bundle  $T^*M$  admits a canonical one-form  $s$ , whose action on  $X \in T_\zeta T^*M$  is given by

$$(3.3.25) \quad \langle s, X \rangle := \langle \zeta, \pi_* X \rangle.$$

One defines the *canonical two-form*  $\sigma$  on  $T^*M$  through  $\sigma = -ds$ . Taking vectors  $X_i = (v_i, \eta_i) \in T_\zeta T^*M \cong E \times E^*$ , ( $i = 1, 2$ ) one can show that

$$(3.3.26) \quad \sigma(X_1, X_2) = \langle \eta_2, v_1 \rangle - \langle \eta_1, v_2 \rangle.$$

We refer the reader to Lang [61] for the precise details of this calculation. The two-form  $\sigma$  induces a map  $J : TT^*M \rightarrow T^*T^*M$  as follows: given  $X_1 \in T_\zeta T^*M$ , let  $J(X_1) \in T_\zeta^* T^*M$  act on  $X_2 \in T_\zeta T^*M$  by  $\langle J(X_1), X_2 \rangle = \sigma(X_1, X_2)$ . Under the assumption that  $E$  is reflexive we may identify  $T_\zeta T^*M$  with  $E \times E^*$  and  $T_\zeta^* T^*M$

with  $E^* \times E$ . One may then check that  $J$  admits the following matrix of operators representation:

$$(3.3.27) \quad J = \begin{pmatrix} 0 & -Id_{E^*} \\ Id_E & 0 \end{pmatrix}.$$

It then follows that  $J$  is an isomorphism with inverse given by

$$(3.3.28) \quad J^{-1} = \begin{pmatrix} 0 & Id_E \\ -Id_{E^*} & 0 \end{pmatrix}.$$

Given a  $C^k$ -smooth function  $H : T^*M \rightarrow \mathbb{R}$ , one may employ the map  $J$  to define a  $C^{k-1}$ -smooth vector field  $\vec{H}$  on  $T^*M$  through  $\vec{H} = J^{-1}(dH)$ . By (3.3.28), the local coordinate expression for  $\vec{H}$ , in coordinates  $(x, p)$ , is given by

$$(3.3.29) \quad \vec{H}(x, p) = (H_p(x, p), -H_x(x, p)).$$

The vector field  $\vec{H}$  is called the *Hamiltonian lift* of  $H$  and generalizes the familiar notion of a *Hamiltonian system* on  $\mathbb{R}^n$ , given by

$$(3.3.30) \quad \begin{cases} \dot{x} &= H_p(x, p) \\ \dot{p} &= -H_x(x, p) \end{cases}$$

When  $H$  has a time dependence, so that  $H : [0, T] \times T^*M \rightarrow \mathbb{R}$ , we will write  $\vec{H}(t, \zeta)$  for the lift defined through  $dH$ , with the exterior derivative taken only with respect to  $q \in M$  and not time  $t$ .

Such Hamiltonian lifts arise naturally from pullback flows. Suppose, for example, that  $\zeta : [0, T] \rightarrow T^*M$  is defined through (3.3.19) and let  $H : [0, T] \times T^*M \rightarrow \mathbb{R}$  be defined by

$$(3.3.31) \quad H(t, \zeta) = \langle \zeta, V_t(q) \rangle - L(t, q).$$

Comparing (3.3.29) with either Proposition 3.3.4 or Equation (3.3.24) we obtain the following proposition:

PROPOSITION 3.3.5. *If  $M$  is modeled over a reflexive Banach space  $E$  then the map  $\zeta : [0, T] \rightarrow T^*M$  defined by (3.3.19) is a solution to*

$$(3.3.32) \quad \dot{\zeta}(t) = \vec{H}(t, \zeta(t))$$

with  $\zeta(T) = \zeta_T$ .

### 3.4. Lagrangian Charts

In problems of dynamic optimization where the state evolves on a manifold, a common obstacle in the derivation of necessary optimality conditions is that the state is not confined to the domain of a single chart. In this dissertation we address this problem by introducing the technique of *Lagrangian charts*. This technique is similar to a technique developed by Bismut [11] and Montgomery [72], which makes use of specially constructed affine control systems.

**3.4.1. Definition and Basic Properties.** We present a technique which uses the flow of a single vector field to construct a kind of time-varying coordinate chart. A strength of our approach is that the semigroup properties of flows are compatible with these charts in a useful manner. In addition, for problems in which a control system is being studied, this vector field may often be chosen to be the control system itself and the construction is then in some sense natural to the problem.

The construction of Lagrangian charts is in direct analogy with the notion of Lagrangian coordinates in fluid dynamics, in which fluid particles are given time-invariant coordinates through their initial conditions, see [8, 9].

Let  $P_{s,t}$  denote the flow of a nonautonomous vector field  $V_t$  and suppose that  $P_{0,t}(q)$  is defined for all  $t \in [0, T]$ , for all  $q$  in a neighborhood of  $q_0$ . Let  $(\mathcal{O}, \varphi)$  be a

coordinate chart with  $q_0 \in \mathcal{O}$ . Shrinking  $\mathcal{O}$  if necessary we may assume that  $P_{0,t}(q)$  is defined for all  $q \in \mathcal{O}$  and all  $t \in [0, T]$ . We define a map  $\psi_t : P_{0,t}(\mathcal{O}) \rightarrow \varphi(\mathcal{O})$  through

$$\psi_t = \varphi \circ P_{t,0}.$$

DEFINITION 3.4.1. We refer to the map  $\psi_t$  as the *Lagrangian coordinates* associated with the vector field  $V_t$  and coordinate chart  $(\mathcal{O}, \varphi)$ .

It can be shown that the collection of continuous maps  $q : [0, T] \rightarrow M$  for which  $q(t) \in P_{0,t}(\mathcal{O})$  for all  $t$  is a compact-open neighborhood of  $q^0$  and so the collection of such maps is a natural object of study in the theory of dynamic optimization.

Let  $q : [0, T] \rightarrow M$  be absolutely continuous and suppose that for all  $t \in [0, T]$  we have  $q(t) \in P_{0,t}(\mathcal{O})$ . Let  $x : [0, T] \rightarrow \varphi(\mathcal{O})$  be defined through  $x(t) = \psi_t(q(t))$ . Equation (3.1.1) yields the following:

PROPOSITION 3.4.2. *For almost all  $t$  we have*

$$(3.4.1) \quad \dot{x}(t) = \psi_{t*}(q(t))\dot{q}(t) - \psi_{t*}(q(t))V_t(q(t)).$$

Informally, we may think of  $\psi_{t*}$  as “translating” velocities from  $T_{q(t)}M$  to  $E$ , Proposition 3.4.2 may be interpreted as the claim that in Lagrangian coordinates the velocity of  $q(t)$  is given by the translation of  $\dot{q}(t)$  corrected by the “velocity of the chart.” Note that  $q(t)$  evolves according to the flow of  $V_t$  if and only if  $\psi_t(q(t))$  is constant. Herein lies the analogy with Lagrangian coordinates of fluid dynamics [8].

Now we introduce some properties of Lagrangian charts in relation to control systems. Let  $\mathbb{U}$  be a set. A control system  $f : [0, T] \times M \times \mathbb{U} \rightarrow TM$  is a map for which  $f(t, q, u) \in T_qM$  for any choice of  $(t, q, u)$ . Later we will introduce assumptions on  $f$  and  $\mathbb{U}$  and definitions such as measurability of control  $u : [0, T] \rightarrow \mathbb{U}$ . For



the moment we suppose that for a given control  $u : [0, T] \rightarrow \mathbb{U}$  one obtains a nonautonomous vector field  $(t, q) \mapsto f(t, q, u(t))$ .

Fix a control  $u^0$  and suppose that the associated vector field  $(t, q) \mapsto f(t, q, u^0(t))$  has a well-defined flow  $P_{s,t}$  which is  $C^1$ -smooth with Lipschitz derivative. Let  $\psi_t : P_{0,t}(\mathcal{O}) \rightarrow \varphi(\mathcal{O})$  be the associated Lagrangian coordinates. In accordance with (3.4.1) we define a control system  $g : [0, T] \times \varphi(\mathcal{O}) \times \mathbb{U} \rightarrow E$  through

$$(3.4.2) \quad g(t, x, u) = (\psi_{t*}f)(t, x, u) - (\psi_{t*}f)(t, x, u^0(t)).$$

We immediately obtain the following:

PROPOSITION 3.4.3. *Let  $u : [0, T] \rightarrow \mathbb{U}$  be a control and  $q(t, u)$  the solution to*

$$(3.4.3) \quad \dot{q}(t, u) = f(t, q(t, u), u(t))$$

*with  $q(0, u) = q_0$ . Suppose  $q(t; u) \in P_{0,t}(\mathcal{O})$  for all  $t \in [0, T]$  and let  $R_{s,t}$  be the flow of  $f$  associated with  $u$ . Let  $Q_{s,t}$  be the flow of  $g$  corresponding to  $u$ . For  $q$  sufficiently close to  $q_0$  we have*

$$(3.4.4) \quad \psi_t \circ R_{0,t}(q) = Q_{0,t} \circ \varphi(q).$$

In addition, we can prove

PROPOSITION 3.4.4. *Let  $u : [0, T] \rightarrow \mathbb{U}$  be a control and  $q(t, u)$  the solution to*

$$(3.4.5) \quad \dot{q}(t, u) = f(t, q(t, u), u(t))$$

*with  $q(0, u) = q_0$ . Suppose  $q(t; u) \in P_{0,t}(\mathcal{O})$  for all  $t \in [0, T]$  and let  $R_{s,t}$  be the flow of  $f$  corresponding to  $u$ . Let  $Q_{s,t}$  be the flow of  $g$  corresponding to  $u$ . There holds*

$$(3.4.6) \quad \psi_t^* \circ Q_{t,T}^* = R_{t,T}^* \circ \psi_T^*.$$

PROOF. By (3.4.4) we have  $Q_{0,t} = \psi_t \circ R_{0,t} \circ \varphi^{-1}$ . Since  $Q_{t,T} = Q_{0,T} \circ Q_{t,0}$  we obtain

$$\begin{aligned}
(3.4.7) \quad Q_{t,T}^* &= (\varphi \circ R_{t,0} \circ \psi_t^{-1})^* \circ (\psi_T \circ R_{0,T} \circ \varphi^{-1})^* \\
&= \psi_t^{-1*} \circ R_{t,0}^* \circ \varphi^* \circ \varphi^{-1*} \circ R_{0,T}^* \circ \psi_T^* \\
&= \psi_t^{-1*} \circ R_{t,T}^* \circ \psi_T^*,
\end{aligned}$$

which completes the proof.  $\square$

Proposition 3.4.4 demonstrates the inherit utility of defining charts through with flows of vector fields. We conclude this section by demonstrating that a Lagrangian chart  $\psi_t$  induces a chart  $\Psi_t$  on  $T^*M$  which respects the underlying symplectic structure.

**3.4.2. Symplectic Structure and Lagrangian Charts.** In this section we clarify some important details regarding the canonical symplectic structures on  $T^*M$  and  $T^*E \cong E \times E^*$  and their relation to Lagrangian charts. We assume in this section that  $E$  is a reflexive Banach space. Let  $\sigma_M$  and  $\sigma_E$  denote the canonical symplectic two-forms for  $T^*M$  and  $T^*E$ , respectively. Let  $J_M$  and  $J_E$  denote the corresponding isomorphisms, so that  $J_M : TTM \rightarrow TT^*M$  and  $J_E : E \times E^* \rightarrow E^* \times E$ .

A coordinate chart  $(\mathcal{O}, \varphi)$  induces a chart  $\theta$  on  $T^*\mathcal{O}$  through  $\theta(\zeta) = (\varphi(q), \varphi^{-1*}\zeta)$  and since  $\sigma_E$  is the local coordinate expression for  $\sigma_M$  we obtain the relation  $\theta^*\sigma_E = \sigma_M$ . Our goal in this subsection is to demonstrate that an analogue of this formula holds for Lagrangian charts.

Consider a Lagrangian chart  $\psi_t : P_{0,t}(\mathcal{O}) \rightarrow \varphi(\mathcal{O})$  corresponding to a vector field  $V_t$ . This chart induces a chart  $\Psi_t : T^*P_{0,t}(\mathcal{O}) \rightarrow E \times E^*$  through

$$(3.4.8) \quad \Psi_t(\zeta) = (\psi_t(q), (\psi_t^{-1})^*\zeta)$$

for  $\zeta \in T_q^*M$ .

PROPOSITION 3.4.5. *Let  $E$  be a reflexive Banach space. Then the map  $\Psi_t$  is structure-preserving in the sense that  $\Psi_t^* \sigma_E = \sigma_M$ .*

PROOF. Introduce a function  $H_V : [0, T] \times T^*M \rightarrow \mathbb{R}$  through

$$(3.4.9) \quad H_V(t, \zeta) = \langle \zeta, V_t(q) \rangle$$

for  $\zeta \in T_q^*M$ . Let  $Q_{s,t} : T^*M \rightarrow T^*M$  denote the flow of the vector field  $\vec{H}_V$ . By Proposition 3.3.5  $Q_{s,t}$  is pullback flow associated with  $V_t$  so that

$$(3.4.10) \quad Q_{0,t}(\zeta) = P_{t,0}^* \zeta.$$

We first claim that the map  $\Psi_t^{-1}$  decomposes as

$$(3.4.11) \quad \Psi_t^{-1}(x, p) = Q_{0,t} \circ \theta^{-1}(x, p)$$

To see this note that because  $\theta^{-1}(x, p) = (\varphi^{-1}(x), \varphi^*p)$  we have

$$(3.4.12) \quad \Psi_t^{-1}(x, p) = (\psi_t^{-1}(x), \psi_t^*p) = (P_{0,t} \circ \varphi^{-1}(x), P_{t,0}^* \varphi^*p) = Q_{0,t} \circ \theta^{-1}(x, p).$$

Next we claim that the map  $Q_{0,t}$  is structure-preserving on  $T^*M$ . That is, for all  $t$  there holds

$$(3.4.13) \quad \sigma_M = Q_{0,t}^* \sigma_M.$$

Indeed this is a standard result and is typically shown to be true using the following formula of Cartan [61]:

$$(3.4.14) \quad \mathcal{L}_X = d \circ i_X + i_X \circ d.$$

We carry out the necessary computations. First, (3.4.13) holds for  $t = 0$ . Further,

$$(3.4.15) \quad \frac{d}{dt} Q_{0,t}^* \sigma_M = \mathcal{L}_{\vec{H}_{V_t}} \sigma_M = \left( d \circ i_{\vec{H}_{V_t}} + i_{\vec{H}_{V_t}} \circ d \right) \sigma_M = 0.$$

Finally we note that

$$(3.4.16) \quad \sigma_E = \theta^{-1*} \sigma_M = \theta^{-1*} \circ Q_{0,t}^* \sigma_M = (Q_{0,t} \circ \theta^{-1})^* \sigma_M = \Psi_t^{-1*} \sigma_M$$

and this completes the proof.  $\square$

We conclude this introduction to Lagrangian charts by demonstrating that the induced charts  $\Psi_t$  on  $T^*M$  are compatible with the isomorphisms  $J_M$  and  $J_E$  in the following way:

PROPOSITION 3.4.6. *Let  $E$  be a reflexive Banach space. Then the following diagram commutes:*

$$(3.4.17) \quad \begin{array}{ccc} TT^*E & \xrightarrow{(\Psi_t^{-1})^*} & TT^*M \\ J_E \downarrow & & \downarrow J_M \\ T^*T^*E & \xrightarrow{\Psi_t^*} & T^*T^*M \end{array}$$

That is,  $\Psi_t^* J_E = J_M \circ (\Psi_t^{-1})^*$ . Equivalently,  $J_E^{-1} \circ (\Psi_t^{-1})^* = \Psi_t^* J_M^{-1}$ .

PROOF. Let  $X \in TT^*M$  and  $Y \in TT^*E$ . We will show that

$$(3.4.18) \quad \langle J_M \circ (\Psi_t^{-1})^* Y, X \rangle = \langle \Psi_t^* J_E Y, X \rangle$$

and since  $X, Y$  are arbitrary this will be enough. On the left we have

$$(3.4.19) \quad \begin{aligned} \langle J_M \circ (\Psi_t^{-1})^* Y, X \rangle &= \sigma_M \left( ((\Psi_t^{-1})^* Y, X) \right) = \Psi_t^{-1*} \sigma_M (Y, \Psi_{t*} X) \\ &= \sigma_E (Y, \Psi_{t*} X) = \langle J_E Y, \Psi_{t*} X \rangle = \langle \Psi_t^* J_E Y, X \rangle \end{aligned}$$

and this completes the proof.

□

## CHAPTER 4

# The Fundamental Lemma for Calculus of Variations and Optimal Control on Manifolds

In this chapter we begin a study of dynamic optimization in nonlinear and infinite dimensional spaces that will continue through the remainder of the dissertation. The central goal of this chapter is the derivation of necessary optimality conditions for optimal control on Banach manifolds. Such techniques are of interest, for example, from the point of view of control of partial differential equations [87].

We further demonstrate that all first-order necessary conditions from the classical Calculus of Variations can be derived from the Pontryagin Maximum Principle, which we derive here for problems on Banach manifolds. Thus optimal control plays a central role in the Calculus of Variations and dynamic optimization as a whole.

Let us reiterate a definition of the proceeding chapter.

DEFINITION 4.0.7. A control system on a manifold  $M$  is a mapping  $f : [0, T] \times M \times \mathbb{U} \rightarrow TM$  which satisfies  $f(t, q, u) \in T_q M$  for any choice of  $(t, q, u)$ .

We fix an initial state  $q_0 \in M$  and for a given a mapping  $u : [0, T] \rightarrow \mathbb{U}$  we write  $q(t; u)$  for the Carathéodory solution to

$$(4.0.20) \quad \dot{q}(t; u) = f(t, q(t; u), u(t))$$

satisfying  $q(0; u) = q_0$ . Additional assumptions on  $\mathbb{U}$ ,  $f$ , and  $u$  which ensure this solution is well-defined will be given below.

The dynamic optimization problem with which we concern ourselves in this chapter is the following problem of Bolza. Given a control system  $f : [0, T] \times M \times \mathbb{U} \rightarrow TM$  and functions  $L : [0, T] \times TM \times \mathbb{U} \rightarrow \mathbb{R}$  and  $\ell : M \rightarrow \mathbb{R}$ , find a control  $u^0$  whose trajectory  $q^0$  provides a local minimum for the following cost:

$$(4.0.21) \quad \Lambda_0(u) = \ell(q(T; u)) + \int_0^T L(t, q(t; u), u(t)) dt.$$

Throughout this chapter we assume that  $u^0$  is an optimal control and we write  $q^0(t)$  for the corresponding optimal trajectory. In case  $M = E$  we write  $x^0(t)$ .

The term *local minimum* here refers to strong local minimum. Formally, this means that there is a compact-open neighborhood  $\mathcal{N}$  of  $q^0$  such that if  $u$  is any control for which  $q(\cdot; u) \in \mathcal{N}$ , then  $\Lambda_0(u) \geq \Lambda_0(u^0)$ . Perhaps more familiarly, if  $M = E$  then there exists  $\varepsilon > 0$  such that for any control  $u$  such that

$$(4.0.22) \quad \max_{t \in [0, T]} \|x(t; u) - x^0(t)\|_E < \varepsilon$$

we have  $\Lambda(u) \geq \Lambda(u^0)$  and similar definitions can be made for Riemannian manifolds.

A classical condition for characterizing optimal controls in problems such as this is the *Pontryagin Maximum Principle* [74], a central result in the field of dynamic optimization. Since its appearance in the 1950's, the Maximum Principle has inspired considerable effort in the study of optimal control problems in  $\mathbb{R}^n$ . During the past few decades, optimal control problems on manifolds have also been studied intensively. Statements of the Maximum Principle for such problems can be found in [4, 7, 12, 17, 54, 79, 84]. However, of these, only a handful (see e.g. [4, 7, 17]) offer a full proof of the Maximum Principle for problems on general manifolds and these papers are limited to special cases.

In this chapter we provide a new proof of the Pontryagin Maximum Principle. This work follows the approach taken by Clarke [23, 29, 30], demonstrating that

a broad class of optimal control problems can be reduced to a kind of generalized problem in the nonsmooth Calculus of Variations, removing the dynamic constraint (4.0.20) through a technique of exact penalization. In particular, in our derivation of the Maximum Principle for the case  $M = E$  we use the functional

$$\Lambda(x, \mu) = \ell(x(T)) + \int_0^T \widehat{L}(t, x(t), \mu(t)) dt + C_f \int_0^T \left\| \widehat{f}(t, x(t), \mu(t)) - \dot{x}(t) \right\|_E dt,$$

where  $\mu$  is a type of relaxed control,  $x : [0, T] \rightarrow E$  is an absolutely continuous mapping, and  $C_f$  is a sufficiently large constant. Part of our contribution for problems on Banach manifolds is the application of Lagrangian charts to define such a penalty function on a neighborhood of optimal trajectory  $q^0$ .

Classically, derivation of necessary optimality conditions in the Calculus of Variations relies on the following *fundamental lemma*:

LEMMA 4.0.8. *Let  $E$  be a Banach space and  $p \in L^b([0, T], E^*)$ ,  $1 \leq b \leq \infty$ . Let  $\frac{1}{a} + \frac{1}{b} = 1$  and suppose that, for any  $w \in L^a([0, T], E)$ , we have*

$$(4.0.23) \quad \int_0^T \langle p(t), w(t) \rangle dt \geq 0.$$

*Then  $p(t) = 0$  for almost all  $t$ .*

The spaces  $L^a$  and  $L^b$  above are Lebesgue-Bochner spaces of strongly measurable maps into  $E$  and  $E^*$ , see [34, 92] for a detailed theory of such maps or [42] for a summary. Versions of the fundamental lemma first appeared for finite dimensional spaces during the mid-19<sup>th</sup> century: a 1926 paper by Arnold Dresden [35] dates the Fundamental Lemma to the year 1854. Modern versions of the lemma, in which  $p$  is assumed merely Lebesgue integrable, appeared as early as 1913 [50] and some fairly recent generalizations of the lemma can be found, for example in [76]. However, in spite of its age and importance in the linear theory, there does not appear to be an analogue of this lemma for problems of smooth manifolds, even of finite dimension.



The main goal of this chapter is to derive a geometric version of the fundamental lemma and apply it to the study of problems of dynamic optimization on Banach manifolds. This goal is realized below in Lemma 4.2.2, a geometric form of the fundamental lemma, and in Theorem 4.3.5, a version of the Maximum Principle for Bolza problems on Banach manifolds.

Another classical lemma that is often used in classical Calculus of Variations to derive necessary optimality conditions is the *duBois-Reymond lemma*. This lemma appeared first in 1879 [36]. Again we can find no modern, geometric analogue of this lemma in the literature. This chapter will conclude with the derivation of a geometric duBois-Reymond lemma and an application to the Calculus of Variations on Banach manifolds.

We feel it is worth noting that, in a general Banach space, one must take some care in interpreting the fundamental lemma. Though it is very nearly a statement about the dual space for  $L^a([0, T], E)$ , in a general Banach space there does not hold

$$(4.0.24) \quad L^a([0, T], E)^* = L^b([0, T], E^*)$$

as one might hope. However, in the case in which  $E^*$  is separable, then (4.0.24) does hold for  $1 \leq a < \infty$  and we will restrict ourselves to this case. This in turn forces on us the assumption that  $E$  is separable. In 1964 it was shown that a separable Banach space admits a smooth renorm if and only if its dual is separable [77] and so the Banach spaces we work with in this chapter are also smooth. We will reserve a study of the Maximum Principle for a more general state space until a later time<sup>1</sup>.

Among the spaces with separable duals are the Sobolev spaces  $W^{m,p}$  for  $1 < p < \infty$  [1]. These spaces are given careful treatment in [1] and are of interest in the

---

<sup>1</sup>However, we note that many results of this chapter will hold under the weaker assumption that  $E$  is *Gelfand* – a sufficient condition for (4.0.24) that is more general than requiring  $E^*$  to be separable [42].

theory of partial differential equations [69] and optimal control of systems governed by such equations [87].

The remainder of this chapter is arranged as follows. In the first section we establish the Maximum Principle for Banach spaces with separable duals under the assumption that control  $u$  takes values in a metric space  $\mathbb{U}$ . In this section we demonstrate that all first-order necessary conditions from the Calculus of Variations are contained in the Maximum Principle. We also use the linear case to outline the theory needed for a study of the problem on Banach manifolds.

Following this first section we further develop the theory of maps into the vector bundles  $TM$  and  $T^*M$  introduced in the previous chapter. This section includes an integration-by-parts formula as well as a geometric version of the fundamental and duBois-Reymond lemmas. These techniques are then used to derive the Pontryagin Maximum Principle for Bolza problems in which the state evolves on a Banach manifold and for which the terminal point is free. Problems with terminal constraints are addressed in the chapters following. Applications to the Calculus of Variations are then provided and we demonstrate that all first-order necessary conditions from the calculus of variations on manifolds can be derived from Maximum Principle. The chapter concludes with a short application of the duBois-Reymond lemma to the geometric Calculus of Variations.

#### 4.1. Pontryagin Maximum Principle in Linear Spaces

In this section we illustrate the basic technique of the chapter by establishing a version of the Maximum Principle for Banach spaces with separable duals. We assume the state  $x$  evolves in such a space  $E$  and that controls take their values in a metric space  $\mathbb{U}$  with metric  $d_{\mathbb{U}}$ . We say that a mapping  $u : [0, T] \rightarrow \mathbb{U}$  is *simple* if it takes on finitely many values, the preimage of each being a Lebesgue measurable subset of  $[0, T]$ . We say that  $u : [0, T] \rightarrow \mathbb{U}$  is *measurable* if it is the pointwise limit

of simple maps. We fix an integer  $1 \leq p < \infty$  and consider the set  $\mathcal{U}$  of measurable controls  $u$  for which there exists  $\bar{u} \in \mathbb{U}$  such that the function  $t \mapsto d_{\mathbb{U}}(\bar{u}, u(t))^p$  is integrable. Define a metric  $d$  on  $\mathcal{U}$  through

$$(4.1.1) \quad d(u, v) = \left( \int_0^T d_{\mathbb{U}}(u(t), v(t))^p \right)^{1/p}.$$

**4.1.1. Relaxed Controls.** In our derivation of the maximum principle we will make use of a minimax theorem due to Ky Fan, a precise statement of which appears below. This theorem will require our space of controls to be convex. We therefore work with a notion of *relaxed control*. This technique dates back to the work of L.C. Young [93] in the Calculus of Variations and of Gamkrelidze [46] and J. Warga [91, 92] in control. In many settings a relaxed control is a weakly\*-measurable map into  $\mathcal{P}(\mathbb{U})$ , the space of probability measures on  $\mathbb{U}$ . In this definition  $\mathbb{U}$  is taken to be a compact, or at least locally compact, metric space. In the next chapter we will work such a space and will write  $\mathcal{M}$  for the set of relaxed controls. In this chapter we work with a smaller space  $\mathcal{M}_0$  of relaxed controls which correspond to a finite convex combinations of Dirac measures. This will allow us to work with a general metric space  $\mathbb{U}$  without requiring a full theory of measures for such a space.

We introduce a space  $\mathcal{P}_f(\mathbb{U})$  consisting of finite, convex combinations of Dirac measures  $\delta_u$ . Given any linear space  $X$  we can define an integral of arbitrary mappings  $g : \mathbb{U} \rightarrow X$  with respect to a Dirac measure  $\delta_u$  through

$$(4.1.2) \quad \int_{\mathbb{U}} g d\delta_u = g(u).$$

If  $\mu \in \mathcal{P}_f(\mathbb{U})$  then we can write

$$(4.1.3) \quad \mu = \sum_{j=1}^n \lambda_j \delta_{u_j}$$

for convex coefficients  $\lambda_j$  and points  $u_j \in \mathbb{U}$ . Hence for any  $g : \mathbb{U} \rightarrow X$  we obtain

$$(4.1.4) \quad \int_{\mathbb{U}} g d\mu = \sum_{j=1}^n \lambda_j g(u_j).$$

Each control  $u \in \mathcal{U}$  corresponds to a mapping of  $[0, T]$  into  $\mathcal{P}_f(\mathbb{U})$  through  $t \mapsto \delta_{u(t)}$ . We refer to a finite, convex combination of such controls as a *relaxed control* and we write  $\mathcal{M}_0$  for the set of all relaxed controls. Thus if  $\mu \in \mathcal{M}_0$  then there exists an integer  $n$ , controls  $u_j \in \mathcal{U}$ , and convex coefficients  $\lambda_j$  for which

$$(4.1.5) \quad \mu(t) = \sum_{j=1}^n \lambda_j \delta_{u_j(t)}.$$

Let us assume that  $u^0$  is an optimal control for the Bolza problem and define

$$\nu^0 := \delta_{u^0(t)}.$$

Below we will see that, under appropriate assumptions, relaxed control  $\nu^0$  is locally optimal in  $\mathcal{M}_0$ .

In lieu of defining a topology on  $\mathcal{M}_0$  it will suffice define a neighborhood of the control  $\nu^0$ . Given  $\varepsilon > 0$  let  $\mathcal{B}(\varepsilon)$  denote the set of all controls  $\mu \in \mathcal{M}_0$  with representation (4.1.5) which satisfy

$$(4.1.6) \quad \left( \sum_{j=1}^n \lambda_j d(u_j, u^0)^p \right)^{1/p} < \varepsilon.$$

We extend functions  $f : [0, T] \times E \times \mathbb{U} \rightarrow E$  and  $L : [0, T] \times E \times \mathbb{U} \rightarrow \mathbb{R}$  to functions  $\widehat{f} : [0, T] \times E \times \mathcal{P}_f(\mathbb{U}) \rightarrow E$  and  $\widehat{L} : [0, T] \times E \times \mathcal{P}_f(\mathbb{U}) \rightarrow \mathbb{R}$  through

$$(4.1.7) \quad \widehat{f}(t, x, \mu) = \int_{\mathbb{U}} f(t, x, u) d\mu \quad \widehat{L}(t, x, \mu) = \int_{\mathbb{U}} L(t, x, u) d\mu.$$

With these definitions in mind, we turn to the assumptions for the problem of Bolza described above, stating them for functions  $\widehat{f}$  and  $\widehat{L}$ .

**4.1.2. Assumptions for the Bolza Problem.** Let  $E$  be a Banach space with separable dual and fix an initial state  $x_0 \in E$ . Let  $\ell : E \rightarrow \mathbb{R}$  be a  $C^1$ -smooth function and suppose functions  $f : [0, T] \times E \times \mathbb{U} \rightarrow E$  and  $L : [0, T] \times E \times \mathbb{U} \rightarrow \mathbb{R}$  are Carathéodory maps: measurable in  $t$  and continuous in  $(x, u)$ . We further suppose that each map is differentiable in  $x$  and that there are constants  $\varepsilon_0, \varepsilon_1, M_f, M_L$  such that for any  $\mu \in \mathcal{B}(\varepsilon_0)$ :

(i) The differential equation

$$(4.1.8) \quad \dot{x}(t) = \widehat{f}(t, x(t), \mu(t))$$

admits a solution  $x(t; \mu)$  with  $x(0; \mu) = x_0$  which is defined for all  $t \in [0, T]$  and which satisfies

$$(4.1.9) \quad \max_{t \in [0, T]} \|x^0(t) - x(t; \mu)\|_E < \varepsilon_1;$$

(ii) There are  $L^1$  functions  $k_{f, \mu}$  and  $m_{f, \mu}$  with  $L^1$ -norms bounded by  $M_f$  such that for almost all  $t \in [0, T]$ , for any  $x_1, x_2$  satisfying  $\|x^0(t) - x_i\|_E < \varepsilon_1$  ( $i = 1, 2$ ) we have

$$(4.1.10) \quad \begin{aligned} \left\| \widehat{f}(t, x_1, \mu(t)) - \widehat{f}(t, x_2, \mu(t)) \right\|_E &\leq k_{f, \mu}(t) \|x_1 - x_2\|_E \\ \left\| \widehat{f}_x(t, x_1, \mu(t)) - \widehat{f}_x(t, x_2, \mu(t)) \right\|_E &\leq m_{f, \mu}(t) \|x_1 - x_2\|_E \end{aligned}$$

as well as

$$(4.1.11) \quad \left\| \widehat{f}(t, x_1, \mu(t)) \right\|_E \leq m_{f, \mu}(t) \quad \left\| \widehat{f}_x(t, x_1, \mu(t)) \right\|_E \leq m_{f, \mu}(t);$$

(iii) There are  $L^1$  functions  $k_{L, \mu}$  and  $m_{L, \mu}$  with  $L^1$ -norms bounded by  $M_L$  such that for almost all  $t \in [0, T]$ , for any  $x_1, x_2$  satisfying  $\|x^0(t) - x_i\|_E < \varepsilon_1$  ( $i = 1, 2$ )

we have

$$(4.1.12) \quad \begin{aligned} \left| \widehat{L}(t, x_1, \mu(t)) - \widehat{L}(t, x_2, \mu(t)) \right| &\leq k_{L,\mu}(t) \|x_1 - x_2\|_E \\ \left| \widehat{L}_x(t, x_1, \mu(t)) - \widehat{L}_x(t, x_2, \mu(t)) \right| &\leq k_{L,\mu}(t) \|x_1 - x_2\|_E \end{aligned}$$

as well as

$$(4.1.13) \quad \left| \widehat{L}(t, x_1, \mu(t)) \right| \leq m_{L,\mu}(t) \quad \left| \widehat{L}_x(t, x_1, \mu(t)) \right| \leq m_{L,\mu}(t).$$

We pause to consider examples of control systems satisfying these assumptions.

EXAMPLE 4.1.1. If  $\mathbb{U}$  is the space  $E$  with metric  $d_E(x, y) = \|x - y\|_E$  and  $f(t, x, u) = u$  then  $\mathcal{U}$  is the space  $L^1([0, T], E)$ . The neighborhood  $\mathcal{B}(\varepsilon_0)$  of an optimal control  $u^0$  is the set of all  $\mu$  with representation (4.1.5) which satisfy

$$(4.1.14) \quad \sum_{j=1}^n \lambda_j \|u_j - u^0\|_{L^1} < \varepsilon_0.$$

A function  $L : E \times E \rightarrow \mathbb{R}$  which is  $C^1$  with locally Lipschitz derivative will then satisfy (4.1.12) and (4.1.13).

EXAMPLE 4.1.2. A particular functional of interest may not be continuous in the  $L^1$  sense, and for such functionals it may be appropriate to work with  $p > 1$ . For example, consider the case  $\mathbb{U} = E$  with functional

$$(4.1.15) \quad \Lambda_0(u) = \int_0^T \|u\|_E^2 dt.$$

One may check that in an arbitrary neighborhood of 0 there are  $L^1$  functions for which  $\Lambda_0(u) = +\infty$ . In this case we may take  $p = 2$  and work with the space  $\mathcal{U} = L^2([0, T], E)$ . In this example,  $\mathcal{B}$  is the set of controls  $\mu \in \mathcal{M}_0$  with presentation

(4.1.5) that satisfy

$$(4.1.16) \quad \left( \sum_{j=1}^n \lambda_j d(u_j, u^0)^2 \right)^{1/2} < \varepsilon_0.$$

EXAMPLE 4.1.3. Suppose that  $\mathbb{U} = \mathbb{R}^\infty$ , the space of real-valued sequences with finitely many nonzero terms. Let

$$(4.1.17) \quad d_{\mathbb{U}}(u, v) = \sum_{j=1}^{\infty} |u_j - v_j|.$$

Let  $\{V_i\}_{i=1}^{\infty}$  be a countable family of vector fields which are uniformly locally Lipschitz with uniformly locally Lipschitz derivatives. That is, for each  $x$  there exists a neighborhood  $\mathcal{O}$  of  $x$  on which each vector field  $V_i$  and its derivative  $V_i'$  is Lipschitz rank  $K_V$ . Suppose also that the vector fields  $V_i$  and their derivatives can be uniformly locally bounded. Then the generalized affine control system

$$(4.1.18) \quad f(t, x, u) = \sum_{i=1}^{\infty} u_i V_i(x)$$

satisfies our assumptions.

**4.1.3. Approximation of Relaxed Controls.** An important quality of relaxed controls is that, for free terminal point problems, no loss occurs if we expand our investigation to include relaxed controls. That is, a control  $u^0 \in \mathcal{U}$  that is optimal for the original problem will provide us with an optimal control  $\nu^0 := \delta_{u^0}$  among the relaxed controls. We make this precise in the following:

PROPOSITION 4.1.4. *Let  $\mu \in \mathcal{B}(\varepsilon_0)$  be arbitrary. For any  $\varepsilon > 0$  there is a control  $u \in \mathcal{U}$  satisfying  $d(u, u^0) < \varepsilon_0$  whose trajectory uniformly approximates that of  $\mu$ :*

$$(4.1.19) \quad \max_{t \in [0, T]} \|x(t; u) - x(t; \mu)\| < \varepsilon.$$

PROOF. The proof uses a technique similar to Gamkrelidze's *chattering control* [46]. Similar techniques may be found in [10]. Consider a relaxed control  $\mu \in \mathcal{M}_0$  with presentation (4.1.5) and let  $\pi$  be a partition of  $[0, T]$  into uniform intervals  $[t_i, t_{i+1}]$ . Let  $\Delta_\pi$  denote the diameter of  $\pi$ , so that  $t_i + \Delta_\pi = t_{i+1}$ . Define subintervals  $I_{i,j} \subset [t_i, t_{i+1}]$  through

$$\begin{aligned}
 I_{i,1} &= [t_i, t_i + \lambda_1 \Delta_\pi] \\
 I_{i,2} &= [t_i + \lambda_1 \Delta_\pi, t_i + (\lambda_1 + \lambda_2) \Delta_\pi] \\
 &\vdots \\
 I_{i,n} &= [t_i + (\lambda_1 + \cdots + \lambda_{n-1}) \Delta_\pi, t_{i+1}].
 \end{aligned}
 \tag{4.1.20}$$

Thus

$$I_{i,j} = \left[ t_i + \sum_{k < j} \lambda_k \Delta_\pi, t_i + \sum_{k \leq j} \lambda_k \Delta_\pi \right].
 \tag{4.1.21}$$

Note that  $[t_i, t_{i+1}] = \cup_{j=1}^n I_{i,j}$  and  $m(I_{i,j}) = \lambda_j \Delta_\pi$ .

LEMMA 4.1.5. *Let  $X$  be a Banach space and  $h \in L^1([0, T], X)$ . Let  $\varepsilon > 0$  be given and let  $I_{i,j}$  be defined through (4.1.21). Let  $I_j = \cup_i I_{ij}$ ,  $j = 1, \dots, n$ . If  $\Delta_\pi$  is sufficiently small then there holds*

$$\left\| \lambda_j \int_0^t h(s) ds - \int_{I_j \cap [0, t]} h(s) ds \right\|_X < \varepsilon.
 \tag{4.1.22}$$

for any  $t \in [0, T]$ .

PROOF. First note that because continuous functions are dense in  $L^1([0, T], X)$  it will suffice to prove this result under the assumption that  $h$  is continuous. We further assume without loss of generality that  $t = t_k$  for some integer  $k$ . Let  $\varepsilon > 0$



be given. We will show that for a continuous function  $h$ ,

$$(4.1.23) \quad \left\| \lambda_j \int_0^k h(s) ds - \int_{I_j \cap [0, t_k]} h(s) ds \right\|_X < \varepsilon,$$

provided  $\Delta_\pi$  is sufficiently small.

Note that

$$(4.1.24) \quad \int_{I_j \cap [0, t_k]} h(s) ds = \sum_{i=1}^k \int_{I_{i,j}} h(s) ds.$$

By uniform continuity we may choose our partition sufficiently fine that  $\|h(t_i) - h(s)\|_X < \varepsilon$  for all  $s \in [t_i, t_{i+1}]$ . Hence

$$(4.1.25) \quad \begin{aligned} & \left\| \lambda_j \int_0^{t_k} h(s) ds - \sum_{i=1}^k \int_{I_{i,j}} h(s) ds \right\|_X \\ & \leq \left\| \lambda_j \int_0^{t_k} h(s) ds - \sum_{i=1}^k \int_{I_{i,j}} h(t_i) ds \right\|_X + \left\| \sum_{i=1}^k \int_{I_{i,j}} h(s) - h(t_i) ds \right\|_X \\ & \leq \left\| \lambda_j \sum_{i=1}^k \int_{t_i}^{t_{i+1}} h(s) ds - \sum_{i=1}^k h(t_i) \lambda_j \Delta_\pi \right\|_X + \lambda_j \sum_{i=1}^k \varepsilon \Delta_\pi \\ & \leq \left\| \lambda_j \sum_{i=1}^k \int_{t_i}^{t_{i+1}} h(s) - h(t_i) ds \right\|_X + \lambda_j T \varepsilon \leq 2\lambda_j T \varepsilon. \end{aligned}$$

Since  $\varepsilon > 0$  is arbitrary this completes the proof of the lemma.  $\square$

Now let  $u(t)$  be the control which takes on the value  $u_j(t)$  for  $t \in I_{i,j}$ . We first claim that  $d(u, u^0) < \varepsilon_0$  for sufficiently fine partition  $\pi$ . To see this, we can apply Lemma 4.1.5 to find that

$$(4.1.26) \quad d(u, u^0)^p = \sum_{j=1}^n \int_{I_j} d_{\mathbb{U}}(u_j(t), u^0(t))^p dt = \sum_{j=1}^n \lambda_j \int_0^T d_{\mathbb{U}}(u_j(t), u^0(t))^p dt + o(1) < \varepsilon_0^p$$

when  $\Delta_\pi$  is sufficiently small, because  $\mu \in \mathcal{B}(\varepsilon_0)$ .

Now since  $d(u, u^0) < \varepsilon_0$  we have  $\delta_u \in \mathcal{B}$ . Consider

(4.1.27)

$$\begin{aligned} \|x(t; \mu) - x(t; u)\| &= \left\| \int_0^t \sum_{j=0}^n \lambda_j f(s, x(s; \mu), u_j(s)) - f(s, x(s; u), u(s)) ds \right\|_E \\ &\leq \left\| \int_0^t \sum_{j=0}^n \lambda_j f(s, x(s; \mu), u_j(s)) - f(s, x(s; \mu), u(s)) ds \right\|_E \\ &\quad + \left\| \int_0^t f(s, x(s; \mu), u(s)) - f(s, x(s; u), u(s)) ds \right\|_E. \end{aligned}$$

By Lemma 4.1.5 we can choose  $\Delta_\pi$  sufficiently small that

(4.1.28)

$$\begin{aligned} &\left\| \int_0^t \sum_{j=1}^n \lambda_j f(s, x(s; \mu), u_j(s)) - f(s, x(s; \mu), u(s)) ds \right\|_E \\ &= \left\| \sum_{j=1}^n \lambda_j \int_0^t f(s, x(s; \mu), u_j(s)) ds - \sum_{j=1}^n \int_{[0, t] \cap I_j} f(s, x(s; \mu), u_j(s)) ds \right\|_E \\ &\leq \sum_{j=1}^n \left\| \lambda_j \int_0^t f(s, x(s; \mu), u_j(s)) ds - \int_{[0, t] \cap I_j} f(s, x(s; \mu), u_j(s)) ds \right\|_E < \varepsilon e^{-M_f}. \end{aligned}$$

Hence, by (4.1.27) and (4.1.10), we find

$$(4.1.29) \quad \|x(t; \mu) - x(t; u)\| \leq \varepsilon e^{-M_f} + \int_0^t k_f(s) \|x(s; \mu) - x(s; u)\|_E ds.$$

Gronwall's lemma now implies that

$$(4.1.30) \quad \max_{t \in [0, T]} \|x(t; \mu) - x(t; u)\| \leq \varepsilon e^{\|k_f\|_{L^1} - M_f} \leq \varepsilon,$$

which completes the proof.  $\square$

We extend  $\Lambda_0$  to a functional  $\widehat{\Lambda}_0$  on  $\mathcal{M}_0$  through

$$(4.1.31) \quad \widehat{\Lambda}_0(\mu) = \ell(x(T; \mu)) + \int_0^T \widehat{L}(t, x(t; \mu), \mu(t)) dt$$

and consider the problem of minimizing  $\widehat{\Lambda}_0$  over  $\mathcal{M}_0$ .

