A geometric study of abnormality in optimal control problems for control and mechanical control systems

María Barbero Liñán

Supervisor: Dr. Miguel C. Muñoz Lecanda

A thesis submitted for the degree of Doctor of Applied Mathematics
Technical University of Catalonia
September 2008

Abstract

For the last forty years, differential geometry has provided a means of understanding optimal control theory. Usually the best strategy to solve a difficult problem is to transform it into a different problem that can be dealt with more easily. Pontryagin's Maximum Principle provides the optimal control problem with a Hamiltonian structure. The solutions to the Hamiltonian problem, satisfying particular conditions, are candidates to be solutions to the optimal control problem. These candidates are called extremals. Thus, Pontryagin's Maximum Principle lifts the original problem to the cotangent bundle.

In this thesis, we develop a complete geometric proof of Pontryagin's Maximum Principle. We investigate carefully the crucial points in the proof such as the perturbations of the controls, the linear approximation of the reachable set and the separation condition.

Among all the solutions to an optimal control problem, there exist the abnormal curves. These do not depend on the cost function we want to minimize, but only on the geometry of the control system. Some work has been done in the study of abnormality, although only for control–linear and control–affine systems with mainly control–quadratic cost functions. Here we present a novel geometric method to characterize all the different kinds of extremals—not only the abnormal ones—in general optimal control problems. This method is an adaptation of the presymplectic constraint algorithm. Our interest in the abnormal curves is with the strict abnormal minimizers. These last minimizers can be characterized by the geometric algorithm presented in this thesis.

As an application of the above–mentioned method, we characterize the extremals for the free optimal control problems that include, in particular, the time–optimal control problem. Moreover, an example of an strict abnormal extremal for a control–affine system is found using the geometric method.

Furthermore, we focus on the description of abnormality for optimal control problems for mechanical control systems, because no results about the existence of strict abnormal minimizers are known for these problems. Results about the abnormal extremals are given when the cost function is control–quadratic or the time must be minimized. In this dissertation, the abnormality is characterized in particular cases through geometric constructions such as vector–valued quadratic forms that appear as a result of applying the previous geometric procedure.

The optimal control problems for mechanical control systems are also tackled taking advantage of the equivalence between nonholonomic control systems and kinematic control systems. In this thesis, it is found an equivalence between time–optimal control problems for both control systems. The results allow us to give an example of a local strict abnormal minimizer in a time–optimal control problem for a mechanical control system.

Finally, setting aside the abnormality, the non–autonomous optimal control problem is described geometrically using the Skinner–Rusk unified formalism. This approach is valid for

implicit control systems that arise in optimal control problems for the controlled Lagrangian systems and for descriptor systems. Both systems are common in engineering problems.

Certification

Dr. Miguel Carlos Muñoz Lecanda, chair in Applied Mathematics,

CERTIFIES

that this report with the title

A geometric study of abnormality in optimal control problems for control and mechanical control systems

has been composed and originated in the Department of Applied Mathematics IV at the Technical University of Catalonia under his supervision, by María Barbero Liñán and gives his permission so as to be qualified as a Ph.D. dissertation.

Barcelona, 26 September 2008.

Acknowledgements

This dissertation would not have been possible without the support and the guidance provided by my supervisor, Dr. Miguel C. Muñoz Lecanda at every single moment. These last fours years working under his supervision have made me grow as a researcher and as a person. I am also grateful to Professor Andrew D. Lewis as a non–official supervisor for encouraging me to work on a deep understanding of optimal control theory during his sabbatical stay in Barcelona. I would also like to thank him for accepting me as a visiting student in his department and for continuously showing interest in my work. He is always willing to share his eye–opening new ideas with others.

I would like to thank the anonymous referees of journals and conferences because they helped to improve my work with their comments.

I am grateful to have become part of the research group DGDSA in the Department of Applied Mathematics IV at the Technical University of Catalonia. Narciso, Arturo and Xavier gave me a warm welcome from the first day and always show their interest in my work. I would like to thank especially Narciso for sharing his teaching duties with me and taking care of my bamboo.