PROPOSITION 4.1.6. *There is no loss in switching to relaxed controls. That is,*

$$(4.1.32) \quad \inf \{ \Lambda_0(u) : d(u, u^0) < \varepsilon_0 \} = \inf \{ \widehat{\Lambda}_0(\mu) : \mu \in \mathcal{B}(\varepsilon_0) \}.$$

PROOF. If  $u \in \mathcal{U}$  satisfies  $d(u, u^0) < \varepsilon_0$  then  $\mu(t) := \delta_{u(t)}$  is an element of  $\mathcal{B}(\varepsilon_0)$ . Therefore

$$(4.1.33) \quad \inf \{ \Lambda_0(u) : d(u, u^0) < \varepsilon_0 \} \geq \inf \{ \widehat{\Lambda}_0(\mu) : \mu \in \mathcal{B}(\varepsilon_0) \}.$$

Now suppose there exists  $\delta > 0$  such that

$$(4.1.34) \quad \inf \{ \Lambda_0(u) : d(u, u^0) < \varepsilon_0 \} \geq \inf \{ \widehat{\Lambda}_0(\mu) : \mu \in \mathcal{B}(\varepsilon_0) \} + \delta.$$

Choose control  $\mu \in \mathcal{B}(\varepsilon_0)$  with

$$(4.1.35) \quad \widehat{\Lambda}_0(\mu) < \inf \{ \widehat{\Lambda}_0(\mu) : \mu \in \mathcal{B}(\varepsilon_0) \} + \delta \leq \inf \{ \Lambda_0(u) : d(u, u^0) < \varepsilon_0 \}.$$

Introduce an auxiliary problem on  $E \times \mathbb{R}$  with control system

$$(4.1.36) \quad \begin{pmatrix} \dot{x}(t) \\ \dot{z}(t) \end{pmatrix} = g(t, x, z, u) := \begin{pmatrix} f(t, x, u) \\ L(t, x, u) \end{pmatrix},$$

fixed initial condition  $(x_0, 0)$ , and cost  $(x(T), z(T)) \mapsto \ell(x(T)) + z(T)$ . The system  $g$  satisfies the assumptions of Proposition 4.1.4 and so we may choose a control  $u \in \mathcal{U}$  with  $d(u, u^0) < \varepsilon_0$  whose trajectory satisfies

$$(4.1.37) \quad k_\ell \|x(T; u) - x(T; \mu)\|_E + |z(T; u) - z(T; \mu)| < \delta,$$

where  $k_\ell$  is a local Lipschitz constant for  $\ell$ . Thus

$$(4.1.38) \quad \ell(x(T; u)) + \int_0^T L(t, x(t; u), u(t)) dt \leq \ell(x(T; \mu)) + \int_0^T L(t, x(t; \mu), \mu(t)) dt + \delta.$$

This proves that

$$(4.1.39) \quad \Lambda_0(u) < \inf \left\{ \widehat{\Lambda}_0(\mu) : \mu \in \mathcal{B} \right\} + \delta \leq \inf \left\{ \Lambda_0(u) : d(u, u^0) < \varepsilon_0 \right\}$$

and this contradiction completes the proof.  $\square$

Thus we may consider the following relaxed problem of Bolza. Minimize  $\widehat{\Lambda}_0$  among controls  $\nu \in \mathcal{M}_0$  subject to fixed initial condition  $x_0$  and dynamic constraint

$$(4.1.40) \quad \dot{x}(t; \nu) = \widehat{f}(t, x(t; \nu), \nu(t)).$$

By Proposition 4.1.6, control  $\nu^0 := \delta_{u^0}$  is an optimal control for this problem among relaxed controls  $\mu \in \mathcal{B}(\varepsilon_0)$ .

**4.1.4. Decoupling the Dynamics.** In this section we demonstrate that the dynamic constraint (4.1.40) may be removed through the addition of a suitable penalty function. This approach first appeared in the work of Francis Clarke [29, 23] although the penalty function itself appeared earlier, in Filippov's approximation lemma for differential inclusions [44]. Filippov's approximation lemma is given a clear statement in Chapter Seven, where it is used extensively for the study of differential inclusions. The following lemma is a special case of this lemma, particular to our current setting, and admits a very short proof:

PROPOSITION 4.1.7. *Let  $y : [0, T] \rightarrow E$  be an absolutely continuous map for which*

$$(4.1.41) \quad \max_{t \in [0, T]} \|x^0(t) - y(t)\|_E < \varepsilon_1$$

and let  $\nu \in \mathcal{B}(\varepsilon_0)$ . Then we have

$$(4.1.42) \quad \max_{t \in [0, T]} \|y(t) - x(t; \nu)\|_E \leq e^{M_f} \int_0^T \|\dot{y}(s) - f(s, y(s), \nu(s))\|_E ds.$$

PROOF. For any  $t \in [0, T]$  we have

$$\begin{aligned}
(4.1.43) \quad \|y(t) - x(t; \nu)\|_E &\leq \int_0^t \left\| \dot{y}(s) - \widehat{f}(s, y(s), \nu(s)) \right\|_E \\
&\quad + \int_0^t \left\| \widehat{f}(s, y(s), \nu(s)) - \widehat{f}(s, x(s; \nu), \nu(s)) \right\|_E ds \\
&\leq \int_0^T \left\| \dot{y}(s) - \widehat{f}(s, y(s), \nu(s)) \right\|_E \\
&\quad + \int_0^t k_{f, \nu}(s) \|y(s) - x(s; \nu)\|_E dt.
\end{aligned}$$

Gronwall's lemma now implies (4.1.42).  $\square$

Proposition 4.1.7 will allow us to vary trajectories and controls independently. Introduce a space  $W$  of absolutely continuous mappings  $x : [0, T] \rightarrow E$  and define a functional  $\Lambda : W \times \mathcal{M}_0 \rightarrow \mathbb{R}$  by

$$(4.1.44) \quad \Lambda(x, \nu) = \ell(x(T)) + \int_0^T \widehat{L}(t, x(t), \nu(t)) dt + C_f \int_0^T \left\| \widehat{f}(t, x(t), \nu(t)) - \dot{x}(t) \right\|_E dt,$$

where  $C_f = (k_\ell + M_L) e^{M_f}$ . Here  $k_\ell$  is the local Lipschitz rank for  $\ell$  in a neighborhood of  $x^0(T)$ . This functional agrees with  $\widehat{\Lambda}_0(\nu)$  whenever  $\dot{x}(t) = \widehat{f}(t, x(t), \nu(t))$  and penalizes any discrepancy. We see below that for this choice of  $C_f$  (or larger) the penalization is exact.

PROPOSITION 4.1.8. *Suppose that  $\mu \in \mathcal{B}(\varepsilon_0)$  and that  $y : [0, T] \rightarrow E$  is an absolutely continuous mapping which satisfies (4.1.41). Then we have  $\Lambda(y, \mu) \geq \Lambda(x^0, \nu^0)$ .*

PROOF. The Lipschitz assumptions (4.1.12) imply that

$$\begin{aligned}
(4.1.45) \quad \Lambda(y, \mu) &= \ell(y(T)) + \int_0^T \widehat{L}(t, y(t), \mu(t)) dt + C_f \int_0^T \left\| \widehat{f}(t, y(t), \mu(t)) - \dot{y}(t) \right\|_E dt \\
&\geq \ell(x(T; \mu)) - k_\ell \|y(T) - x(T; \mu)\|_E + \int_0^T \widehat{L}(t, x(t; \mu), \mu(t)) dt \\
&\quad - \|k_{L, \mu}\|_{L^1} \max_{t \in [0, T]} \|y(t) - x(t; \mu)\|_E \\
&\quad + C_f \int_0^T \left\| \widehat{f}(t, y(t), \mu(t)) - \dot{y}(t) \right\|_E dt.
\end{aligned}$$

By Proposition 4.1.7 we have

$$(4.1.46) \quad \max_{t \in [0, T]} \|y(t) - x(t; \mu)\|_E \leq e^{M_f} \int_0^T \left\| \dot{y}(s) - \widehat{f}(s, y(s), \mu(s)) \right\|_E ds$$

and so (4.1.45) implies that  $\Lambda(y, \mu) \geq \Lambda(x(\cdot; \mu), \mu)$ . By Proposition 4.1.6 we also have

$$(4.1.47) \quad \Lambda(x(\cdot; \mu), \mu) = \widehat{\Lambda}_0(\mu) \geq \widehat{\Lambda}_0(\nu^0) = \Lambda(x^0, \nu^0),$$

completing the proof. □

**4.1.5. Necessary Optimality Conditions.** We are now in a position to derive the Maximum Principle. A central technique in our derivations of the Maximum Principle, throughout this dissertation, is a minimax theorem due to Ky Fan [41]. We provide a statement of this theorem for the convenience of the reader and suggest [15] for a modern statement and elegant proof. The theorem is given for functions which are *concave-convex-like* and we first define this notion:

DEFINITION 4.1.9. Let  $X$  and  $Y$  be nonempty sets. A function  $f : X \times Y \rightarrow \overline{\mathbb{R}}$  is *concave-convex-like* if for any  $\lambda \in [0, 1]$ ,

(i) For any  $x_1, x_2 \in X$  there exists  $x_3 \in X$  such that

$$(4.1.48) \quad f(x_3, y) \geq \lambda f(x_1, y) + (1 - \lambda)f(x_2, y)$$

for all  $y \in Y$ ;

(ii) For any  $y_1, y_2 \in Y$  there exists  $y_3 \in Y$  such that

$$(4.1.49) \quad f(x, y_3) \leq \lambda f(x, y_1) + (1 - \lambda)f(x, y_2)$$

for all  $x \in X$ .

Ky Fan's theorem is the following:

**THEOREM 4.1.10 (Fan).** *Suppose that  $X$  and  $Y$  are nonempty sets,  $X$  is a compact topological space,  $f : X \times Y \rightarrow \mathbb{R}$  concave-convex-like on  $X \times Y$ , and  $x \mapsto f(x, y)$  is upper semicontinuous for each  $y \in Y$ . Then*

$$(4.1.50) \quad \max_{x \in X} \inf_{y \in Y} \varphi(x, y) = \inf_{y \in Y} \max_{x \in X} \varphi(x, y).$$

With this in mind we now introduce variations that will lead to an application of this theorem. Let  $w \in L^1([0, T], E)$  be arbitrary and set

$$(4.1.51) \quad y(t) = \int_0^t w(s) ds.$$

Define a variation of  $x^0$  through

$$(4.1.52) \quad x^\lambda(t) = x^0(t) + \lambda y(t).$$

For arbitrary  $\nu \in \mathcal{M}_0$  we define a variation of  $\nu^0$  through

$$(4.1.53) \quad \nu^\lambda(t) = (1 - \lambda)\nu^0(t) + \lambda\nu(t).$$

One may check that, for  $\nu \in \mathcal{M}_0$ , we will have  $\nu^\lambda \in \mathcal{B}(\varepsilon_0)$  when  $\lambda > 0$  is sufficiently small. Moreover, for  $\lambda$  small enough, we will have

$$(4.1.54) \quad \max_{t \in [0, T]} \|x^\lambda(t) - x^0(t)\|_E < \varepsilon_1.$$

Proposition 4.1.8 then implies that for sufficiently small  $\Lambda$  we have

$$(4.1.55) \quad \Lambda(x^\lambda, \nu^\lambda) \geq \Lambda(x^0, \nu^0)$$

and consequently

$$(4.1.56) \quad \limsup_{\lambda \downarrow 0} \frac{\Lambda(x^\lambda, \nu^\lambda) - \Lambda(x^0, \nu^0)}{\lambda} \geq 0.$$

A careful analysis of this upper directional derivative will lead us directly to the Maximum Principle. We first make the following claim:

PROPOSITION 4.1.11. *For any  $w \in L^1([0, T], E)$  and  $\nu \in \mathcal{M}_0$  there exists mapping  $p \in L^\infty([0, T], E^*)$  with  $\|p\|_{L^\infty} \leq C_f$  for which*

$$(4.1.57) \quad \begin{aligned} & \limsup_{\lambda \downarrow 0} \frac{\Lambda(x^\lambda, \nu^\lambda) - \Lambda(x^0, \nu^0)}{\lambda} \\ & \leq \int_0^T \left\langle \ell'(x^0(T)) + \int_t^T \widehat{L}_x(s, x^0(s), \nu^0(s)) + \widehat{f}_x(s, x^0(s), \nu^0(s))^* p(s) ds - p(t), w(t) \right\rangle dt \\ & \quad + \int_0^T \widehat{L}(t, x^0(t), \nu(t)) - \widehat{L}(t, x^0(t), \nu^0(t)) dt \\ & \quad + \int_0^T \left\langle p(t), \widehat{f}(t, x^0(t), \nu(t)) - \widehat{f}(t, x^0(t), \nu^0(t)) \right\rangle dt, \end{aligned}$$

where  $\widehat{f}_x(s, x^0(s), \nu^0(s))^* : E^* \rightarrow E^*$  is the adjoint of the bounded linear operator  $\widehat{f}_x(s, x^0(s), \nu^0(s)) : E \rightarrow E$ .

PROOF. Let  $\lambda_n \downarrow 0$  be a sequence for which the limit supremum in (4.1.56) is attained. For each  $n$ , for each  $t$ , there exists  $p_n(t) \in E^*$  with  $\|p_n(t)\|_{E^*} \leq C_f$  for



which

$$(4.1.58) \quad \begin{aligned} C_f \left\| \widehat{f}(t, x^{\lambda_n}(t), \nu^{\lambda_n}(t)) - \dot{x}^{\lambda_n}(t) \right\|_E \\ = \left\langle p_n(t), \widehat{f}(t, x^{\lambda_n}(t), \nu^{\lambda_n}(t)) - \dot{x}^{\lambda_n}(t) \right\rangle. \end{aligned}$$

Since  $E^*$  is separable we may find a measurable selection  $p_n \in L^\infty([0, T], E^*)$  for which  $\|p_n\|_{L^\infty} \leq C_f$  and such that

$$(4.1.59) \quad \begin{aligned} \int_0^T \left\langle p_n(t), \widehat{f}(t, x^{\lambda_n}(t), \nu^{\lambda_n}(t)) - \dot{x}^{\lambda_n}(t) \right\rangle dt \\ = C_f \int_0^T \left\| \widehat{f}(t, x^{\lambda_n}(t), \nu^{\lambda_n}(t)) - \dot{x}^{\lambda_n}(t) \right\|_E dt. \end{aligned}$$

Since  $L^\infty([0, T], E^*) = L^1([0, T], E)^*$  and  $\|p_n\|_{L^\infty} \leq C_f$ , the Banach-Aloaglu theorem allows us to pass to a subsequence for which  $p_n \xrightarrow{w^*} p \in L^\infty$  with  $\|p\|_{L^\infty} \leq C_f$ .

For each  $n$ ,

$$(4.1.60) \quad \begin{aligned} \left\langle p_n(t), \widehat{f}(t, x^{\lambda_n}(t), \nu^{\lambda_n}(t)) - \dot{x}^{\lambda_n}(t) \right\rangle \\ = \lambda_n \left\langle p_n(t), \widehat{f}(t, x^{\lambda_n}, \nu(t)) - \widehat{f}(t, x^{\lambda_n}(t), \nu^0(t)) - w(t) \right\rangle \\ + \left\langle p_n(t), \widehat{f}(t, x^{\lambda_n}(t), \nu^0(t)) - \widehat{f}(t, x^0(t), \nu^0(t)) \right\rangle. \end{aligned}$$

Moreover, (4.1.10) implies that the sequence of functions

$$(4.1.61) \quad t \mapsto \frac{\widehat{f}(t, x^{\lambda_n}(t), \nu^0(t)) - \widehat{f}(t, x^0(t), \nu^0(t))}{\lambda_n}$$

converges in  $L^1([0, T], E)$  to  $f_x(t, x^0(t), \nu^0(t))y(t)$ . We then have

$$\begin{aligned}
(4.1.62) \quad & \frac{\Lambda(x^{\lambda_n}, \nu^{\lambda_n}) - \Lambda(x^0, \nu^0)}{\lambda_n} = \frac{\ell(x^{\lambda_n}(T)) - \ell(x^0(T))}{\lambda_n} \\
& + \frac{1}{\lambda_n} \int_0^T \widehat{L}(t, x^{\lambda_n}(t), \nu^0(t)) - \widehat{L}(t, x^0(t), \nu^0(t)) dt \\
& + \int_0^T \widehat{L}(t, x^{\lambda_n}(t), \nu(t)) - \widehat{L}(t, x^{\lambda_n}(t), \nu^0(t)) dt \\
& + \int_0^T \left\langle p_n(t), \widehat{f}(t, x^{\lambda_n}, \nu(t)) - \widehat{f}(t, x^{\lambda_n}(t), \nu^0(t)) - w(t) \right\rangle dt \\
& + \int_0^T \left\langle p_n(t), \frac{\widehat{f}(t, x^{\lambda_n}(t), \nu^0(t)) - \widehat{f}(t, x^0(t), \nu^0(t))}{\lambda_n} \right\rangle dt
\end{aligned}$$

In the limit we obtain

$$\begin{aligned}
(4.1.63) \quad & \limsup_{\lambda \downarrow 0} \frac{\Lambda(x^\lambda, \nu^\lambda) - \Lambda(x^0, \nu^0)}{\lambda} \leq \langle \ell'(x^0(T)), y(T) \rangle \\
& + \int_0^T \left\langle \widehat{L}_x(t, x^0(t), \nu^0(t)), y(t) \right\rangle \\
& + \int_0^T \widehat{L}(t, x^0(t), \nu(t)) - \widehat{L}(t, x^0(t), \nu^0(t)) dt \\
& + \int_0^T \left\langle p(t), \widehat{f}(t, x^0, \nu(t)) - \widehat{f}(t, x^0(t), \nu^0(t)) - w(t) \right\rangle dt \\
& + \int_0^T \left\langle p(t), \widehat{f}_x(t, x^0(t), \nu^0(t))y(t) \right\rangle dt.
\end{aligned}$$

Now note that  $\langle \ell'(x^0(T)), y(T) \rangle = \int_0^T \langle \ell'(x^0(T)), w(t) \rangle dt$  and

$$(4.1.64) \quad \int_0^T \langle L_x(t, x^0(t), \nu^0(t)), y(t) \rangle dt = \int_0^T \left\langle L_x(t, x^0(t), \nu^0(t)), \int_0^t w(s) ds \right\rangle dt.$$

We can integrate (4.1.64) by parts to find

$$(4.1.65) \quad \int_0^T \langle L_x(t, x^0(t), \nu^0(t)), y(t) \rangle dt = \int_0^T \left\langle \int_t^T L_x(s, x^0(s), \nu^0(s)) ds, w(t) \right\rangle dt.$$

Similarly,

$$(4.1.66) \quad \int_0^T \langle p(t), f_x(t, x^0(t), \nu^0(t))y(t) \rangle dt = \int_0^T \left\langle \int_t^T f_x(s, x^0(s), \nu^0(s))^* p(s) ds, w(t) \right\rangle dt.$$

Rearranging (4.1.63) now gives us (4.1.57).  $\square$

As mentioned above, this directional derivative of the functional  $\Lambda$  implies the Maximum Principle. In particular we obtain the following:

**THEOREM 4.1.12.** *Suppose that  $u^0$  is an optimal control for the problem of minimizing*

$$(4.1.67) \quad \ell(x(T)) + \int_0^T L(t, x(t), u(t)) dt$$

*subject to fixed initial condition  $x(0) = x_0$  and dynamic constraint*

$$(4.1.68) \quad \dot{x}(t) = f(t, x(t), u(t)).$$

*There exists an absolutely continuous map  $p : [0, T] \rightarrow E^*$  which satisfies for almost all  $t$*

$$(4.1.69) \quad \dot{p}(t) = L_x(t, x^0(t), u^0(t)) - f_x(t, x^0(t), u^0(t))^* p(t),$$

*where  $f_x(t, x^0(t), u^0(t))^* : E^* \rightarrow E^*$  denotes the adjoint of the bounded linear operator  $f_x(t, x^0(t), u^0(t)) : E \rightarrow E$ . We further have  $p(T) = -\ell'(x^0(T))$  and*

$$(4.1.70) \quad \int_0^T H(t, x^0(t), p(t), u^0(t)) dt = \max_{u \in \mathcal{U}} \int_0^T H(t, x^0(t), p(t), u(t)) dt,$$

*where  $H : [0, T] \times E \times E^* \times \mathbb{U} \rightarrow \mathbb{R}$  is the Pontryagin function*

$$(4.1.71) \quad H(t, x, p, u) = \langle p, f(t, x, u) \rangle - L(t, x, u).$$

In the case where  $\mathbb{U}$  is separable then for almost all  $t$ ,

$$(4.1.72) \quad H(t, x^0(t), p(t), u^0(t)) = \max_{u \in \mathbb{U}} H(t, x^0(t), p(t), u),$$

We note that an integral form of the Maximum Principle of the type in (4.1.70), under more restrictive assumptions, may be found in L.C. Young's lectures on the subject [94].

PROOF. Denoting the left-hand side of (4.1.57) by  $\Phi(p, \nu, w)$  and letting  $\mathcal{K} \subset L^\infty([0, T], E^*)$  be the set of all  $p$  with  $\|p\|_{L^\infty} \leq C_f$  we see that

$$(4.1.73) \quad \inf_{\nu \in \mathcal{M}_0, w \in L^1} \max_{p \in \mathcal{K}} \Phi(p, \nu, w) \geq 0.$$

Since  $\mathcal{K}$  is weakly\*-compact we may apply Fan's minimax theorem to find

$$(4.1.74) \quad \max_{p \in \mathcal{K}} \inf_{\nu \in \mathcal{M}_0, w \in L^1} \Phi(p, \nu, w) \geq 0.$$

Choose  $p \in \mathcal{K}$  which attains this maximum. For this  $p$ , the inequality (4.1.57) holds for all  $w \in L^1$  and  $\nu \in \mathcal{M}_0$ .

In order to derive a Maximum Principle instead of the corresponding (and equivalent) Minimum Principle, we replace  $p$  with  $-p$ . Restating (4.1.57) in terms of controls  $u \in \mathcal{U}$  we see that for all  $w \in L^1$  and  $u \in \mathcal{U}$  there holds

$$(4.1.75) \quad \int_0^T \left\langle \ell'(x^0(T)) + \int_t^T L_x(s, x^0(s), u^0(s)) - f_x(s, x^0(s), u^0(s))^* p(s) ds + p(t), w(t) \right\rangle dt \\ + \int_0^T H(t, x^0(t), p(t), u^0(t)) - H(t, x^0(t), p(t), u(t)) dt \geq 0.$$

First take  $u = u^0$  to see that for any  $w \in L^1$ ,

$$(4.1.76) \quad \int_0^T \left\langle \ell'(x^0(T)) + \int_t^T L_x(s, x^0(s), u^0(s)) - f_x(s, x^0(s), u^0(s))^* p(s) ds + p(t), w(t) \right\rangle dt \geq 0.$$

The fundamental lemma now implies that

$$(4.1.77) \quad p(t) = -\ell'(x^0(T)) - \int_t^T L_x(s, x^0(s), u^0(s)) - f_x(s, x^0(s), u^0(s))^* p(s) ds.$$

This implies that  $p$  is absolutely continuous and satisfies (4.1.69) for almost all  $t$ , along with the boundary condition  $p(T) = -\ell'(x^0(T))$ .

On the other hand, taking  $w = 0$  we see that for any control  $u$  there holds

$$(4.1.78) \quad \int_0^T H(t, x^0(t), p(t), u^0(t)) - H(t, x^0(t), p(t), u(t)) dt \geq 0,$$

and this is (4.1.70).

The proof is completed through the following lemma:

LEMMA 4.1.13. *If  $\mathbb{U}$  is separable then (4.1.78) implies that (4.1.72) holds for almost all  $t$ .*

PROOF. Let  $\{u_n\}_{n \in \mathbb{N}} \subset \mathbb{U}$  be a countable dense subset. Choose a particular  $u_n$  and fix any time  $t_0 \in [0, T]$ . Given  $\delta > 0$  we define a control  $u^\delta \in \mathcal{U}$  through

$$(4.1.79) \quad u^\delta(t) = \begin{cases} u_n & t \in [t_0, t_0 + \delta] \\ u^0(t) & \text{otherwise} \end{cases}$$

For this control (4.1.78) implies

$$(4.1.80) \quad \int_{t_0}^{t_0 + \delta} H(t, x^0(t), p(t), u^0(t)) - H(t, x^0(t), p(t), u_n) dt \geq 0.$$

This implies that  $H(t_0, x^0(t_0), p(t_0), u^0(t_0)) \geq H(t_0, x^0(t_0), p(t_0), u_n)$ . Let  $I_n \subset [0, T]$  denote the set of all  $t_0 \in [0, T]$  for which this inequality holds and note that  $m(I_n) = T$ . Taking  $I = \cup_{n \in \mathbb{N}} I_n$  we see that for all  $t \in I$  and any  $n \in \mathbb{N}$  we have

$$(4.1.81) \quad H(t, x^0(t), p(t), u^0(t)) \geq H(t, x^0(t), p(t), u_n).$$

Continuity now implies that (4.1.72) holds for all  $t \in I$ , a set of full measure.  $\square$

This completes the proof of Theorem 4.1.12. □

Let us pause to make some comments regarding this proof. Note that the pair  $(x^0(t), p(t))$  evolves in  $E \times E^*$  according to the symplectic system

$$(4.1.82) \quad \begin{cases} \dot{x}^0(t) &= H_p(t, x^0(t), p(t), u^0(t)) \\ \dot{p}(t) &= -H_x(t, x^0(t), p(t), u^0(t)) \end{cases}$$

On a manifold, this system corresponds to the evolution of a map  $\omega : [0, T] \rightarrow T^*M$  according to the vector field  $\vec{H}$  on  $T^*M$  which generalizes the above notion of symplectic vector field, as discussed in the previous chapter.

Note also that our proof uses the derivative  $f_x$ . This can complicate the situation on a manifold, as one cannot differentiate vector fields without some notion of a *connection*, see e.g. [19]. Though connections play an important role in the geometric theory of control [12] we have made an effort in this dissertation to maintain a close analogy between the general manifold case and the linear case and so have avoided the theory of connections. In the following sections we will develop techniques which allow us to avoid taking this derivative – or perhaps more accurately to conceal this derivative within the construction of the symplectic vector field  $\vec{H}$ .

**4.1.6. Applications to the Calculus of Variations.** Before beginning work on the theory of optimal control for Banach manifolds, we wish to emphasize that all first order necessary conditions from the Calculus of Variations are contained in the Maximum Principle. Thus optimal control, though very useful in engineering applications, also plays an important role in the realm of pure mathematics.

Consider the following problem from the Calculus of Variations: *Minimize*

$$(4.1.83) \quad \ell(x(T)) + \int_0^T L(t, x(t), \dot{x}(t)) dt$$

with fixed initial condition  $x_0$  and free terminal point. Let us suppose that  $L$  is measurable in time and  $C^1$ -smooth in  $(x, v)$  with locally integrable Lipschitz first derivative. Let  $x^0$  be a strong local minimizer for this problem.

Introduce the control system  $f(t, x, u) = u$ , where control  $u$  takes values in  $\mathbb{U} := E$ . Because  $E$  is separable, the Maximum Principle implies the existence of an absolutely continuous curve  $p : [0, T] \rightarrow E^*$  with  $p(T) = -\ell'(x^0(T))$  that satisfies, for almost all  $t$

$$(4.1.84) \quad \dot{p}(t) = L_x(t, x^0(t), \dot{x}^0(t)).$$

In addition we have for any  $y \in E$  the inequality

$$(4.1.85) \quad H(x^0(t), p(t), \dot{x}^0(t)) \geq H(x^0(t), p(t), y).$$

Rearranging we find that for almost all  $t \in [0, T]$ , for any  $y \in E$ ,

$$(4.1.86) \quad L(t, x^0(t), y) \geq L(t, x^0(t), \dot{x}^0(t)) + \langle p(t), y - \dot{x}^0(t) \rangle.$$

We note that (4.1.86) implies that

$$(4.1.87) \quad p(t) = L_v(t, x^0(t), \dot{x}^0(t)).$$

As a consequence, (4.1.84) gives us the classical Euler equation. Equation (4.1.87) implies that  $L_v(t, x^0(t), \dot{x}^0(t))$  is continuous, which is the first Erdmann condition. Finally, (4.1.86) is itself the classical Weierstrass condition. We suggest Clarke's recent book [25] or the classic [48] for more in this direction. Additional useful references for the Calculus of Variations and optimal control are the texts by Bloch [12] and Cesari [21].

## 4.2. Integration by Parts, Fundamental and duBois-Reymond Lemmas

We now turn to the development of the theory necessary for carrying out the above proof on a manifold, including geometric versions of the integration-by-parts formula and the fundamental lemma for Banach manifolds. We remind the reader that if  $q : [0, T] \rightarrow M$  is continuous, a mapping  $\zeta : [0, T] \rightarrow T^*M$  along  $q$  is merely a map for which  $\pi(\zeta(t)) = q(t)$ . We also caution the reader that we are writing  $\pi$  both for the projection  $\pi_{TM} : TM \rightarrow M$  and the projection  $\pi_{T^*M} : T^*M \rightarrow M$ .

PROPOSITION 4.2.1 (Integration by Parts). *Let  $P_{s,t}$  denote the flow of a nonautonomous  $C^1$ -smooth vector field  $V_t$ . Fix  $q_0 \in M$  and let  $q(t) = P_{0,t}(q_0)$ . Suppose that we have  $L^1$ -bounded mappings  $\zeta : [0, T] \rightarrow T^*M$  and  $v : [0, T] \rightarrow TM$  along  $q$ . Then the following integration by parts formula holds:*

$$(4.2.1) \quad \int_0^T \left\langle \zeta(t), \int_0^t P_{s,t*} v(s) ds \right\rangle dt = \int_0^T \left\langle \int_t^T P_{t,s}^* \zeta(s) ds, v(t) \right\rangle dt.$$

PROOF. The pairing  $t \mapsto \left\langle \int_t^T P_{t,s}^* \zeta(s) ds, \int_0^t P_{s,t*} v(s) ds \right\rangle$  is absolutely continuous on  $[0, T]$  and so we obtain

$$(4.2.2) \quad \int_0^T \frac{d}{dt} \left\langle \int_t^T P_{t,s}^* \zeta(s) ds, \int_0^t P_{s,t*} v(s) ds \right\rangle dt = 0.$$

The integrand in (4.2.2) may be written as

$$\frac{d}{d\varepsilon} \Big|_{\varepsilon=0} \left\langle \int_{t+\varepsilon}^T P_{t+\varepsilon,s}^* \zeta(s) ds, \int_0^{t+\varepsilon} P_{s,t+\varepsilon*} v(s) ds \right\rangle$$

Using the semigroup property of flows we may remove dependence on  $\varepsilon$  from the integrands. First write:

$$(4.2.3) \quad \begin{aligned} & \left\langle \int_{t+\varepsilon}^T (P_{t,s} \circ P_{t+\varepsilon,t})^* \zeta(s) ds, \int_0^{t+\varepsilon} (P_{t,t+\varepsilon} \circ P_{s,t})_* v(s) ds \right\rangle \\ &= \left\langle \int_{t+\varepsilon}^T P_{t+\varepsilon,t}^* \circ P_{t,s}^* \zeta(s) ds, \int_0^{t+\varepsilon} P_{t,t+\varepsilon*} \circ P_{s,t*} v(s) ds \right\rangle \end{aligned}$$



These integrals take place in  $T_{q(t+\varepsilon)}^*M$  and  $T_{q(t+\varepsilon)}M$ . Now write

$$\begin{aligned}
(4.2.4) \quad &= \left\langle P_{t+\varepsilon,t}^* \circ P_{t,t+\varepsilon}^* \int_{t+\varepsilon}^T P_{t+\varepsilon,t}^* \circ P_{t,s}^* \zeta(s) ds, \int_0^{t+\varepsilon} P_{t,t+\varepsilon} \circ P_{s,t} v(s) ds \right\rangle \\
&= \left\langle \int_{t+\varepsilon}^T P_{t,t+\varepsilon}^* \circ P_{t+\varepsilon,t}^* \circ P_{t,s}^* \zeta(s) ds, \int_0^{t+\varepsilon} P_{s,t} v(s) ds \right\rangle \\
&= \left\langle \int_{t+\varepsilon}^T P_{t,s}^* \zeta(s) ds, \int_0^{t+\varepsilon} P_{s,t} v(s) ds \right\rangle.
\end{aligned}$$

The integrals appearing in the last expression take place entirely in  $T_{q(t)}^*M$  and  $T_{q(t)}M$ . Classic integration theory now yields, for almost all  $t \in [0, T]$ ,

$$\begin{aligned}
\frac{d}{dt} \left\langle \int_t^T P_{t,s}^* \zeta(s) ds, \int_0^t P_{s,t} v(s) ds \right\rangle &= \frac{d}{d\varepsilon} \Big|_{\varepsilon=0} \left\langle \int_{t+\varepsilon}^T P_{t+\varepsilon,s}^* \zeta(s) ds, \int_0^{t+\varepsilon} P_{s,t+\varepsilon} v(s) ds \right\rangle \\
&= - \left\langle \zeta(t), \int_0^t P_{s,t} v(s) ds \right\rangle + \left\langle \int_t^T P_{t,s}^* \zeta(s) ds, v_t \right\rangle
\end{aligned}$$

Substituting into (4.2.2) and rearranging gives us the result.  $\square$

LEMMA 4.2.2 (The Fundamental Lemma). *Suppose that  $\zeta : [0, T] \rightarrow T^*M$  is an  $L^a$ -bounded mapping along an absolutely continuous curve  $q(t)$  and that for any  $L^b$ -bounded mapping  $w$  along  $q$  we have*

$$(4.2.5) \quad \int_0^T \langle \zeta(t), w(t) \rangle dt \geq 0.$$

*Then for almost all  $t$ ,  $\zeta(t) = 0$ .*

PROOF. Choose  $t_0 \in [0, T]$  and let  $(\mathcal{O}, \varphi)$  be a coordinate chart with  $q(t_0) \in \mathcal{O}$ . Choose  $\delta > 0$  so that for any  $t \in (t_0 - \delta, t_0 + \delta)$  we have  $q(t) \in \mathcal{O}$ . Let  $t_1 \in (t_0 - \delta, t_0 + \delta)$  be a Lebesgue point for  $\zeta$ .

Let  $u \in E$  be arbitrary and define, for  $0 < \lambda < t_0 - t_1 + \delta$ ,

$$(4.2.6) \quad w^\lambda(t) = \begin{cases} \varphi_*^{-1}(\varphi(q(t)))u & t \in [t_1, t_1 + \lambda] \\ 0 & \text{elsewhere} \end{cases}$$

This map is locally  $L^\infty$ -bounded, hence  $L^b$ -bounded, and so we must have

$$(4.2.7) \quad \int_0^T \langle \zeta(t), w^\lambda(t) \rangle dt \geq 0.$$

Hence

$$(4.2.8) \quad \int_{t_1}^{t_1+\lambda} \langle \varphi^{-1*}\zeta(t), u \rangle dt \geq 0.$$

Since  $t_1$  is a Lebesgue point we obtain  $\langle \varphi^{-1*}\zeta(t_1), u \rangle \geq 0$  for all  $u \in E$ . Recall that  $\varphi^{-1*}\zeta(t_1)$  is the local coordinate representation of  $\zeta(t_1)$ . The inequality just stated implies that  $\|\varphi^{-1*}\zeta(t_1)\|_E = 0$  and hence  $\zeta(t_1) = 0$ . Since almost all times are Lebesgue the proof is complete.  $\square$

There is another classical lemma that is often used in the calculus of variations called the *duBois-Reymond lemma*. We provide a statement for the linear case below:

LEMMA 4.2.3 (duBois-Reymond). *Suppose that  $p \in L^\infty([0, T], E^*)$  is such that for any  $w \in L^1([0, T], E)$  satisfying*

$$(4.2.9) \quad \int_0^T w(t) dt = 0$$

*there holds*

$$(4.2.10) \quad \int_0^T \langle p(t), w(t) \rangle dt = 0.$$

*Then there exists  $p_T \in E^*$  such that for almost all  $t$ ,  $p(t) = p_T$ .*

The question of how to extend this result to manifolds is an interesting one, since we cannot expect the same conclusion, that  $p(t)$  is almost everywhere equal to a constant map. We find instead that, for any manifold on which an extension lemma such as Lemma 3.2.8 holds, the duBois-Reymond lemma has a natural analogue given in terms of flows and that flow invariance replaces the claim that  $p$  is constant. For the remainder of this subsection we will assume that  $M$  admits a locally  $L^1$ -bounded,  $C^1$ -smooth extension  $V_t$  of  $\dot{q}(t)$  and write  $P_{s,t}$  for the flow of  $V_t$ .

We define a class of mappings  $w : [0, T] \rightarrow TM$  which generalize the property expressed in (4.2.9):

DEFINITION 4.2.4. An  $L^1$ -bounded mapping  $w : [0, T] \rightarrow TM$  along  $q(t)$  is said to be *endpoint-preserving* if

$$(4.2.11) \quad \int_0^T P_{t,T*}(q(t))w(t) dt = 0.$$

Momentarily we will show that, under a mild assumption on the smoothness of  $E$ , any endpoint preserving map  $w$  along  $q$  can be extended to a vector field  $W_t$  with the property that the perturbed flow  $V_t + \lambda W_t$  preserves  $q(T)$  for small  $\lambda$ , justifying the name. Before we do this we suggest the following generalization of the duBois-Reymond lemma:

LEMMA 4.2.5 (duBois-Reymond Lemma). *Suppose that  $\zeta : [0, T] \rightarrow T^*M$  is a mapping along  $q(t)$  and suppose that for each endpoint-preserving map  $w : [0, T] \rightarrow TM$  along  $q$  there holds*

$$(4.2.12) \quad \int_0^T \langle \zeta(t), w(t) \rangle dt = 0.$$

*Then there exists  $\zeta_T \in T_{q(T)}^*M$  such that for almost all  $t$  we have*

$$(4.2.13) \quad \zeta(t) = P_{t,T}^* \zeta_T.$$

Thus  $\zeta$  is flow-invariant for  $P_{s,t}$ .

PROOF. Choose times  $t_0, t_1$  which are Lebesgue times for  $\zeta$ . Let  $v \in T_{q(T)}M$  be arbitrary and consider the following map:

$$(4.2.14) \quad w(t) = \begin{cases} P_{T,t}^*v & t \in [t_0, t_0 + \delta] \\ -P_{T,t}^*v & t \in [t_1, t_1 + \delta] \\ 0 & \text{elsewhere} \end{cases}$$

One can check that  $w$  is endpoint-preserving. As a consequence we have

$$(4.2.15) \quad \int_{t_0}^{t_0+\delta} \langle \zeta(t), P_{T,t}^*v \rangle dt = \int_{t_1}^{t_1+\delta} \langle \zeta(t), P_{T,t}^*v \rangle dt.$$

Thus

$$(4.2.16) \quad \int_{t_0}^{t_0+\delta} \langle P_{T,t}^* \zeta(t), v \rangle dt = \int_{t_1}^{t_1+\delta} \langle P_{T,t}^* \zeta(t), v \rangle dt.$$

Since the times  $t_0, t_1$  are Lebesgue times we arrive at

$$(4.2.17) \quad \langle P_{T,t_0}^* \zeta(t_0), v \rangle = \langle P_{T,t_1}^* \zeta(t_1), v \rangle$$

for any  $v \in T_{q(T)}M$ . Defining  $\zeta_T \in T_{q(T)}^*M$  through  $\zeta_T := P_{T,t_0}^* \zeta(t_0)$  we see that for almost all  $t$ ,  $P_{t,T}^* \zeta(t) = \zeta_T$ .  $\square$

We now justify the name *endpoint-preserving*:

LEMMA 4.2.6. *If  $E$  admits a  $C^1$ -smooth bump function and  $V_t$  is  $C^2$ -smooth then to each endpoint-preserving map  $w : [0, T] \rightarrow TM$  we can assign a  $C^1$ -smooth vector field  $W_t$  for which  $W_t(q(t)) = w(t)$  and such that for sufficiently small  $\lambda > 0$ , the flow  $P_{s,t}^\lambda$  of the vector field  $V_t + \lambda W_t$  satisfies*

$$(4.2.18) \quad P_{0,T}^\lambda(q_0) = q(T).$$

PROOF. Recall that, by Proposition 3.1.5, there holds

$$(4.2.19) \quad P_{0,t}^\lambda = P_{0,t} \circ C_{0,t}^\lambda$$

where  $C_{0,t}^\lambda$  is the flow of the vector field  $\lambda P_{t,0*} W_t$ . We will construct  $W_t$  so that, for small  $\lambda$ , there will hold  $C_{0,T}^\lambda(q_0) = q_0$ .

To do this, let  $(\varphi, \mathcal{O})$  be a coordinate chart with  $q_0 \in \mathcal{O}$  and let  $b : M \rightarrow [0, 1]$  be a  $C^1$ -smooth bump function which is identically equal to 1 in a neighborhood  $\mathcal{O}_0 \subset \overline{\mathcal{O}_0} \subset \mathcal{O}$  of  $q_0$ . Define a curve  $\tilde{w} \in E$  by

$$(4.2.20) \quad \tilde{w}(t) = \varphi_* \circ P_{t,0*}(q(t))w(t).$$

Consider the vector field

$$(4.2.21) \quad \widetilde{W}_t(q) = b(q)\varphi_*^{-1}(\varphi(q))\tilde{w}(t),$$

extended smoothly to zero outside of  $\mathcal{O}$ . Let  $W_t = P_{0,t*}\widetilde{W}_t$  and note that  $W_t$  is  $C^1$ -smooth.

We claim that for small  $\lambda$ , the flow  $C_{s,t}^\lambda$  of the vector field  $\lambda P_{t,0*} W_t$  satisfies  $C_{0,T}^\lambda(q_0) = q_0$ . To see this, note that for small time we have  $C_{0,t}^\lambda(q_0) \in \mathcal{O}_0$ . For these times there holds

$$(4.2.22) \quad \frac{d}{dt}\varphi \circ C_{0,t}^\lambda(q_0) = \lambda\varphi_* \circ (P_{t,0*} W_t)(C_{0,t}^\lambda(q_0)) = \lambda\tilde{w}(t).$$

When  $\lambda$  is sufficiently small we will then have

$$(4.2.23) \quad \varphi \circ C_{0,t}^\lambda(q_0) = \varphi(q_0) + \lambda \int_0^t \tilde{w}(s) ds \in \varphi(\mathcal{O}_0)$$

for all  $t \in [0, T]$ . Finally, note that

$$(4.2.24) \quad \int_0^T \tilde{w}(t) dt = \int_0^T \varphi_* P_{t,0*} w(t) dt = \varphi_* P_{T,0*} \int_0^T P_{t,T*} w(t) dt = 0.$$

Thus, for small  $\lambda$  we have  $C_{0,T}^\lambda(q_0) = q_0$  and hence  $P_{0,T}^\lambda(q_0) = q(T)$ .  $\square$

We now turn to a study of Pontryagin Maximum Principle in a fully geometric setting.

### 4.3. Pontryagin Maximum Principle

We now return to our Bolza problem of minimizing the functional

$$(4.3.1) \quad \Lambda_0(u) = \ell(q(T; u)) + \int_0^T L(t, q(t; u), u(t)) dt$$

subject to fixed initial condition  $q_0 \in M$  and dynamic constraint

$$(4.3.2) \quad \dot{q}(t; u) = f(t, q(t; u), u(t)).$$

We suppose that  $u^0$  is an optimal control with trajectory  $q^0$ . As before we assume that measurable controls  $u$  take values in a metric space  $\mathbb{U}$ . We work with the same class  $\mathcal{M}_0$  of relaxed controls and define  $\nu^0(t) := \delta_{u^0(t)}$ .

We make the following assumptions on this problem:

ASSUMPTION 4.3.1. We assume that  $f : [0, T] \times M \times \mathbb{U} \rightarrow TM$  and  $L : [0, T] \times M \times \mathbb{U} \rightarrow \mathbb{R}$  are Carathéodory. We further assume there exist constants  $\varepsilon_0, M_f, M_L$  and a coordinate chart  $(\mathcal{O}, \varphi)$  with  $q_0 \in \mathcal{O}$  such that for any relaxed control  $\mu \in \mathcal{B}(\varepsilon_0)$  we have:

- (i) The nonautonomous vector field  $f(t, q, u^0(t))$  induced by the optimal control is a locally  $L^1$ -bounded,  $C_{lip}^1$ -smooth vector field with a well-defined flow  $P_{0,t}(q)$  for all  $q \in \mathcal{O}$  for all  $t \in [0, T]$ ;
- (ii) The differential equation

$$(4.3.3) \quad \dot{q}(t) = \widehat{f}(t, q(t), \mu(t))$$

admits a solution  $q(t; \mu)$  with  $q(0; \mu) = q_0$  which is defined for all  $t \in [0, T]$  and which satisfies  $q(t; \mu) \in P_{0,t}(\mathcal{O})$  for all  $t$ ;

- (iii) The Lagrangian coordinates  $\psi_t : P_{0,t}(\mathcal{O}) \rightarrow \varphi(\mathcal{O})$  corresponding to the nonautonomous vector field  $f(t, q, u^0(t))$  are such that for any  $\mu \in \mathcal{B}(\varepsilon_0)$  there are  $L^1$  functions  $k_f, k_L$  with  $L^1$ -norms bounded by  $M_f$  and  $M_L$ , respectively such that

$$(4.3.4) \quad \left\| (\psi_{t*} \widehat{f})(t, x_1, \mu(t)) - (\psi_{t*} \widehat{f})(t, x_2, \mu(t)) \right\|_E \leq k_{f,\mu}(t) \|x_1 - x_2\|_E$$

and

$$(4.3.5) \quad \left| \widehat{L}(t, \psi_t^{-1}(x_1), \mu(t)) - \widehat{L}(t, \psi_t^{-1}(x_2), \mu(t)) \right| \leq k_{L,\mu}(t) \|x_1 - x_2\|_E$$

Following the route from the linear case, we establish an approximation lemma for relaxed controls and employ it to implement exact penalization.

**PROPOSITION 4.3.2.** *Suppose that  $\mu \in \mathcal{B}(\varepsilon_0)$ . Then for any  $\varepsilon > 0$  there exists a control  $u \in \mathcal{U}$  such that  $d(u, u^0) < \varepsilon_0$  and*

$$(4.3.6) \quad \max_{t \in [0, T]} \|\psi_t(q(t; u)) - \psi_t(q(t; \mu))\|_E < \varepsilon.$$

**PROOF.** Introduce a control system  $g : [0, T] \times \varphi(\mathcal{O}) \times \mathbb{U} \rightarrow \varphi(\mathcal{O})$  through

$$(4.3.7) \quad g(t, x, w) = (\psi_{t*} f)(t, x, w) - (\psi_{t*} f)(t, x, u^0(t)).$$

Let  $\mu \in \mathcal{B}(\varepsilon_0)$  and set  $x(t; \mu) = \psi_t(q(t; \mu))$ , so that for almost all  $t$  there holds

$$(4.3.8) \quad \frac{d}{dt} x(t; \mu) = \widehat{g}(t, x(t; \mu), \mu(t)).$$

Our assumptions on  $f$  assure us that for any  $\mu \in \mathcal{B}(\varepsilon_0)$  there exists an  $L^1$  function  $k_{f,\mu}$  such that for any  $x_1, x_2 \in \varphi(\mathcal{O})$  we have

$$(4.3.9) \quad \|\widehat{g}(t, x_1, \mu(t)) - \widehat{g}(t, x_2, \mu(t))\|_E \leq k_{f,\mu} \|x_1 - x_2\|_E.$$

Moreover, we can bound  $\|k_{f,\mu}\|_{L^1}$  independently of  $\mu \in \mathcal{B}(\varepsilon_0)$ . The proof of Proposition 4.1.4 required only this Lipschitz property and so we may repeat the same argument, replacing control system  $f$  with  $g$  to find, for any  $\varepsilon > 0$ , a control  $u$  such that the solution  $x(t; u)$  to the system

$$(4.3.10) \quad \dot{x}(t; u) = g(t, x(t; u), u(t))$$

satisfies

$$(4.3.11) \quad \max_{t \in [0, T]} \|x(t; u) - x(t; \mu)\|_E < \varepsilon.$$

By Proposition 3.4.3 we have  $x(t; u) = \psi_t(q(t; u))$  and because  $\varepsilon > 0$  was arbitrary the proof is complete.  $\square$

In the next two propositions we carry out the decoupling technique that was successful in  $E$ . Define a functional  $\widehat{\Lambda}_0 : \mathcal{M}_0 \rightarrow \mathbb{R}$  through

$$(4.3.12) \quad \widehat{\Lambda}_0(\mu) = \ell(q(T; \mu)) + \int_0^T \widehat{L}(t, q(t; \mu), \mu(t)) dt.$$

PROPOSITION 4.3.3. *Suppose that  $q : [0, T] \rightarrow M$  is an absolutely continuous map such that  $q(t) \in P_{0,t}(\mathcal{O})$  for all  $t$  and  $\mu \in \mathcal{B}(\varepsilon_0)$  is a relaxed control. Then we have*

$$(4.3.13) \quad \max_{t \in [0, T]} \|\psi_t(q(t)) - \psi_t(q(t; \mu))\|_E < e^{2M_f} \int_0^T \left\| \psi_{t*} \dot{q}(t) - \psi_{t*} \widehat{f}(t, q(t), \mu(t)) \right\|_E dt.$$

*Note that our assumptions on  $f$  imply that  $\psi_t(q(t; \mu))$  is well-defined for all  $t$ .*



PROOF. Consider

$$\begin{aligned}
(4.3.14) \quad & \|\psi_t(q(t)) - \psi_t(q(t; \mu))\|_E \leq \int_0^t \left\| \frac{d}{ds} \psi_s(q(s)) - \frac{d}{ds} \psi_s(q(s; \mu)) \right\|_E ds \\
& \leq \int_0^t \left\| \frac{d}{ds} \psi_s(q(s)) - \psi_{s*} \widehat{f}(s, q(s), \mu(s)) + \psi_{s*} \widehat{f}(s, q(s), \nu^0(s)) \right\|_E ds \\
& \quad + \int_0^t \left\| \psi_{s*} \widehat{f}(s, q(s), \mu(s)) - \psi_{s*} \widehat{f}(s, q(s), \nu^0(s)) - \frac{d}{ds} \psi_s(q(s; \mu)) \right\|_E ds.
\end{aligned}$$

In the first term we have

$$\begin{aligned}
(4.3.15) \quad & \frac{d}{ds} \psi_s(q(s)) - \psi_{s*} \widehat{f}(s, q(s), \widehat{\mu}(s)) + \psi_{s*} \widehat{f}(s, q(s), \nu^0(s)) \\
& = \psi_{s*} \dot{q}(s) - \psi_{s*} \widehat{f}(s, q(s), \widehat{\mu}(s)).
\end{aligned}$$

The final term in (4.3.14) can be bounded above as follows:

$$\begin{aligned}
(4.3.16) \quad & \int_0^t \left\| \psi_{s*} \widehat{f}(s, q(s), \mu(s)) - \psi_{s*} \widehat{f}(s, q(s), \nu^0(s)) - \frac{d}{ds} \psi_s(q(s; \mu)) \right\|_E ds \\
& \leq \int_0^t \left\| \psi_{s*} \widehat{f}(s, q(s), \mu(s)) - \psi_{s*} \widehat{f}(s, q(s; \mu), \mu(s)) \right\|_E ds \\
& \quad + \int_0^t \left\| \psi_{s*} \widehat{f}(s, q(s), \nu^0(s)) - \psi_{s*} \widehat{f}(s, q(s; \mu), \nu^0(s)) \right\|_E ds
\end{aligned}$$

Employing our Lipschitz assumption (4.3.4) we arrive at

$$\begin{aligned}
(4.3.17) \quad & \|\psi_t(q(t)) - \psi_t(q(t; \mu))\|_E \leq \int_0^t \left\| \psi_{s*} \dot{q}(s) - \psi_{s*} \widehat{f}(s, q(s), \widehat{\mu}(s)) \right\|_E \\
& \quad + \int_0^t (k_{f, \mu}(s) + k_{f, \nu^0}(s)) \|\psi_s(q(s)) - \psi_s(q(s; \mu))\|_E ds.
\end{aligned}$$

The proof is completed through Gronwall's lemma.  $\square$

Let  $AC_0$  denote the collection of absolutely continuous curves  $q : [0, T] \rightarrow M$  for which  $q(0) = q_0$ . Introduce a functional  $\Lambda : AC_0 \times \mathcal{M}_0 \rightarrow \mathbb{R}$  through

$$(4.3.18) \quad \begin{aligned} \Lambda(q, \mu) := & \ell(q(T)) + \int_0^T \widehat{L}(t, q(t), \mu(t)) dt \\ & + C_f \int_0^T \left\| \psi_{t*}(q(t)) \left( \widehat{f}(t, q(t), \mu(t)) - \dot{q}(t) \right) \right\|_E dt, \end{aligned}$$

where  $C_f = (k_\ell + M_L) e^{2M_f}$  and  $k_\ell$  is a local Lipschitz rank for the function  $\ell \circ \psi_T^{-1}$ .

PROPOSITION 4.3.4. *Let  $q : [0, T] \rightarrow M$  be an absolutely continuous map for which  $q(t) \in P_{0,t}(\mathcal{O})$  for all  $t$  and suppose  $\mu \in \mathcal{B}(\varepsilon_0)$ . Then we have  $\Lambda(q, \mu) \geq \Lambda(q^0, \nu^0)$ .*

PROOF. Let  $k_\ell$  be a Lipschitz constant for  $\ell$  with respect to the local coordinate chart  $(P_{0,T}(\mathcal{O}), \psi_T)$ . We have

$$(4.3.19) \quad \begin{aligned} \Lambda(q, \mu) &= \ell(q(T)) + \int_0^T L(t, q(t), \mu(t)) dt \\ &+ C_f \int_0^T \left\| \psi_{t*}(q(t)) \left( \widehat{f}(t, q(t), \mu(t)) - \dot{q}(t) \right) \right\|_E dt \\ &= \ell(q(T; \mu)) - k_\ell \|\psi_T(q(T; \mu)) - \psi_T(q(T))\|_E + \int_0^T \widehat{L}(t, q(t; \mu), \mu(t)) dt \\ &\quad - \|k_{L, \mu}\|_{L^1} \max_{t \in [0, T]} \|\psi_t(q(t; \mu)) - \psi_t(q(t))\|_E dt \\ &\quad + C_f \int_0^T \left\| \psi_{t*}(q(t)) \left( \widehat{f}(t, q(t), \mu(t)) - \dot{q}(t) \right) \right\|_E dt \\ &\geq \Lambda(q(\cdot; \mu), \mu) \end{aligned}$$

where the final inequality follows from Proposition 4.3.3. Since  $\Lambda(q(\cdot; \mu), \mu) = \widehat{\Lambda}_0(\mu)$  we find

$$(4.3.20) \quad \Lambda(q, \mu) \geq \Lambda(q(\cdot; \mu), \mu) = \widehat{\Lambda}_0(\mu) \geq \widehat{\Lambda}_0(\nu^0) = \Lambda(q^0, \nu^0).$$

□

An interesting form of the exact penalization result present in Proposition 4.3.4 can be obtained as follows. For each  $t$ , there exists  $p(t) \in E^*$  with  $\|p(t)\|_{E^*} = 1$  for which

$$(4.3.21) \quad \begin{aligned} & C_f \left\| \psi_{t*}(q(t)) \left( \widehat{f}(t, q(t), \mu(t)) - \dot{q}(t) \right) \right\|_E \\ &= C_f \left\langle p(t), \psi_{t*}(q(t)) \left( \widehat{f}(t, q(t), \mu(t)) - \dot{q}(t) \right) \right\rangle. \end{aligned}$$

As a consequence, we may decouple the dynamics by considering a functional of the form

$$(4.3.22) \quad \begin{aligned} \Lambda(q, \mu) &:= \ell(q(T)) + \int_0^T \widehat{L}(t, q(t), \mu(t)) dt \\ &+ C_f \int_0^T \max_{\|p\|_{E^*} \leq 1} \left\langle p, \psi_{t*}(q(t)) \left( \widehat{f}(t, q(t), \mu(t)) - \dot{q}(t) \right) \right\rangle dt. \end{aligned}$$

Further, if we introduce a weakly\*-compact set  $\mathcal{A}_t \subset T_{q(t)}^*M$  through

$$(4.3.23) \quad \mathcal{A}_t := \{\psi_t^* p : \|p\|_E \leq 1\}$$

then we see that it suffices to consider the penalized functional

$$(4.3.24) \quad \begin{aligned} \Lambda(q, \mu) &:= \ell(q(T)) + \int_0^T \widehat{L}(t, q(t), \mu(t)) dt \\ &+ C_f \int_0^T \max_{\zeta \in \mathcal{A}_t} \left\langle \zeta, \widehat{f}(t, q(t), \mu(t)) - \dot{q}(t) \right\rangle dt. \end{aligned}$$

**4.3.1. Maximum Principle.** We now turn to the derivation of the Maximum Principle for problems on Banach manifolds. We first introduce variations of  $q^0$  and  $\nu^0$ . Let  $\nu \in \mathcal{M}_0$  be arbitrary and choose any  $L^1$ -bounded map  $w : [0, T] \rightarrow TM$  along  $q^0$ . Extend  $w$  to a  $C^1$ -smooth,  $L^1$ -bounded vector field  $W_t$  and define  $q^\lambda \in AC_0$  through

$$(4.3.25) \quad \dot{q}^\lambda(t) = f(t, q^\lambda(t), u^0(t)) + \lambda W_t(q^\lambda(t)),$$

$q^\lambda(0) = q_0$ . Define  $\nu^\lambda \in \mathcal{M}_0$  through

$$(4.3.26) \quad \nu^\lambda = (1 - \lambda)\nu^0 + \lambda\nu.$$

**THEOREM 4.3.5.** *Suppose that  $u^0$  is an optimal control with trajectory  $q^0$ . There exists an absolutely continuous map  $\zeta : [0, T] \rightarrow T^*M$  satisfying  $\zeta(T) = -d\ell(q^0(T))$  which satisfies, for almost all  $t$ ,*

$$(4.3.27) \quad \dot{\zeta}(t) = \overrightarrow{H}(t, \zeta(t), u^0(t)),$$

with  $H(t, \zeta, u) = \langle \zeta, f(t, q, u) \rangle - L(t, q, u)$ . In addition there holds

$$(4.3.28) \quad \int_0^T H(t, \zeta(t), u^0(t)) dt = \max_{u \in \mathcal{U}} \int_0^T H(t, \zeta(t), u(t)) dt.$$

When  $\mathbb{U}$  is separable, then for almost all  $t$ ,

$$(4.3.29) \quad H(t, \zeta(t), u^0(t)) = \max_{u \in \mathbb{U}} H(t, \zeta(t), u).$$

**PROOF.** By Proposition 4.3.4 we will have

$$(4.3.30) \quad \limsup_{\lambda \downarrow 0} \frac{\Lambda(q^\lambda, \nu^\lambda) - \Lambda(q^0, \nu^0)}{\lambda} \geq 0.$$

We bound this derivative from above. Let  $\lambda_n \downarrow 0$  be a sequence which attains this limit supremum.

For each  $n$  we have

$$(4.3.31) \quad \widehat{f}(t, q^{\lambda_n}(t), \nu^{\lambda_n}(t)) - \dot{q}^{\lambda_n}(t) = \lambda_n \widehat{f}(t, q^{\lambda_n}(t), \nu(t)) - \lambda_n \widehat{f}(t, q^{\lambda_n}(t), \nu^0(t)) - \lambda_n W_t(q^{\lambda_n}(t)).$$

Choose  $p_n \in L^\infty([0, T], E^*)$  with  $\|p_n\|_{L^\infty} \leq C_f$  for which

$$(4.3.32) \quad \begin{aligned} & \int_0^T \left\langle p_n(t), \psi_{t*} \left( \widehat{f}(t, q^{\lambda_n}(t), \nu(t)) - \widehat{f}(t, q^{\lambda_n}(t), \nu^0(t)) - W_t(q^{\lambda_n}(t)) \right) \right\rangle dt \\ & = C_f \int_0^T \left\| \psi_{t*} \left( \widehat{f}(t, q^{\lambda_n}(t), \nu(t)) - \widehat{f}(t, q^{\lambda_n}(t), \nu^0(t)) - W_t(q^{\lambda_n}(t)) \right) \right\|_E dt. \end{aligned}$$

We find

$$(4.3.33) \quad \begin{aligned} & \frac{\Lambda(q^{\lambda_n}, \nu^{\lambda_n}) - \Lambda(q^0, \nu^0)}{\lambda_n} = \frac{1}{\lambda_n} \left( \ell(q^{\lambda_n}(T)) - \ell(q^0(T)) \right) \\ & + \frac{1}{\lambda_n} \int_0^T \widehat{L}(t, q^{\lambda_n}(t), \nu^0(t)) - \widehat{L}(t, q^0(t), \nu^0(t)) dt \\ & + \int_0^T \widehat{L}(t, q^{\lambda_n}(t), \nu(t)) - \widehat{L}(t, q^0(t), \nu^0(t)) dt \\ & + \int_0^T \left\langle p_n(t), \psi_{t*} \left( \widehat{f}(t, q^{\lambda_n}(t), \nu(t)) - \widehat{f}(t, q^{\lambda_n}(t), \nu^0(t)) - W_t(q^{\lambda_n}(t)) \right) \right\rangle dt. \end{aligned}$$

Pass to a subsequence such that  $p_n \xrightarrow{w^*} p$  for some  $p \in L^\infty([0, T], E^*)$  with  $\|p\|_{L^\infty} \leq C_f$  to find that

$$(4.3.34) \quad \begin{aligned} & \left\langle d\ell(q^0(T)), \frac{\partial q^\lambda(T)}{\partial \lambda} \Big|_{\lambda=0} \right\rangle + \int_0^T \left\langle d\widehat{L}(t, q^0(t), \nu^0(t)), \frac{\partial q^\lambda(t)}{\partial \lambda} \Big|_{\lambda=0} \right\rangle dt \\ & + \int_0^T \widehat{L}(t, q^0(t), \nu(t)) - \widehat{L}(t, q^0(t), \nu^0(t)) dt \\ & + \int_0^T \left\langle p(t), \psi_{t*} \left( \widehat{f}(t, q^0(t), \nu(t)) - \widehat{f}(t, q^0(t), \nu^0(t)) - W_t(q^0(t)) \right) \right\rangle dt \geq 0. \end{aligned}$$

Recall that by Proposition 3.1.4 we have

$$(4.3.35) \quad \frac{\partial q^\lambda(t)}{\partial \lambda} \Big|_{\lambda=0} = \int_0^t P_{s,t*}(q^0(s)) W_s(q^0(s)) ds.$$

Since  $W_t$  extends  $w$  we have

$$\begin{aligned}
(4.3.36) \quad & \int_0^T \langle d\ell(q^0(T)), P_{t,T*}(q^0(t))w(t) \rangle dt \\
& + \int_0^T \left\langle d\widehat{L}(t, q^0(t), \nu^0(t)), \int_0^t P_{s,t*}(q^0(s))w(s) ds \right\rangle dt \\
& + \int_0^T \widehat{L}(t, q^0(t), \nu(t)) - \widehat{L}(t, q^0(t), \nu^0(t)) dt \\
& + \int_0^T \left\langle p(t), \psi_{t*} \left( \widehat{f}(t, q^0(t), \nu(t)) - \widehat{f}(t, q^0(t), \nu^0(t)) - w(t) \right) \right\rangle dt \geq 0.
\end{aligned}$$

Using our integration by parts formula (4.2.1) we come to

$$\begin{aligned}
(4.3.37) \quad & \int_0^T \left\langle P_{t,T}^* d\ell(q^0(T)) + \int_t^T P_{t,s}^* d\widehat{L}(s, q^0(s), u^0(s)) ds, w(t) \right\rangle dt \\
& + \int_0^T \widehat{L}(t, q^0(t), \nu(t)) - \widehat{L}(t, q^0(t), u^0(t)) dt \\
& + \int_0^T \left\langle p(t), \psi_{t*} \left( \widehat{f}(t, q^0(t), \nu(t)) - \widehat{f}(t, q^0(t), u^0(t)) - w(t) \right) \right\rangle dt \geq 0.
\end{aligned}$$

Denoting the left-hand side by  $\Phi(p, \nu, w)$ , we see that

$$(4.3.38) \quad \inf_{\nu, w} \max_p \Phi(p, \nu, w) \geq 0,$$

where the infimum is taken over relaxed controls  $\nu$  and  $L^1$ -bounded mappings  $w : [0, T] \rightarrow TM$  along  $q^0$  and the maximum is over  $p \in L^\infty([0, T], E^*)$  for which  $\|p\|_{L^\infty} \leq C_f$ . By Fan's minimax theorem we can find  $p \in L^\infty([0, T], E^*)$  for which  $\|p\|_{L^\infty} \leq C_f$  and such that for any  $\nu \in \mathcal{M}_0$ , for any  $L^1$ -bounded mapping  $w :$

$[0, T] \rightarrow TM$  along  $q^0$  we have,

$$(4.3.39) \quad \int_0^T \left\langle P_{t,T}^* d\ell(q^0(T)) + \int_t^T P_{t,s}^* d\widehat{L}(s, q^0(s), \nu^0(s)) ds, w(t) \right\rangle dt \\ + \int_0^T \widehat{L}(t, q^0(t), \nu(t)) - \widehat{L}(t, q^0(t), \nu^0(t)) dt \\ + \int_0^T \left\langle p(t), \psi_{t*} \left( \widehat{f}(t, q^0(t), \nu(t)) - \widehat{f}(t, q^0(t), \nu^0(t)) - w(t) \right) \right\rangle dt \geq 0.$$

Defining a curve  $\zeta : [0, T] \rightarrow T^*M$  through  $\zeta(t) = -\psi_t^* p(t)$  we find that for any  $w$  and any control  $u$ ,

$$(4.3.40) \quad \int_0^T \left\langle P_{t,T}^* d\ell(q^0(T)) + \int_t^T P_{t,s}^* dL(s, q^0(s), u^0(s)) ds + \zeta(t), w(t) \right\rangle dt \\ + \int_0^T H(t, \zeta(t), u^0(t)) - H(t, \zeta(t), u(t)) dt \geq 0.$$

Taking  $u = u^0$  we find that for any  $w$  there holds

$$(4.3.41) \quad \int_0^T \left\langle P_{t,T}^* d\ell(q^0(T)) + \int_t^T P_{t,s}^* dL(s, q^0(s), u^0(s)) ds + \zeta(t), w(t) \right\rangle dt \geq 0$$

and the fundamental lemma now implies that

$$(4.3.42) \quad \zeta(t) = -P_{t,T}^* d\ell(q^0(T)) - \int_t^T P_{t,s}^* dL(s, q^0(s), u^0(s)) ds.$$

This implies that  $\zeta$  is absolutely continuous, satisfies  $\zeta(T) = -d\ell(q^0(T))$  and, by Proposition 3.3.5,

$$(4.3.43) \quad \dot{\zeta}(t) = \overrightarrow{H}(t, \zeta(t), u^0(t))$$

with  $H(t, \zeta, u) = \langle \zeta, f(t, q, u) \rangle - L(t, q, u)$ .

Now taking  $w = 0$  we find that for any  $u \in \mathcal{U}$  there holds

$$(4.3.44) \quad \int_0^T H(t, \zeta(t), u^0(t)) - H(t, \zeta(t), u(t)) dt \geq 0,$$

and this is (4.3.28).

In the case where  $\mathbb{U}$  is separable, Lemma 4.1.13 implies (4.3.29) and this completes the proof.  $\square$

#### 4.4. Calculus of Variations

In this section we consider two applications of our work to problems of the Calculus of Variations on Banach manifolds. The first is an application of the Maximum Principle for problems on sufficiently smooth spaces. The second is an application of the duBois-Reymond lemma. For coordinate-free descriptions of necessary conditions it will help to recall that for a fixed  $t, q$  the map  $L$  induces a mapping  $\mathbb{F}L : T_q M \rightarrow T_q^* M$  through

$$(4.4.1) \quad \langle \mathbb{F}L(t, q, v), w \rangle = \left. \frac{d}{ds} \right|_{s=0} L(t, q, v + sw).$$

This mapping is called the *fiber derivative* of  $L$  and plays an important role in nonholonomic mechanics and control [12]. In local coordinates,  $\mathbb{F}L$  is merely the derivative  $L_v(t, x, v)$ .

**4.4.1. Application: Maximum Principle.** As before we demonstrate the Maximum Principle contains in it all first-order necessary conditions for Calculus of Variations. We will suppose in this subsection that  $M$  is a  $C^2$ -smooth Banach manifold modeled over a space  $E$  which admits  $C^2$ -smooth bump functions whose second derivative is locally Lipschitz. This can be a strong assumption, but certainly holds for Hilbert manifolds. See Proposition 3.2.10 for some other Banach spaces in which this assumption will hold. We consider the problem of minimizing

$$(4.4.2) \quad \ell(q(T)) + \int_0^T L(t, q(t), \dot{q}(t)) dt$$



over absolutely continuous maps  $q : [0, T] \rightarrow M$  with fixed initial condition  $q_0$  and free terminal point  $q(T)$ . We suppose that  $\ell$  is  $C^1$ -smooth and that  $L$  satisfies Assumption 4.3.1. Suppose that  $q^0$  provides a strong local minimum and choose an extension  $V_t$  of  $\dot{q}^0(t)$ . Let  $P_{s,t}$  be the flow of  $V_t$ . Let  $(\varphi, \mathcal{O})$  be a coordinate chart with  $q_0 \in \mathcal{O}$ . Define, for  $u \in E$ , a vector field

$$(4.4.3) \quad X(q, u) = b(q)\varphi_*^{-1}(\varphi(q))u$$

for  $q \in \mathcal{O}$ , extended smoothly to zero.

Consider the following control system:

$$(4.4.4) \quad f(t, q, u) = V_t(q) + (P_{0,t} * X)(q, u)$$

and the running cost

$$(4.4.5) \quad \tilde{L}(t, q, u) = L(t, q, f(t, q, u)).$$

Under our assumptions,  $f$  and  $\tilde{L}$  satisfy the standing hypotheses.

Further, the control  $u^0 \equiv 0$  is an optimal control for the control system  $\dot{q}(t; u) = f(t, q(t; u), u(t))$  with cost

$$(4.4.6) \quad \ell(q(T)) + \int_0^T \tilde{L}(t, q(t), u(t)) dt.$$

By the Maximum Principle there exists an absolutely continuous map  $\zeta : [0, T] \rightarrow T^*M$  for which

$$(4.4.7) \quad \dot{\zeta}(t) = \overrightarrow{H}(t, \zeta(t), 0)$$

and such that

$$(4.4.8) \quad H(t, \zeta(t), 0) = \max_{u \in \mathbb{U}} H(t, \zeta(t), u).$$

There holds for almost all  $t$ , for any  $u \in E$ ,

$$(4.4.9) \quad \begin{aligned} \langle \zeta(t), \dot{q}^0(t) \rangle - L(t, q^0(t), \dot{q}^0(t)) &\geq \langle \zeta(t), \dot{q}^0(t) + P_{0,t*}(q_0)\varphi_*^{-1}(\varphi(q_0))u \rangle \\ &- L(t, q^0(t), \dot{q}^0(t) + P_{0,t*}(q_0)\varphi_*^{-1}(\varphi(q_0))u). \end{aligned}$$

Since  $P_{0,t*} : T_{q_0}M \rightarrow T_{q^0(t)}M$  is an isomorphism we obtain, for any  $v \in T_{q(t)}M$ ,

$$(4.4.10) \quad L(t, q^0(t), v) \geq L(t, q^0(t), \dot{q}^0(t)) + \langle \zeta(t), v - \dot{q}^0(t) \rangle.$$

This is the classical Weierstrass condition. Moreover, this shows that  $\zeta(t)$  is the fiber derivative  $\mathbb{F}L(t, q^0(t), \dot{q}^0(t))$ . It follows that  $\mathbb{F}L(t, q^0(t), \dot{q}^0(t))$  is absolutely continuous – a coordinate-free version of the Erdmann corder condition.

To arrive at the Euler-Lagrange equations, fix any time  $t_0 \in [0, T]$ . Let  $(x^0(t), p(t))$  denote local coordinates for the curve  $\zeta(t)$ . One may check that, in the  $(x, p)$  coordinates there holds:

$$(4.4.11) \quad \dot{p}(t) = -H_x(t, \zeta(t), 0) = -V_t'(x_t^0)^*p(t) + L_x(t, x^0(t), \dot{x}^0(t)) + V_t'(x_t^0)^*L_v(t, x^0, \dot{x}^0(t)).$$

But since  $p(t) = L_v(t, q^0(t), \dot{q}^0(t))$  we see that (4.4.11) simplifies to

$$(4.4.12) \quad \dot{p}(t) = L_x(t, x^0(t), \dot{x}^0(t)),$$

the Euler-Lagrange equation.

**4.4.2. Application: duBois-Reymond Lemma.** In this subsection we study an application of the duBois-Reymond lemma to problems in the geometric Calculus of Variations in which the terminal point  $q(T)$  is fixed. Consider the cost

$$(4.4.13) \quad \int_0^T L(q(t), \dot{q}(t)) dt,$$

where  $L : TM \rightarrow \mathbb{R}$  is  $C^1$ -smooth. Suppose that  $q^0 : [0, T] \rightarrow M$  is a strong local minimizer for the cost (4.4.13) subject to fixed boundary conditions  $q(0) = q_0$  and  $q(T) = q_1$ .

We assume that  $M$  and  $E$  are sufficiently smooth that a  $C^2$ -smooth extension  $V_t$  of  $\dot{q}^0(t)$  can be found.

Suppose that  $w : [0, T] \rightarrow TM$  is an end-point preserving map along  $q^0$ . By Lemma 4.2.6 we can find an extension  $W_t$  of  $w$  such that for small values of  $\lambda > 0$  the flow  $P_{s,t}^\lambda$  of  $V_t + \lambda W_t$  satisfies  $P_{0,T}^\lambda(q_0) = q^0(T) = q_1$ . Since  $q^0$  is optimal we find that, for small  $\lambda > 0$ ,

$$(4.4.14) \quad \int_0^T L(P_{0,t}^\lambda(q_0), V_t(P_{0,t}^\lambda(q_0)) + \lambda W_t(P_{0,t}^\lambda(q_0))) - L(q^0(t), \dot{q}^0(t)) dt \geq 0.$$

Dividing by  $\lambda > 0$  and taking the limit we find that

$$(4.4.15) \quad \int_0^T \frac{d}{d\lambda} \Big|_{\lambda=0} L(P_{0,t}^\lambda(q_0), V_t(P_{0,t}^\lambda(q_0))) + \langle \mathbb{F}L(t, q^0(t), \dot{q}^0(t)), w(t) \rangle dt \geq 0.$$

We also have

$$(4.4.16) \quad \begin{aligned} \frac{d}{d\lambda} \Big|_{\lambda=0} L(P_{0,t}^\lambda(q_0), V_t(P_{0,t}^\lambda(q_0))) &= \frac{d}{d\lambda} \Big|_{\lambda=0} (L \circ V_t)(P_{0,t}^\lambda(q_0)) \\ &= \left\langle V_t^* dL, \int_0^t P_{s,t*}(q^0(s)) w(s) \right\rangle. \end{aligned}$$

Integrating by parts we find

$$(4.4.17) \quad \int_0^T \left\langle \int_t^T P_{t,s}^* V_s^* dL(q^0(s), \dot{q}^0(s)) ds + \mathbb{F}L(t, q^0(t), \dot{q}^0(t)), w(t) \right\rangle dt \geq 0.$$

By the duBois-Reymond lemma, there exists  $\zeta_T \in T_{q^0(T)}^* M$  such that for almost all  $t$ ,

$$(4.4.18) \quad \mathbb{F}L(t, q^0(t), \dot{q}^0(t)) = P_{t,T}^* \zeta_T - \int_t^T P_{t,s}^* V_s^* dL(q^0(s), \dot{q}^0(s)) ds$$

This proves that  $\zeta(t) := \mathbb{F}L(t, q^0(t), \dot{q}^0(t))$  is absolutely continuous. Moreover, by Proposition 3.3.5 for almost all  $t$  we have

$$(4.4.19) \quad \dot{\zeta}(t) = \vec{H}(t, \zeta(t)),$$

where  $H(t, \zeta) = \langle \zeta, V_t(q) \rangle - L(t, q, V_t(q))$  for  $\zeta \in T_q^*M$ .

Fix a time  $t_0 \in [0, T]$  and choose local coordinates  $x$  for  $M$ , defined on a neighborhood of  $q^0(t_0)$ . Let  $(x, p)$  and  $(x, v)$  denote the induced coordinates on  $T^*M$  and  $TM$ , respectively. In these coordinates,  $\zeta(t) = (x(t), p(t))$  and we have, for almost all  $t$ ,

$$(4.4.20) \quad \begin{aligned} \dot{p}(t) &= -H_x(t, x(t), p(t)) \\ &= -V_t'(x(t))^* p(t) + L_x(t, x(t), \dot{x}(t)) + V_t'(x(t))^* L_v(t, x(t), \dot{x}(t)). \end{aligned}$$

But  $L_v(t, x(t), \dot{x}(t)) = p(t)$  and so we arrive at the Euler-Lagrange equations

$$(4.4.21) \quad \dot{p}(t) = L_x(t, x(t), \dot{x}(t)).$$

## CHAPTER 5

### Optimality Conditions via Nonsmooth Analysis

In this chapter we continue our study of geometric optimal control, turning our attention to problems subject to a constraint of the form  $q(T) \in S$ . We recall that while statements of the Maximum Principle can be found in [4, 7, 12, 17, 54, 79, 84], of these, only [4, 7, 17] offer a full proof of the Maximum Principle for problems on general manifolds. The cases covered by these particular sources cover only the possibilities that the set  $S$  constraining the terminal point  $q(T)$  is an immersed submanifold or a singleton and that the terminal cost  $\ell$  is at least  $C^1$ -smooth. In this chapter we use methods of nonsmooth analysis to prove a general statement of the Maximum Principle which allows for terminal costs  $\ell$  which are merely locally Lipschitz and terminal constraints of the form  $q(T) \in S$ , where  $S \subset M$  is merely assumed closed.

We will focus on Mayer problems in which a cost  $\ell(q(T); u)$  is to be minimized. There is very little loss of generality in this choice, since under mild assumptions the more general problems of Bolza considered in the previous chapter may be reduced to such Mayer problems. We will restrict our attention in this chapter to the case in which  $M$  is of finite dimension. The main goal of this chapter is to establish the following version of the Pontryagin Maximum Principle:

**THEOREM 5.0.1.** *Suppose that  $u^0$  is an optimal control with trajectory  $q^0$ . There exist  $\lambda^0 \in \{0, 1\}$  and  $-\zeta_T \in \lambda^0 \partial_L \ell(q^0(T)) + N_S^L(q^0(T))$  such that if  $\zeta : [0, T] \rightarrow T^*M$*

is the solution to

$$(5.0.22) \quad \dot{\zeta}(t) = \overrightarrow{H}(t, \zeta(t), u^0(t)) \quad \zeta(T) = \zeta_T$$

then for almost all  $t$  there holds

$$(5.0.23) \quad H(t, \zeta(t), u^0(t)) = \max_{u \in \mathbb{U}} H(t, \zeta(t), u).$$

Further, either  $\lambda^0 = 1$  or  $\zeta(t) \neq 0$  for all  $t$ .

A second goal of this chapter is to establish sufficient conditions for exact penalization of the terminal constraint  $q(T; u) \in S$ . In particular, we prove the following:

**THEOREM 5.0.2.** *Let  $u^0$  be an optimal control with trajectory  $q^0$  and let  $(\mathcal{O}, \theta)$  be a coordinate chart with  $q^0(T) \in \mathcal{O}$ . If for all nonzero  $-\zeta_T \in N_S^L(q^0(T))$  the solution  $\zeta(t)$  to (5.0.22) fails to satisfy (5.0.23) then there is a constant  $K$  such that  $u^0$  is a local minimizer for the unconstrained Mayer problem with cost  $\ell + Kd_{\theta(S)} \circ \theta$ .*

Finally, we hope to demonstrate in this chapter that nonsmooth analysis can be useful for understanding general problems of optimal control, even when all of the data for a problem are smooth.

Our methods are based on nonsmooth analysis techniques which were developed originally by Clarke in his studies of dynamic optimization problems for differential inclusions [23, 30]. The methods introduced by Clarke provided a foundation for further developments in the application of nonsmooth analysis to problems of dynamic optimization. For monograph expositions, see [24, 25, 73, 89].

Let us state the problem we wish to study. Consider a smooth manifold  $M$  of finite dimension, a compact metric space  $\mathbb{U}$ , and a control system  $f : [0, T] \times M \times \mathbb{U} \rightarrow TM$ . Let  $\mathcal{U}$  denote the set of all measurable mappings  $u : [0, T] \rightarrow \mathbb{U}$ . As before we

write  $q(t; u)$  for the absolutely continuous solution to

$$(5.0.24) \quad \dot{q}(t; u) = f(t, q(t; u), u(t))$$

with  $q(0; u) = q_0$ .

Given a locally Lipschitz function  $\ell : M \rightarrow \mathbb{R}$  and closed set  $S \subset M$ , we consider the following Mayer problem: *minimize*  $\ell(q(T; u))$  *over solutions*  $q(t; u)$  *to* (5.0.24) *subject to terminal constraint*  $q(T; u) \in S$  *and dynamic constraint* (5.0.24).

Problems with the more general cost  $\ell(q(T)) + \int_0^T c(t, q(t), u(t)) dt$  are easily converted to this problem, provided that  $c$  satisfies the same assumptions as  $f$ , which we now make explicit. We remind the reader that the local coordinate representation of the control system  $f$ , in coordinates  $\varphi$ , is given by

$$(5.0.25) \quad (\varphi_* f)(t, x, u) = \varphi_*(\varphi^{-1}(x))f(t, \varphi^{-1}(x), u).$$

ASSUMPTION 5.0.3. We assume that for each  $q \in M$  there exists a coordinate chart  $(\mathcal{O}, \varphi)$  in which control system (5.0.25) is measurable in  $t$ , differentiable with respect to  $x$ , and continuous in  $u$ . Further, we suppose there exist functions  $m_\varphi, k_\varphi \in L^1([0, T], \mathbb{R})$  such that for almost all  $t$ , for any  $x, y \in \varphi(\mathcal{O})$  and  $u \in \mathbb{U}$ , there hold the inequalities

$$(5.0.26) \quad \|(\varphi_* f)(t, x, u)\|_{\mathbb{R}^n} \leq m_\varphi(t) \quad \|(\varphi_* f)_x(t, x, u)\|_{\mathbb{R}^n} \leq m_\varphi(t)$$

along with

$$(5.0.27) \quad \begin{aligned} \|(\varphi_* f)(t, x, u) - (\varphi_* f)(t, y, u)\|_{\mathbb{R}^n} &\leq k_\varphi(t) \|x - y\|_{\mathbb{R}^n} \\ \|(\varphi_* f)_x(t, x, u) - (\varphi_* f)_x(t, y, u)\|_{\mathbb{R}^n} &\leq k_\varphi(t) \|x - y\|_{\mathbb{R}^n}. \end{aligned}$$

Finally, we assume that for any control  $u$  there exists a neighborhood  $\mathcal{O}$  of  $q_0$  such that a solution to (5.0.24) with initial condition  $q \in \mathcal{O}$  exists for all  $t \in [0, T]$ .

We note the following consequence of Assumption 5.0.3 for control systems in  $\mathbb{R}^n$ :

PROPOSITION 5.0.4. *Suppose that  $M = \mathbb{R}^n$ . For any control  $v$  there exists  $\varepsilon_0 > 0$  and  $L^1$  functions  $m_f, k_f$  such that for any continuous maps  $y, z : [0, T] \rightarrow \mathbb{R}^n$  satisfying*

$$(5.0.28) \quad \max_{t \in [0, T]} \|x(t; v) - y(t)\|_{\mathbb{R}^n} < \varepsilon_0 \quad \max_{t \in [0, T]} \|x(t; v) - z(t)\|_{\mathbb{R}^n} < \varepsilon_0$$

*there holds for almost all  $t \in [0, T]$  and for any  $u \in \mathbb{U}$*

$$(5.0.29) \quad \|f(t, y(t), u)\|_{\mathbb{R}^n} \leq m_f(t) \quad \|f_x(t, y(t), u)\|_{\mathbb{R}^n} \leq m_f(t)$$

*along with*

$$(5.0.30) \quad \begin{aligned} \|f(t, y(t), u) - f(t, z(t), u)\|_{\mathbb{R}^n} &\leq k_f(t) \|y(t) - z(t)\|_{\mathbb{R}^n} \\ \|f_x(t, y(t), u) - f_x(t, z(t), u)\|_{\mathbb{R}^n} &\leq k_f(t) \|y(t) - z(t)\|_{\mathbb{R}^n} . \end{aligned}$$

PROOF. The result follows from the compactness of the image of  $x(\cdot; v)$  along with the standing assumption on  $f$ .  $\square$

## 5.1. Relaxed Controls and Sliding Variations

As in the previous chapter, we will find it useful to employ *relaxed controls* [46, 91, 92, 93]. We have assumed for this chapter that  $\mathbb{U}$  is a compact metric space and for finite-dimensional systems with controls taking values in such spaces there is a well-developed theory of relaxed controls as weakly\*-measurable maps into  $\mathcal{P}(\mathbb{U})$ . This theory is developed in [92]. We recall here the basic definitions and results.

### 5.1.1. Definitions and Background Results.



DEFINITION 5.1.1. Let  $\mathcal{P}(\mathbb{U})$  denote the set of Borel probability measures on  $\mathbb{U}$ . A *relaxed control* is a mapping  $\mu : [0, T] \rightarrow \mathcal{P}(\mathbb{U})$  with the property that for each continuous function  $g : \mathbb{U} \rightarrow \mathbb{R}$  the function  $t \mapsto \int_{\mathbb{U}} g d\mu(t)$  is measurable.

Such a mapping is said to be “weakly\*-measurable” [34, 42, 92].

We write  $\mathcal{M}$  for the set of relaxed controls. As before, control system  $f$  induces a function  $\widehat{f} : [0, T] \times M \times \mathcal{P}(\mathbb{U}) \rightarrow TM$  by

$$(5.1.1) \quad \widehat{f}(t, q, \mu) = \int_{\mathbb{U}} f(t, q, u) d\mu(u)$$

and each traditional control  $u \in \mathcal{U}$  can be realized through  $t \mapsto \delta_{u(t)}$ , where  $\delta_u$  is the Dirac mass concentrated at  $u$ .

Relaxed controls are convenient in part because the set  $\mathcal{M}$  of such controls is weakly\*-compact [92]:

PROPOSITION 5.1.2. *If  $\nu_n$  is any sequence of relaxed controls then there exists a subsequence  $\nu_{n_k}$  and a relaxed control  $\nu$  such that for any Carathéodory function  $g : [0, T] \times \mathbb{U} \rightarrow \mathbb{R}^n$  we have*

$$(5.1.2) \quad \lim_{k \rightarrow \infty} \int_0^T \int_{\mathbb{U}} g(t, u) d\nu_{n_k}(t) dt = \int_0^T \int_{\mathbb{U}} g(t, u) d\nu(t) dt.$$

**5.1.2. Sliding Variations and Infinitesimal Perturbations.** The variations employed in the proceeding chapter, of the form

$$(5.1.3) \quad (1 - \lambda)\mu + \lambda\nu$$

are known as *sliding variations*. The system velocity associated with such a variation is

$$(5.1.4) \quad \widehat{f}(t, q, \mu(t)) + \lambda \left( \widehat{f}(t, q, \nu(t)) - \widehat{f}(t, q, \mu(t)) \right)$$

and if  $q^\lambda(t)$  is the trajectory for (5.1.4), then by Proposition 3.1.4 we have

$$(5.1.5) \quad \left. \frac{\partial q^\lambda(T)}{\partial \lambda} \right|_{\lambda=0} = \int_0^T P_{t,T^*}(q(t)) \left( \widehat{f}(t, q(t), \nu(t)) - \widehat{f}(t, q(t), \mu(t)) \right) dt,$$

where  $P_{s,t}$  is the flow of the vector field  $(t, q) \mapsto \widehat{f}(t, q, \mu(t))$  and  $q(t) = P_{0,t}(q_0)$ .

The infinitesimal perturbation given by (5.1.5) plays a central role in our study of control on manifolds and we denote by  $E(\mu)$  the set of all such tangent vectors  $v \in T_{q(T;\mu)}M$ . Since the set of relaxed controls is weakly\*-compact,  $E(\mu)$  is a compact set. The set is also convex, as can be seen from (5.1.5). These properties allow us to establish the following proposition, which will allow us to replace a nonlinear Dini lower derivative with a linear pairing involving the limiting subgradient.

**PROPOSITION 5.1.3.** *Let  $\ell : M \rightarrow \mathbb{R}$  be locally Lipschitz. For any control  $u(t)$  there exists  $\zeta \in \partial_L \ell(q(T; u))$  such that*

$$(5.1.6) \quad \min_{v \in E(u(t))} \langle \zeta, v \rangle \geq \min_{v \in E(u(t))} D\ell(q(T; \mu); v).$$

**PROOF.** Since  $\partial_L \ell$  and  $D\ell$  are coordinate-free, we need only prove the proposition for the case  $M = \mathbb{R}^n$ .

We recall for the following *multidirection mean-value inequality*, originally formulated by Clarke and Ledyaev [26]:

**THEOREM 5.1.4.** *Fix  $x \in \mathbb{R}^n$  and let  $Y \subset \mathbb{R}^n$  be a compact, convex set. Let  $\ell : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be lower semicontinuous. Then for any  $r < \min_{y \in Y} \ell(y) - \ell(x)$  and any  $\varepsilon > 0$  there exists  $z \in [x, Y] + \varepsilon \mathbb{B}$  and  $\zeta \in \partial_F \ell(z)$  such that for any  $y \in Y$  there holds*

$$(5.1.7) \quad r < \langle \zeta, y - x \rangle.$$

*We may further choose  $z$  so that  $\ell(z) < \ell(x) + \max\{0, r\} + \varepsilon$ .*

Choose a sequence  $\lambda_n$  of positive real numbers which converges to zero and for each  $n$ , let

$$(5.1.8) \quad r_n = \min_{y \in q(T; u) + \lambda_n E} \ell(y) - \ell(q(T; u)) - \lambda_n^2.$$

By the mean value inequality we can choose  $z_n \in q(T; u) + \lambda_n E + \lambda_n \mathbb{B}$  and  $\zeta_n \in \partial_F \ell(z_n)$  for which

$$(5.1.9) \quad r_n < \min_{y \in q(T; u) + \lambda_n E} \langle \zeta_n, y - q(T; u) \rangle = \min_{v \in E} \lambda_n \langle \zeta_n, v \rangle.$$

Choose  $v_n \in E$  such that  $r_n = \ell(q(T; u) + \lambda_n v_n) - \ell(q(T; u)) - \lambda_n^2$ . Since  $\ell$  is locally Lipschitz the  $\zeta_n$  are bounded in norm and we may pass to a subsequence for which  $\zeta_n \rightarrow \zeta \in \partial_L \ell(q(T; u))$  and  $v_n \rightarrow \bar{v} \in E$ . Now

$$(5.1.10) \quad \begin{aligned} \min_{v \in E} \langle \zeta, v \rangle &= \lim_{n \rightarrow \infty} \min_{v \in E} \langle \zeta_n, v \rangle \geq \liminf_{n \rightarrow \infty} \frac{r_n}{\lambda_n} \\ &= \liminf_{n \rightarrow \infty} \frac{\ell(q(T; u) + \lambda_n v_n) - \ell(q(T; u)) - \lambda_n^2}{\lambda_n} \\ &\geq D\ell(q(T; u); \bar{v}) \geq \min_{v \in E(u(t))} D\ell(q(T; u); v). \end{aligned}$$

□

A useful form of inequality (5.1.6) is given by the following:

PROPOSITION 5.1.5. *Let  $\ell : M \rightarrow \mathbb{R}$  be locally Lipschitz. For any control  $u \in \mathcal{U}$  there exists  $-\zeta_T \in \partial_L \ell(q(T; u))$  such that*

$$(5.1.11) \quad \min_{\nu \in \mathcal{M}} \int_0^T \widehat{H}(t, \zeta(t), \delta_{u(t)}) - \widehat{H}(t, \zeta(t), \nu(t)) dt \geq \min_{v \in E(\delta_u)} D\ell(q(T; u); v)$$

where  $\zeta(t)$  is the solution to

$$(5.1.12) \quad \dot{\zeta}(t) = \overrightarrow{H}(t, \zeta(t), u(t)) \quad \zeta(T) = \zeta_T.$$

PROOF. If  $v \in E(u)$  then there exists a relaxed control  $\nu$  such that

$$(5.1.13) \quad v = \int_0^T P_{t,T*}(q(t)) \left( \widehat{f}(t, q(t), \nu(t)) - \widehat{f}(t, q(t), \mu(t)) \right) dt.$$

By Proposition 5.1.3 there exists  $\zeta \in \partial_L \ell(q(T; u))$  such that

$$(5.1.14) \quad \begin{aligned} \min_{\nu \in \mathcal{M}} \int_0^T \left\langle \zeta, P_{t,T*}(q(t; u)) \left( \widehat{f}(t, q(t; u), \nu(t)) - \widehat{f}(t, q(t; u), u(t)) \right) \right\rangle dt \\ \geq \min_{v \in E} D\ell(q(T; u); v). \end{aligned}$$

Define  $\zeta_T = -\zeta$  and  $\zeta(t) = P_{t,T}^* \zeta_T$ . By Proposition 3.3.5,  $\zeta(t)$  is the solution to (5.1.12) and the definition of  $H$  implies (5.1.11).  $\square$

Proposition 5.1.5 suggests a relatively short proof of the Maximum Principle for nonsmooth Mayer problems in which the terminal point is free.

## 5.2. Maximum Principle for Nonsmooth Mayer Problem with Free Terminal Point

In this section we provide a proof of the Maximum Principle under the assumption that  $q(T; u)$  is free and  $\ell$  is locally Lipschitz. We first claim that for such a problem there is no loss in working with relaxed controls.

PROPOSITION 5.2.1. *Let  $\nu \in \mathcal{M}$  be arbitrary. Let  $(\mathcal{O}, \varphi)$  be a coordinate chart defined on a neighborhood of  $q_0$  and let  $\psi_t : P_{0,t}(\mathcal{O}) \rightarrow \varphi(\mathcal{O})$  be the Lagrangian coordinates associated with the nonautonomous vector field  $(t, q) \mapsto \widehat{f}(t, q, \nu(t))$ . Given any  $\varepsilon > 0$  there exists a control  $u \in \mathcal{U}$  for which*

$$(5.2.1) \quad \max_{t \in [0, T]} \|\psi_t(q(t; \nu)) - \psi_t(q(t; u))\|_{\mathbb{R}^n} < \varepsilon.$$

PROOF. Consider the control system

$$(5.2.2) \quad g(t, x, w) = (\psi_{t*} f)(t, x, w) - (\psi_{t*} \widehat{f})(t, x, \nu(t)).$$

By the assumptions on  $f$ , this control system is locally integrable Lipschitz. If  $x(t; \nu)$  is the trajectory for  $g$  corresponding to  $\nu$  then we clearly have  $x(t; \nu) \equiv \varphi(q_0)$ . By a standard approximation theorem for relaxed controls [46, 92] there exists a control  $u$  such that the trajectory  $x(t; u)$  for  $g$  satisfies

$$(5.2.3) \quad \max_{t \in [0, T]} \|x(t; u) - x(t; \nu)\|_{\mathbb{R}^n} < \varepsilon.$$

Since  $x(t; u) = \psi_t(q(t; u))$  and  $x(t; \nu) = \psi_t(q(t; \nu))$  the proof is complete.  $\square$

PROPOSITION 5.2.2. *If  $q(T)$  is free then we have*

$$(5.2.4) \quad \inf \{\ell(q(T; u)) : u \in \mathcal{U}\} = \inf \{\ell(q(T; \nu)) : \nu \in \mathcal{M}\}.$$

PROOF. We need only prove that

$$(5.2.5) \quad \inf \{\ell(q(T; u)) : u \in \mathcal{U}\} \leq \inf \{\ell(q(T; \nu)) : \nu \in \mathcal{M}\}.$$

Suppose that this is not true, so that for some  $\delta > 0$  there exists  $\nu \in \mathcal{M}$  with

$$(5.2.6) \quad \inf \{\ell(q(T; u)) : u \in \mathcal{U}\} \geq \ell(q(T; \nu)) + \delta.$$

Since  $\ell$  is continuous, Proposition 5.2.1 implies the existence of a control  $u \in \mathcal{U}$  such that  $\ell(q(T; \nu)) + \delta \geq \ell(q(T; u))$  and this contradiction completes the proof.  $\square$

With this we turn to the proof of the Maximum Principle for nonsmooth Mayer problems with free terminal points. Let  $\ell : M \rightarrow \mathbb{R}$  be locally Lipschitz and suppose that  $u^0$  is locally optimal for the problem of minimizing  $\ell(q(T; u))$  in the absence of terminal constraints. Let  $q^0$  be the trajectory for  $u^0$ .

THEOREM 5.2.3 (Pontryagin Maximum Principle). *There exists  $-\zeta_T \in \partial_L \ell(q^0(T))$  such that the solution  $\zeta(t)$  to (5.1.12) corresponding to  $u^0$  and  $\zeta_T$  satisfies*

$$(5.2.7) \quad H(t, \zeta(t), u^0(t)) = \max_{u \in \mathbb{U}} H(t, \zeta(t), u)$$

for almost all  $t \in [0, T]$ .

PROOF. By Proposition 5.2.2, control  $\nu^0(t) := \delta_{u^0(t)}$  is optimal among relaxed controls  $\mathcal{M}$ . Choose  $-\zeta_T \in \partial_L \ell(q^0(T))$  such that (5.1.11) holds for the solution  $\zeta(t)$  to (5.1.12) corresponding to  $u^0$  and  $\zeta_T$ . Let  $q^\lambda(t)$  be the trajectory for a sliding variation  $\delta_{u^0(t)} + \lambda(\nu(t) - \delta_{u^0(t)})$  and let  $v = \left. \frac{\partial q^\lambda(T)}{\partial \lambda} \right|_{\lambda=0}$ . Since  $q^0(t)$  is optimal we have  $\ell(q^0(T)) \leq \ell(q^\lambda(T))$  and this gives us  $D\ell(q^0(T); v) \geq 0$ . It follows that  $\min_{v \in E(u^0)} D\ell(q^0(T); v) \geq 0$  and so

$$(5.2.8) \quad \min_{\nu \in \mathcal{M}} \int_0^T \widehat{H}(t, \zeta(t), u^0(t)) - \widehat{H}(t, \zeta(t), \nu(t)) dt \geq 0.$$

By Filippov's lemma, there is a control  $u^{\max}$  such that for almost all  $t$ ,

$$(5.2.9) \quad H(t, \zeta(t), u^{\max}(t)) = \max_{u \in \mathbb{U}} H(t, \zeta(t), u).$$

By (5.2.8) we find that

$$(5.2.10) \quad \int_0^T H(t, \zeta(t), u^0(t)) - \max_{u \in \mathbb{U}} H(t, \zeta(t), u) dt \geq 0.$$

Since the integrand is almost everywhere nonpositive, we obtain (5.7.13).  $\square$

In the case of nonconvex dynamics with terminal constraint  $q(T) \in S$  the problem becomes more difficult. In this case  $u^0$  may not remain optimal among relaxed trajectories, a phenomenon sometimes called a *relaxation gap*. To progress in this case we introduce a pseudometric  $\rho$  on the set  $\mathcal{U}$  of measurable controls  $u : [0, T] \rightarrow \mathbb{U}$  and study conditions under which the associated constraints are *metrically regular*, a

term explained below. Due to its importance in applications, we develop our results in  $\mathbb{R}^n$  before turning to the case of a general manifold.

### 5.3. Metric Regularity and Penalization in $\mathbb{R}^n$

In this section we assume that  $M = \mathbb{R}^n$  and we write  $x$  rather than  $q$  to emphasize the underlying linear structure. We introduce the following pseudometric on the set of controls  $\mathcal{U}$ :

$$(5.3.1) \quad \begin{aligned} \rho(u, v) = & \int_0^T \|f(t, x(t; u), u(t)) - f(t, x(t; v), v(t))\| dt \\ & + \int_0^T \|f_x(t, x(t; u), u(t)) - f_x(t, x(t; v), v(t))\| dt. \end{aligned}$$

This pseudometric is central to our study of metric regularity. We next establish some elementary properties of this pseudometric.

**5.3.1. Elementary Properties and Estimates.** The following propositions are consequences of the definition and of the integral form Gronwall's lemma. We leave their proofs to the reader.

PROPOSITION 5.3.1. *Suppose that  $u, v \in \mathcal{U}$ . Then*

$$(5.3.2) \quad \max_{t \in [0, T]} \|x(t; u) - x(t; v)\| \leq \rho(u, v).$$

As a consequence, the function  $u \mapsto \ell(x(T; u))$  is locally Lipschitz with respect to  $\rho$ . Further, for a fixed control  $u^0$  if  $\varepsilon_0 > 0$  is sufficiently small we may assume the functions  $m_f, k_f$  providing inequalities (5.0.26) and (5.0.27) are defined on an open set containing all  $x, y$  attainable through controls  $u$  with  $\rho(u, u^0) < \varepsilon_0$ .

PROPOSITION 5.3.2. *Fix control  $u^0$  and  $\varepsilon_0 > 0$  as above. There exists a constant  $D_0$  such that if  $Q_{s,t}$  is the flow of any control  $u$  with  $\rho(u, u^0) < \varepsilon_0$  and  $\zeta \in \mathbb{R}^n$  is*

arbitrary, then

$$(5.3.3) \quad \max_{t \in [0, T]} \|Q_{t, T}^* \zeta\| \leq D_0 \|\zeta\|.$$

PROPOSITION 5.3.3. Fix control  $u^0$  and  $\varepsilon_0 > 0$  as above. Let  $Q_{s, t}^i$ ,  $i = 1, 2$  denote the flows of controls  $u^i$  with  $\rho(u^i, u^0) < \varepsilon_0$ . There exists a constant  $D_1$ , depending only on  $m_f$  such that for  $\zeta^1, \zeta^2 \in \mathbb{R}^n$

$$(5.3.4) \quad \max_{t \in [0, T]} \|Q_{t, T}^{1*} \zeta^1 - Q_{t, T}^{2*} \zeta^2\|_{\mathbb{R}^n} \leq D_1 \left( \|\zeta^1 - \zeta^2\|_{\mathbb{R}^n} + \rho(u^1, u^2) \max_i \|\zeta^i\|_{\mathbb{R}^n} \right).$$

As a first consequence of Proposition 5.3.3, consider the following. Suppose that  $u^k \xrightarrow{\rho} u$  and  $\zeta^k \rightarrow \zeta$ . Let  $R_{s, t}^k$  denotes the flow for  $u^k$  and  $R_{s, t}$  the flow corresponding to  $u$ . Then the curves  $R_{t, T}^{k*} \zeta^k$  converge uniformly to  $R_{t, T}^* \zeta$ .

**5.3.2. Completeness.** In 1974, Ekeland introduced in [39] his *variational principle*, now a well-established tool in optimization theory. In [63] it was shown that this principle holds for complete pseudometric spaces. The paper [16] includes several proofs of the original variational principle, one of which we find can be adapted to prove the following version for complete pseudometric space:

PROPOSITION 5.3.4. Let  $(X, \rho)$  be a complete pseudometric space and let  $\ell : X \rightarrow (-\infty, +\infty]$  be lower semicontinuous. If  $\bar{x}$  satisfies

$$(5.3.5) \quad \ell(\bar{x}) < \inf_{x \in X} \ell(x) + \varepsilon$$

then for any  $\lambda > 0$  there exists  $y \in X$  with  $\rho(\bar{x}, y) < \lambda$  which provides a global minimizer of the perturbed function

$$(5.3.6) \quad z \mapsto \ell(z) + \frac{\varepsilon}{\lambda} \rho(y, z).$$

This principle can be applied with our pseudometric  $\rho$  because of the following:



PROPOSITION 5.3.5. *The pseudometric  $\rho$  is complete.*

PROOF. Let  $\{u^k\}$  be Cauchy in  $\rho$  and let  $x^k$  be the trajectory for  $u^k$ . The sequence of functions  $(v^k, A^k) \in L^1([0, T], \mathbb{R}^n \times \mathbb{R}^{n \times n})$  defined by

$$(5.3.7) \quad (v^k(t), A^k(t)) = \left( f(t, x^k(t), u^k(t)), f_x(t, x^k(t), u^k(t)) \right)$$

is Cauchy in  $L^1$  and so converges to some function  $(v(t), A(t))$ . Let

$$(5.3.8) \quad x(t) = x_0 + \int_0^t v(\tau) d\tau.$$

We claim that there is a control  $u \in \mathcal{U}$  such that

$$(5.3.9) \quad (v(t), A(t)) = (f(t, x(t), u(t)), f_x(t, x(t), u(t)))$$

for almost all  $t$ . To see this, let  $\Gamma : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n \times \mathbb{R}^{n \times n}$  be the set-valued map given by

$$(5.3.10) \quad \Gamma(t, x) = \{(f(t, x, u), f_x(t, x, u)) : u \in \mathbb{U}\}.$$

Passing to a subsequence we may assume that  $(v^k(t), A^k(t))$  which converges pointwise almost everywhere to  $(v(t), A(t))$ . For almost all  $t$  we have  $(v^k(t), A^k(t)) \in \Gamma(t, x^k(t))$  for all  $k$  and so, because  $\Gamma$  is continuous in  $x$ , we find  $(v(t), A(t)) \in \Gamma(t, x(t))$  for these  $t$ . By Lemma 1.2.10 there is a control  $u$  such that (5.3.9) holds and thus  $\rho$  is complete.  $\square$

**5.3.3. Approximation in the Pseudometric.** In this subsection we show that the approximation of relaxed trajectories using traditional controls can be carried out in a manner that is compatible with  $\rho$ . These results will enable us to make use of sliding variations even in problems with nonconvex dynamics. We begin with the following approximation result:

PROPOSITION 5.3.6. Consider  $\nu \in \mathcal{M}$ ,  $u \in \mathcal{U}$ , and a fixed real number  $\lambda \in [0, 1]$ . Let  $x^\lambda$  be the trajectory for the variation  $(1 - \lambda)\delta_{u(t)} + \lambda\nu(t)$ . For any  $\varepsilon > 0$  there exists a control  $v \in \mathcal{U}$  and disjoint measurable sets  $A, B \subset [0, T]$  satisfying  $m(A) = (1 - \lambda)T$  and  $m(B) = \lambda T$  such that the control  $w(t) = \chi_A(t)u(t) + \chi_B(t)v(t)$  satisfies

$$(5.3.11) \quad \max_{t \in [0, T]} \left\| x^\lambda(t) - x(t; w) \right\| < \varepsilon.$$

PROOF. As before we follow Gamkrelidze [46] and approach this problem through chattering controls. The following lemma will be of use:

LEMMA 5.3.7. Suppose that  $x : [0, T] \rightarrow \mathbb{R}^n$  is continuous and  $\nu$  is a relaxed control. If  $f$  is integrable Lipschitz then for any  $\varepsilon > 0$  there exists a control  $v \in \mathcal{U}$  such that

$$(5.3.12) \quad \max_{t \in [0, T]} \left\| \int_0^t \widehat{f}(\tau, x(\tau), \nu(\tau)) - \widehat{f}(\tau, x(\tau), \delta_{v(\tau)}) d\tau \right\| < \varepsilon.$$

PROOF. Let  $\pi$  be a partition of  $[0, T]$  with diameter small enough that

$$(5.3.13) \quad \int_{t_i}^{t_{i+1}} \max_{u \in \mathbb{U}} \|f(\tau, x(\tau), u)\| d\tau < \frac{\varepsilon}{2}.$$

By Aumann's theorem, we may choose a control  $v$  on  $[t_i, t_{i+1}]$  so that

$$(5.3.14) \quad \int_{t_i}^{t_{i+1}} \widehat{f}(\tau, x(\tau), \nu(\tau)) d\tau = \int_{t_i}^{t_{i+1}} \widehat{f}(\tau, x(\tau), \delta_{v(\tau)}) d\tau.$$

Let  $t \in [0, T]$  be arbitrary, say  $t \in [t_i, t_{i+1}]$ . Then we have

$$(5.3.15) \quad \begin{aligned} & \left\| \int_0^t \widehat{f}(\tau, x(\tau), \nu(\tau)) - \widehat{f}(\tau, x(\tau), \delta_{v(\tau)}) d\tau \right\| \\ &= \left\| \int_{t_i}^t \widehat{f}(\tau, x(\tau), \nu(\tau)) - \widehat{f}(\tau, x(\tau), \delta_{v(\tau)}) d\tau \right\| \leq \varepsilon \end{aligned}$$

and this completes the proof.  $\square$

Let  $\varepsilon > 0$  be given. By Lemma 5.3.7, we may choose a control  $v \in \mathcal{U}$  so that

$$(5.3.16) \quad \max_{t \in [0, T]} \left\| \int_0^t \widehat{f}(\tau, x^\lambda(\tau), \nu(\tau)) - \widehat{f}(\tau, x^\lambda(\tau), \delta_{v(\tau)}) d\tau \right\| < \varepsilon.$$

We will apply Lemma 4.1.5. Partition  $[0, T]$  uniformly into intervals  $[t_i, t_{i+1}]$  and let

$$(5.3.17) \quad A = \bigcup_i [t_i, t_i + \lambda \operatorname{diam}(\pi)] \quad B = \bigcup_i (t_i + \lambda \operatorname{diam}(\pi), t_{i+1}].$$

Notice that  $A$  and  $B$  satisfy the necessary assumptions on measure. Lemma 4.1.5, applied to the functions  $g(\tau) = \widehat{f}(\tau, x(\tau), \delta_{u(\tau)})$  and  $h(\tau) = \widehat{f}(\tau, x(\tau), \delta_{v(\tau)})$ , implies that we may choose  $\operatorname{diam}(\pi)$  sufficiently small so that for any  $t \in [0, T]$ ,

$$(5.3.18) \quad \begin{aligned} & \left\| (1 - \lambda) \int_0^t \widehat{f}(\tau, x(\tau), \delta_{u(\tau)}) d\tau - \int_{A \cap [0, t]} \widehat{f}(\tau, x(\tau), \delta_{u(\tau)}) d\tau \right\| \leq \varepsilon \\ & \left\| \lambda \int_0^t \widehat{f}(\tau, x(\tau), \delta_{v(\tau)}) d\tau - \int_{B \cap [0, t]} \widehat{f}(\tau, x(\tau), \delta_{v(\tau)}) d\tau \right\| \leq \varepsilon. \end{aligned}$$

Now define  $w = \chi_A(t)u(t) + \chi_B(t)v(t)$ . For any  $t \in [0, T]$ , we have

$$(5.3.19) \quad \begin{aligned} \left\| x^\lambda(t) - x(t; w) \right\| & \leq \left\| (1 - \lambda) \int_0^t \widehat{f}(\tau, x^\lambda(\tau), \delta_{u(\tau)}) d\tau - \int_{A \cap [0, t]} \widehat{f}(\tau, x^\lambda(\tau), \delta_{u(\tau)}) d\tau \right\| \\ & + \left\| \lambda \int_0^t \widehat{f}(\tau, x^\lambda(\tau), \nu(\tau)) d\tau - \lambda \int_0^t \widehat{f}(\tau, x^\lambda(\tau), \delta_{v(\tau)}) d\tau \right\| \\ & + \left\| \lambda \int_0^t \widehat{f}(\tau, x^\lambda(\tau), \delta_{v(\tau)}) d\tau - \int_{B \cap [0, t]} \widehat{f}(\tau, x^\lambda(\tau), \delta_{v(\tau)}) d\tau \right\| \\ & + \int_{A \cap [0, t]} \left\| \widehat{f}(\tau, x^\lambda(\tau), \delta_{u(\tau)}) - \widehat{f}(\tau, x(\tau; w), \delta_{u(\tau)}) \right\| d\tau \\ & + \int_{B \cap [0, t]} \left\| \widehat{f}(\tau, x^\lambda(\tau), \delta_{v(\tau)}) - \widehat{f}(\tau, x(\tau; w), \delta_{v(\tau)}) \right\| d\tau \\ & \leq \varepsilon + \lambda \varepsilon + \varepsilon + \int_0^t k_f(\tau) \left\| x^\lambda(\tau) - x(\tau; w) \right\| d\tau. \end{aligned}$$

By Gronwall's lemma, we find that for any  $t \in [0, T]$ ,

$$(5.3.20) \quad \|x(t) - x(t; w)\|_{\mathbb{R}^n} \leq 3\varepsilon \exp(\|k_f\|_{L^1})$$

and since  $\varepsilon > 0$  was arbitrary this completes the proof.  $\square$

Our main approximation result, which will later allow us to use sliding variations without convexity assumptions, is the following:

PROPOSITION 5.3.8. *There exist constants  $C_1, C_2$  with the following property. Let control  $u \in \mathcal{U}$ , relaxed control  $\nu \in \mathcal{M}$ ,  $\lambda \in [0, 1]$ , and  $\varepsilon > 0$  be arbitrary and let  $x^\lambda$  be the trajectory for the sliding variation  $(1 - \lambda)\delta_{u(t)} + \lambda\nu(t)$ . There exists a control  $w^\lambda$  such that*

$$(5.3.21) \quad \max_{t \in [0, T]} \|x^\lambda(t) - x(t; w^\lambda)\| < \varepsilon$$

and

$$(5.3.22) \quad \rho(w^\lambda, u) \leq C_1\lambda + C_2\varepsilon.$$

PROOF. By Proposition 5.3.6 we may choose disjoint measurable sets  $A_\lambda, B_\lambda \subset [0, T]$  with  $m(A_\lambda) = (1 - \lambda)T$  and  $m(B_\lambda) = \lambda T$  and control  $v^\lambda$  such that the trajectory for  $w_t^\lambda = \chi_{A_\lambda}(t)u(t) + \chi_{B_\lambda}(t)v^\lambda(t)$  satisfies (5.3.21). We need only verify

(5.3.22). First,

$$\begin{aligned}
\rho(u, w^\lambda) &= \int_{A_\lambda} \left\| f(t, x(t; u), u(t)) - f(t, x(t; w^\lambda), u(t)) \right\| dt \\
&\quad + \int_{B_\lambda} \left\| f(t, x(t; u), u(t)) - f(t, x(t; w^\lambda), v(t)) \right\| dt \\
(5.3.23) \quad &+ \int_{A_\lambda} \left\| f_x(t, x(t; u), u(t)) - f_x(t, x(t; w^\lambda), u(t)) \right\| dt \\
&\quad + \int_{B_\lambda} \left\| f_x(t, x(t; u), u(t)) - f_x(t, x(t; w^\lambda), v(t)) \right\| dt \\
&\leq 2 \int_0^T k_f(t) \left\| x(t; u) - x(t; w^\lambda) \right\| dt + 4\lambda \|m_f\|_{L^1}.
\end{aligned}$$

We require a bound on  $\|x(t; u) - x(t; w^\lambda)\|$ , which we will obtain through a bound on  $\|x(t; u) - x^\lambda(t)\|$ . Consider

$$\begin{aligned}
(5.3.24) \quad \left\| x(t; u) - x^\lambda(t) \right\| &\leq (1 - \lambda) \int_0^t \left\| \widehat{f}(\tau, x(\tau; u), \delta_{u(\tau)}) - \widehat{f}(\tau, x^\lambda(\tau), \delta_{u(\tau)}) \right\| d\tau \\
&\quad + \lambda \int_0^t \left\| \widehat{f}(\tau, x(\tau; u), \delta_{u(\tau)}) - \widehat{f}(\tau, x^\lambda(\tau), \nu(\tau)) \right\| d\tau \\
&\leq (1 - \lambda) \int_0^t k_f(\tau) \left\| x(\tau; u) - x^\lambda(\tau) \right\| d\tau + 2\lambda \int_0^t m_f(\tau) d\tau.
\end{aligned}$$

By Gronwall's lemma, we find

$$(5.3.25) \quad \max_{t \in [0, T]} \left\| x(t; u) - x^\lambda(t) \right\| \leq 2\lambda \|m_f\|_{L^1} \exp(\|k_f\|_{L^1}).$$

Returning to (5.3.23) with (5.3.25) we find

$$\begin{aligned}
(5.3.26) \quad \rho(u, w^\lambda) &\leq 2 \int_0^T k_f(t) \left\| x^\lambda(t) - x(t; w^\lambda) \right\| dt \\
&\quad + 4\lambda \|m_f\|_{L^1} \|k_f\|_{L^1} \exp(\|k_f\|_{L^1}) + 4\lambda \|m_f\|_{L^1}.
\end{aligned}$$

Taking

$$(5.3.27) \quad \begin{aligned} C_1 &= 4 \|m_f\|_{L^1} \|k_f\|_{L^1} \exp(\|k_f\|_{L^1}) + 4 \|m_f\|_{L^1} \\ C_2 &= 2 \|k_f\|_{L^1} \end{aligned}$$

will yield (5.3.22).  $\square$

**5.3.4. Metric Regularity.** In this section we apply the pseudometric  $\rho$  to establish a sufficient condition for metric regularity of the constraint  $x(T) \in S$ . Such metric regularity will allow us to remove the terminal constraint  $x(T) \in S$  by penalizing our cost with a nonsmooth function. We denote by  $\mathcal{A}$  the set of controls  $u \in \mathcal{U}$  for which  $x(T; u) \in S$  and we introduce the following constraint qualification:

**Condition C:** There exist  $\varepsilon_0 > 0$  and  $\Delta > 0$  such that if control  $u$  satisfies  $\rho(u, u^0) < \varepsilon_0$  and  $d_S(x(T; u)) > 0$  then for any  $-\zeta_T \in \partial_L d_S(x(T; u))$  we have

$$(5.3.28) \quad \int_0^T H(t, \zeta(t), u(t)) - \max_{u \in \mathcal{U}} H(t, \zeta(t), u) dt \leq -\Delta,$$

where  $\zeta(t)$  is a solution to (5.1.12) corresponding to  $u$  and  $\zeta_T$ .

Condition C implies the following decrease principle:

**PROPOSITION 5.3.9.** *If Condition C holds then for any  $u \in \mathcal{U}$  satisfying  $\rho(u, u^0) < \varepsilon_0$  and  $d_S(x(T; u)) > 0$  there exists a relaxed control  $\nu$  such that if  $x^\lambda$  corresponds to the sliding variation  $(1 - \lambda)\delta_{u(t)} + \lambda\nu(t)$  there holds*

$$(5.3.29) \quad \liminf_{\lambda \downarrow 0} \frac{d_S(x^\lambda(T)) - d_S(x(T; u))}{\lambda} \leq -\Delta.$$

**PROOF.** We prove the contrapositive: suppose that the proposition is false. Then there exists a control  $u$  with  $\rho(u, u^0) < \varepsilon_0$  and  $d_S(x(T; u)) > 0$  such that for any  $v \in E(u)$  we have

$$(5.3.30) \quad Dd_S(x(T; u); v) > -\Delta.$$

By Proposition 5.1.5 we can find  $-\zeta_T \in \partial_L d_S(x(T; u))$  such that

$$(5.3.31) \quad \min_{\nu \in \mathcal{M}} \int_0^T \widehat{H}(t, \zeta(t), \delta_{u(t)}) - \widehat{H}(t, \zeta(t), \nu(t)) dt > -\Delta,$$

where  $\zeta(t)$  is a solution to (5.1.12) corresponding to  $u$  and  $\zeta_T$ . Choosing, through Filippov's selection lemma, a control  $u^{\max}$  such that  $H(t, \zeta(t), u^{\max}(t)) = \max_{u \in \mathcal{U}} H(t, \zeta(t), u)$  we find that

$$(5.3.32) \quad \int_0^T H(t, \zeta(t), u(t)) - \min_{u \in \mathcal{U}} H(t, \zeta(t), u) dt > -\Delta,$$

and this shows that Condition  $C$  cannot hold.  $\square$

Thus Condition  $C$  implies that any control  $u$  which is near  $u^0$  in the pseudometric  $\rho$  and is not admissible can be varied in a manner such that the terminal point is driven toward  $S$  in a manner that decreases the distance to  $S$  with linear rate not less than  $\Delta$ .

This decrease principle, coupled with the variational principle of Proposition 5.3.4 allows us to establish the following metric regularity result:

PROPOSITION 5.3.10. *Suppose that Condition  $C$  holds and let  $C_1$  and  $C_2$  be as in Proposition 5.3.8. There exists  $\varepsilon_1 > 0$  such that for  $\rho(u, u^0) < \varepsilon_1$ ,*

$$(5.3.33) \quad \inf_{w \in \mathcal{A}} \rho(u, w) \leq 4 \frac{C_1}{\Delta} d_S(x(T; u)).$$

PROOF. Choose  $0 < \varepsilon_1 < \frac{1}{2}\varepsilon_0$  so that if  $\rho(u, u^0) < \varepsilon_1$  we have  $4 \frac{C_1}{\Delta} d_S(x(T; u)) < \frac{1}{2}\varepsilon_0$ . Suppose by way of contradiction that there exists  $u \in \mathcal{U}$  with  $\rho(u, u^0) < \varepsilon_1$  for which

$$(5.3.34) \quad \inf_{w \in \mathcal{A}} \rho(u, w) > 4 \frac{C_1}{\Delta} d_S(x(T; u)).$$

Set  $\varepsilon = 2d_S(x(T; u))$  and  $\lambda = 4\frac{C_1}{\Delta}d_S(x(T; u))$ . Note that  $u$  is an  $\varepsilon$ -minimizer of the function  $w \mapsto d_S(x(T; w))$ . By the Ekeland variational principle, there exists  $v$  with

$$(5.3.35) \quad \rho(u, v) \leq 4\frac{C_1}{\Delta}d_S(x(T; u))$$

such that the function

$$(5.3.36) \quad w \mapsto d_S(x(T; w)) + \frac{\Delta}{2C_1}\rho(v, w)$$

attains a minimum over  $\mathcal{U}$  at  $v$ . By (5.3.34) and (5.3.35) we cannot have  $v \in \mathcal{A}$  and hence  $d_S(x(T; v)) > 0$ .

Our choice of  $\varepsilon_1$  assures us that

$$(5.3.37) \quad \rho(u^0, v) \leq \rho(u^0, u) + \rho(u, v) < \varepsilon_0$$

and so, by Proposition 5.3.9 we may choose a relaxed control  $\nu$  such that if  $x^\lambda$  is the trajectory for the sliding variation  $(1 - \lambda)\delta_{v(t)} + \lambda\nu(t)$  there holds

$$(5.3.38) \quad \liminf_{\lambda \downarrow 0} \frac{d_S(x^\lambda(T)) - d_S(x(T; v))}{\lambda} \leq -\Delta.$$

By Proposition 5.3.8 we may choose control  $w^\lambda$  such that  $\rho(v, w^\lambda) \leq C_1\lambda + C_2\lambda^2$  and  $\max_{t \in [0, T]} \|x^\lambda(T) - x(T; w^\lambda)\| < \lambda^2$ . Since  $v$  is optimal for (5.3.36) we find

$$(5.3.39) \quad \begin{aligned} 0 &\leq d_S(x(T; w^\lambda)) + \frac{\Delta}{2C_1}\rho(v, w^\lambda) - d_S(x(T; v)) \\ &\leq d_S(x^\lambda(T)) + \lambda^2 + \frac{\Delta}{2}\lambda + \frac{\Delta C_2}{2C_1}\lambda^2 - d_S(x(T; v)). \end{aligned}$$

Dividing by  $\lambda$  and letting  $\lambda \downarrow 0$  we find that  $0 \leq -\Delta + \frac{\Delta}{2} < 0$  and this contradiction proves (5.3.33).  $\square$



Inequality (5.3.33) is sometimes called *metric regularity* of the constraint  $x(T) \in S$ . In the case where this constraint is metrically regular we obtain the following penalization result:

PROPOSITION 5.3.11. *If Condition C holds then there exists  $\varepsilon_1 > 0$ ,  $K > 0$  such that  $u^0$  minimizes  $\ell(x(T; u)) + Kd_S(x(T; u))$  among controls with  $\rho(u, u^0) < \varepsilon_1$ .*

PROOF. Let  $k_\ell$  be a local Lipschitz constant for  $\ell$  with respect to  $\rho$ . By Proposition 5.3.10 we may choose  $\varepsilon_1, K > 0$  such that for  $\rho(u, u^0) < \varepsilon_1$  we have  $Kd_S(x(T; u)) \geq K_\ell \inf_{w \in \mathcal{A}} \rho(u, w)$ . Fix any  $u$  with  $\rho(u, u^0) < \varepsilon_1$ .

Given any  $\delta > 0$ , choose  $w \in \mathcal{A}$  for which  $\rho(u, w) < \inf_{w \in \mathcal{A}} \rho(u, w) + \delta$ . We have

$$\begin{aligned}
 \ell(x(T; u)) + Kd_S(x(T; u)) &\geq \ell(x(T; u)) + k_\ell \inf_{w \in \mathcal{A}} \rho(u, w) \\
 (5.3.40) \qquad \qquad \qquad &\geq \ell(x(T; w)) - k_\ell \rho(u, w) + k_\ell \inf_{w \in \mathcal{A}} \rho(u, w) \\
 &\geq \ell(x^0(T)) - k_\ell \delta,
 \end{aligned}$$

the last line following from optimality of  $u^0$  among controls in  $\mathcal{A}$ . Letting  $\delta \downarrow 0$  gives us the result.  $\square$

#### 5.4. Maximum Principle in $\mathbb{R}^n$ – Nonsmooth Mayer Problem with Terminal Constraints

We are now in a position to prove Theorem 5.0.1 in the case  $M = \mathbb{R}^n$ . We begin by studying the case in which Condition C fails to hold.

PROPOSITION 5.4.1. *Suppose that  $M = \mathbb{R}^n$  and that Condition C fails. Then Theorem 5.0.1 holds with  $\lambda^0 = 0$ .*

PROOF. In this case we may choose sequences  $\varepsilon_k, \Delta_k \downarrow 0$  and controls  $u^k \in \mathcal{U}$  with  $\rho(u^0, u^k) < \varepsilon_k$  whose trajectories  $x^k$  satisfy  $d_S(x^k(T)) > 0$  and for which there

exist  $-\zeta_T^k \in \partial_L d_S(x^k(T))$  such that

$$(5.4.1) \quad \int_0^T H(t, \zeta^k(t), u^k(t)) - \max_{u \in \mathbb{U}} H(t, \zeta^k(t), u) dt \geq -\Delta_k,$$

where  $\zeta^k$  is a solution to (5.1.12) corresponding to  $u^k$  and  $\zeta_T^k$ . Since  $x^k(T) \notin S$  we have  $\|\zeta_T^k\|_{\mathbb{R}^n} = 1$  and so we may pass to a subsequence which converges to  $-\zeta_T \in \partial_L d_S(x^0(T))$  with  $\|\zeta_T\|_{\mathbb{R}^n} = 1$ .

By Proposition 3.3.5 each arc  $\zeta^k$  can be written  $\zeta(t)^k = P_{t,T}^{k*} \zeta_T^k$ , where  $P_{s,t}^k$  is the flow for  $f$  corresponding to control  $u^k$ . Let  $P_{s,t}$  denote the flow of  $f$  corresponding to control  $u^0$  and set  $\zeta(t) = P_{t,T}^* \zeta_T$ . Note that  $\zeta$  is a solution to (5.1.12) for  $u^0$  and  $\zeta_T$ .

Since  $u^k \xrightarrow{\rho} u^0$  the comments following Proposition 5.3.3 imply that  $\zeta^k(t) \rightarrow \zeta(t)$  uniformly. And because  $P_{t,T}^*$  is an isomorphism and  $\|\zeta_T\|_{\mathbb{R}^n} = 1$  we see that  $\zeta(t) \neq 0$  for all  $t$ .

Finally, taking the limit in (5.6.1) we obtain

$$(5.4.2) \quad \int_0^T H(t, \zeta(t), u^0(t)) - \max_{u \in \mathbb{U}} H(t, \zeta(t), u) dt \geq 0.$$

Since the integrand is almost everywhere nonpositive, we see that  $H(t, \zeta(t), u^0(t)) = \max_{u \in \mathbb{U}} H(t, \zeta(t), u)$  for almost all  $t$ , completing the proof.  $\square$

We note that Theorem 5.0.2, in the case  $M = \mathbb{R}^n$  and  $\theta = Id_{\mathbb{R}^n}$ , is an immediate consequence of Propositions 5.3.11 and 5.4.1.

Thus in the case where Condition  $C$  fails the control  $u^0$  is an *abnormal minimizer* for our Mayer problem. In the following Proposition we see that when Condition  $C$  holds,  $u^0$  is a *normal minimizer*.

**PROPOSITION 5.4.2.** *Suppose that  $M = \mathbb{R}^n$  and that Condition  $C$  holds. Then Theorem 5.0.1 holds with  $\lambda^0 = 1$ .*

PROOF. Let  $v \in E(u^0)$  and choose a relaxed control  $\nu \in \mathcal{M}$  such that if  $x^\lambda$  is the trajectory for  $(1 - \lambda)\delta_{u^0(t)} + \lambda\nu(t)$  then

$$(5.4.3) \quad v = \lim_{\lambda \downarrow 0} \frac{x^\lambda(T) - x^0(T)}{\lambda}.$$

By Proposition 5.3.8 we may construct a family of controls  $w^\lambda \in \mathcal{U}$  with

$$(5.4.4) \quad \lim_{\lambda \downarrow 0} \frac{x^\lambda(T) - x(T; w^\lambda)}{\lambda} = 0.$$

For this family of controls we obtain

$$(5.4.5) \quad v = \lim_{\lambda \downarrow 0} \frac{x(T; w^\lambda) - x^0(T)}{\lambda}.$$

It now follows that  $D\ell(q^0(T); v) \geq 0$ .

Thus by Proposition 5.1.5 there exists

$$(5.4.6) \quad -\zeta_T \in \partial_L(\ell + Kd_S)(q^0(T)) \subset \partial_L\ell(q^0(T)) + N_S^L(q^0(T))$$

such that the solution to (5.1.12) corresponding to  $u^0$  and  $\zeta_T$  satisfies

$$(5.4.7) \quad \min_{u \in \mathcal{U}} \int_0^T H(t, \zeta(t), u^0(t)) - H(t, \zeta(t), u(t)) dt \geq 0.$$

We have seen that (5.4.7) implies the Maximum Principle and so the proof is complete.  $\square$

## 5.5. Metric Regularity and Penalization for Manifolds

In this section we use Lagrangian charts to provide a definition of the pseudo-metric  $\rho$  for problems posed on a general finite dimensional manifold  $M$ . As before we suppose that  $u^0$  is an optimal control with trajectory  $q^0$ . By Proposition 3.2.8 we may choose a compactly supported, locally  $L^1$ -bounded,  $C^\infty$ -smooth vector field  $V_t$  with flow  $P_{s,t}$  which extends  $\dot{q}^0$  in the sense that  $\dot{q}^0(t) = V_t(q^0(t))$ . Choose a

coordinate chart  $(\mathcal{O}, \varphi)$  with  $q_0 \in \mathcal{O}$  and let  $\psi_t := \varphi \circ P_{t,0}$  be the Lagrangian chart associated with vector field  $V_t$  and coordinate chart  $(\mathcal{O}, \varphi)$ .

We introduce a control system  $g : [0, T] \times \varphi(\mathcal{O}) \times \mathbb{U} \rightarrow \mathbb{R}^n$  by

$$(5.5.1) \quad g(t, x, u) = (\psi_{t*}f)(t, x, u) - (\psi_{t*}V_t)(x).$$

Our assumptions on  $f$  and  $V_t$  ensure that we may choose  $\mathcal{O}$  small enough that  $g$  satisfies the standing assumptions on  $f$ . In particular, one can establish the following:

PROPOSITION 5.5.1. *There exists  $\varepsilon_0 > 0$  and  $L^1$  functions  $m_g, k_g$  such that for any maps  $y, z : [0, T] \rightarrow \mathbb{R}^n$  satisfying*

$$(5.5.2) \quad \max_{t \in [0, T]} \|\varphi(q^0) - y(t)\| < \varepsilon_0 \quad \max_{t \in [0, T]} \|\varphi(q^0) - z(t)\| < \varepsilon_0$$

*there holds for almost all  $t \in [0, T]$  and for any  $u \in \mathbb{U}$*

$$(5.5.3) \quad \|g(t, y(t), u)\| \leq m_g(t) \quad \|g_x(t, y(t), u)\| \leq m_g(t)$$

*along with*

$$(5.5.4) \quad \begin{aligned} \|g(t, y(t), u) - g(t, z(t), u)\| &\leq k_g(t) \|y(t) - z(t)\| \\ \|g_x(t, y(t), u) - g_x(t, z(t), u)\| &\leq k_g(t) \|y(t) - z(t)\|. \end{aligned}$$

Introduce a neighborhood  $\mathcal{O}_0$  of  $q_0$  for which  $\overline{\mathcal{O}_0} \subset \mathcal{O}$ . Let  $\mathcal{U}_0$  denote the set of controls  $u$  for which  $q(t; u) \in P_{0,t}(\overline{\mathcal{O}_0})$  for all  $t$ . We define on  $\mathcal{U}_0$  the following pseudometric:

$$(5.5.5) \quad \begin{aligned} \rho(u, v) = &\int_0^T \|g(t, x(t; u), u(t)) - g(t, x(t; v), v(t))\|_{\mathbb{R}^n} \\ &+ \|g_x(t, x(t; u), u(t)) - g_x(t, x(t; v), v(t))\|_{\mathbb{R}^n} dt. \end{aligned}$$

Following the same proof technique as that for Proposition 5.3.5, one may check that  $\rho$  is a complete pseudometric. Further, because  $\psi_t(q(t; u))$  evolves according to control system  $g$ , the following generalization of Proposition 5.3.8 can be established:

**PROPOSITION 5.5.2.** *There exist constants  $C_1, C_2$  with the following property. Let control  $u \in \mathcal{U}$ , relaxed control  $\nu \in \mathcal{M}$ ,  $\lambda \in [0, 1]$ , and  $\varepsilon > 0$  be arbitrary and let  $q^\lambda(t)$  be the trajectory for  $\hat{f}$  corresponding to the sliding variation  $(1-\lambda)\delta_{u(t)} + \lambda\nu(t)$ . There exists a control  $w^\lambda$  for which*

$$(5.5.6) \quad \max_{t \in [0, T]} \left\| \psi_t(q^\lambda(t)) - \psi_t(q(t; w^\lambda)) \right\| < \varepsilon$$

and

$$(5.5.7) \quad \rho(w^\lambda, u) \leq C_1\lambda + C_2\varepsilon$$

when  $\lambda$  is sufficiently small.

**5.5.1. Metric Regularity.** In this subsection we derive sufficient conditions for exact penalization of the terminal constraint  $q(T) \in S$  through the locally defined nonsmooth penalty function  $d : P_{0, T}(\mathcal{O}) \rightarrow \mathbb{R}$  given by

$$(5.5.8) \quad d(q) = d_{\psi_T(S)} \circ \psi_T(q).$$

We introduce the following variant of Condition  $C$ :

**Condition  $C$ :** There exist  $\varepsilon_0 > 0$  and  $\Delta > 0$  such that if control  $u$  satisfies  $\rho(u, u^0) < \varepsilon_0$  and  $d(q(T; u)) > 0$  then for any  $-\zeta_T \in \partial_L d(q(T; u))$  there holds (5.3.28) where  $\zeta$  is the solution to (5.1.12) corresponding to  $u$  and  $\zeta_T$ .

Condition  $C$  again implies a decrease principle:

**PROPOSITION 5.5.3.** *If Condition  $C$  holds then for all  $u \in \mathcal{U}_0$  satisfying  $\rho(u, u^0) < \varepsilon_0$  and  $d(q(T; u)) > 0$ , then there exists a relaxed control  $\nu$  such that if  $q^\lambda$  corresponds*

to the sliding variation  $(1 - \lambda)\delta_{u(t)} + \lambda\nu(t)$ , then

$$(5.5.9) \quad \liminf_{\lambda \downarrow 0} \frac{d(q^\lambda(T)) - d(q(T; u))}{\lambda} \leq -\Delta.$$

PROOF. The proof is the same as that of Proposition 5.3.9 – a proof by contradiction employing Proposition 5.1.5.  $\square$

Now let  $\mathcal{A}_0 \subset \mathcal{U}_0$  be the set of controls  $u \in \mathcal{U}_0$  for which  $q(T; u) \in S$ . The decrease principle represented by Proposition 5.5.3 is enough to establish the following metric regularity result:

PROPOSITION 5.5.4. *Suppose that Condition C holds and let  $C_1$  and  $C_2$  be as in Proposition 5.5.2. There exists  $\varepsilon_1 > 0$  such that if  $\rho(u, u^0) < \varepsilon_1$ , then*

$$(5.5.10) \quad \inf_{w \in \mathcal{A}_0} \rho(u, w) \leq 4 \frac{C_1}{\Delta} d(q(T; u)).$$

PROOF. First, if necessary, we ask that  $\varepsilon_0$  is small enough that when  $\rho(u, u^0) < \varepsilon_0$  we have  $q(t; u) \in P_{0,t}(\mathcal{O}_0)$  for all  $t$  and hence  $u \in \mathcal{U}_0$ . Choose  $0 < \varepsilon_1 < \frac{1}{2}\varepsilon_0$  such that for  $\rho(u^0, u) < \varepsilon_1$  there holds  $4 \frac{C_1}{\Delta} d(q(T; u)) < \frac{1}{2}\varepsilon_0$ .

Now suppose by way of contradiction that there exists  $u \in \mathcal{U}$  with  $\rho(u, u^0) < \varepsilon_1$  for which

$$(5.5.11) \quad \inf_{w \in \mathcal{A}_0} \rho(u, w) > 4 \frac{C_1}{\Delta} d(q(T; u)).$$

Setting  $\varepsilon = 2d(q(T; u))$  and  $\lambda = 4 \frac{C_1}{\Delta} d(q(T; u))$ , we note that  $u$  is an  $\varepsilon$ -minimizer of  $w \mapsto d(q(T; w))$ . By the Ekeland variational principle, there exists  $v$  with

$$(5.5.12) \quad \rho(u, v) \leq 4 \frac{C_1}{\Delta} d(q(T; u))$$

such that the function

$$(5.5.13) \quad w \mapsto d(q(T; w)) + \frac{\Delta}{2C_1} \rho(v, w)$$

attains a minimum over  $\mathcal{U}_0$  at  $v$ . By (5.5.12) we find  $v \in \mathcal{U}_0$  and (5.5.11) implies  $v \notin \mathcal{A}_0$ , hence  $d(q(T; v)) > 0$ . By Proposition 5.5.3 we may choose a relaxed control  $\nu$  such that the trajectory  $q^\lambda$  for  $(1 - \lambda)\delta_{v(t)} + \lambda\nu(t)$  satisfies

$$(5.5.14) \quad \liminf_{\lambda \downarrow 0} \frac{d(q^\lambda(T)) - d(q(T; v))}{\lambda} \leq -\Delta.$$

By Lemma 5.5.2 we may choose control  $w^\lambda$  such that  $\rho(v, w^\lambda) \leq C_1\lambda + C_2\lambda^2$  and

$$(5.5.15) \quad \max_{t \in [0, T]} \left\| \psi_t(q^\lambda(t)) - \psi_t(q(t; w^\lambda)) \right\| < \lambda^2.$$

Now consider

$$(5.5.16) \quad d(q(T; w^\lambda)) = d_{\psi_T(S)} \circ \psi_T(q(T; w^\lambda)) \leq d_{\psi_T(S)} \circ \psi_T(q^\lambda(T)) + \lambda^2 = d(q^\lambda(T)) + \lambda^2.$$

Since  $v$  is optimal for (5.5.13) we have

$$(5.5.17) \quad d(q(T; v)) \leq d(q(T; w^\lambda)) + \frac{\Delta}{2C_1}\rho(v, w^\lambda) \leq d(q^\lambda(T)) + \lambda^2 + \frac{\Delta}{2}\lambda + \frac{\Delta C_2}{2C_1}\lambda^2.$$

This leads, as in the proof of Proposition 5.3.10, to the contradiction  $0 < 0$ .  $\square$

It now follows, just as in Proposition 5.3.11, that the following is true:

**PROPOSITION 5.5.5.** *If Condition C holds with  $d(q) = d_{\psi_T(S)} \circ \psi_T(q)$ . Then there exists  $\varepsilon_1 > 0$ ,  $K > 0$  such that  $u^0$  is optimal among controls with  $\rho(u, u^0) < \varepsilon_1$  for the problem of minimizing  $\ell(q(T; u)) + Kd(q(T; u))$ .*

This proposition may seem at first glance to be of limited use. After all, in practice it will be difficult or impossible to construct  $\psi_T$  explicitly and in this case our penalty function  $d(q) = d_{\psi_T(S)} \circ \psi_T(q)$  is also not explicitly computable. We will see in the following section that the function  $d$ , as defined, nonetheless holds considerable theoretical strength. For applied problems, Proposition 5.5.5 can be strengthened using the following:

PROPOSITION 5.5.6. *Suppose that  $\psi, \theta : M \rightarrow \mathbb{R}^n$  are diffeomorphisms defined in a neighborhood of  $q^0(T) \in S$ . There exists a constant  $K_{\psi, \theta}$  such that*

$$(5.5.18) \quad d_{\psi(S)} \circ \psi(q) \leq K_{\psi, \theta} d_{\theta(S)} \circ \theta(q)$$

for all  $q$  near  $q^0(T)$ .

PROOF. The map  $\psi \circ \theta^{-1}$  is locally Lipschitz, say with constant  $K_{\psi, \theta}$  on a neighborhood  $\mathcal{O}$  of  $q^0(T)$ . Let  $q \in \mathcal{U}$  be given and choose  $s_0 \in S$  such that  $\|\theta(q) - \theta(s_0)\| = d_{\theta(S)}(\theta(q))$ . We have

$$(5.5.19) \quad d_{\psi(S)}(\psi(q)) \leq \|\psi(q) - \psi(s_0)\|_{\mathbb{R}^n} \leq K_{\psi, \theta} \|\theta(q) - \theta(s_0)\| = K_{\psi, \theta} d_{\theta(S)}(\theta(q)).$$

□

As a consequence, the penalty function  $d$  may be replaced by  $d(q) = d_{\theta(S)} \circ \theta(q)$  for any coordinate chart  $\theta$ , provided that  $K$  is chosen sufficiently large.

## 5.6. Maximum Principle on Manifolds – Nonsmooth Mayer Problem with Terminal Constraints

We are now able to provide a general proof of the Maximum Principle in a fully geometric setting. Again we will find that the cases where Condition  $C$  holds or fails to hold correspond to the normality or abnormality of  $u^0$ , respectively. Let us start by considering the case in which Condition  $C$  fails:

PROPOSITION 5.6.1. *If Condition  $C$  fails then Theorem 5.0.1 holds with  $\lambda^0 = 0$ .*

PROOF. In this case we may choose sequences  $\varepsilon_k, \Delta_k \downarrow 0$  and controls  $u^k \in \mathcal{U}$  with  $\rho(u^0, u^k) < \varepsilon_k$  whose trajectories satisfy  $d(q^k(T)) > 0$  and for which there exists  $-\zeta_T^k \in \partial_L d(q^k(T))$  such that the solution  $\zeta^k$  to (5.1.12) for  $u^k$  and  $\zeta_T^k$  satisfies

$$(5.6.1) \quad \int_0^T H(t, \zeta^k(t), u^k(t)) - \max_{u \in \mathcal{U}} H(t, \zeta^k(t), u) dt \geq -\Delta_k.$$



Since the subdifferential is invariant under  $C^1$ -diffeomorphism [66] we have

$$(5.6.2) \quad \partial_L d(q) = \partial_L(d_{\psi_T(S)} \circ \psi_T)(q) = \psi_T^* \partial_L d_{\psi_T(S)}(x),$$

where  $x = \psi_T(q)$ .

Since  $-\zeta_T^k \in \partial_L d(q(T)^k)$ , we have  $\zeta^k = \psi_T^* \tilde{\zeta}^k$  for some  $-\tilde{\zeta}^k \in \partial_L d_{\psi_T(S)}(\psi_T(q^k(T)))$ . Since  $q^k(T) \notin S$  we have  $\|\tilde{\zeta}_T^k\|_{\mathbb{R}^n} = 1$  and so we may pass to a subsequence which converges to  $-\tilde{\zeta}_T \in \partial_L d_{\psi_T(S)}(\psi_T(q^0(T)))$  with  $\|\tilde{\zeta}_T\|_{\mathbb{R}^n} = 1$ .

Let  $R_{s,t}^k$  and  $Q_{s,t}^k$  denote flows corresponding to  $f$  and  $g$  for control  $u^k$ . The solutions to (5.1.12) for controls  $u^k$  and covectors  $\zeta^k$  are given by  $\zeta^k(t) = R_{t,T}^{k*} \zeta_T^k$ . By Proposition 3.4.4,

$$(5.6.3) \quad \zeta^k(t) = R_{t,T}^{k*} \psi_T^* \tilde{\zeta}_T^k = \psi_t^* Q_{t,T}^{k*} \tilde{\zeta}_T^k.$$

Since  $u^k \xrightarrow{\rho} u^0$ , the maps  $Q_{t,T}^{k*} \tilde{\zeta}_T^k$  converge uniformly to  $Q_{t,T}^* \tilde{\zeta}_T$ , where  $Q_{s,t}$  is the flow of  $g$  corresponding to  $u^0$ . Letting  $R_{s,t}$  denote the flow of  $f$  for  $u^0$  we find that, for  $\zeta_T = \psi_T^* \tilde{\zeta}_T$ ,

$$(5.6.4) \quad \lim_{k \rightarrow \infty} \zeta^k(t) = \lim_{k \rightarrow \infty} \psi_t^* Q_{t,T}^{k*} \tilde{\zeta}_T^k = \psi_t^* Q_{t,T}^* \tilde{\zeta}_T = R_{t,T}^* \psi_T^* \tilde{\zeta}_T = R_{t,T}^* \zeta_T,$$

the convergence being uniform. The adjoint equations (5.1.12) follow from  $\zeta(t) = R_{t,T}^* \zeta_T$  and the nondegeneracy condition  $\zeta(t) \neq 0$  follows from  $\zeta_T \neq 0$  along with the fact that  $R_{t,T}^* : T_{q^0(T)}^* M \rightarrow T_{q^0(t)}^* M$  is an isomorphism.

Finally, we may take the limit in (5.6.1) to obtain

$$(5.6.5) \quad \int_0^T H(t, \zeta(t), u^0(t)) - \max_{u \in \mathbb{U}} H(t, \zeta(t), u) dt \geq 0.$$

Since the integrand is almost everywhere nonpositive, we obtain (5.7.13).  $\square$

We note that Theorem 5.0.2 is now an immediate consequence of Propositions 5.5.5, 5.5.6, and 5.6.1. Finally, we obtain the following:

PROPOSITION 5.6.2. *If Condition C holds then Theorem 5.0.1 holds with  $\lambda^0 = 1$ .*

PROOF. In this case,  $u^0$  is optimal for the free terminal point problem with cost  $\ell + Kd$ . Further, if  $\zeta \in \partial_L d(q^0(T))$  then

(5.6.6)

$$\zeta \in \partial_L (d_{\psi_T(S)} \circ \psi_T) (q^0(T)) = \psi_T^* \partial_L d_{\psi_T(S)} (\psi_T(q^0(T))) = \psi_T^* N_{\psi_T(S)}^L (\psi_T(q^0(T)))$$

and so  $\zeta$  can be represented as  $\psi_T^* \tilde{\zeta}$  for some  $\tilde{\zeta} \in N_{\psi_T(S)}^L (\psi_T(q^0(T)))$ . Since the limiting normal is invariant under  $C^1$  diffeomorphisms, we have  $\zeta \in N_S^L (q^0(T))$  and the result reduces to the proof of the free-terminal point problem, which can be carried out just as in Proposition 5.4.2.  $\square$

## 5.7. Smooth Constraints

We conclude this chapter by studying the particular case in which  $S \subset M$  is given through smooth constraints for the form

$$(5.7.1) \quad \begin{aligned} \alpha_i(q) &\leq 0 \quad 1 \leq i \leq r \\ \beta_j(q) &= 0 \quad 1 \leq j \leq s, \end{aligned}$$

where  $\alpha_i : M \rightarrow \mathbb{R}$  and  $\beta_j : M \rightarrow \mathbb{R}$  are  $C^1$ -smooth functions.

Under an additional constraint qualification, the penalty function  $d_{\theta(S)} \circ \theta$  (where  $\theta$  is a coordinate chart defined on a neighborhood of  $q^0(T)$ ) can be replaced by a max-type function of the constraints. Such functions may be easier to work with in applications and can provide a useful characterization for the boundary condition  $\zeta_T$  in equation (5.1.12).

First check that (5.7.1) may be reduced to a single nonsmooth constraint  $\Phi(q) \leq 0$ , where

$$(5.7.2) \quad \Phi(q) = \max \left\{ \sum_{i=1}^r \lambda^i \alpha_i(q) + \sum_{j=1}^s \mu^j \beta_j(q) : \lambda_i \geq 0 \text{ and } \sum_{i=1}^r \lambda_i + \sum_{j=1}^s |\mu^j| \leq 1 \right\}.$$

Let  $\mathcal{K} \subset \mathbb{R}^r \times \mathbb{R}^s$  denote the set of all  $(\lambda, \mu)$  for which

- (i)  $\lambda^i \geq 0$  for  $1 \leq i \leq r$ ;
- (ii) The  $\lambda^i$  satisfy the complementary slackness condition  $\lambda^i \alpha_i(q^0(T)) = 0$  for each fixed  $i$ ,  $1 \leq i \leq r$ ;
- (iii)  $(\lambda, \mu)$  satisfies the nondegeneracy condition  $\sum_{i=1}^r \lambda^i + \sum_{j=1}^s |\mu^j| = 1$

**Condition S:** Suppose that for all  $(\lambda, \mu) \in \mathcal{K}$  there holds

$$(5.7.3) \quad \sum_{i=1}^r \lambda^i d\alpha_i(q^0(T)) + \sum_{j=1}^s \mu^j d\beta_j(q^0(T)) \neq 0.$$

When Condition S holds, then for any coordinate chart  $(\tilde{\mathcal{O}}, \theta)$  with  $q^0(T) \in \tilde{\mathcal{O}}$  there exists  $\varepsilon_0 > 0$  and  $\bar{\Delta} > 0$  such that for all  $x \in \theta(\tilde{\mathcal{O}})$  with  $d_{\theta(S)}(x) > 0$  there holds

$$(5.7.4) \quad \min_{(\lambda, \mu) \in \mathcal{K}} \left\{ \left\| \sum_{i=1}^r \lambda^i (\alpha_i \circ \theta^{-1})'(x) + \sum_{j=1}^s \mu^j (\beta_j \circ \theta^{-1})'(x) \right\|_{\mathbb{R}^n} \right\} \geq \bar{\Delta},$$

In this case one can show that for some  $K_\theta > 0$ , for all  $x \in \theta(\tilde{\mathcal{O}})$ ,

$$(5.7.5) \quad d_{\theta(S)}(x) \leq K_\theta \Phi(\theta^{-1}(x)).$$

Proofs of this and more general results can be found, for example, in [28].

Under this assumption, for any  $q \in \tilde{\mathcal{O}}$  we have  $d_{\theta(S)} \circ \theta(q) \leq K_\theta \Phi(q)$  and so for a sufficiently large constant  $\tilde{K}$  a control  $u^0$  which is locally optimal for the cost  $\ell(q) + K d_{\psi_T(S)} \circ \psi_T(q)$  will be locally optimal for the cost  $\ell(q) + \tilde{K} \Phi(q)$ . Further,

we have

$$(5.7.6) \quad N_S^L(q) = \theta^* N_{\theta(S)}^L(\psi_T(q)) = \text{cone } \theta^* \partial_L d_{\theta(S)}(\theta(q)) \subset \text{cone } \partial_L \Phi(q).$$

It remains only to characterize  $\partial_L \Phi(q)$ .

This can be done using techniques introduced in [65]. One may show that if  $\xi \in \partial_L \Phi(q)$  then there exist coefficients  $\lambda^i, \mu^j$  for which

$$(5.7.7) \quad \xi = \sum_{i=1}^r \lambda^i d\alpha_i(q) + \sum_{j=1}^s \mu^j d\beta_j(q).$$

Further,  $\lambda^i \geq 0$  and for a fixed  $i$  (no summation implied)  $\lambda^i \alpha_i(q) = 0$ .

The boundary conditions for the adjoint equations are then

$$(5.7.8) \quad -\zeta(T) \in \lambda_0 \partial_L \ell(q^0(T)) + \text{cone } \partial_L \Phi(q^0(T)).$$

Hence there exists  $\eta \in \partial_L \ell(q^0(T))$  and coefficients  $\lambda^i, \mu^j$ , satisfying the complementary slackness condition, such that

$$(5.7.9) \quad -\zeta(T) = \eta + \sum_{i=1}^r \lambda^i d\alpha_i(q^0(T)) + \sum_{j=1}^s \mu^j d\beta_j(q^0(T)).$$

We have established the following version of Theorem 5.0.1:

**THEOREM 5.7.1.** *Suppose that  $u^0$  is an optimal control with trajectory  $q^0$  for the problem of minimizing  $\ell(q(T))$  subject to the smooth constraints (5.7.1) and suppose that Condition S holds. Then there exist coefficients  $\lambda^i, \mu^j$  with  $\lambda_i \geq 0$  satisfying the complementary slackness assumption such that the solution  $\zeta$  to (5.1.12) corresponding to  $u^0$  and*

$$(5.7.10) \quad \zeta_T = -\lambda^0 d\ell(q^0(T)) - \sum_{i=1}^r \lambda^i d\alpha_i(q) - \sum_{j=1}^s \mu^j d\beta_j(q)$$

satisfies

$$(5.7.11) \quad H(t, \zeta(t), u^0(t)) = \max_{u \in \mathbb{U}} H(t, \zeta(t), u)$$

for almost all  $t$ . Further, either  $\lambda^0 = 1$  or  $\zeta(t) \neq 0$  for all  $t$ .

We also have the following:

**THEOREM 5.7.2.** *Suppose that  $u^0$  is an optimal control with trajectory  $q^0$  for the problem of minimizing  $\ell(q(T))$  subject to the smooth constraints (5.7.1) and suppose that Condition S holds. If for any coefficients  $\lambda^i, \mu^j$  with  $\lambda_i \geq 0$  satisfying the complementary slackness assumption such that the solution  $\zeta$  to (5.1.12) corresponding to  $u^0$  and*

$$(5.7.12) \quad \zeta_T = - \sum_{i=1}^r \lambda^i d\alpha_i(q) - \sum_{j=1}^s \mu^j d\beta_j(q)$$

*fails to satisfy the Maximum Principle*

$$(5.7.13) \quad H(t, \zeta(t), u^0(t)) = \max_{u \in \mathbb{U}} H(t, \zeta(t), u)$$

*on some set  $I \subset [0, T]$  of positive measure, then  $u^0$  is an unconstrained local minimizer for the penalized function  $\ell + K\Phi$ ,  $K$  sufficiently large.*

## Generic Existence and Uniqueness of Optimal Trajectories in Geometric Optimal Control

In this chapter we demonstrate that under, appropriate assumptions, there exists a solution to the following problem of Bolza: *Minimize*

$$(6.0.14) \quad \ell(q(T)) + \int_0^T L(t, q(t), u(t)) dt$$

subject to  $q(0) = q_0$  and

$$(6.0.15) \quad \dot{q}(t) = f(t, q(t), u(t)).$$

We further show that for almost any initial condition  $q$ , this optimal control will be unique and can be expressed through a feedback law  $u : [0, T] \times T^*M \rightarrow \mathbb{U}$ .

In the theory of existence and uniqueness for optimal control [10, 21, 25, 43], one often makes assumptions on convexity of associated sets. In the Mayer problem, a common assumption is that the set

$$(6.0.16) \quad f(t, q, \mathbb{U}) = \{f(t, q, u) : u \in \mathbb{U}\}$$

is convex and in the Bolza problem it is enough for the set

$$(6.0.17) \quad \{(f(t, q, u), \alpha) : \alpha \geq L(t, q, u), u \in \mathbb{U}\}$$

to be convex.

Such convexity assumptions are not strictly necessary. For example, consider a problem in which  $M = \mathbb{R}$ ,  $x(0) = x_0$ , and  $\dot{x}(t) = u(t)$ . Suppose that control  $u$  is required to take values in the set  $\mathbb{U} = \{\pm 1\}$  and that we are to minimize  $x(1)$ . This problem is non-convex and yet admits a unique optimal control given by  $u(t) = -1$  for almost all  $t$ .

The relevant feature of this problem is the following: for any  $(t, x) \in [0, T] \times \mathbb{R}$  and any nonzero  $p \in \mathbb{R}$  there is a unique  $u(t, x, p) \in \mathbb{U}$  which maximizes the Hamiltonian. It was shown in [62] that this is enough (under some assumptions) to guarantee the existence of a solution to the Mayer problem for any initial condition at which the value function<sup>1</sup> is differentiable. Here the value function is  $v(x) = x - 1$  and so is differentiable everywhere.

For this particular Mayer problem the adjoint arc given by  $p(t) = -1$  for all  $t$  and thus the optimal control may itself be expressed through the feedback law

$$(6.0.18) \quad u^0(t) := u(t, x^0(t), p(t)).$$

This is another salient feature of [62], in which optimal controls are given through (6.0.18), where  $(x(t), p(t))$  is the solution to the adjoint equations with  $-p(0) = v'(x(0))$ .

The central goal of this chapter is to generalize the work of [62] to the case of Bolza problems on manifolds.

The chapter is organized as follows. In the first section we introduce the value function for Mayer problems posed on manifolds, providing sufficient conditions for its Lipschitz continuity and characterizing its subdifferential. In the second section we prove the main result of the chapter, a generic existence theorem for Mayer problems with  $C^1$ -smooth cost functions and free terminal points. In the third section we demonstrate that this result can be generalized to the Bolza problem

---

<sup>1</sup>This function is defined in the following section.

discussed above. The final section is then devoted to the careful study of several convergence theorems for differential equations and control systems on manifolds which are required for the first sections.

We continue to assume that controls take values in a compact metric space  $\mathbb{U}$  and we write  $\mathcal{U}$  for the set of measurable controls  $u : [0, T] \rightarrow \mathbb{U}$ . Equipped with the metric

$$(6.0.19) \quad d(u, w) = \int_0^T d_{\mathbb{U}}(u(t), w(t)) dt,$$

$\mathcal{U}$  is a complete metric space.

We suppose that control system  $f : [0, T] \times M \times \mathbb{U} \rightarrow TM$  is measurable in time, continuous in  $u$ , and that for any  $q \in M$  there exists a coordinate chart  $(\mathcal{O}, \varphi)$  with  $q \in \mathcal{O}$  along with  $L^1$  functions  $k_\varphi$  and  $m_\varphi$  such that for almost all  $t$ , for all  $x, y \in \varphi(\mathcal{O})$  and  $u \in \mathbb{U}$  there holds

$$(6.0.20) \quad \begin{aligned} \|(\varphi_* f)(t, x, u) - (\varphi_* f)(t, y, u)\|_{\mathbb{R}^n} &\leq k_\varphi(t) \|x - y\|_{\mathbb{R}^n} \\ \|(\varphi_* f)_x(t, x, u) - (\varphi_* f)_x(t, y, u)\|_{\mathbb{R}^n} &\leq k_\varphi(t) \|x - y\|_{\mathbb{R}^n} \end{aligned}$$

along with

$$(6.0.21) \quad \|(\varphi_* f)(t, x, u)\|_{\mathbb{R}^n} \leq m_\varphi(t) \quad \|(\varphi_* f)_x(t, x, u)\|_{\mathbb{R}^n} \leq m_\varphi(t).$$

That is, we assume that  $f$  is  $C^1$  in  $q$  with locally integrable Lipschitz and integrable bounded derivative.

In this chapter we will also make an additional assumption on  $f$  which is akin to a growth assumption. This assumption is the following:

**ASSUMPTION 6.0.3.** Given any initial condition  $q_0 \in M$ , there is a compact set  $K \subset M$ , depending on  $q_0$ , which contains the reachable set  $\mathcal{R}_T(q_0)$ .



This assumption rules out many topological obstacles to existence. For example, when  $M = \mathbb{R}^n \setminus \{0\}$ , many seemingly innocuous control problems fail to admit solutions. Interestingly, the map  $x \mapsto x/\|x\|$  is a diffeomorphism of this manifold under which such control systems will often be seen to fail a growth condition.

### 6.1. The Value Function

The generic existence and uniqueness theorems presented in [62] rely on a careful study of the value function and its subdifferentials. In this section we define the value function, establish sufficient conditions for its Lipschitz continuity, and characterize its subdifferential.

We restrict our attention for the moment to the Mayer problem with locally Lipschitz cost  $\ell : M \rightarrow \mathbb{R}$  and we define a function  $v : M \times \mathcal{U} \rightarrow \mathbb{R}$  by

$$(6.1.1) \quad v(q, u) = \ell(q(T; q, u)).$$

Thus  $v(q, u)$  is the cost of running the system with initial condition  $q$  and control  $u$ . The *value function*  $v : M \rightarrow \mathbb{R}$  is defined as

$$(6.1.2) \quad v(q) = \inf_{u \in \mathcal{U}} v(q, u).$$

Notice that  $u^0$  is an optimal control if and only  $v(q) = v(q, u^0)$ .

The study of value function in dynamic optimization has a long history which we will not attempt to summarize here. It should be mentioned that one often accounts for initial time  $t$  and defines

$$(6.1.3) \quad v(t, q) = \inf_{u \in \mathcal{U}} \ell(q(T; t, q, u)),$$

with  $q(\cdot; t, q, u)$  the solution to (6.0.15) satisfying  $q(t; t, q, u) = q$ . In this case the function  $v(t, q)$  is the solution, in an appropriately chosen nonsmooth sense, to the Hamilton-Jacobi equations. For a modern presentation of in this direction

we suggest [28] and also the work of Subbotin [81]. A classical presentation in the theory of calculus of variations is given in [48]. For the purposes of this dissertation, we will restrict our attention to the case given by (6.1.2).

**6.1.1. Lipschitz Continuity.** We prove that the value function  $v : M \rightarrow \mathbb{R}$  is locally Lipschitz. Fix a point  $q_0 \in M$ . By Assumption 6.0.3 we may choose a compact set  $K$  with the property that  $\mathcal{R}_T(q_0) \subset K$  and this compact set will facilitate the construction of a finite collection of coordinate charts  $(\mathcal{C}_i, \varphi_i)$  with the property that if  $q$  is sufficiently close to  $q_0$  then, for any control  $u \in \mathcal{U}$ , the trajectories  $q(t; q_0, u)$  and  $q(t; q, u)$  always lie in a common coordinate domain  $\mathcal{C}_i$ . We may further control the number of chart changes required to track these two trajectories and this will allow us to study the endpoint map  $q \mapsto q(T; q, u)$  using only finitely many changes of charts.

We make these claims precise through the following propositions.

**PROPOSITION 6.1.1.** *There exists a finite collection of charts  $(\mathcal{C}_i, \varphi_i)_{i=1}^r$ , a collection of subsets  $\mathcal{B}_i \subset \overline{\mathcal{B}}_i \subset \mathcal{C}_i$ , and a number  $\delta > 0$  such that:*

- (i) *Each set  $\overline{\mathcal{C}}_i$  is compact and  $\varphi_i$  is defined on  $\overline{\mathcal{C}}_i$ ;*
- (ii) *Given any control  $u \in \mathcal{U}$  and any interval  $[a, b] \subset [0, T]$  with  $|a - b| < \delta$ , we may choose an index  $i$  for which  $q(t; 0, q_0, u) \in \mathcal{B}_i$  for all  $t \in [a, b]$ ;*
- (iii) *There exist  $m_i$  and  $k_i$  for which the bounds (6.0.20) and (6.0.21) hold in the local coordinates  $\varphi_i$  on  $\mathcal{C}_i$ ;*
- (iv)  *$K \subset \cup_{1 \leq i \leq r} \mathcal{C}_i$ .*

**PROOF.** For each  $q \in K$  choose a chart  $(\varphi_q, \mathcal{C}_q)$  on which the bounds (6.0.20) and (6.0.21) hold for  $L^1$  functions  $k_q, m_q$ . Without loss of generality we assume that  $\overline{\mathcal{C}}_q$  is compact. Choose open sets  $\mathcal{A}_q$  and  $\mathcal{B}_q$  for which

$$(6.1.4) \quad q \in \mathcal{A}_q \subset \overline{\mathcal{A}}_q \subset \mathcal{B}_q \subset \overline{\mathcal{B}}_q \subset \mathcal{C}_q$$

and let  $\delta_q > 0$  such that for any  $q' \in \mathcal{A}_q$ , for any  $t \in [0, T]$  and control  $u \in \mathcal{U}$ , we have  $q(s; t, q', u(\cdot)) \in \mathcal{B}_q$  whenever  $|s - t| < \delta$ . Such  $\delta$  exists because the local coordinate representation of  $f$  is bounded in norm by an integrable function.

The collection  $\{\mathcal{A}_q : q \in K\}$  forms an open cover for  $K$ . By compactness we may choose a finite collection of charts  $(\varphi_i, \mathcal{C}_i)_{1 \leq i \leq r}$  so that  $K \subset \cup_{1 \leq i \leq r} \mathcal{A}_i$ . Let  $\delta > 0$  denote the minimum of the corresponding  $\delta_i$ . We claim that  $(\mathcal{C}_i, \varphi_i)_{1 \leq i \leq r}$  and  $\delta$  have the desired properties.

Indeed, properties i, iii and iv are immediate from our construction. To check property ii choose any control  $u(\cdot)$  and  $[a, b] \subset [0, T]$  with  $|a - b| < \delta$ . Fix  $i$  so that  $q(a; 0, q_0, u) \in \mathcal{A}_i$ . Recalling that

$$(6.1.5) \quad q(t; 0, q_0, u) = q(t; a, q(a; 0, q_0, u), u),$$

we have by construction  $q(t; a, q(a; 0, q_0, u), u) \in \mathcal{B}_i$  for  $t \in [a, b]$ .  $\square$

As a first application of Proposition 6.1.1 we prove that the endpoint map  $q \mapsto q(T; q, u)$  is locally Lipschitz in  $q$  and that the constant may be chosen independently of  $u$ . The constant, of course, depends on the charts chosen in Proposition 6.1.1.

**PROPOSITION 6.1.2.** *Fix  $q_0 \in M$  and let  $(\mathcal{C}_i, \varphi_i)_{i=1}^r$  and  $\delta > 0$  be as in Proposition 6.1.1. Let  $\pi$  be a partition  $0 = t_0 < t_1 < \dots < t_n = T$  of  $[0, T]$  with  $\text{diam}(\pi) < \delta$ . There exists a neighborhood  $\mathcal{O}_\pi$  of  $q_0$  such that:*

- (i) *For any interval  $[t_i, t_{i+1}]$ , any control  $u$ , and any  $q_1, q_2 \in \mathcal{O}_\pi$  there exists an index  $j_i$  such that  $q(t; 0, q_1, u), q(t; 0, q_2, u) \in \mathcal{C}_{j_i}$  for all  $t \in [t_i, t_{i+1}]$ ;*
- (ii) *The endpoint map  $q \mapsto q(T; 0, q, u)$  is locally Lipschitz on  $\mathcal{O}_\pi$  with rank independent of  $u$ .*

*Note that the first property implies that for any  $q \in \mathcal{O}_\pi$  we have  $\mathcal{R}_T(q) \subset \cup_{i=1}^r \mathcal{C}_i$ .*

PROOF. For each  $1 \leq i \leq r$  let  $k_i$  denote the  $L^1$  Lipschitz rank of  $\varphi_i * f$ . Set  $k(t) = \max_{1 \leq i \leq r} k_i(t)$ . Let  $L_\varphi$  be an upper bound on the Lipschitz ranks of  $\varphi_i \circ \varphi_j^{-1}$  for indices  $i, j$  such that  $\mathcal{O}_i \cap \mathcal{O}_j \neq \emptyset$ . Since the sets  $\varphi_i(\overline{\mathcal{C}_i})$  are all compact,  $L_\varphi$  is finite.

Now choose a number  $\varepsilon > 0$  so that for any  $1 \leq i \leq r$  we have

$$(6.1.6) \quad \varphi_i(\mathcal{B}_i) + \varepsilon\mathbb{B} \subset \varphi_i(\mathcal{C}_i).$$

We will prove that for some neighborhood  $\mathcal{O}_\pi$  of  $q_0$ , for any control  $u \in \mathcal{U}$  and any  $q_1 \in \mathcal{O}_\pi$ , the trajectory  $q(t; 0, q_1, u)$  lies within the  $\varepsilon$ -tube about  $q(t; 0, q_0, u)$  for any of the local coordinates  $\varphi_i$ . This will prove the first property.

Regardless of which control we choose, we will have  $q_0 \in \mathcal{A}_{i_1}$  for some  $i_1$ . Fix  $i_1$  and let  $\mathcal{O}_\pi$  be the set of all  $q_1 \in \mathcal{A}_1$  for which

$$(6.1.7) \quad \|\varphi_{i_1}(q_0) - \varphi_{i_1}(q_1)\| < \frac{\varepsilon}{L_\varphi^{r-1}} e^{-\|k\|_{L^1}}.$$

Let  $q_1 \in \mathcal{O}_\pi$  and control  $u \in \mathcal{U}$  be arbitrary. Define

$$(6.1.8) \quad q(t) = q(t; 0, q_0, u) \quad \tilde{q}(t) = q(t; 0, q_1, u).$$

By Proposition 6.1.1 we may choose indices  $i_j$ ,  $2 \leq j \leq r$  such that  $q(t) \in \mathcal{B}_{i_j}$  for all  $t \in [t_j, t_{j+1}]$ . For  $t \in [0, t_1]$  we have

$$(6.1.9) \quad \begin{aligned} \|\varphi_{i_1}(q(t)) - \varphi_{i_1}(\tilde{q}(t))\| &\leq \|\varphi_{i_1}(q_0) - \varphi_{i_1}(q_1)\| \\ &\quad + \int_0^t k(\tau) \|(\varphi_{i_1} * f)(\tau, \varphi_{i_1}(q(\tau)), u(\tau)) - (\varphi_{i_1} * f)(\tau, \varphi_{i_1}(\tilde{q}(\tau)), u(\tau))\| d\tau \\ &\leq \|\varphi_{i_1}(q_0) - \varphi_{i_1}(q_1)\| + \int_0^t k(\tau) \|\varphi_{i_1}(q(\tau)) - \varphi_{i_1}(\tilde{q}(\tau))\| d\tau, \end{aligned}$$

provided that  $\tilde{q}(t) \in \mathcal{C}_{i_1}$  for all  $t \in [0, t_1]$ . By the Gronwall lemma,

$$(6.1.10) \quad \|\varphi_{i_1}(q(t)) - \varphi_{i_1}(\tilde{q}(t))\| \leq \|\varphi_{i_1}(q_0) - \varphi_{i_1}(q_1)\| e^{\int_0^t k(\tau) d\tau} < \varepsilon$$

and by (6.1.6) we see that indeed  $\tilde{q}(t) \in \mathcal{C}_{i_1}$  for all  $t \in [0, t_1]$ .

For  $t \in [t_1, t_2]$  we have by the same argument

$$\begin{aligned}
(6.1.11) \quad & \|\varphi_{i_2}(q(t)) - \varphi_{i_2}(\tilde{q}(t))\|_{\mathbb{R}^n} \leq \|\varphi_{i_2}(q(t_1)) - \varphi_{i_2}(\tilde{q}(t_1))\| e^{\int_{t_1}^t k(\tau) d\tau} \\
& \leq L_\varphi \|\varphi_{i_1}(q(t_1)) - \varphi_{i_1}(\tilde{q}(t_1))\| e^{\int_{t_1}^t k(\tau) d\tau} \\
& \leq L_\varphi \|\varphi_{i_1}(q_0) - \varphi_{i_1}(q_1)\| e^{\int_0^{t_1} k(\tau) d\tau} e^{\int_{t_1}^t k(\tau) d\tau} \\
& \leq L_\varphi \|\varphi_{i_1}(q_0) - \varphi_{i_1}(q_1)\| e^{\int_0^{t_2} k(\tau) d\tau} < \varepsilon.
\end{aligned}$$

Arguing inductively we see that

$$(6.1.12) \quad \|\varphi_{i_j}(q(t)) - \varphi_{i_j}(\tilde{q}(t))\|_{\mathbb{R}^n} \leq L_\varphi^{r-1} \|\varphi_i(q_0) - \varphi_i(q_1)\| e^{\|k\|_{L^1}} < \varepsilon$$

for any  $t \in [t_j, t_{j+1}]$ ,  $j = 1, \dots, r$ . This proves the first property and implies the second.  $\square$

**PROPOSITION 6.1.3.** *If  $\ell$  is locally Lipschitz and  $f$  satisfies Assumption 6.0.3 and local estimates (6.0.20) and (6.0.20) then there exists a coordinate neighborhood  $(\mathcal{O}, \varphi)$  of  $q_0$  on which the functions  $q \mapsto v(q, u)$  are locally Lipschitz with rank independent of  $u$ .*

**PROOF.** This is a consequence of Proposition 6.1.2, along with our assumption that  $\ell$  is locally Lipschitz.  $\square$

**PROPOSITION 6.1.4.** *If  $\ell$  is locally Lipschitz and  $f$  satisfies Assumption 6.0.3 and local estimates (6.0.20) and (6.0.20) then there exists a coordinate neighborhood  $(\mathcal{O}, \varphi)$  of  $q_0$  on which the function  $v$  is Lipschitz.*

**PROOF.** The inf-envelope of a uniformly locally Lipschitz family of functions is again locally Lipschitz and so this Proposition follows from Proposition 6.1.3.  $\square$

As a consequence  $v$  has a well-defined subdifferential, toward which we now turn our attention.

**6.1.2. Subdifferential.** The value function is defined through (6.1.2) as the inf-envelope of a family of functions. In [65] the multidirectional mean value inequality was used to characterize the subdifferentials of such functions. Here we apply this result to study the subdifferential of  $v$  in terms of subdifferentials of the functions  $q \mapsto v(q, u)$ . In particular, we show that if  $\zeta \in \partial_F v(q)$  then there are sequences  $q_n \rightarrow q$  and  $\zeta_n \in \partial_F v(\cdot, u_n)(q_n)$  for which  $\zeta_n \rightarrow \zeta$ . Further, controls  $u_n$  may be chosen to satisfy  $\lim v(q_0, u_n) = v(q_0)$ . Since  $\partial_F v(\cdot, u_n)(q_n) = P_{0,T}^* \partial_F \ell(q(T); q_n, u_n)$ , with  $P_{s,t}$  the flow of  $u_n$ , the covectors  $\zeta_n$  may be described in terms of the adjoint equations and this relationship is the essential tool in the existence theorem proved in the next section.

We begin by recalling the needed result from [65], stated here for  $\mathbb{R}^n$ :

PROPOSITION 6.1.5. *Let  $\{g_\gamma\}_{\gamma \in \Gamma}$  be a family of lower semicontinuous functions  $g_\gamma : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ . Suppose function  $g$  is defined through*

$$(6.1.13) \quad g(x) = \inf_{\gamma \in \Gamma} g_\gamma(x),$$

*suppose that  $g$  is lower semicontinuous at  $x_0$  and that  $\zeta \in \partial_F g(x_0)$ .*

*There exists a function  $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  with  $\lim_{\varepsilon \downarrow 0} \psi(\varepsilon) = 0$  such that for any small  $\varepsilon$  and any  $(\gamma_\varepsilon, x_\varepsilon)$  satisfying*

$$(6.1.14) \quad x_\varepsilon \in x_0 + \varepsilon \mathbb{B} \quad g_{\gamma_\varepsilon}(x_\varepsilon) < g(x_0) + \varepsilon$$

*there exists  $z_\varepsilon \in x_0 + 2\sqrt{\varepsilon} \mathbb{B}$  such that*

$$(6.1.15) \quad g(z_\varepsilon) < g(x_0) + O(\sqrt{\varepsilon}) \quad \zeta \in \partial_F g_{\gamma_\varepsilon}(z_\varepsilon) + \varphi(\varepsilon) \mathbb{B}.$$

In [65] one may also find the following:

PROPOSITION 6.1.6. *If, in the definition of  $g$  (6.1.13), each  $g_\gamma$  is  $C^1$  smooth, then  $\partial_F v(x)$  is a singleton whenever it is nonempty.*

In our situation, the pair  $(\gamma_\varepsilon, x_\varepsilon)$  is replaced by a pair  $(u_\varepsilon, q_0)$  with  $u_\varepsilon$  a control satisfying  $v(q_\varepsilon, u_\varepsilon) < v(q_\varepsilon) + \varepsilon$ . We follow [62] and introduce the following set of controls:

$$(6.1.16) \quad \mathcal{U}(q; \varepsilon) := \{u \in \mathcal{U} : v(q, u) < v(q) + \varepsilon\}.$$

The following proposition gives us a characterization of  $\partial_F v(q)$  in terms of subgradients for the functions  $v(q, u)$  and the sets  $\mathcal{U}(q; \varepsilon)$ :

**PROPOSITION 6.1.7.** *There exists  $C_0 > 0$  such that for any  $\zeta \in \partial_F v(q_0)$  and any sequences  $\varepsilon_n \downarrow 0$ ,  $u_n \in \mathcal{U}(q_0; \varepsilon_n)$ , there exist sequences  $q_n \in M$ ,  $\zeta_n \in \partial_F v(\cdot, u_n)(q_n)$ , and  $\delta_n > 0$  such that  $q_n \rightarrow q_0$ ,  $\zeta_n \rightarrow \zeta$  in the topology of  $T^*M$ , and  $\delta_n \rightarrow 0$  for which*

$$(6.1.17) \quad |v(q_n) - v(q_0)| \leq C_0 \delta_n$$

and

$$(6.1.18) \quad |v(q_n, u_n) - v(q_0, u_n)| \leq C_0 \delta_n.$$

**PROOF.** Choose a coordinate chart  $(\varphi, \mathcal{O})$  for which  $q_0 \in \mathcal{O}$  and set  $x_0 = \varphi(q_0)$ . Define functions  $g : \varphi(\mathcal{O}) \times \mathcal{U} \rightarrow \mathbb{R}$  through

$$(6.1.19) \quad g(x, u) = v(\varphi^{-1}(x), u)$$

and set

$$(6.1.20) \quad g(x) = \inf_{u \in \mathcal{U}} g(x, u) = v(\varphi^{-1}(x)).$$

By Proposition 6.1.2 we may let  $C_0$  be a bound on the Lipschitz ranks of these functions.

Suppose that  $\zeta \in \partial_F v(q_0)$  and that we are given sequences  $\varepsilon_n \downarrow 0$  and  $u_n \in \mathcal{U}(q_0; \varepsilon_n)$ . Choose  $\tilde{\zeta} \in \partial_F g(x_0)$  for which  $\varphi^* \tilde{\zeta} = \zeta$ . By Proposition 6.1.5 there exists a function  $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that for  $n$  sufficiently large there exists  $z_n \in x_0 + 2\sqrt{\varepsilon_n} \mathbb{B}$  and  $\tilde{\zeta}^n \in \partial_F g(\cdot; u_n)(z_n)$  for which  $\|\tilde{\zeta}^n - \tilde{\zeta}\| < \psi(\varepsilon_n)$ . Let  $\delta_n = \|z_n - x_0\|$  and note that  $\delta_n \downarrow 0$ . Further, we have

$$(6.1.21) \quad |g(z_n) - g(x_0)| < C_0 \delta_n,$$

and

$$(6.1.22) \quad |g(z_n; u_n) - g(x_0; u_n)| \leq C_0 \delta_n.$$

The proof is completed by setting  $q_n = \varphi^{-1}(z_n)$  and  $\zeta^n = \varphi^* \tilde{\zeta}^n$ .  $\square$

## 6.2. Existence and Uniqueness for Mayer Problems

In this section we establish the main result of the chapter, which is an existence theorem for the following generic Mayer problem: minimize  $\ell(q(T; q, u))$  subject to (6.0.15) and  $q(0; q, u) = q$ .

We first introduce the assumptions on  $f$ .

ASSUMPTION 6.2.1. Let control system  $f$  satisfy Assumption 6.0.3 as well as local estimates (6.0.20) and (6.0.21). Suppose in addition that the associated Hamiltonian  $H : [0, T] \times T^*M \times \mathbb{U} \rightarrow \mathbb{R}$  is such that

- (i) For any nonzero  $\zeta \in T^*M$ , there is a unique  $u(t, \zeta)$  for which

$$(6.2.1) \quad H(t, \zeta, u(t, \zeta)) = \max_{u \in \mathbb{U}} H(t, \zeta, u)$$

holds;

- (ii) The differential equation

$$(6.2.2) \quad \dot{\zeta}(t) = \vec{H}(t, \zeta(t), u(t, \zeta(t))) \quad \zeta(0) = \zeta_0$$



admits unique solutions for any  $\zeta_0 \in T^*M$  on the time interval  $[0, T]$ .

We will see that these assumptions are enough to demonstrate existence of an optimal control whenever  $\partial_F v(q)$  is nonempty and uniqueness whenever  $v$  is differentiable. We further show that the optimal control  $u^0$  can be expressed in the feedback form

$$(6.2.3) \quad u^0(t) = u(t, \zeta(t)),$$

where  $\zeta : [0, T] \rightarrow T^*M$  is a solution to (6.2.2) satisfying  $-\zeta(0) \in \partial_F v(q)$ .

Let us note that the feedback map  $\zeta \mapsto u(t, \zeta)$  is continuous for fixed  $t$ . Applying Filippov's measurable selection theorem one finds that for any continuous map  $\zeta : [0, T] \rightarrow T^*M$  the control  $u(t) = u(t, \zeta(t))$  is measurable.

**6.2.1. Approximate Maximum Principle.** The generic existence and uniqueness result established in this chapter relies on the following approximate maximum principle:

**PROPOSITION 6.2.2.** *Suppose that  $u \in \mathcal{U}$  minimizes the perturbed functional  $w \mapsto v(q_0, w) + \Delta d(u, w)$  for a fixed  $\Delta > 0$ . Then there exists  $-\zeta_T \in \partial_L \ell(q(T; u))$  such that*

$$(6.2.4) \quad \int_0^T H(t, \zeta(t), u(t)) - \max_{u \in \mathbb{U}} H(t, \zeta(t), u) dt \geq -\Delta T \text{diam}(\mathbb{U}),$$

where  $\zeta : [0, T] \rightarrow T^*M$  is the solution to the adjoint equations

$$(6.2.5) \quad \dot{\zeta}(t) = \overrightarrow{H}(t, \zeta(t), u(t))$$

satisfying  $\zeta(T) = \zeta_T$ .

**PROOF.** The proof is similar to that of Theorem 5.2.3. Fix a coordinate chart  $\varphi$  at  $q(T; u)$  and let  $k_\ell$  denote the local Lipschitz rank of the function  $\ell \circ \varphi^{-1}$ . Let

$P_{s,t}$  denote the flow associated with  $u$  and let

$$(6.2.6) \quad E = \left\{ \int_0^T P_{\tau,T^*} \left( \widehat{f}(\tau, q(\tau; u), \nu(\tau)) - \widehat{f}(\tau, q(\tau; u), \delta_{u(\tau)}) \right) d\tau : \nu \in \mathcal{M} \right\},$$

where  $\mathcal{M}$  is the set of relaxed controls  $\nu : [0, T] \rightarrow \mathcal{P}(\mathbb{U})$  introduced in Chapter Five. The set  $E \subset T_{q(T;u)}M$  is a compact, convex set. Let  $v \in E$  be arbitrary and let  $\nu$  be the corresponding relaxed control. Let  $q^\lambda$  be the trajectory for  $\delta_{u(t)} + \lambda(\nu(t) - \delta_{u(t)})$ . As in Proposition 5.5.2 we may choose disjoint measurable sets  $A_\lambda, B_\lambda$  with  $m(A_\lambda) = (1 - \lambda)T$  and  $m(B_\lambda) = \lambda T$  and a control  $v^\lambda \in \mathcal{U}$  such that the trajectory for the control

$$(6.2.7) \quad w^\lambda(t) = \chi_{A_\lambda}(t)u(t) + \chi_{B_\lambda}(t)v^\lambda(t)$$

satisfies

$$(6.2.8) \quad \left\| \varphi(q^\lambda(T)) - \varphi(q(T; w^\lambda)) \right\| < \lambda^2.$$

Consider

$$(6.2.9) \quad \ell(q^\lambda(T)) \geq \ell(q(T; w^\lambda)) - k_\ell \left\| \varphi(q^\lambda(T)) - \varphi(q(T; w^\lambda)) \right\| \geq \ell(q(T; w^\lambda)) - k_\ell \lambda^2.$$

Further,

$$(6.2.10) \quad d(w^\lambda, u) = \int_0^T d_{\mathbb{U}}(w^\lambda(t), u(t)) dt \leq \int_{B_\lambda} d_{\mathbb{U}}(v(t), u(t)) dt \leq \lambda T \text{diam}(\mathbb{U}).$$

It follows

$$(6.2.11) \quad \begin{aligned} D\ell(q(T; u); v) &= \liminf_{\lambda \downarrow 0} \frac{\ell(q^\lambda(T)) - \ell(q(T; u))}{\lambda} \\ &\geq \liminf_{\lambda \downarrow 0} \frac{\ell(q(T; w^\lambda)) - \ell(q(T; u))}{\lambda} - \lim_{\lambda \downarrow 0} \frac{k_\ell \lambda^2}{\lambda} \\ &= \liminf_{\lambda \downarrow 0} \frac{\ell(q(T; w^\lambda)) - \ell(q(T; u)) + \Delta d(w^\lambda, u) - \Delta d(w^\lambda, u)}{\lambda}. \end{aligned}$$

Note that  $\ell(q(T; w^\lambda)) = v(q_0, w^\lambda)$ . We have

$$(6.2.12) \quad D\ell(q(T; u); v) = \liminf_{\lambda \downarrow 0} \frac{v(q_0, w^\lambda) \Delta d(w^\lambda, u) - v(q_0, u) - \Delta d(w^\lambda, u)}{\lambda}.$$

Since  $u$  is optimal for the functional  $w \mapsto v(q_0, w) + \Delta d(u, w)$  and

$$(6.2.13) \quad d(w^\lambda, u) = \int_0^T d_{\mathbb{U}}(w^\lambda(t), u(t)) dt \leq T \operatorname{diam}(\mathbb{U})$$

we obtain

$$(6.2.14) \quad \begin{aligned} D\ell(q(T; u); v) &\geq \liminf_{\lambda \downarrow 0} \frac{-\Delta d(w^\lambda, u)}{\lambda} \\ &\geq \liminf_{\lambda \downarrow 0} \frac{-\Delta \lambda T \operatorname{diam}(\mathbb{U})}{\lambda} = -\Delta T \operatorname{diam}(\mathbb{U}). \end{aligned}$$

By Proposition 5.1.3 we may now choose  $-\zeta_T \in \partial_L \ell(q(T; u))$  such that for any  $v \in E$  there holds  $\langle -\zeta_T, v \rangle \geq -\Delta T \operatorname{diam}(\mathbb{U})$ . Define a map  $\zeta(t) : [t_0, T] \rightarrow T^*M$  through  $\zeta(t) = P_{t,T}^* \zeta$ , where  $P$  is the flow associated with control  $u$ . By Proposition 3.3.5,  $\zeta(t)$  is a solution to the adjoint equations (6.2.5) and for any relaxed control  $\nu$  we have

$$(6.2.15) \quad \int_0^T \left\langle \zeta(t), \widehat{f}(t, q(T; u), \delta_{u(t)}) - \widehat{f}(t, q(T; u), \nu(t)) \right\rangle dt \geq -\Delta \operatorname{diam}(\mathbb{U}).$$

By Filippov's lemma, we may take a control  $u^{\max}$  for which

$$(6.2.16) \quad H(t, \zeta(t), u^{\max}(t)) = \max_{u \in \mathbb{U}} H(t, \zeta(t), u)$$

for almost all  $t$ . Taking  $\nu(t) = \delta_{u^{\max}(t)}$  in (6.2.15) we obtain (6.2.4).  $\square$

The uniqueness of the feedback control, along with the approximate maximum principle, gives us the following useful convergence result:

**PROPOSITION 6.2.3.** *Suppose we are given sequences  $w_n \in \mathcal{U}$ ,  $\xi_0^n \rightarrow \xi_0 \in T^*M$ , and  $\Delta_n \downarrow 0$ . Define curves  $\xi^n : [0, T] \rightarrow T^*M$  through  $\dot{\xi}^n(t) = \overrightarrow{H}(t, \xi^n(t), w_n(t))$*

and  $\xi^n(0) = \xi_0^n$ . Suppose that for each  $n$ ,

$$(6.2.17) \quad \int_0^T H(t, \xi^n(t), w_n(t)) - \max_{u \in \mathbb{U}} H(t, \xi^n(t), u) dt \geq -\Delta_n \text{diam}(\mathbb{U}).$$

If  $u_n$  is the control  $u_n(t) := u(t, \xi^n(t))$  then we have  $\lim_{n \rightarrow \infty} d(w_n, u_n) = 0$ .

PROOF. Choose a subsequence for which

$$(6.2.18) \quad \lim_{k \rightarrow \infty} d(w_{n_k}, u_{n_k}) = \limsup_{n \rightarrow \infty} d(w_n, u_n)$$

and fix  $\varepsilon > 0$ . We claim that  $\lim_{n_k \rightarrow \infty} d(w_{n_k}, u_{n_k}) \leq \varepsilon$ . Since the sequence  $d(w_{n_k}, u_{n_k})$  converges it will suffice to prove this for a subsequence.

For each  $n$  introduce sets

$$(6.2.19) \quad I_n = \left\{ t : |H(t, \xi^n(t), w_n(t)) - \max_{u \in \mathbb{U}} H(t, \xi^n(t), u)| \geq \sqrt{\Delta_n} \right\}$$

and  $J_n = [0, T] \setminus I_n$ . Inequality (6.2.17) implies that

$$(6.2.20) \quad m(I_n) \leq \frac{1}{\sqrt{\Delta_n}} \int_0^T \max_{u \in \mathbb{U}} H(t, \xi^n(t), u) - H(t, \xi^n(t), w_n(t)) dt \leq \sqrt{\Delta_n} \text{diam}(\mathbb{U}).$$

Passing to a subsequence, we may therefore assume that  $\cup_k I_{n_k}$  has measure at most  $\varepsilon/\text{diam}(\mathbb{U})$ . Let  $I = \cup_k I_{n_k}$  and  $J = [0, T] \setminus I = \cap_k J_{n_k}$ .

Now,

$$(6.2.21) \quad \begin{aligned} d(w_{n_k}, u_{n_k}) &= \int_I d_{\mathbb{U}}(w_{n_k}(t), u(t, \xi^{n_k}(t))) dt + \int_J d_{\mathbb{U}}(w_{n_k}(t), u(t, \xi^{n_k}(t))) dt \\ &\leq \varepsilon + \int_J d_{\mathbb{U}}(w_{n_k}(t), u(t, \xi^{n_k}(t))) dt. \end{aligned}$$

We claim that, for any fixed  $t_* \in J$ , there holds  $\lim_{k \rightarrow \infty} d_{\mathbb{U}}(w_{n_k}(t_*), u(t_*, \xi^{n_k}(t_*))) = 0$ .

Indeed, suppose this is not the case. Choose  $\varepsilon_0 > 0$  and a subsequence such that  $d_{\mathbb{U}}(w_{n_k}(t_*), u(t_*, \xi^{n_k}(t_*))) \geq \varepsilon_0$  for all  $k$ . We prove below, in Proposition 6.2.4, that we may pass to a subsequence for which  $\xi^{n_k}(t_*) \rightarrow \xi$  for some  $\xi \in T^*M$ . Likewise,

because  $\mathbb{U}$  is compact, we may assume that  $w_{n_k}(t_*) \rightarrow w \in \mathbb{U}$ . But now because  $t_* \in J$  we see that  $H(t_*, \xi, w) = \max_{u \in \mathbb{U}} H(t_*, \xi, u)$  and this forces  $w = u(t_*, \xi)$ , which is a contradiction. It follows that  $d_{\mathbb{U}}(w_{n_k}(t), u(t, \xi^{n_k}(t))) \rightarrow 0$  for all  $t \in J$ .

Applying the Lebesgue dominated convergence theorem in (6.2.21) we then obtain

$$(6.2.22) \quad \limsup_{n \rightarrow \infty} d(w_n, u_n) = \lim_{k \rightarrow \infty} d(w_{n_k}, u_{n_k}) \leq \varepsilon.$$

This holds for any  $\varepsilon > 0$  and so completes the proof.  $\square$

**PROPOSITION 6.2.4.** *Given a sequence of controls  $w_n \in \mathcal{U}$ , points  $q_n \rightarrow q_0$ , and covectors  $-\xi_0^n \in \partial_F v(q_n; w_n)$  which satisfy  $\xi_0^n \rightarrow \xi_0$  in the topology of  $T^*M$ . Define a family of arcs  $\xi^n(t)$  satisfying  $\dot{\xi}^n(t) = \vec{H}(t, \xi^n(t), w_n(t))$  and possessing initial conditions  $\xi_0^n$ . The sequence  $\{\xi^n\}$  admits a uniformly convergent subsequence of the maps  $\xi^n(t)$ .*

**PROOF.** Choose a partition  $\pi$  and open set  $\mathcal{O}_\pi$  as in Proposition 6.1.2. For each  $n$ , the trajectory  $q^n$  for  $w_n$  takes an interval  $[t_i, t_{i+1}]$  into one of the sets  $\mathcal{C}_{i_n}$ . Since there are only finitely many intervals and finitely many sets  $\mathcal{C}_i$ , we first pass to a subsequence such that each  $q^n$  takes each interval  $[t_i, t_{i+1}]$  to the *same*  $\mathcal{C}_{j_i}$ .

One may check that in the local coordinates for  $\mathcal{C}_{i_0}$  the sequence  $\xi^n(t)$  is equicontinuous. Since the maps  $q \mapsto v(q; w)$  are locally Lipschitz with uniformly bounded rank, the sequence  $\xi^n(t)$  is bounded in local coordinates. By the Arzelá-Ascoli theorem we may pass to a subsequence which converges uniformly on  $[t_0, t_1]$ . Likewise we may then use local coordinates on  $\mathcal{C}_{i_1}$  and pass to a subsequence which converges uniformly for  $t \in [t_1, t_2]$ . Continuing in this way we may choose a subsequence which converges uniformly on  $[0, T]$ .  $\square$

**6.2.2. Generic Existence and Uniqueness.** We are now in a position to present a generalization of the existence and uniqueness result given in [62].

THEOREM 6.2.5. *Suppose that control system  $f$  satisfies Assumption 6.2.1 and that the cost function  $\ell$  is  $C^1$ . Then for each  $-\zeta_0 \in \partial_F v(q_0)$  there exists a unique optimal control  $u^0$  for the Mayer problem. Moreover, if  $\zeta : [0, T] \rightarrow T^*M$  is a solution to (6.2.2) with  $\zeta(0) = \zeta_0$  then control  $u^0$  satisfies the feedback law  $u^0 = u(t, \zeta(t))$ .*

PROOF. Let  $\varepsilon_n \downarrow 0$  be an arbitrary sequence and for each  $n$  choose a control  $u_n \in \mathcal{U}(q_0; \varepsilon_n)$ . By Proposition 6.1.7 we can choose sequences  $\delta_n \downarrow 0$ ,  $q_n \rightarrow q_0$ ,  $-\zeta^n \in \partial_F v(q_n; u_n)$  such that  $\zeta^n \rightarrow \zeta_0$  and such that

$$(6.2.23) \quad \begin{aligned} v(q_n, u_n) &\leq v(q_0, u_n) + C_0 \delta_n \leq v(q_0) + \varepsilon_n + C_0 \delta_n \\ &\leq v(q_n) + C_0 \delta_n + C_0 \delta_n + \varepsilon_n. \end{aligned}$$

Set  $\Delta_n = 2C_0 \delta_n + \varepsilon_n$  so that  $u_n \in \mathcal{U}(q_n; \Delta_n)$ .

Control  $u_n$  is a  $\Delta_n$ -minimizer of the function  $\tilde{u} \mapsto v(q_n; \tilde{u})$ . By the Ekeland variational principle, we can find  $w_n \in \mathcal{U}(q_n; \Delta_n)$  with  $d(w_n, u_n) < \sqrt{\Delta_n}$  and such that the function

$$(6.2.24) \quad \tilde{u} \mapsto v(q_n; \tilde{u}) + \sqrt{\Delta_n} d(\tilde{u}, w_n)$$

attains a minimum over  $\mathcal{U}$  at  $w_n$ . By Proposition 6.2.2 the solution  $\xi^n$  to  $\dot{\xi}^n(t) = \overrightarrow{H}(t, \xi^n(t), w_n(t))$  with  $-\xi^n(T) = d\ell(q(T); q_n, w_n)$  satisfies

$$(6.2.25) \quad \int_0^T H(t, \xi(t), w_n(t)) - \max_{u \in \mathbb{U}} H(t, \xi^n(t), u) dt \geq -\sqrt{\Delta_n} \text{diam}(\mathbb{U}).$$

Let  $u_n^f(t) := u(t, \xi^n(t))$ . By Proposition 6.2.3 we have  $\lim_{n \rightarrow \infty} d(w_n, u_n^f) = 0$ . It is shown below in Proposition 6.4.3 that the sequences  $\xi^n$  converge uniformly to  $\zeta$ , the solution to (6.2.2) satisfying  $\zeta(0) = \zeta_0$ .

Set  $u^0(t) := u(t, \zeta(t))$  and note that

$$(6.2.26) \quad v(q_0, u^0) = \lim_{n \rightarrow \infty} v(q_n, u_n^f) = \lim_{n \rightarrow \infty} v(q_n, w_n) = \lim_{n \rightarrow \infty} v(q_n, u_n) = v(q_0).$$

It follows that the control  $u^0(t) := u(t, \zeta(t))$  is an optimal control.

Further, if  $\tilde{u}$  is any other optimal control then we may take the sequence  $u_n$  to be  $u_n := \tilde{u}$  for all  $n$  and in doing so obtain

$$(6.2.27) \quad d(\tilde{u}, u^0) = \lim_{n \rightarrow \infty} d(u_n, u^0) = 0,$$

implying that  $u^0$  is the unique optimal control for  $-\zeta_0 \in \partial_F v(q_0)$ . By Proposition 6.1.6,  $\zeta_0$  is the only element of  $\partial_F v(q_0)$  and so optimal trajectories are also unique.  $\square$

**PROPOSITION 6.2.6.** *If  $f$  satisfies Assumption 6.2.1 and the cost function  $\ell$  is  $C^1$  then the set of points  $q \in M$  for which a unique optimal control exists has full measure in  $M$ . That is, if  $(\mathcal{O}, \varphi)$  is any coordinate chart then the image of the complement of this set under  $\varphi$  has measure zero in  $\mathbb{R}^n$ .*

**PROOF.** Since  $v$  is locally Lipschitz,  $\partial_F v$  is empty only on a set of measure zero.  $\square$

### 6.3. Existence and Uniqueness for Bolza Problems

In this section apply Theorem 6.2.5 to the study of the following problem of Bolza: *Minimize the cost*

$$(6.3.1) \quad \ell(q(T; q, u)) + \int_0^T L(t, q(T; q, u), u(t)) dt$$

*subject to* (6.0.15). Let us suppose that  $L$  is locally integrable Lipschitz and locally integrable bounded. That is, for any  $q \in M$  there exists a coordinate chart  $(\mathcal{O}, \varphi)$  and  $L^1$  functions  $k_\varphi, m_\varphi$  such that for almost all  $t$ , for all  $x, y \in \varphi(\mathcal{O})$  we have

$$(6.3.2) \quad \begin{aligned} \|(\varphi_*L)(t, x, u) - (\varphi_*L)(t, y, u)\|_{\mathbb{R}^n} &\leq k_\varphi(t) \|x - y\|_{\mathbb{R}^n} \\ \|(\varphi_*L)_x(t, x, u) - (\varphi_*L)_x(t, y, u)\|_{\mathbb{R}^n} &\leq k_\varphi(t) \|x - y\|_{\mathbb{R}^n} \end{aligned}$$

along with

$$(6.3.3) \quad \|(\varphi_*L)(t, x, u)\|_{\mathbb{R}^n} \leq m_\varphi(t) \quad \|(\varphi_*L)_x(t, x, u)\|_{\mathbb{R}^n} \leq m_\varphi(t),$$

where the function  $\varphi_*L$  is defined through  $\varphi_*L(t, x, u) = L(t, \varphi^{-1}(x), u)$ .

The Hamiltonian associated with this problem is given by

$$(6.3.4) \quad H(t, \zeta, u) = \langle \zeta, f(t, q, u) \rangle - L(t, q, u).$$

We make the following assumptions on this problem:

ASSUMPTION 6.3.1. Let control system  $f$  satisfy Assumption 6.0.3 as well as local estimates (6.0.20) and (6.0.21). Suppose in addition that the associated Hamiltonian  $H : [0, T] \times T^*M \times \mathbb{U}$  is such that

(i) For any  $\zeta \in T^*M$ , there is a unique  $u(t, \zeta)$  for which

$$(6.3.5) \quad H(t, \zeta, u(t, \zeta)) = \max_{u \in \mathbb{U}} H(t, \zeta, u)$$

holds;

(ii) The differential equation

$$(6.3.6) \quad \dot{\zeta}(t) = \vec{H}(t, \zeta(t), u(t, \zeta(t))) \quad \zeta(0) = \zeta_0$$

admits unique solutions for any  $\zeta_0 \in T^*M$ .



We reduce this to a Mayer problem using a standard technique. Introduce a variable  $z$  and consider control system  $g$  defined on  $\widetilde{M} := M \times \mathbb{R}$  defined by

$$(6.3.7) \quad g(t, q, z, u) = \begin{pmatrix} f(t, q, u) \\ L(t, q, u) \end{pmatrix}.$$

Fix initial condition  $(q_0, 0)$ . The Bolza problem is now equivalent to the Mayer problem of minimizing

$$(6.3.8) \quad \widetilde{\ell}(q(T; u), z(T; u)) ::= \ell(q(T; u)) + z(T; u).$$

Note that the exterior derivative of  $\widetilde{\ell}$  is

$$(6.3.9) \quad d\widetilde{\ell} = d\ell + dz.$$

Thus if  $(z, p)$  denote global coordinates on  $T^*\mathbb{R} \cong \mathbb{R}^2$ , then  $H_g : T^*M \times T^*\mathbb{R} \rightarrow \mathbb{R}$  corresponding to (6.3.7) is given by

$$(6.3.10) \quad H_g(t, q, z, \zeta, p) = \langle \zeta, f(t, q, u) \rangle + pL(t, q, u).$$

If  $(\zeta, p) : [0, T] \rightarrow T^*M \times T^*\mathbb{R}$  is defined through

$$(6.3.11) \quad \left( \dot{\zeta}(t), \dot{p}(t) \right) = \overrightarrow{H}_g(t, \zeta(t), p(t), u(t))$$

then we have  $\dot{p} = \frac{\partial H_g}{\partial z}$  and hence if  $(\zeta(T), p(T)) = -d\widetilde{\ell}$  then  $p(t) = -1$  for all  $t$ .

It follows that if for any  $\zeta \in T^*M$  there is a unique  $u(t, \zeta)$  which maximizes

$$(6.3.12) \quad \widetilde{H}_g(t, q, \zeta) := \langle \zeta, f(t, q, u) \rangle - L(t, q, u)$$

then Theorem 6.2.5 may be applied to the control system  $g$ , thus proving the following:

THEOREM 6.3.2. *Suppose that Assumption 6.3.1 is satisfied and let*

$$(6.3.13) \quad v(q) = \inf \left\{ \ell(q(T; q, u)) + \int_0^T L(t, q(t; q, u), u(t)) dt : u \in \mathcal{U} \right\}.$$

*For each  $-\zeta_0 \in \partial_F v(q_0)$  there is a unique optimal control  $u^0$  for the associated Bolza problem.*

In the same manner as before (Proposition 6.2.6) we obtain

PROPOSITION 6.3.3. *If  $f$  satisfies Assumption 6.3.1 then the set of points  $q \in M$  for which a unique optimal control exists has full measure in  $M$ . That is, if  $(\mathcal{O}, \varphi)$  is any coordinate chart then the image of the complement of this set under  $\varphi$  has measure zero in  $\mathbb{R}^n$ .*

#### 6.4. Convergence Theorems

Finally, we turn our attention to a careful proof of the convergence results used in the proceeding sections. Before proving our main convergence result, we remind the reader of two closely related results for  $\mathbb{R}^n$ . Proofs of similar theorems can be found, for example, in [49]. We include a proof of the first for completeness.

Given a continuous map  $x : [0, T] \rightarrow \mathbb{R}^d$  let us agree to write

$$(6.4.1) \quad \|x\|_C := \max_{t \in [0, T]} \|x(t)\|_{\mathbb{R}^d}.$$

LEMMA 6.4.1. *Let  $g : [0, T] \times \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^d$  be a Carathéodory function. Suppose there is a feedback control  $u : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^m$  which is measurable in  $t$  and continuous in  $x$  and for which the function  $G(t, x) = g(t, x, u(t, x))$  admits unique solutions. Suppose that  $x_0^n \rightarrow x_0$  and  $x^n : [0, T] \rightarrow \mathbb{R}^d$  is defined through*

$$(6.4.2) \quad \dot{x}^n(t) = g(t, x^n(t), w_n(t))$$

for a sequence  $w_n \in \mathcal{U}$ . Define a sequence of controls  $u_n \in \mathcal{U}$  through  $u_n(t) = u(t, x^n(t))$  and suppose that  $d(w_n, u_n) \rightarrow 0$ . If there are nonnegative functions  $a(t), b(t)$  such that

$$(6.4.3) \quad \|g(t, x^n(t), w_n(t))\| \leq a(t) + \|x^n(t)\| b(t)$$

for almost all  $t$  and for all  $n$ , then the sequence  $x^n$  converges uniformly to an absolutely continuous mapping  $x : [0, T] \rightarrow \mathbb{R}^d$  which satisfies  $\dot{x}(t) = g(t, x(t), u(t, x(t)))$ .

PROOF. First we claim that the sequence  $x^n$  is uniformly bounded and equicontinuous. To see this, let  $n$  be arbitrary and note that

$$(6.4.4) \quad \|x^n(t)\| \leq \|x_0^n\| + \int_0^t a(\tau) + \|x^n(\tau)\| b(\tau) d\tau.$$

By Gronwall's lemma we find

$$(6.4.5) \quad \max_{t \in [0, T]} \|x^n(t)\| \leq (\|x_0^n\| + \|a\|_{L^1}) e^{\|b\|_{L^1}}.$$

Since  $x_0^n$  is a convergent sequence we see that  $x^n$  is uniformly bounded and (6.4.3) implies the family is equicontinuous.

We claim next that

$$(6.4.6) \quad \limsup_{n \rightarrow \infty} \|x^n(t) - x(t)\|_C = 0.$$

Choose a subsequence of the  $x^n$  which attains this limsup. By the Arzelá-Ascoli theorem we may pass to a subsequence which converges uniformly to a continuous function  $y : [0, T] \rightarrow \mathbb{R}^d$ . We further pass to a subsequence for which  $w_n(t) - u(t, x^n(t)) \rightarrow 0$  pointwise for almost all  $t$ .

Let  $K$  be a compact set containing the images of the  $x^n(t)$  and define  $\Omega = K \times \mathbb{U}$ . Because  $\Omega$  is compact, for any fixed time  $t$ , the continuous function  $(x, u) \mapsto g(t, x, u)$

is uniformly continuous on  $\Omega$ . Therefore

$$(6.4.7) \quad \lim_{n \rightarrow \infty} g(t, x^n(t), w_n(t)) - g(t, x^n(t), u(t, x^n(t))) = 0.$$

We consider now

$$(6.4.8) \quad \begin{aligned} y(t) &= x_0 + \lim_{n \rightarrow \infty} \int_0^t g(\tau, x^n(\tau), w_n(\tau)) d\tau \\ &= x_0 + \lim_{n \rightarrow \infty} \int_0^t g(\tau, x^n(\tau), u(\tau, x^n(\tau))) d\tau \\ &\quad + \lim_{n \rightarrow \infty} \int_0^t g(\tau, x^n(\tau), w_n(\tau)) - g(\tau, x^n(\tau), u(\tau, x^n(\tau))) d\tau \\ &= x_0 + \int_0^t g(\tau, y(\tau), u(\tau, y(\tau))) d\tau \end{aligned}$$

where we have used Lebesgue dominated convergence theorem for the last equality.

This proves that  $y(t) = x(t)$  and since  $x^n(t) \rightarrow y(t)$  uniformly we have established (6.4.6).  $\square$

In the same way one can prove the following:

LEMMA 6.4.2. *Let  $g : [0, T] \times \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^d$  be a Carathéodory function. Suppose we are given initial conditions  $x^n(0)$  and  $y^n(0)$  satisfying  $\lim x^n(0) = \lim y^n(0)$  and sequences  $u_n, w_n \in \mathcal{U}$  for which  $d(w_n, u_n) \rightarrow 0$ . Define arcs  $x^n, y^n : [0, T] \rightarrow \mathbb{R}^d$  through*

$$(6.4.9) \quad \dot{x}^n(t) = g(t, x^n(t), w_n(t)) \quad \dot{y}^n(t) = g(t, y^n(t), u_n(t)).$$

*If there are nonnegative functions  $a(t), b(t)$  such that*

$$(6.4.10) \quad \|g(t, x^n(t), w_n(t))\| \leq a(t) + \|x^n(t)\| b(t)$$

*and*

$$(6.4.11) \quad \|g(t, y^n(t), u_n(t))\| \leq a(t) + \|y^n(t)\| b(t)$$

for almost all  $t$  and for all  $n$  then we have

$$(6.4.12) \quad \lim_{n \rightarrow \infty} \|x^n(t) - y^n(t)\|_C = 0.$$

With these results we turn to our main convergence theorem:

PROPOSITION 6.4.3. *Suppose that we have sequences  $u_n, w_n \in \mathcal{U}$  for which  $\lim_{n \rightarrow \infty} d(u_n, w_n) = 0$  and a sequence  $q_n \rightarrow q_0 \in M$  for which  $\lim_{n \rightarrow \infty} dv(q_n, u_n) = \zeta_0 \in T_{q_0}^*M$ . Define a sequence of arcs  $\xi^n : [0, T] \rightarrow T^*M$  through*

$$(6.4.13) \quad \dot{\xi}^n(t) = \overrightarrow{H}(t, \xi^n(t), w_n(t)),$$

with initial condition  $-\xi^n(0) = dv(q_n, w_n)$  and define controls  $u_n^{\text{feedback}} := u(t, \xi^n(t))$ .

If  $\lim_{n \rightarrow \infty} d(w_n, u_n^{\text{feedback}}) = 0$  then  $\xi^n$  converges uniformly to  $\zeta$ , the solution to

$$(6.4.14) \quad \dot{\zeta}(t) = \overrightarrow{H}(t, \zeta(t), u(t, \zeta(t)))$$

with initial condition  $-\zeta_0$ .

PROOF. We first claim that the conditions  $\lim_{n \rightarrow \infty} d(u_n, w_n) = 0$  and  $\lim_{n \rightarrow \infty} dv(q_n, u_n) = \zeta_0 \in T_{q_0}^*M$  together imply that  $\lim_{n \rightarrow \infty} dv(q_n, w_n) = \zeta_0 \in T_{q_0}^*M$ . Indeed we have

$$(6.4.15) \quad dv(q_n, w_n) = P_{0,T}^{n*} d\ell(q(T; q_n, w_n)),$$

where  $P_{0,T}^n$  is the flow associated with  $w_n$ . For each index  $n$ ,  $q(T; q_n, w_n)$  lies in the domain for some chart  $\varphi_{j_n}$ . By Proposition 6.1.1 we may apply Lemma 6.4.2 finitely many times to find that for  $n$  sufficiently large  $n$ ,  $q(T; q_n, u_n)$  lies in the same domain and that for any of the (finitely many possible) coordinate charts  $\varphi_{j_n}$  we have

$$(6.4.16) \quad \lim_{n \rightarrow \infty} \|\varphi_{j_n}(q(T; q_n, u_n)) - \varphi_{j_n}(q(T; q_n, w_n))\|_{\mathbb{R}^n} = 0.$$

Thus, in any of the finitely many local coordinates on  $T^*M$  induced by the maps  $\varphi_j$  we will have

$$(6.4.17) \quad \lim_{n \rightarrow \infty} \varphi_{j_n}^{-1*} d\ell(q(T; q_n, u_n)) - \varphi_{j_n}^{-1*} d\ell(q(T; q_n, w_n)) = 0.$$

Again we may apply Lemma 6.4.2 finitely many times to find that

$$(6.4.18) \quad \lim_{n \rightarrow \infty} Q_{0,T}^{n*} d\ell(q(T; q_n, u_n)) - P_{0,T}^{n*} d\ell(q(T; q_n, w_n)) = 0,$$

where  $Q_{s,t}^n$  is the flow associated with  $u_n$  and  $P_{s,t}^n$  is the flow associated with  $w_n$ . Since  $Q_{0,T}^{n*} d\ell(q(T; q_n, u_n)) = dv(q_n, u_n)$  we find that

$$(6.4.19) \quad \lim_{n \rightarrow \infty} P_{0,T}^{n*} d\ell(q(T; q_n, w_n)) = \lim_{n \rightarrow \infty} dv(q_n, u_n) = \zeta_0.$$

Now by Lemma 6.4.1 the Proposition is true for  $t \in [0, t_1]$ . It then holds inductively on  $[t_1, t_2]$  and so on, and this completes the proof.  $\square$

## CHAPTER 7

### Differential Inclusions

In this final chapter we turn our attention to problems of dynamic optimization on manifolds which are subject to a dynamic constraint of the form

$$(7.0.20) \quad \dot{q}(t) \in F(t, q(t)),$$

where  $F : [0, T] \times M \rightrightarrows TM$  is a set-valued map. Such differential inclusions arise naturally in problems of pure mathematics. An obvious example is the geodesic problem, in which one may take

$$(7.0.21) \quad F(t, q) = \left\{ v \in T_q M : \|v\|_g \leq 1 \right\}.$$

More generally, if  $\mathcal{H}$  is a subriemannian distribution on  $M$  with metric  $g$  then a set-valued map of interest is given by

$$(7.0.22) \quad F(t, q) = \left\{ v \in \mathcal{H}_q : \|v\|_g \leq 1 \right\}.$$

Differential inclusions also arise naturally optimal control problems which are subject to mixed constraints of the form

$$g_j(t, q(t), u(t)) \leq 0 \quad 1 \leq j \leq r.$$

Early studies of such problems in the context of optimal control include [37].

In this chapter we establish necessary optimality conditions for differential inclusions on manifolds, including a version of our exact penalization result in Chapter

Five. The problem addressed in this chapter is the following. Fix an initial condition  $q_0 \in M$ , let  $\ell : M \rightarrow \mathbb{R}$  be a locally Lipschitz function and  $S \subset M$  a closed set. Suppose that  $q^0 : [0, T] \rightarrow M$  minimizes  $\ell(q^0(T))$  over absolutely continuous curves  $q : [0, T] \rightarrow M$  which satisfy

$$(7.0.23) \quad \dot{q}(t) \in F(t, q(t)),$$

$q(0) = q_0$  and  $q(T) \in S$ . The main goal of this chapter is to prove the following necessary condition:

**THEOREM 7.0.4.** *Suppose that  $q^0 : [0, T] \rightarrow M$  provides a local minimum for the Mayer problem described above and that  $F$  satisfies Assumption 7.2.1. Then there is a  $\lambda^0 \in \{0, 1\}$  and an arc  $\zeta : [0, T] \rightarrow T^*M$  which satisfies  $-\zeta(T) \in \lambda^0 \partial_L \ell(q^0(T)) + N_S^L(q^0(T))$  as well as*

$$(7.0.24) \quad \dot{\zeta}(t) \in \overrightarrow{H}(t, \zeta(t)).$$

The set-valued map  $\overrightarrow{H} : [0, T] \times T^*M \rightarrow TT^*M$  is a generalization of the Hamiltonian lift and is described below, as will Assumption 7.2.1. In the process we will also prove the following theorem:

**THEOREM 7.0.5.** *Suppose  $q^0 : [0, T] \rightarrow M$  provides a local minimum for the Mayer problem described above and  $F$  satisfies Assumption 7.2.1. If there are no solutions to (7.0.24) satisfying  $-\zeta(T) \in N_S^L(q(T))$  with  $\zeta(T) \neq 0$  then there exists a constant  $K > 0$  such that  $q^0$  provides a local minimum to the free terminal point Mayer problem of minimizing*

$$(7.0.25) \quad q \mapsto \ell(q) + Kd(q),$$

where  $d : M \rightarrow \mathbb{R}$  is the locally defined function  $d_{\theta(S)} \circ \theta$  for a coordinate chart  $\theta$  whose domain includes  $q^0(T)$ .



Theorem 7.0.5 is new in  $\mathbb{R}^n$  and so this chapter begins by studying the case  $M = \mathbb{R}^n$ . We then use the method of Lagrangian charts to transfer our results to manifolds.

Our work in this chapter relies on tools made use of by Clarke in [23]. In this section we provide the three main background results needed for the following sections. We recommend the book of Clarke, Ledyaev, Stern, and Wolenski [28] or the classic [23] for a broader introduction to this material.

Because  $L^2([0, T], \mathbb{R}^n)$  is a smooth Banach space in which the multidirectional mean value inequality<sup>1</sup> holds, we formulate our results for absolutely continuous mappings whose derivative is in  $L^2$ . Our problems have a fixed initial condition  $x_0$ . Given an element  $v$  of  $L^2$  we define

$$(7.0.26) \quad x(t; v) = x_0 + \int_0^t v(\tau) d\tau.$$

We will have need of the following convergence result, a version of which can be found in [28].

**PROPOSITION 7.0.6.** *Suppose that  $\Gamma : [0, T] \times \mathbb{R}^n \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is a set-valued map with convex, compact values, is measurable in  $t$ , and has a support function*

$$(7.0.27) \quad \sigma_\Gamma(t, \alpha, \beta, p) = \sup_{v \in \Gamma(t, \alpha, \beta)} \langle p, v \rangle$$

*which is continuous in  $(\alpha, \beta)$ . Suppose that we have sequences of  $L^2$  functions  $\alpha^k, \beta^k, \gamma^k, e^k$  such that  $\alpha^k \rightarrow \alpha$  pointwise,  $\beta^k \rightarrow \beta$  pointwise,  $e^k \xrightarrow{L^2} 0$ , and for almost all  $t$*

$$(7.0.28) \quad \gamma^k(t) + e^k(t) \in \Gamma(t, \alpha^k(t), \beta^k(t)).$$

---

<sup>1</sup>See Theorem 5.1.4

Suppose also that there is an  $L^1$  function  $m_\Gamma$  such that for all  $k$ , for almost all  $t$

$$(7.0.29) \quad \Gamma(t, \alpha^k(t), \beta^k(t)) \subset m_\Gamma(t)\mathbb{B}.$$

Then if  $\gamma^k \xrightarrow{w} \gamma$  we have  $\gamma(t) \in \Gamma(t, \alpha(t), \beta(t))$  for almost all  $t$ .

PROOF. Let  $\{p_n\}_{n=1}^\infty$  be a countable dense subset of  $\mathbb{R}^n$ . In order to show that  $\gamma(t) \in \Gamma(t, \alpha(t), \beta(t))$  for a particular time  $t$  it will suffice to show that

$$(7.0.30) \quad \langle p_n, \gamma(t) \rangle \leq \sigma_\Gamma(t, \alpha(t), \beta(t), p_n)$$

for any  $n$ .

Fix an integer  $n$  and let  $\mathcal{T}_n \subset [0, T]$  denote the collection of times  $t_0$  which are Lebesgue points for both  $\gamma(t)$  and for the function  $t \mapsto \sigma_\Gamma(t, \alpha(t), \beta(t), p_n)$ . Given  $t_0 \in \mathcal{T}_n$  and  $\delta > 0$  consider the map  $p_n^\delta(t)$  defined by

$$(7.0.31) \quad p_n^\delta(t) = \begin{cases} p_n & t \in [t_0, t_0 + \delta] \\ 0 & \text{elsewhere} \end{cases}.$$

Equation (7.0.28) implies that, for each  $k$ ,

$$(7.0.32) \quad \int_0^T \langle p_n^\delta(t), \gamma^k(t) + e^k(t) \rangle dt \leq \int_0^T \sigma_\Gamma(t, \alpha^k(t), \beta^k(t), p_n^\delta(t)) dt.$$

The integrand on the right is dominated by integrable function  $m_F(t) \|p_n\|$  and so by the Lebesgue theorem we obtain in the limit

$$(7.0.33) \quad \int_0^T \langle p_n^\delta(t), \gamma(t) \rangle dt \leq \int_0^T \sigma_\Gamma(t, \alpha(t), \beta(t), p_n^\delta(t)) dt.$$

That is,

$$(7.0.34) \quad \int_{t_0}^{t_0+\delta} \langle p_n, \gamma(t) \rangle dt \leq \int_{t_0}^{t_0+\delta} \sigma_\Gamma(t, \alpha(t), \beta(t), p_n) dt.$$

Dividing by  $\delta > 0$  and taking the limit as  $\delta \downarrow 0$  we find that for any  $t_0 \in \mathcal{T}_n$ ,

$$(7.0.35) \quad \langle p_n, \gamma(t_0) \rangle \leq \sigma_\Gamma(t_0, \alpha(t_0), \beta(t_0), p_n)$$

The set  $\mathcal{T} = \bigcap_{n=1}^\infty \mathcal{T}_n \subset [0, T]$  has full measure and for any  $t_0 \in \mathcal{T}$  the inequality (7.0.35) holds for any  $n$ , completing the proof.  $\square$

We also require a necessary condition from the nonsmooth calculus of variations. Suppose that  $L : [0, T] \times \mathbb{R}^n \times \mathbb{R}^n$  satisfies the following hypotheses:

- (i) For fixed  $x, v$  the map  $t \mapsto L(t, x, v)$  is measurable;
- (ii) For any continuous map  $x : [0, T] \rightarrow \mathbb{R}^n$  there exists  $\varepsilon > 0$  and an  $L^2$  function  $k_L$  such that for any  $(t, y_1)$  and  $(t, y_2)$  in the set

$$(7.0.36) \quad \Omega := \{(t, y) : \|y - x(t)\| < \varepsilon\}$$

there holds

$$(7.0.37) \quad |L(t, y_1, v) - L(t, y_2, v)| \leq k_L(t) \|y_1 - y_2\|;$$

- (iii)  $L$  is globally Lipschitz rank  $k_v$  in  $v$ .

Define a functional  $J : L^2 \rightarrow \mathbb{R}$  by

$$(7.0.38) \quad J(v) = \int_0^T L(t, x(t; v), v(t)) dt.$$

One can check that  $J$  is a locally Lipschitz functional. Its Clarke subgradient will play a central role in what follows and we provide the following characterization.

**PROPOSITION 7.0.7.** *Suppose that  $\zeta \in \partial_C J(v)$ . Then there exists a selection  $(\xi(t), \eta(t)) \in \partial_C L(t, x(t; v), v(t))$  such that*

$$(7.0.39) \quad \zeta(t) = \eta(t) + \int_t^T \xi(\tau) d\tau.$$

Finally, we have need of Filippov's approximation lemma for differential inclusions. This lemma first appeared in [44] and is proved, under slightly more general hypotheses, in [23]. Suppose that  $F : [0, T] \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is measurable in  $t$  and takes on nonempty, closed values. Suppose also that for any given continuous map  $x : [0, T] \rightarrow \mathbb{R}^n$  there exists  $\delta > 0$  and  $L^2$  functions  $k_F$  and  $m_F$  such that for all  $t, y_1, y_2$  which satisfy  $\|x(t) - y_i\| < \delta$  we have

$$(7.0.40) \quad F(t, y_1) \subset F(t, y_2) + k_F(t) \|y_1 - y_2\| \mathbb{B}$$

and

$$(7.0.41) \quad F(t, y_1) \subset m_F(t) \mathbb{B}.$$

Fix a particular map  $x^0$  and let  $\delta, k_F, m_F$  be as above.

LEMMA 7.0.8 (Filippov). *If  $x : [0, T] \rightarrow \mathbb{R}^n$  is absolutely continuous and satisfies  $\|x^0 - x\|_C < \frac{1}{2}\delta$  and*

$$(7.0.42) \quad \int_0^T d(F(t, x(t)), \dot{x}(t)) < \frac{1}{2}\delta e^{-\|k_F\|_{L^1}}$$

*then there exists an absolutely continuous function  $y : [0, T] \rightarrow \mathbb{R}^n$  with  $x_0 = y_0$  and  $\dot{y}(t) \in F(t, y(t))$  such that*

$$(7.0.43) \quad \|\dot{x}(t) - \dot{y}(t)\|_{L^1} \leq e^{\|k_F\|_{L^1}} \int_0^T d(F(t, x(t)), \dot{x}(t)) dt.$$

## 7.1. Necessary Optimality Conditions and Exact Penalization in $\mathbb{R}^n$

We now turn to a study of necessary optimality conditions for a differential inclusion of the form

$$(7.1.1) \quad \dot{x}(t) \in F(t, x(t))$$

with fixed initial condition  $x(0) = x_0$  and terminal constraint  $x(T) \in S \subset \mathbb{R}^n$ . We will formulate our problems in the space  $L^2([0, T], \mathbb{R}^n)$  and so (7.1.1) becomes the requirement that

$$(7.1.2) \quad v(t) \in F(t, x(t; v))$$

for almost all  $t \in [0, T]$ .

Throughout this section we make the following assumptions on the set-valued map  $F : [0, T] \times M \rightarrow TM$ .

ASSUMPTION 7.1.1. The map  $F : [0, T] \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  is measurable in  $t$  and takes on nonempty, convex, compact values. In addition:

- (i)  $F$  is locally integrable Lipschitz: For a given continuous map  $x : [0, T] \rightarrow \mathbb{R}^n$  there exists  $\varepsilon > 0$  and an  $L^2$  function  $k_F$  such that for almost all  $t$ , for any  $y_1, y_2$  satisfying  $\|x(t) - y_i\| < \varepsilon$  there holds

$$(7.1.3) \quad F(t, y_1) \subset F(t, y_2) + k_F(t) \|y_1 - y_2\| \mathbb{B};$$

- (ii)  $F$  is integrable bounded: For a given continuous map  $x : [0, T] \rightarrow \mathbb{R}^n$  there exists  $\varepsilon > 0$  and an  $L^2$  function  $m_F$  such that for almost all  $t$ , for any  $y$  satisfying  $\|x(t) - y\| < \varepsilon$ , there holds

$$(7.1.4) \quad F(t, y) \subset m_F(t) \mathbb{B}.$$

Notice that because of (7.1.4), *any* absolutely continuous curve satisfying  $\dot{x}(t) \in F(t, x(t))$  will satisfy  $\dot{x}(t) \in L^2$  and so we lose nothing by working in this space. We will make use of a penalty function of the form

$$(7.1.5) \quad v \mapsto \int_0^T d(F(t, x(t; v)), v(t)) dt.$$

The reader may check that the function  $L(t, x, v) := d(F(t, x), v)$  satisfies the assumptions needed for Proposition 7.0.7. Following Clarke's own route, detailed in [23], we study the perturbed problem

$$(7.1.6) \quad \dot{x}(t) \in F(t, x(t)) + \alpha \bar{\mathbb{B}}.$$

A proof of the following can be found in [23], page 127.

**PROPOSITION 7.1.2.** *If  $(\xi, \eta) \in \partial_C Kd(F(t, x) + \alpha \bar{\mathbb{B}}, v)$  and  $v \in F(t, x) + \alpha \bar{\mathbb{B}}$  then  $(-\xi, v) \in \partial_C H(t, x, \eta) + \alpha \bar{\mathbb{B}}$ .*

Our approach is analogous to that taken in Chapter Five and again relies on the Ekeland principle. We therefore turn to the metric spaces we will use in the study of this problem. These metric spaces are designed for the study of a particular map  $v^0(t) \in L^2$  and while we will later take  $v^0$  to be a local minimizer for a functional for the moment we consider any fixed  $v^0 \in L^2$  satisfying for almost all  $t$  the inclusion  $v^0(t) \in F(t, x^0(t))$ . Here  $x^0$  is the trajectory for  $v^0$ . Let  $\varepsilon_*$  be chosen so that (7.1.3) and (7.1.4) hold for  $\|y - x^0(t)\| \leq \varepsilon_*$  and define

$$(7.1.7) \quad \mathcal{X}_\alpha = \{v \in L^2 : v(t) \in F(t, x(t; v)) + \alpha \bar{\mathbb{B}} \text{ a.a. } t \text{ and } \|v - v^0\|_{L^1} \leq \varepsilon_*\}.$$

Note the use of the  $L^1$  norm, which has been chosen for compatibility with the Filippov lemma. Note also that the particular map  $v^0$  is not explicit in the notation. We have chosen to write  $\mathcal{X}_\alpha$  instead of  $\mathcal{X}_\alpha(v^0, \varepsilon_*)$  to keep the notation compact. We define a metric  $d$  on  $\mathcal{X}_\alpha$  by

$$(7.1.8) \quad d(v, w) = \|v - w\|_{L^1}$$

and we define  $\mathcal{A}_\alpha = \{v(t) \in \mathcal{X}_\alpha : x(t; v) \in S\}$ .

PROPOSITION 7.1.3. *The space  $\mathcal{X}_\alpha$  is a complete metric space and  $\mathcal{A}_\alpha$  is a closed subset.*

PROOF. Let  $v^n$  be a Cauchy sequence in  $\mathcal{X}_\alpha$ . Then  $v^n$  is Cauchy in  $L^1$ -norm and must converge to an  $L^1$  function  $v$  with  $\|v - v^0\|_{L^1} \leq \varepsilon_*$ . Since  $F$  is integrable bounded, we have  $\|v^n\|_{L^2} \leq \|m_F\|_{L^2}$  for all  $n$  and so we must have  $v \in L^2$ . Passing to a sequence  $v^k$  which converges pointwise to  $v(t)$ , the continuity of  $F$  in  $x$  implies that  $v(t) \in F(t, x(t; v)) + \alpha \overline{\mathbb{B}}$  for almost all  $t$  and thus  $v \in \mathcal{X}_\alpha$ . The proof that  $\mathcal{A}_\alpha$  is closed is similar.  $\square$

We define the distance from  $v$  to  $\mathcal{A}_\alpha$  through the formula

$$(7.1.9) \quad d_{\mathcal{A}_\alpha}(v) = \inf_{w \in \mathcal{A}_\alpha} d(v, w).$$

We next prove this map is lower semicontinuous in  $\alpha$  at  $\alpha = 0$ .

PROPOSITION 7.1.4. *For any  $v \in \mathcal{X}$  there holds*

$$(7.1.10) \quad \liminf_{\alpha \downarrow 0} d_{\mathcal{A}_\alpha}(v) \geq d_{\mathcal{A}_0}(v).$$

PROOF. Fix  $v \in \mathcal{X}$  and choose a sequence  $\alpha_k \downarrow 0$  for which

$$(7.1.11) \quad \lim_{k \rightarrow \infty} d_{\mathcal{A}_{\alpha_k}}(v) = \liminf_{\alpha \downarrow 0} d_{\mathcal{A}_\alpha}(v).$$

For each  $k$ , choose  $w^k \in \mathcal{A}_{\alpha_k}$  such that  $d(v, w^k) < d_{\mathcal{A}_{\alpha_k}}(v) + 1/k$ . The sequence  $w^k$  is bounded in  $L^2$  norm by  $\|m_F\|_{L^2}$  and so we may assume that  $w^k \xrightarrow{w} w$  for some  $w \in L^2$ . Choose  $p \in L^\infty$  with  $\|p\|_{L^\infty} = 1$  for which

$$(7.1.12) \quad d(v, w) = \int_0^T \langle p(t), v(t) - w(t) \rangle dt$$

and note that

$$(7.1.13) \quad d(v, w) = \lim_{k \rightarrow \infty} \int_0^T \langle p(t), v(t) - w(t)^k \rangle dt \leq \liminf_{k \rightarrow \infty} d(v, w^k).$$

If we can now establish that  $w \in \mathcal{A}_0$  then we will have

$$(7.1.14) \quad d_{\mathcal{A}_0}(v) \leq d(v, w) \leq \liminf_{k \rightarrow \infty} d(v, w^k) = \liminf_{\alpha \downarrow 0} d_{\mathcal{A}_\alpha}(v)$$

and this will complete the proof.

We prove that  $w \in \mathcal{A}_0$ . Notice that the trajectories  $x^k$  corresponding to  $w^k$  converge uniformly. Since each  $x^k(T) \in S$  we have  $x(T; w) \in S$ . It remains only to show that  $w(t) \in F(t, x(t; w))$  for almost all  $t$ . This is implied by Proposition 7.0.6 and the standing assumptions on  $F$ .  $\square$

**7.1.1. Metric Regularity.** In this section we introduce a constraint qualification which ensures the metric regularity of this constraint  $x(T) \in S$  and, as in Chapter Five, we prove that failure of our constraint qualification corresponds to a kind of abnormality. The constraint qualification follows:

**Condition C:** There exist  $\varepsilon_0 \in (0, \varepsilon_*)$  and  $\alpha_0, \Delta_0 > 0$  such that for any  $\alpha \in (0, \alpha_0)$ , for any  $v \in \mathcal{X}_\alpha$  satisfying  $d(v, v^0) < \varepsilon_0$  and  $d_S(x(T; v)) > 0$ , and any  $\zeta \in L^2$  with  $\|\zeta\|_{L^2} \leq \Delta_0$ , there is no absolutely continuous map  $p : [0, T] \rightarrow \mathbb{R}^n$  which satisfies  $-p(T) \in \partial_L d_S(x(T; v))$  and

$$(7.1.15) \quad (-\dot{p}(t), v(t)) \in \partial_C H(t, x(t; v), p(t) + \zeta(t)) + \alpha \bar{\mathbb{B}}.$$

If Condition  $C$  holds we obtain a decrease principle analogous to that of Proposition 5.3.9. Set

$$(7.1.16) \quad C_F := \left( \frac{\Delta_0}{2T^{1/2}} + 1 \right) e^{\|k_F\|_{L^1}}$$



and define for  $\alpha \geq 0$  a functional  $\Phi_\alpha : L^2 \rightarrow \mathbb{R}$  through

$$(7.1.17) \quad \Phi_\alpha(v) = d_S(x(T; v)) + C_F \int_0^T d(F(t, x(t; v)) + \alpha \bar{\mathbb{B}}, v(t)) dt.$$

The decrease principle will be given in terms of the *sequential weak lower Dini derivative* introduced by Clarke and Ledyaev in [26]:

$$(7.1.18) \quad D^w \Phi(v; w) = \inf_{\substack{\lambda_n \downarrow 0 \\ w^n \xrightarrow{w} w}} \lim_{n \rightarrow \infty} \frac{\Phi(v + \lambda_n w^n) - \Phi(v)}{\lambda_n}.$$

This lower derivative is useful for studying decrease of nonlinear functionals defined on smooth Banach spaces. It is shown in [26] that if  $E \subset L^2$  is a closed, bounded, and convex set and

$$(7.1.19) \quad D^w \Phi_\alpha(v; w) > -\Delta_0$$

for all  $w \in E$ , then for any  $\varepsilon > 0$  there exists  $v^\varepsilon$  with  $\|v^\varepsilon - v\|_{L^2} < \varepsilon$  and  $\zeta^\varepsilon \in \partial_P \Phi_\alpha(v^\varepsilon)$ , each of which is such that for any  $w \in E$  there holds

$$(7.1.20) \quad \langle \zeta^\varepsilon, w \rangle_{L^2} > -\Delta_0.$$

In fact their result is even more general, but what we have written is enough to establish the following decreasing principle, which must hold whenever Condition C does:

**PROPOSITION 7.1.5.** *Suppose that Condition C holds at  $v^0$ . Then for any  $\alpha \in (0, \alpha_0)$  and  $v \in \mathcal{X}_\alpha$  satisfying  $d(v, v^0) < \varepsilon_0$  and  $x(t; v) \notin S$  there exists  $w \in \mathbb{B}_{L^2}$  such that*

$$(7.1.21) \quad D^w \Phi_\alpha(v; w) \leq -\Delta_0.$$

PROOF. If (7.1.21) does not hold then there exists  $\alpha \in (0, \alpha_0)$  and  $v \in \mathcal{X}_\alpha$  such that

$$(7.1.22) \quad D^w \Phi_\alpha(v; w) > -\Delta_0$$

for all  $w \in \mathbb{B}_{L^2}$ . Fix this value of  $\alpha$  and choose a sequence  $v^k \xrightarrow{L^2} v$  along with proximal subgradients  $\zeta^k \in \partial_P \Phi_\alpha(v^k)$  such that for any  $w \in \mathbb{B}_{L^2}$  we have

$$(7.1.23) \quad \langle \zeta^k, w \rangle > -\Delta_0.$$

Note that each  $\zeta^k$  has norm bounded by  $\Delta_0$ . Passing to a weakly convergent subsequence we find that there exists  $\zeta \in \partial_L \Phi_\alpha(v)$  with  $\|\zeta\|_{L^2} \leq \Delta_0$ . Define functionals  $J_0, J_1 : L^2 \rightarrow \mathbb{R}$  by

$$(7.1.24) \quad J_0(v) = d_S(x(t; v)) \quad J_1(v) = C_F \int_0^T d(F(t, x(t; v)) + \alpha \bar{\mathbb{B}}, v(t)) dt$$

we must have  $\zeta = \zeta^0 + \zeta^1$  for some  $\zeta^i \in \partial_L J_i(v)$ ,  $i = 0, 1$ . We characterize  $\zeta^0, \zeta^1$ , beginning with  $\zeta^0$ .

For  $\zeta^0$ , we recall (see [28]) that there exists an element  $-p(T) \in \partial_L d_S(x(T; v))$  for which  $\zeta^0(t) = -p(T)$  for almost all  $t$ . To characterize  $\zeta^1$ , set

$$(7.1.25) \quad L(t, x(t; v), v(t)) = C_F d(F(t, x(t; v)) + \alpha \bar{\mathbb{B}}, v(t)).$$

Proposition 7.0.7 implies the existence of a selection

$$(7.1.26) \quad (\xi(t), \eta(t)) \in \partial_C C_F d(F(t, x(t; v)) + \alpha \bar{\mathbb{B}}, v(t))$$

for which

$$(7.1.27) \quad \zeta^1(t) = \eta(t) + \int_t^T \xi(\tau) d\tau.$$

By Proposition 7.1.2 we then have

$$(7.1.28) \quad (-\xi(t), v(t)) \in \partial_C H(t, x(t; v), \eta(t)) + \alpha \bar{\mathbb{B}}.$$

Define

$$(7.1.29) \quad p(t) = p(T) - \int_t^T \xi(\tau) d\tau$$

so that  $\dot{p}(t) = \xi(t)$  and note that

$$(7.1.30) \quad p(t) + \zeta(t) = p(T) - \int_t^T \xi(\tau) d\tau + \zeta^0(T) + \zeta^1(t) = \eta(t)$$

for almost all  $t$ . Hence  $p$  satisfies

$$(7.1.31) \quad (-\dot{p}(t), v(t)) \in \partial_C H(t, x(t; v), p(t) + \zeta(t)) + \alpha \bar{\mathbb{B}}$$

for almost all  $t$ . This contradicts condition  $C$  and so the weak decrease principle (7.1.21) must hold.  $\square$

As in Chapter Five, the weak decrease principle can be used to establish a metric regularity-type result:

PROPOSITION 7.1.6. *If Condition C holds, then there exists  $\varepsilon_1 > 0$  such that for any  $v \in \mathcal{X}_0$  with  $d(v, v^0) < \varepsilon_1$  we have*

$$(7.1.32) \quad d_{\mathcal{A}_0}(v) \leq \frac{4T^{1/2}}{\Delta_0} d_S(x(T; v)).$$

PROOF. Choose  $0 < \varepsilon_1 < \frac{1}{2}\varepsilon_0$  small enough that if  $d(v, v^0) < \varepsilon_1$  then  $\frac{4T^{1/2}}{\Delta_0} d_S(x(T; v)) < \frac{1}{2}\varepsilon_0$ . Suppose that for some  $\bar{v} \in \mathcal{X}_0$  with  $d(\bar{v}, v^0) < \varepsilon_1$  there holds

$$(7.1.33) \quad d_{\mathcal{A}_0}(\bar{v}) > \frac{4T^{1/2}}{\Delta_0} d_S(x(T; \bar{v})).$$

Since  $\liminf_{\alpha \downarrow 0} d_{\mathcal{A}_\alpha}(\bar{v}) \geq d_{\mathcal{A}_0}(\bar{v})$  we can pick some  $\alpha \in (0, \alpha_0)$  such that

$$(7.1.34) \quad d_{\mathcal{A}_\alpha}(\bar{v}) > \frac{4T^{1/2}}{\Delta_0} d_S(x(T; \bar{v})).$$

Now set

$$(7.1.35) \quad \varepsilon = 2d_S(x(T; \bar{v})) \quad \sigma = \frac{4T^{1/2}}{\Delta_0} d_S(x(T; \bar{v}))$$

and notice that  $\bar{v}$  is an  $\varepsilon$ -minimizer over  $\mathcal{X}_\alpha$  of the functional  $w \mapsto d_S(x(T; w))$ . By the Ekeland variational principle, we may therefore pick  $v \in \mathcal{X}_\alpha$  with  $d(v, \bar{v}) < \sigma$  which minimizes the functional

$$(7.1.36) \quad w \mapsto d_S(x(T; w)) + \frac{\varepsilon}{\sigma} d(w, v)$$

over  $\mathcal{X}_\alpha$ . Check that  $\varepsilon/\sigma = \Delta_0/(2T^{1/2})$ . Notice that

$$(7.1.37) \quad d(v, v^0) \leq d(v, \bar{v}) + d(\bar{v}, v^0) < \sigma + \varepsilon_1 < \varepsilon_0.$$

Further

$$(7.1.38) \quad d(\bar{v}, v) < \sigma = \frac{4T^{1/2}}{\Delta_0} d_S(x(T; \bar{v})) < d_{\mathcal{A}_\alpha}(\bar{v}).$$

Therefore  $v \notin \mathcal{A}_\alpha$  and so the requirements for Proposition 7.1.5 are met at  $v$ .

Now suppose that  $w \in \mathbb{B}_{L^2}$  is arbitrary. For each  $\lambda > 0$  let  $x^\lambda$  be the trajectory for  $v + \lambda w$ . Note that

$$(7.1.39) \quad \int_0^T d(F(t, x^\lambda(t)) + \alpha \bar{\mathbb{B}}, v(t) + \lambda w(t)) dt \leq \lambda \int_0^T \|w(t)\| dt \leq \lambda T^{1/2}.$$

It follows that there exists  $\lambda_0$  independent of  $w \in \mathbb{B}_{L^2}$  such that for each  $\lambda \in (0, \lambda_0)$  we may choose, through Filippov's approximation lemma, a trajectory  $y^\lambda \in \mathcal{X}_\alpha$  for

which

$$(7.1.40) \quad \left\| \dot{x}^\lambda - \dot{y}^\lambda \right\|_{L^1} \leq e^{\|k_F\|_{L^1}} \int_0^T d(F(t, x^\lambda(t)) + \alpha \bar{\mathbb{B}}, \dot{x}^\lambda(t)) dt.$$

Recalling that  $C_F = \left( \frac{\Delta_0}{2T^{1/2}} + 1 \right) e^{\|k_F\|_{L^1}}$ , we now have

$$(7.1.41) \quad \begin{aligned} \Phi_\alpha(v + \lambda w) + \frac{\Delta_0}{2T^{1/2}} \left\| \dot{x}^\lambda - v \right\|_{L^1} &= d_S(x^\lambda(T)) + C_F \int_0^T d(F(t, x^\lambda(t)) + \alpha \bar{\mathbb{B}}, v(t) + \lambda w(t)) dt + \frac{\Delta_0}{2T^{1/2}} \left\| \dot{x}^\lambda - v \right\|_{L^1} \\ &\geq d_S(y^\lambda(T)) - \left\| \dot{x}^\lambda - \dot{y}^\lambda \right\|_{L^1} + C_F \int_0^T d(F(t, x^\lambda(t)) + \alpha \bar{\mathbb{B}}, \dot{x}^\lambda(t)) dt \\ &\quad + \frac{\Delta_0}{2T^{1/2}} \left\| \dot{y}^\lambda - v \right\|_{L^1} - \frac{\Delta_0}{2T^{1/2}} \left\| \dot{y}^\lambda - \dot{x}^\lambda \right\|_{L^1} \\ &\geq d_S(y^\lambda(T)) + \frac{\Delta_0}{2T^{1/2}} d(\dot{y}^\lambda, v) \geq d_S(x(T; v)) = \Phi_\alpha(v). \end{aligned}$$

The Cauchy-Schwarz inequality implies that  $\|w\|_{L^1} \leq T^{1/2}$  and so

$$(7.1.42) \quad \frac{\Delta_0}{2T^{1/2}} \left\| \dot{x}^\lambda - v \right\|_{L^1} = \frac{\Delta_0}{2T^{1/2}} \lambda \|w\|_{L^1} \leq \frac{\Delta_0}{2} \lambda.$$

Therefore, for any  $w \in \mathbb{B}_{L^2}$  and  $0 < \lambda < \lambda_0$  we have

$$(7.1.43) \quad \frac{\Phi_\alpha(v + \lambda w) - \Phi_\alpha(v)}{\lambda} \geq -\frac{\Delta_0}{2}.$$

But this implies that  $D^w \Phi_\alpha(v; w) \geq -\Delta_0/2$  for all  $w \in \mathbb{B}_{L^2}$ , contradicting Proposition 7.1.5. Thus (7.1.32) must hold.  $\square$

**7.1.2. Optimality Conditions.** In this section we apply Propositions 7.1.11 and 7.1.6 to establish necessary optimality conditions for the constrained Mayer problem described above. Consider the functional

$$(7.1.44) \quad v \mapsto \ell(x(T; v)).$$

Let  $v^0 \in \mathcal{X}_0$  provide a local minimum for the functional (7.1.44) subject to the constraint  $x(T; v) \in S$ .

PROPOSITION 7.1.7. *If Condition C holds at  $v^0$  then  $v^0$  is a local minimum in  $\mathcal{X}_0$  for the functional*

$$(7.1.45) \quad w \mapsto \ell(x(T; w)) + Kd_S(x(T; w)),$$

for  $K = k_\ell \frac{4T^{1/2}}{\Delta_0}$ , where  $k_\ell$  is the local Lipschitz rank for  $\ell$ .

PROOF. By Proposition 7.1.6 there exists  $\varepsilon_1 > 0$  such that for any  $v \in \mathcal{X}_0$  with  $d(v, v^0) < \varepsilon_1$  there holds

$$(7.1.46) \quad d_{\mathcal{A}_0}(v) \leq \frac{4T^{1/2}}{\Delta_0} d_S(x(T; v)).$$

Suppose that  $d(v, v^0) < \varepsilon_1$ . Given any  $\sigma > 0$  choose  $w \in \mathcal{A}_0$  such that  $d(v, w) < d_{\mathcal{A}_0}(v) + \sigma$ . Consider

$$(7.1.47) \quad \ell(x(T; v)) \geq \ell(x(T; w)) - k_\ell \|x(T; w) - x(T; v)\|_{\mathbb{R}^n} \geq \ell(x^0(T)) - k_\ell d(v, w)$$

because  $v^0$  is optimal in  $\mathcal{A}_0$ . Thus we have

$$(7.1.48) \quad \ell(x(T; v)) \geq \ell(x^0(T)) - k_\ell d_{\mathcal{A}_0}(v) - k_\ell \sigma$$

and so

$$(7.1.49) \quad \ell(x(T; v)) + k_\ell \frac{4T^{1/2}}{\Delta_0} d_S(x(t; v)) \geq \ell(x^0(T)) - k_\ell \sigma.$$

Letting  $\sigma \downarrow 0$  completes the proof.  $\square$

Using Filippov's approximation lemma, we may apply a standard technique to remove the dynamic constraint and so reduce our problem to one of nonsmooth calculus of variations.

PROPOSITION 7.1.8. *Suppose  $v^0$  is locally optimal for the functional (7.1.44) and that Condition C holds. Then  $v^0$  is a local (in  $L^1$  norm) minimizer for the functional*

$$(7.1.50) \quad w \mapsto \ell(x(T; w)) + K_1 d_S(x(T; w)) + K_2 \int_0^T d(F(t, x(t; w)), w(t)) dt,$$

*for  $K_1, K_2$  sufficiently large. Since the  $L^2$  norm dominates the  $L^1$  norm,  $v^0$  is also a local minimizer in  $L^2$  norm.*

PROOF. We have already seen that there exists  $\varepsilon_1 > 0$  such that  $v^0$  is optimal among  $v \in \mathcal{X}$  satisfying  $d(v, v^0) < \varepsilon_1$  for the functional

$$(7.1.51) \quad w \mapsto \ell(x(T; w)) + K d_S(x(T; w)),$$

provided that  $K$  is sufficiently large. Let  $\varepsilon_2 > 0$  be defined by

$$(7.1.52) \quad \varepsilon_2 = \min \left\{ \frac{\varepsilon_1}{2}, \frac{\varepsilon_1}{2e^{\|k_F\|_{L^1}} (\|k_F\|_{L^1} + 1)} \right\}$$

and suppose that  $w \in L^2$  satisfies

$$(7.1.53) \quad \|v^0 - w\|_{L^1} < \varepsilon_2.$$

By Filippov's approximation lemma we may choose  $u \in \mathcal{X}_0$  with

$$(7.1.54) \quad \|u - w\|_{L^1} \leq e^{\|k_F\|_{L^1}} \int_0^T d(F(t, x(t; w)), w(t)) dt.$$

We also have the bounds

$$\begin{aligned}
(7.1.55) \quad & \int_0^T d(F(t, x(t; w)), w(t)) dt \\
& \leq \int_0^T d(F(t, x^0(t)), v^0(t)) + k_F(t) \|x^0(t) - x(t; w)\| + \|v^0(t) - w(t)\| dt \\
& \leq (\|k_F\|_{L^1} + 1) \|v^0 - w\|_{L^1}.
\end{aligned}$$

As a consequence,

$$\begin{aligned}
(7.1.56) \quad & d(u, v^0) \leq d(v^0, w) + d(w, u) \\
& \leq \frac{\varepsilon_1}{2} + e^{\|k_F\|_{L^1}} (\|k_F\|_{L^1} + 1) \|v^0 - w\|_{L^1} d(v^0, w) \\
& < \varepsilon_1.
\end{aligned}$$

The local optimality of  $v^0$  then provides us with

$$\begin{aligned}
(7.1.57) \quad & \ell(x(T; w)) + Kd_S(x(T; w)) + (K + k_\ell)e^{\|k_F\|_{L^1}} \int_0^T d(F(t, x(t; w)), w(t)) dt \\
& \geq \ell(x(T; u)) + Kd_S(x(T; u)) - (K + k_\ell) \|u - w\|_{L^1} \\
& \quad + (K + k_\ell)e^{\|k_F\|_{L^1}} \int_0^T d(F(t, x(t; w)), w(t)) dt \\
& \geq \ell(x^0(T)).
\end{aligned}$$

Therefore  $v^0$  is optimal for the unconstrained problem among all  $w \in L^2$  with  $\|w - v^0\|_{L^1} < \varepsilon_1$ .  $\square$

We have shown that Condition  $C$  implies that all constraints may be lifted from the problem through a process of exact penalization and we are now in a position to prove the following theorem originally due to Clarke [**23**, **31**]:



**THEOREM 7.1.9.** *Suppose that  $v^0$  is a local minimizer for the problem of minimizing  $\ell(x(T; v))$  subject to  $v(t) \in F(t, x(t; v))$  and  $x(T; v) \in S$ . We assume that  $\ell$  is locally Lipschitz and that  $F$  satisfies Assumption 7.1.1.*

*There exists  $\lambda^0 \in \{0, 1\}$  and absolutely continuous map  $p : [0, T] \rightarrow \mathbb{R}^n$  for which  $-p(T) \in \lambda^0 \partial_L \ell(x^0(T)) + N_S^L(x^0(T))$  and which satisfies*

$$(7.1.58) \quad (-\dot{p}(t), v(t)) \in \partial_C H(t, x^0(t), p(t)).$$

*Equivalently,*

$$(7.1.59) \quad (-\dot{p}(t), \dot{x}^0(t)) \in \partial_C H(t, x^0(t), p(t)).$$

*Moreover we have  $\|p(T)\| + \lambda^0 \neq 0$ .*

Following a path analogous to that in Chapter Five we prove this theorem in two steps, corresponding to the success or failure of Condition  $C$ .

**PROPOSITION 7.1.10.** *Suppose that Condition  $C$  holds. Then we obtain Theorem 7.1.9 with  $\lambda^0 = 1$ .*

**PROOF.** Let

$$(7.1.60) \quad J(\alpha, w) = \ell(x(T; w)) + K_1 d_S(x(T; w)) + K_2 \int_0^T d(F(t, x(t; w)) + \alpha \bar{\mathbb{B}}, w(t)) dt,$$

with  $K_1, K_2$  defined as in Proposition 7.1.8.

Proposition 7.1.8 provides us with  $\varepsilon > 0$  such that  $v^0$  is optimal, among  $v \in L^2$  with  $\|v - v^0\|_{L^2} < \varepsilon$ , for the functional  $v \mapsto J(0, v)$ .

Let  $E := v^0 + \varepsilon \bar{\mathbb{B}}$  and let  $\chi_E$  be the function which is zero on  $E$  and  $+\infty$  elsewhere. Let  $C = TK_2$ . Note that for  $\alpha > 0$  and  $w \in E$  we have

$$(7.1.61) \quad J(\alpha, w) \geq J(0, w) - C\alpha \geq J(0, v^0) - C\alpha = J(\alpha, v^0) - C\alpha.$$

Therefore, if  $\alpha > 0$ , then

$$(7.1.62) \quad 2C\alpha + \inf_{w \in L^2} \{J(\alpha, w) + \chi_E(w)\} > J(\alpha, v^0).$$

By Ekeland's principle we may choose  $v^\alpha \in L^2$  with

$$(7.1.63) \quad \|v^0 - v^\alpha\|_{L^2} < \sqrt{\alpha}$$

which minimizes the functional

$$(7.1.64) \quad w \mapsto J(\alpha, w) + \chi_E(w) + 2C\sqrt{\alpha} \|w - v^\alpha\|_{L^2}$$

over  $L^2$ .

Take a sequence  $\alpha_k \downarrow 0$  and choose such  $v^\alpha$  for each  $k$ , labelling it  $v^k$ . By (7.1.63) we see that when  $k$  is sufficiently large  $v^k$  will lie in the interior of  $E$  and (7.1.64) then implies that we will have

$$(7.1.65) \quad 0 \in \partial_L J(\alpha_k, v^k) - e^k,$$

where  $e^k$  is an element of  $L^2$  with norm bounded by  $2C\sqrt{\alpha_k}$ .

Define the following functionals:

$$(7.1.66) \quad J_\ell(v) = \ell(x(T; v))$$

$$(7.1.67) \quad J_d(v) = K_1 d_S(x(T; v))$$

$$(7.1.68) \quad J_F(\alpha, v) = K_2 \int_0^T d(F(t, x(t; v)) + \alpha \bar{\mathbb{B}}, v(t)) dt.$$

For each  $k$  we can find  $\zeta_\ell^k \in \partial_L J_\ell(v^k)$ ,  $\zeta_d^k \in \partial_L J_d(v^k)$ , and  $\zeta_F \in \partial_L J_F(\alpha_k, v^k)$  such that

$$(7.1.69) \quad \zeta_\ell^k + \zeta_d^k + \zeta_F^k = e^k.$$

Let  $x^k$  be the arc for  $v^k$ . There are constants  $\theta_\ell^k \in \partial_L \ell(x^k(T))$  and  $\theta_d^k \in \partial_L K_1 d_S(x^k(T))$  such that  $\zeta_\ell^k(t) = \theta_\ell^k$  and  $\zeta_d^k(t) = \theta_d^k$  for almost all  $t$ . Since these sequences are bounded, we may pass to a subsequence for which  $\theta_\ell^k \rightarrow \theta_\ell \in \partial_L \ell(x^0(T))$  and  $\theta_d^k \rightarrow \theta_d \in \partial_L K_1 d_S(x^0(T)) \subset N_S^L(x^0(T))$ . It follows from (7.1.69) that the sequence  $\zeta_F^k$  converges in  $L^2$  norm to the constant function  $-\theta_\ell - \theta_d$ .

For each  $k$  there exist selections  $(\xi^k(t), \eta^k(t)) \in \partial_C K_2 d(F(t, x^k(t)) + \alpha_k \bar{\mathbb{B}}, v^k(t))$  for which

$$(7.1.70) \quad \zeta_F^k(t) = \eta^k(t) + \int_t^T \xi^k(\tau) d\tau.$$

Since  $\alpha_k > 0$  and  $v^k \in \mathcal{X}_{\alpha_k}$  Proposition 7.1.2 implies that

$$(7.1.71) \quad \left(-\xi^k(t), v^k(t)\right) \in \partial_C H(t, x^k(t), \eta^k(t)) + \alpha_k \bar{\mathbb{B}}.$$

Setting

$$(7.1.72) \quad p^k(t) = -\theta_\ell^k - \theta_d^k - \int_t^T \xi^k(\tau) d\tau,$$

we find that

$$(7.1.73) \quad p^k(t) + e^k(t) = -\int_t^T \xi^k(\tau) d\tau + \zeta_F^k(t) = \eta^k(t).$$

Thus

$$(7.1.74) \quad \left(-p^k(t), v^k(t)\right) \in \partial_C H(t, x^k(t), p^k(t) + e^k(t)) + \alpha_k \bar{\mathbb{B}}.$$

As in the proof of Proposition 7.1.11 we may pass to a subsequence for which  $\dot{p}^k \xrightarrow{w} r$  and apply Proposition 7.0.6 to show that

$$(7.1.75) \quad \left(-r(t), v^0(t)\right) \in \partial_C H(t, x^0(t), p(t)),$$

where

$$(7.1.76) \quad p(t) = -\theta_d - \theta_\ell - \int_t^T r(\tau) d\tau.$$

Hence  $p(t)$  satisfies the differential inclusion (7.1.59). Since  $p(T) = -\theta_\ell - \theta_d$  we see that

$$(7.1.77) \quad -p(T) \in \partial_L \ell(x^0(T)) + N_S^L(x^0(T)),$$

completing the proof. □

Now consider the case in which Condition  $C$  fails.

**PROPOSITION 7.1.11.** *Suppose that Condition  $C$  fails at  $v^0$  and let  $x^0$  be the trajectory for  $v^0$ . There exists an arc  $p$  which satisfies  $-p(T) \in \partial_L d_S(x^0(T))$ ,  $\|p(T)\| = 1$ , and for almost all  $t$*

$$(7.1.78) \quad (-\dot{p}(t), v^0(t)) \in \partial_C H(t, x^0(t), p(t)).$$

**PROOF.** If Condition  $C$  fails then we can choose sequences  $\varepsilon_k \downarrow 0$ ,  $\alpha_k \downarrow 0$ , and  $\Delta_k \downarrow 0$  such that, for each  $k$ , there exists  $v^k \in \mathcal{X}_{\alpha_k}$  with  $d(v^k, v^0) < \varepsilon_k$  and  $\zeta^k \in L^2$  with  $\|\zeta^k\|_{L^2} < \Delta_k$  such that the arcs  $x^k$  for  $v^k$  satisfy  $x^k(T) \notin S$  and such that there exist arcs  $p^k$  which satisfy  $-p^k(T) \in \partial_L d_S(x^k(T))$  and

$$(7.1.79) \quad \left(-\dot{p}^k(t), v^k(t)\right) \in \partial_C H(t, x^k(t), p^k(t) + \zeta^k(t)) + \alpha_k \bar{\mathbb{B}}$$

for almost all  $t \in [0, T]$ . Since  $x^k(T) \notin S$  and  $-p(T)^k \in \partial_L d_S(x^k(T))$  we have  $\|p(T)^k\|_{\mathbb{R}^n} = 1$ .

We claim that the maps  $\dot{p}^k$  can be uniformly bounded in  $L^2$ -norm. To see this, first recall [28] that (7.1.79) implies the inequality

$$(7.1.80) \quad \left\| \dot{p}^k(t) \right\|_{\mathbb{R}^n} \leq k_F(t) \|p(t)\|_{\mathbb{R}^n} + \alpha_k.$$

As a consequence,

$$(7.1.81) \quad \begin{aligned} \left\| p^k(T-t) \right\|_{\mathbb{R}^n} &= \left\| p^k(T) + \int_0^t \dot{p}^k(T-\tau) d\tau \right\|_{\mathbb{R}^n} \\ &\leq 1 + \int_0^t k_F(T-\tau) \left\| p^k(T-\tau) \right\|_{\mathbb{R}^n} + \alpha_k d\tau. \end{aligned}$$

By Gronwall's lemma we obtain the bound

$$(7.1.82) \quad \max_{t \in [0, T]} \left\| p^k(t) \right\|_{\mathbb{R}^n} \leq (1 + \alpha_k T) e^{\|k_F\|_{L^1}}.$$

Inequality (7.1.80) now implies that  $\dot{p}^k$  is uniformly bounded in  $L^2$ -norm.

We may therefore pass to a subsequence such that  $p^k \xrightarrow{w} r$  for some  $r \in L^2$ . Since  $\zeta^k \xrightarrow{L^2} 0$  we may also assume that  $\zeta^k \rightarrow 0$  pointwise and because the  $p^k(T)$  are all norm one, we can assume that  $p^k(T) \rightarrow p(T)$  for some  $p(T)$  with norm one. Note that  $p(T)$  satisfies  $-p(T) \in \partial_L d_S(x^0(T))$ .

Setting

$$(7.1.83) \quad p(t) = p(T) - \int_t^T r(\tau) d\tau$$

and applying Proposition 7.0.6 we find that for almost all  $t$  there holds

$$(7.1.84) \quad (-\dot{p}(t), v(t)) \in \partial_C H(t, x^0(t), p(t)),$$

which completes the proof. □

Together Propositions 7.1.10 and 7.1.11 imply 7.1.9. We also have the following:

**THEOREM 7.1.12.** *Suppose there is no map  $p : [0, T] \rightarrow \mathbb{R}^n$  which satisfies*

$$(7.1.85) \quad (-\dot{p}(t), v(t)) \in \partial_C H(t, x^0(t), p(t)),$$

$-p(T) \in N_S^L(x^0(T))$  and  $p(T) \neq 0$ . Then for  $K > 0$  sufficiently large the map  $x^0$  provides a local minimum for the functional

$$(7.1.86) \quad v \mapsto \ell(x(T; v)) + Kd_S(x(T; v)).$$

PROOF. If Condition  $C$  fails then such a map exists. Thus if there are no such maps then Condition  $C$  must hold and the result is implied by Proposition 7.1.8.  $\square$

## 7.2. Necessary Optimality Conditions and Exact Penalization on General Manifolds

In this section we use the method of Lagrangian charts to establish a version of Theorem 7.1.9 which is valid for problems posed on smooth manifolds. We begin by introducing the standing assumptions on  $F$ .

ASSUMPTION 7.2.1. The map  $F : [0, T] \times M \rightrightarrows TM$  is measurable in time for each fixed  $q$  and for any  $(t, q)$  the set  $F(t, q) \subset T_q M$  is nonempty, closed, and convex. In addition, for each  $q \in M$  there exists a coordinate chart  $(\mathcal{O}, \varphi)$  and  $L^2$  functions  $m_\varphi$  and  $k_\varphi$  for which

$$(7.2.1) \quad \varphi_*(q_1)F(t, q_1) \subset \varphi_*(q_2)F(t, q_2) + k_\varphi(t) \|\varphi(q_1) - \varphi(q_2)\|_{\mathbb{R}^n} \mathbb{B}$$

and

$$(7.2.2) \quad \varphi_*(q_1)F(t, q_1) \subset m_\varphi(t) \mathbb{B}$$

for almost all  $t$ , for all  $q_1, q_2 \in \mathcal{O}$ .

Suppose that  $q^0 : [0, T] \rightarrow M$  is optimal for the problem of minimizing  $\ell(q(T))$  subject to  $q(T) \in S$  and

$$(7.2.3) \quad \dot{q}(t) \in F(t, q(t)),$$

where  $\ell$  is locally Lipschitz and  $F$  satisfies 7.2.1. By Proposition 3.2.8 we may let  $V_t$  be a  $C^\infty$ -smooth vector field, compactly supported and measurable in time, which extends  $\dot{q}^0$  in the sense that

$$(7.2.4) \quad V_t(q^0(t)) = \dot{q}^0(t).$$

Let  $P_{s,t}$  be the flow of  $V_t$ . Choose a coordinate neighborhood  $(\varphi, \mathcal{O})$  for which  $q_0 \in \mathcal{O}$  and let  $\psi_t : P_{0,t}(\mathcal{O}) \rightarrow \varphi(\mathcal{O})$  be the associated Lagrangian coordinates  $\psi_t := \varphi \circ P_{t,0}$ .

Define set-valued map  $\psi_{t*}F : \varphi(\mathcal{O}) \rightarrow \mathbb{R}^n$  through

$$(7.2.5) \quad (\psi_{t*}F)(t, x) = \psi_{t*}(\psi_t^{-1}(x))F(t, \psi_t^{-1}(x))$$

and let  $G : [0, T] \times \varphi(\mathcal{O}) \rightrightarrows \mathbb{R}^n$  be the map

$$(7.2.6) \quad G(t, x) = (\psi_{t*}F)(t, x) - (\psi_{t*}V_t)(x).$$

Suppose that  $q : [0, T] \rightarrow M$  is an absolutely continuous map which satisfies  $q(t) \in P_{0,t}(\mathcal{O})$  for all  $t$ . Setting

$$(7.2.7) \quad x(t) := \psi_t(q(t))$$

one may check that we have  $\dot{q}(t) \in F(t, q(t))$  for almost all  $t$  if and only if  $\dot{x}(t) \in G(t, x(t))$  for almost all  $t$ . Indeed this follows from the formula

$$(7.2.8) \quad \frac{d}{ds}P_{s,t}(q) = -P_{s,t*}(q)V_s(q),$$

established in Proposition 3.1.1.

The map  $G$  is behind our reduction to the  $\mathbb{R}^n$  case.

**PROPOSITION 7.2.2.** *The map  $G$  satisfies Assumption 7.1.1 in a neighborhood of  $x^0(t) := \psi_t(q^0(t)) \equiv \varphi(q_0)$ .*

PROOF. Certainly  $G$  is measurable for fixed  $t$  and takes on nonempty, compact, convex values. Partition  $[0, T]$  into intervals  $0 = t_0 < t_1 < \cdots < t_n = T$  which are small enough that for each  $k$ ,  $q^0$  takes  $[t_i, t_{i+1}]$  into a single coordinate chart  $\mathcal{O}_i$  on which (7.2.1) and (7.2.2) hold. Assume without loss of generality that  $\overline{\mathcal{O}}_i$  is compact and that  $\varphi_i$  is defined on a neighborhood of  $\overline{\mathcal{O}}_i$ . Recalling that  $(\mathcal{O}, \varphi)$  is the coordinate chart used to define the Lagrangian coordinates and that  $\psi_t^{-1}(x) = P_{0,t} \circ \varphi^{-1}(x)$ , continuous dependence on initial conditions implies that we may choose  $\varepsilon > 0$  such that for all  $x \in \varphi(\mathcal{O})$  which satisfy  $\|x - \varphi(q_0)\| < \varepsilon$  we have  $\psi_t^{-1}(x) \in \mathcal{O}_i$  for all  $t \in [t_i, t_{i+1}]$ .

Let  $t \in [t_i, t_{i+1}]$  and choose points  $y_1, y_2 \in \varphi(\mathcal{O})$  such that  $\|y_i - \varphi(q_0)\| < \varepsilon$  for  $i = 1, 2$ . Set  $q_i = \psi_t^{-1}(y_i)$ . Let us first prove that (7.1.4) holds for some function depending on  $i$ . We have

(7.2.9)

$$\begin{aligned} G(t, y_1) &= (\psi_{t*}F)(t, y_1) = \psi_{t*}(\psi_t^{-1}(y_1))F(t, \psi_t^{-1}(y_1)) \\ &= \psi_{t*}(q_1)F(t, q_1) = (\psi_t \circ \varphi_i^{-1})'(\varphi_i(q_1)) \varphi_{i*}(q_1)F(t, q_1) \subset A_i m_i(t) \mathbb{B} \end{aligned}$$

where  $A_i$  is a bound on the derivative  $(\psi_t \circ \varphi_i^{-1})'$  for  $x \in \varphi_i(\mathcal{O}_i)$  and  $t \in [t_i, t_{i+1}]$ .

This proves that  $G$  is locally integrable bounded, which is (7.1.4).

To prove  $G$  is locally integrable Lipschitz, which is (7.1.3), consider

$$\begin{aligned} (7.2.10) \quad G(t, y_1) &= \psi_{t*}(q_1)F(t, q_1) - \psi_{t*}(q_1)V_t(q_1) \\ &= (\psi_t \circ \varphi_i^{-1})'(\varphi_i(q_1)) \varphi_{i*}(q_1)F(t, q_1) - \psi_{t*}(q_1)V_t(q_1). \end{aligned}$$

Choose constants  $B_i$  and  $C_i$  such that

$$(7.2.11) \quad \|\varphi_i(q_1) - \varphi_i(q_2)\| \leq B_i \|y_1 - y_2\|$$



and

$$(7.2.12) \quad \left\| (\psi_t \circ \varphi_i^{-1})'(\varphi_i(q_1)) - (\psi_t \circ \varphi_i^{-1})'(\varphi_i(q_2)) \right\| \leq C_i \|y_1 - y_2\|.$$

Such constants exist because  $\psi_t \circ \varphi_i^{-1} = \varphi \circ P_{t,0} \circ \varphi_i^{-1}$  is the local coordinate expression for the flow of a  $C^\infty$ -smooth vector field.

By (7.2.1) we have

$$(7.2.13) \quad \varphi_{i*}(q_1)F(t, q_1) \subset \varphi_{i*}(q_2)F(t, q_2) + k_{\varphi_i}(t) \|\varphi_i(q_1) - \varphi_i(q_2)\| \mathbb{B}$$

and so

$$(7.2.14) \quad \begin{aligned} & (\psi_t \circ \varphi_i^{-1})'(\varphi_i(q_1)) \varphi_{i*}(q_1)F(t, q_1) \subset (\psi_t \circ \varphi_i^{-1})'(\varphi_i(q_1)) \varphi_{i*}(q_2)F(t, q_2) \\ & \quad + k_{\varphi_i}(t) \|\varphi_i(q_1) - \varphi_i(q_2)\| (\psi_t \circ \varphi_i^{-1})'(\varphi_i(q_1)) \mathbb{B} \\ & \subset (\psi_t \circ \varphi_i^{-1})'(\varphi_i(q_2)) \varphi_{i*}(q_2)F(t, q_2) + C_i m_{\varphi_i}(t) \|y_1 - y_2\| \mathbb{B} + B_i A_i k_{\varphi_i}(t) \|y_1 - y_2\| \mathbb{B}. \end{aligned}$$

Finally, choose a function  $C_{V,i}(t)$  for which

$$(7.2.15) \quad \|(\varphi_{i*} V_t)(\varphi_i(q_1)) - (\varphi_{i*} V_t)(\varphi_i(q_2))\| \leq C_{V,i}(t) \|\varphi_i(q_1) - \varphi_i(q_2)\| \leq B_i C_{V,i}(t) \|y_1 - y_2\|.$$

Putting things together we have

$$(7.2.16) \quad G(t, y_1) \subset G(t, y_2) + (C_i m_{\varphi_i}(t) + B_i A_i k_{\varphi_i}(t) + B_i C_{V,i}(t)) \|y_1 - y_2\| \mathbb{B}.$$

□

Thus we may reduce a problem on  $M$  to a problem on  $\mathbb{R}^n$  defined in terms of the set-valued map  $G$ .

**7.2.1. Hamiltonian Lift of the Clarke Subgradient.** Before transferring Theorem 7.1.9 to manifolds we must first interpret it geometrically and reformulate

it as a result in  $T^*\mathbb{R}^n$ . For a smooth manifold,  $H_F$  is defined as a map from  $T^*M$  into  $\mathbb{R}$ , defined for  $\zeta \in T_q^*M$  through

$$(7.2.17) \quad H_F(t, \zeta) = \max_{v \in F(t, q)} \langle \zeta, v \rangle.$$

For a fixed  $t$ , this map is locally Lipschitz and so is differentiable on a dense subset  $\mathcal{D}_t \subset T^*M$ . As in [66] one may define

$$(7.2.18) \quad \partial_C H_F(t, \zeta) = \text{co} \left\{ \lim_{i \rightarrow \infty} dH_F(t, \zeta_i) : \zeta_i \in \mathcal{D}_t, \zeta_i \rightarrow \zeta \right\} \subset T_\zeta^* T^*M.$$

Recall that there is a map  $J_M : TT^*M \rightarrow T^*T^*M$  defined through

$$(7.2.19) \quad \sigma_M(v, w) = \langle J_M v, w \rangle,$$

where  $\sigma_M$  is the natural symplectic structure on  $T^*M$ . Following [82] we make the following definition:

DEFINITION 7.2.3. The *Hamiltonian Lift* of a locally Lipschitz map  $H : [0, T] \times T^*M \rightarrow \mathbb{R}$  is the set  $\vec{H}(t, p) \subset T_p T^*M$  defined through

$$(7.2.20) \quad \vec{H}(t, \zeta) = J_M^{-1} \partial_C H(t, \zeta).$$

Since  $J_{\mathbb{R}^n}^{-1}(a, b) = (-b, a)$  we have shown in Theorem 7.1.9 that if  $x^0$  is an optimal curve for a differential inclusion in  $\mathbb{R}^n$  then there is an arc  $p$  so that  $(x^0, p)$ , interpreted as an arc in  $T^*\mathbb{R}^n \cong \mathbb{R}^n \times \mathbb{R}^n$ , satisfies

$$(7.2.21) \quad (\dot{x}^0(t), \dot{p}(t)) \in \vec{H}(t, x(t), p(t)).$$

Now let  $\Psi_t : T^*M \rightarrow T^*\mathbb{R}^n$  be the map defined, for  $\zeta \in T_q^*M$ , by

$$(7.2.22) \quad \Psi_t(\zeta) = (\psi_t(q), \psi_t^{-1*} \zeta),$$

as studied in Chapter 3.

Define maps  $H_F, H_V : [0, T] \times T^*M \rightarrow \mathbb{R}^n$  through

$$(7.2.23) \quad H_F(t, \zeta) = \max_{v \in F(t, q)} \langle \zeta, v \rangle \quad H_V(t, \zeta) = \langle \zeta, V_t(q) \rangle$$

and  $H_G : [0, T] \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  by

$$(7.2.24) \quad H_G(t, x, p) = \max_{v \in G(t, x)} \langle p, v \rangle.$$

Our next proposition, stated informally, says that the coordinates induced on  $T^*M$  by (7.2.22) take arcs for  $\vec{H}_G$  to those for  $\vec{H}_F$ . Thus our necessary conditions in  $\mathbb{R}^n$ , which are written in terms of the map  $G$ , will translate through (7.2.22) to necessary conditions on  $M$  in terms of the map  $F$ .

PROPOSITION 7.2.4. *Suppose  $(x(t), p(t)) \in \vec{H}_G(t, x(t), p(t))$ . Then the curve  $\zeta : [0, T] \rightarrow T^*M$  defined through  $\zeta(t) = \Psi_t^{-1}(x(t), p(t))$  satisfies*

$$(7.2.25) \quad \dot{\omega}(t) \in \vec{H}_F(t, \zeta(t)).$$

PROOF. Consider

$$(7.2.26) \quad \frac{d}{dt} \zeta(t) = \frac{d}{dt} \Psi_t^{-1}(x(t), p(t)) = \frac{d}{d\varepsilon} \Big|_{\varepsilon=0} \Psi_{t+\varepsilon}^{-1}(x(t), p(t)) + \frac{d}{d\varepsilon} \Big|_{\varepsilon=0} \Psi_t^{-1}(x(t+\varepsilon), p(t+\varepsilon)).$$

We evaluate each of the derivatives on the right. First, if  $Q_{s,t}$  denotes the flow on  $T^*M$  of  $\vec{H}_V$  then we have

$$(7.2.27) \quad \frac{d}{d\varepsilon} \Big|_{\varepsilon=0} \Psi_{t+\varepsilon}^{-1}(x(t), p(t)) = \frac{d}{d\varepsilon} \Big|_{\varepsilon=0} Q_{0,t+\varepsilon} \circ \theta^{-1}(x(t), p(t)) = \vec{H}_V(t, \zeta(t)).$$

For the second,

$$(7.2.28) \quad \frac{d}{d\varepsilon} \Big|_{\varepsilon=0} \Psi_t^{-1}(x(t+\varepsilon), p(t+\varepsilon)) = (\Psi_t^{-1})_*(\dot{x}(t), \dot{p}(t)) \in (\Psi_t^{-1})_* \vec{H}_G(t, x(t), p(t)).$$

Now consider

$$\begin{aligned}
(7.2.29) \quad H_G(t, x, p) &= \max_{v \in F(t, \psi_t^{-1}(x))} \langle p, \psi_{t*} v \rangle - \langle p, (\psi_{t*} V_t)(x) \rangle \\
&= H_F(t, \Psi_t^{-1}(x, p)) - H_V(t, \Psi_t^{-1}(x, p)) \\
&= (H_F - H_V) \circ \Psi_t^{-1}(x, p).
\end{aligned}$$

Since  $V_t$  is differentiable, we see that  $H_G$  is differentiable if and only if  $H_F$  is. At any point of differentiability we have

$$(7.2.30) \quad dH_G = d((H_F - H_V) \circ \Psi_t^{-1}) = \Psi_t^{-1*} dH_F - \Psi_t^{-1*} dH_V.$$

Hence

$$(7.2.31) \quad \vec{H}_G = J_{\mathbb{R}^n}^{-1} \Psi_t^{-1*} dH_F - J_{\mathbb{R}^n}^{-1} \Psi_t^{-1*} dH_V$$

By Proposition 3.4.6 we can write this as

$$(7.2.32) \quad \vec{H}_G = \Psi_{t*} \circ J_M^{-1} dH_F - \Psi_{t*} \circ J_M^{-1} dH_V$$

or

$$(7.2.33) \quad (\Psi_t^{-1})_* \vec{H}_G = \vec{H}_F - \vec{H}_V.$$

Thus we have, by (7.2.28),

$$(7.2.34) \quad \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} \Psi_t^{-1}(x(t+\varepsilon), p(t+\varepsilon)) \in \vec{H}_F(t, \Psi_t^{-1}(x(t), p(t))) - \vec{H}_V(t, \Psi_t^{-1}(x(t), p(t))).$$

Finally, adding (7.2.34) and (7.2.27), we find

$$(7.2.35) \quad \dot{\zeta}(t) \in \vec{H}_V(t, \zeta(t)) + \vec{H}_F(t, \zeta(t)) - \vec{H}_V(t, \zeta(t)) = \vec{H}_F(t, \zeta(t)),$$

which completes the proof. □

We turn to the necessary optimality conditions for the Mayer problem in question:

**THEOREM 7.2.5.** *Suppose that  $q_t^0 : [0, T] \rightarrow M$  provides a local minimum for  $\ell(q(T))$  subject to  $q(T) \in S$  and  $\dot{q}(t) \in F(t, q(t))$ , where  $\ell$  is locally Lipschitz,  $S$  is closed, and  $F$  satisfies Assumption 7.2.1. Then there is a  $\lambda^0 \in \{0, 1\}$  and an arc  $\zeta : [0, T] \rightarrow T^*M$  which satisfies  $-\zeta(T) \in \lambda^0 \partial_L \ell(q^0(T)) + N_S^L(q^0(T))$  as well as*

$$(7.2.36) \quad \dot{\zeta}(t) \in \overrightarrow{H}(t, \zeta(t)).$$

**PROOF.** We first reduce our problem to one in  $\mathbb{R}^n$  through a Lagrangian chart. Let  $(\mathcal{O}, \varphi)$  be a coordinate chart with  $q_0 \in \mathcal{O}$ . The map  $x^0(t) := \psi_t(q^0(t))$  provides a local minimum for the cost  $\tilde{\ell} := \ell \circ \psi_T^{-1}$  subject to  $\dot{x}(t) \in G(t, x(t))$  and  $x(T) \in \tilde{S} := \psi_T(S)$ . By Theorem 7.1.9 there must exist  $\lambda^0 \in \{0, 1\}$  and an arc  $p : [0, T] \rightarrow \mathbb{R}^n$  for which  $-p(T) \in \lambda^0 \partial_L \tilde{\ell}(x^0(T)) + N_{\tilde{S}}^L(x^0(T))$  for which

$$(7.2.37) \quad (-\dot{p}(t), \dot{x}^0(t)) \in \partial_C H_G(t, x(t), p(t)),$$

where

$$(7.2.38) \quad H_G(t, x, p) = \max_{v \in G(t, x)} \langle p, v \rangle.$$

Consider the map  $(x(t), p(t))$  as a map into  $T^*\mathbb{R}^n \cong \mathbb{R}^n \times \mathbb{R}^n$  in order to see that

$$(7.2.39) \quad (\dot{x}^0(t), \dot{p}(t)) \in J_{\mathbb{R}^n}^{-1} \partial_C H_G(t, x^0(t), p(t)) = \overrightarrow{H}(t, x^0(t), p(t)).$$

Define a map  $\zeta : [0, T] \rightarrow T^*M$  through

$$(7.2.40) \quad \zeta = \Psi_t^{-1}(x^0(t), p(t)).$$

Notice that, since  $\zeta(T) = \psi_T^* p(T)$ , we have

$$(7.2.41) \quad -\zeta(T) \in \psi_T^* \left( \lambda^0 \partial_L \tilde{\ell}(x^0(T)) + N_S^L(x^0(T)) \right).$$

Because the subdifferential is invariant under  $C^1$ -diffeomorphism [66] we have

$$(7.2.42) \quad \psi_T^* \partial_L (\ell \circ \psi_T^{-1})(\psi_T(q)) = \partial_L \ell(q).$$

Similarly we obtain

$$(7.2.43) \quad \psi_T^* N_S^L(x^0(T)) = N_S^L(q^0(T)).$$

Thus  $\omega$  satisfies the condition  $-\zeta(T) \in \lambda^0 \partial_L \ell(q^0(T)) + N_S^L(q^0(T))$  and there remains only to show that  $\zeta$  satisfies the Hamiltonian inclusion. This has been done in Proposition 7.2.4 and so the proof is complete.  $\square$

We conclude with a proof of Theorem 7.0.5.

PROOF. If there are no solutions to (7.0.24) satisfying  $-\zeta(T) \in N_S^L(q(T))$  with  $\zeta(T) \neq 0$  then there are no solutions to

$$(7.2.44) \quad (\dot{x}(t), \dot{p}(t)) \in \overrightarrow{H}_G(t, x(t), p(t))$$

satisfying  $-p(T) \in N_{\psi_T(S)}^L(x(T))$  and  $p(T) \neq 0$ , for then

$$(7.2.45) \quad \zeta(t) := \Psi_t^{-1}(x(t), p(t))$$

will be a solution to  $\dot{\zeta} \in \overrightarrow{H}_F(t, \zeta)$  with  $-\zeta(T) \in N_S^L(q(T))$  and  $\zeta(T) \neq 0$ , where  $q(t) := \pi(\zeta(t))$ .

By Proposition 7.1.11, Condition  $C$  must be satisfied for the problem in  $\mathbb{R}^n$  corresponding to cost  $\ell \circ \psi_T^{-1}$ , terminal constraint set  $\psi_T(S)$ , and set-valued map  $G$  given by (7.2.6). As a consequence, Proposition 7.1.7 implies that  $x^0(t) := \psi_t(q^0(t))$

is an unconstrained local minimizer for the cost

$$(7.2.46) \quad \ell \circ \psi_T^{-1}(x) + Kd_{\psi_T(S)}(x),$$

$K$  sufficiently large. Consequently  $q^0$  is an unconstrained local minimizer for the cost

$$(7.2.47) \quad \ell(q) + Kd(q),$$

where  $d$  is the locally defined penalty function  $d_{\psi_T(S)} \circ \psi_T$ . By Proposition 5.5.6, we may replace  $d$  with the penalty function  $d_{\theta(S)} \circ \theta$  (possibly with a corresponding increase in  $K$ ). □

## Bibliography

- [1] R. A. Adams. *Sobolev Spaces*. Academic Press, New York, NY, 1975.
- [2] A. Agrachev and R. Gamkrelidze. Exponential representation of flows and a chronological calculus. *Mat. Sb. (N.S.)*, 149, 1978.
- [3] A. Agrachev and R. Gamkrelidze. Chronological algebras and nonstationary vector fields. *Journal of Mathematical Sciences*, 17(1), 1981.
- [4] A. Agrachev and Y. Sachkov. *Control Theory from the Geometric Viewpoint*. Springer-Verlag, Heidelberg, Germany, 2004.
- [5] A. Agrachev and A. Sarychev. On abnormal extremals for lagrange variational problems. *Journal of Mathematical Systems Estimation and Control*, 8:87–118, 1998.
- [6] A. A. Agrachev and R. V. Gamkrelidze. Local controllability for families of diffeomorphisms. *Systems & control letters*, 20(1):67–76, 1993.
- [7] M. Barbero-Liñán and M. C. Muñoz-Lecanda. Geometric approach to Pontryagin's maximum principle. *Acta applicandae mathematicae*, 108(2):429–485, 2009.
- [8] G. K. Batchelor. *An introduction to fluid dynamics*. Cambridge University Press, 2000.
- [9] A. Bennett. *Lagrangian fluid dynamics*. Cambridge University Press, 2006.
- [10] L. D. Berkovitz. *Optimal control theory*. Springer-Verlag, New York, NY, 1974.
- [11] J.-M. Bismut. *Large Deviations and the Malliavin Calculus*. Birkhäuser, Boston, MA, 1984.
- [12] A. Bloch. *Nonholonomic Mechanics and Control*. Springer, 2003.
- [13] R. Bonic and J. Frampton. Smooth functions on banach manifolds. *J. Math. Mech*, 15(5):877–898, 1966.
- [14] J. Borwein and Q. J. Zhu. *Techniques of Variational Analysis*. CMS Books in Mathematics. Springer, New York, NY, 2010.
- [15] J. Borwein and D. Zhuang. On fan's minimax theorem. *Mathematical Programming*, 34:232 – 234, 1986.



- [16] J. M. Borwein. Stability and regular points of inequality systems. *Journal of Optimization Theory and Applications*, 48(1):9 – 52, 1986.
- [17] F. Bullo and A. Lewis. Supplementary chapters for geometric control of mechanical systems [18]. <http://motion.mee.ucsb.edu/book-gcms/>. Retrieved October, 2013.
- [18] F. Bullo and A. Lewis. *Geometric Control of Mechanical Systems: Modeling, Analysis, and Design for Simple Mechanical Control Systems*. Springer, New York, NY, 2005.
- [19] W. L. Burke. *Applied differential geometry*. Cambridge University Press, 1985.
- [20] S. A. Caiado, M.I. Remarks on stability of inverted pendula. *Rendiconti del Seminario Matematico*, 63(4):333–347, 2005.
- [21] L. Cesari. *Optimization-theory and applications: problems with ordinary differential equations*, volume 17. Springer New York, 1983.
- [22] W.-L. Chow. Über systeme von linearen partiellen differentialgleichungen erster ordnung. *Mathematische Annalen*, 117:98 – 105, 1940.
- [23] F. Clarke. *Optimization and Nonsmooth Analysis*. SIAM Classics in Applied Mathematics. John Wiley and Sons, New York, NY, 1983. Reprinted as vol. 5 of Classics in Applied Mathematics, SIAM, Philadelphia, PA, 1990; Russian translation, Nauka, Moscow, 1988.
- [24] F. Clarke. *Necessary conditions in dynamic optimization*. American Mathematical Society, Providence, RI, 2005.
- [25] F. Clarke. *Functional Analysis, Calculus of Variations and Optimal Control*. Springer, New York, NY, 2013.
- [26] F. Clarke and Y. S. Ledyaev. Mean value inequalities in hilbert space. *Transactions of the American Mathematical Society*, 344:307 – 324, July 1994.
- [27] F. Clarke, Y. S. Ledyaev, and M. Radulescu. Approximate invariance and differential inclusions in hilbert spaces. *Journal of Dynamical and Control Systems*, 3:493 – 518, 1997.
- [28] F. Clarke, Y. S. Ledyaev, R. Stern, and P. Wolenski. *Nonsmooth Analysis and Control Theory*. Springer-Verlag, New York, NY, 1998.
- [29] F. H. Clarke. *Necessary conditions for nonsmooth problems in optimal control and the calculus of variations*. PhD thesis, University of Washington., 1973.
- [30] F. H. Clarke. The maximum principle under minimal hypotheses. *SIAM Journal on Control and Optimization*, 14(6):1078–1091, 1976.

- [31] F. H. Clarke. Optimal solutions to differential inclusions. *Journal of Optimization Theory and Applications*, 19(3):469–478, 1976.
- [32] J.-M. Coron. *Control and nonlinearity*. American Mathematical Society, Providence, RI, 2009.
- [33] K. Deimling. *Ordinary Differential Equations in Banach Spaces*, volume 596 of *Lecture Notes in Mathematics*. Springer-Verlag, New York, NY, 1977.
- [34] J. Diestel and J. Uhl. *Vector Measures*. Number 15 in Mathematical Surveys. American Mathematical Society, Providence, RI, 1977.
- [35] A. Dresden. Some recent work in the calculus of variations. *Bulletin of the American Mathematical Society*, 32(5):475 – 521, 1926.
- [36] P. du Bois-Reymond. Erläuterungen zu den anfangsgründen der variationsrechnung. *Mathematische Annalen*, 15(2):283–314, 1879.
- [37] A. Y. Dubovitskii and A. A. Milyutin. Necessary conditions for a weak extremum in optimal control problems with mixed constraints of the inequality type. *USSR Computational Mathematics and Mathematical Physics*, 8(4):24–98, 1968.
- [38] N. Dunford and J. T. Schwarz. *Linear Operators: Part I*, volume 7 of *Pure and Applied Mathematics*. Wiley Interscience, New York, NY, 1958.
- [39] I. Ekeland. On the variational principle. *Journal of Mathematical Analysis and Applications*, 47(2):324–353, 1974.
- [40] M. Fabian, J. Whitfield, and V. Zizler. Norms with locally lipschitzian derivatives. *Israel Journal of Mathematics*, 44(3):262–276, 1983.
- [41] K. Fan. Minimax theorems. *Proceedings of the National Academy of Sciences of the United States of America*, 39(1):42 – 47, 1953.
- [42] H. O. Fattorini. *Infinite dimensional optimization and control theory*, volume 54. Cambridge University Press, 1999.
- [43] A. Filippov. On certain questions in the theory of optimal control. *Journal of SIAM Control, Series A*, 1(1):76–84, 1962.
- [44] A. Filippov. Classical solutions of differential equations with multivalued right-hand side. *SIAM Journal on Control*, 5(4):609–621, 1967.
- [45] R. Fry and S. McManus. Smooth bump functions and the geometry of banach spaces: a brief survey. *Expositiones Mathematicae*, 20(2):143–183, 2002.
- [46] R. Gamkrelidze. *Principles of Optimal Control Theory*. Plenum Press, New York, NY, 1978.

- [47] R. Gamkrelidze, A. Agrachev, and S. A. Vakhrameev. Ordinary differential equations on vector bundles, and chronological calculus. *Itogi Nauki i Tekhniki. Seriya "Sovremennye Problemy Matematiki. Noveishie Dostizheniya"*, 35:3–107, 1989.
- [48] I. Gelfand and S. Fomin. *Calculus of Variations*. Prentice-Hall, Inc., 1963.
- [49] P. Hartman. *Ordinary Differential Equations*. Classics in Applied Mathematics. SIAM, Philadelphia, PA, 2002.
- [50] E. Hobson. On the fundamental lemma of the calculus of variations, and on some related theorems. *Proceedings of the London Mathematical Society*, 2(1):17–28, 1913.
- [51] L. Hormander. *The Analysis of Linear Partial Differential Operators I. Distributions Theory and Fourier Analysis*. Springer-Verlag, Berlin, 1983.
- [52] A. Ioffe. Necessary conditions in nonsmooth optimization. *Mathematics of Operations Research*, 9(2):159–189, 1984.
- [53] A. Isidori. *Nonlinear Control Systems V*. Springer-Verlag, London, 1989.
- [54] V. Jurdjevic. *Geometric Control Theory*, volume 52 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, UK, 1997.
- [55] M. Kawski. Controllability via chronological calculus. In *Decision and Control, 1999. Proceedings of the 38th IEEE Conference on*, volume 3, pages 2920–2925. IEEE, 1999.
- [56] M. Kawski. Calculating the logarithm of the chen flies series. In *Proc. MTNS*, 2000.
- [57] M. Kawski. Chronological algebras: combinatorics and control. *Itogi Nauki i Tekhniki*, 68:144–178, 2000.
- [58] M. Kawski and H. J. Sussmann. Noncommutative power series and formal lie-algebraic techniques in nonlinear control theory. pages 111–128. Teubner, 1997.
- [59] I. Kolar, P. W. Michor, and J. Slovák. *Natural Operations in Differential Geometry*. Springer-Verlag, New York, NY, 2010.
- [60] G. Köthe. *Topological Vector Spaces I*. Springer-Verlag, New York, NY.
- [61] S. Lang. *Fundamentals of Differential Geometry*. Springer-Verlag, New York, NY, 1999.
- [62] Y. S. Ledyev. On generic existence and uniqueness in nonconvex optimal control problems. *Set-Valued Analysis*, 12(1):147–162.
- [63] Y. S. Ledyev. *Extremal Problems in Differential Game Theory*. Dr. Sc. Dissertation, Steklov Institute of Mathematics, Moscow, 1990.

- [64] Y. S. Ledyayev. On an infinite-dimensional variant of the rashevski-chow theorem. *Dokl. Akad. Nauk Rus*, 398:735–737, 2004.
- [65] Y. S. Ledyayev and J. S. Treiman. Sub- and supergradients of envelopes, semicontinuous closures, and limits of sequences of functions. *Russian Mathematical Surveys*, 67(2):345–373, 2012.
- [66] Y. S. Ledyayev and Q. J. Zhu. Nonsmooth analysis on smooth manifolds. *Transactions of the American Mathematical Society*, 359(8):3687–3732, 2007.
- [67] J. Lee. *Introduction to Smooth Manifolds*. Springer-Verlag, New York, NY, 2006.
- [68] M. Mauhart and P. Michor. Commutators of flows and fields. *Arch. Math. (Brno)*, 28:229–236, 1992.
- [69] R. McOwen. *Partial Differential Equations*. Pearson Education, Inc., Upper Saddle River, NJ, 2003.
- [70] S. Monaco, D. Normand-Cyrot, and C. Califano. Exponential representations of volterra-like expansions: an explicit computation of the exponent series. In *Decision and Control, 2002, Proceedings of the 41st IEEE Conference on*, volume 3, pages 2708–2713. IEEE, 2002.
- [71] S. Monaco, D. Normand-Cyrot, and C. Califano. From chronological calculus to exponential representations of continuous and discrete-time dynamics: a lie-algebraic approach. *Automatic Control, IEEE Transactions on*, 52(12):2227–2241, 2007.
- [72] R. Montgomery. *A tour of subriemannian geometries, their geodesics and applications*. The American Mathematical Society, Providence, RI, 2002.
- [73] B. Mordukhovich. *Variational analysis and generalized differentiation, II: Applications*. Springer-Verlag, New York, NY, 2006.
- [74] L. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and E. Mishchenko. *The Mathematical Theory of Optimal Processes*. John Wiley & Sons, New York, NY, 1962.
- [75] P. Rashevskii. About connecting two points of a completely nonholonomic space by admissible curve. *Uch. Zapiski Ped. Inst. Libknechta*, 2:83 – 94, 1938.
- [76] W. T. Reid. Ramifications of the fundamental lemma of the calculus of variations. *Houston Journal of Mathematics*, 4(2):249–262, 1978.
- [77] G. Restrepo. Differentiable norms in banach spaces. *Bulletin of the American Mathematical Society*, 70(3):413–414, 1964.
- [78] A. V. Sarychev. Lie-and chronologico-algebraic tools for studying stability of time-varying systems. *Systems & Control Letters*, 43(1):59–76, 2001.

- [79] H. Schättler and U. Ledzewicz. *Geometric Optimal Control: Theory, Methods and Examples*. Springer-Verlag, New York, NY, 2010.
- [80] W. Schirotzek. *Nonsmooth Analysis*. Springer-Verlag, New York, NY, 2007.
- [81] A. I. Subbotin. *Generalized solutions of first-order PDEs: the dynamical optimization perspective*. Birkhäuser Boston, 1995.
- [82] H. Sussmann. Symmetries and integrals of motion in optimal control. *Geometry in nonlinear control and differential inclusions. Mathematics institute of Polish academy of sciences. Banach center publications*, 32:379 – 393, 1995.
- [83] H. J. Sussmann. Two new methods for motion planning for controllable systems without drift. In *In European Control Conference*. Citeseer, 1991.
- [84] H. J. Sussmann. An introduction to the coordinate-free maximum principle. *Pure and Applied Mathematics - Marcel Dekker Incorporated*, 207:463–557, 1998.
- [85] A. Tretyak. Sufficient conditions for local controllability and high-order necessary conditions for optimality. a differential-geometric approach. *Journal of Mathematical Sciences*, 85(3):1899–2001, 1997.
- [86] A. Tretyak. Chronological calculus, high-order necessary conditions for optimality, and perturbation methods. *Journal of dynamical and control systems*, 4(1):77–126, 1998.
- [87] F. Tröltzsch. *Optimal control of partial differential equations: Theory, methods and applications*, volume 112. AMS Bookstore, 2010.
- [88] P. A. Vela and J. W. Burdick. A general averaging theory via series expansions. In *Proceedings of the American Control Conference*, volume 2, pages 1530–1535, 2003.
- [89] R. Vinter. *Optimal Control*. Birkhauser, Boston, MA, 2000.
- [90] R. B. Vinter and G. Pappas. A maximum principle for nonsmooth optimal-control problems with state constraints. *Journal of Mathematical Analysis and Applications*, 89(1):212–232, 1982.
- [91] J. Warga. Relaxed variational problems. *Journal of Mathematical Analysis and Applications*, 4:111–128, 1962.
- [92] J. Warga. *Optimal Control of Differential and Functional Equations*. Academic Press, New York, NY, 1972.

- [93] L. C. Young. Generalized curves and the existence of an attained absolute minimum in the calculus of variations. *Comptes Rendus de la Société des Sciences et des Lettres de Varsovie, classe III*, 30:212–234, 1937.
- [94] L. C. Young. *Lectures on the calculus of variations and optimal control theory*. American Mathematical Society, Providence, RI, 1980.