I would like to thank the research network Geometry, Mechanics and Control that has given me the opportunity to attend international conferences and to interact with expert researchers like David Martín de Diego (who even invited me to go to Madrid to make me go into the world of Lie algebroids), Eduardo Martínez, and Juan Carlos Marrero who were always willing to discuss my research and help me with different viewpoints. I do not forget the young people in this research network who always create a great environment for everything, also thanks to them. A special thank goes to Edith who always succeeds in coordinating all the issues in an unbeatable way, it is just astonishing.

The financial support of *Ministerio de Educación y Ciencia*—Project MTM2005-04947, the Network Project MTM2006-27467-E/ and an FPU grant— and of *Generalitat de Catalunya* with an FI grant have been essential to be able to devote my energy to research.

I would like to thank the staff and colleagues in Barcelona and in the Department of Mathematics and Statistics at Queen's University in Kingston (Ontario, Canada) for their continuous help. In Kingston I had the pleasure to join Professor Lewis' group where I found PhD students with common grounds, above all, Bahman and Cesar. I have always enjoyed the discussions and seminars we had together.

I am grateful to my friends and to my family in Barcelona because they have been there to listen to me, to cheer me up and to give me different opinions. All of them have made this period of my life enjoyable, even when I was forced to take a few months off. I would like to express my gratitude to all of them because they were there taking care of me when I most

needed. Special thanks to Ciara, Anna, Bernat, Verónicas, María, Óscar, Oriol, Marina, Puri, Mary Cármenes, Carmen and Elena.

Finally, to my parents because I know it has not been easy to accompany me through this process. Not all the news I had for you were the ones you have liked to listen. In spite of that, you always give me your love and unconditional support. To my brother, you are always in the lead and so I can benefit from your experience, professionally and personally. To James, you have given a different meaning to my life since we met. To be able to type the first draft of this thesis next to you made it pleasant.

a mis padres y a mi hermano

Contents

List of Figures

Acronyms

ACCS	Affine connection control systems
FHP	Free Hamiltonian problem
FOCP	Free optimal control problem
$\widehat{\text{FOCP}}$	Extended free optimal control problem
FPMP	Free Pontryagin's Maximum Principle
HP	Hamiltonian problem
OCP	Optimal control problem
$\widehat{\text{OCP}}$	Extended optimal control problem
PMP	Pontryagin's Maximum Principle
STLC	Small–time locally controllable

List of Symbols

$a.e.$	almost everywhere
$a.c.$	absolutely continuous
M	m–dimensional manifold
U	Control set in $\mathbb{R}^k$
TM	Tangent bundle of M
T^*M	Cotangent bundle of M
τ_M	Canonical tangent projection
π_M	Canonical cotangent projection
$\mathfrak{X}(M)$	Set of vector fields on M
$\Omega^1(M)$	Set of 1–forms on M
f^*	Pullback of the $\mathcal{C}^\infty$–function f
f_*	Pushforward of a diffeomorphism $f \colon M \to M$
d	Exterior derivative
i_X	Inner or interior product of a vector field X on M
I	Compact interval in $\mathbb{R}$
$\phi_{x_0}^X$	Integral curve of the vector field X with initial condition x_0 at time zero
$\Gamma(\pi)$	Set of sections of the fiber or the vector bundle π
$\mathfrak{X}^V(M,\pi)$	Set of vertical vector fields on M with respect to π
$\mathfrak{X}(\pi)$	Set of vector fields along a projection π
Φ^X	Time–dependent flow or evolution operator of the time–dependent vector field X on M
X^T	Complete or tangent lift of the vector field X
κ_M	Canonical involution of TTM
$X^{\{u\}}$	Time–dependent vector field on M such that $X^{\{u\}}(t,x) = X(x,u(t))$.
X^{T^*}	Cotangent lift of the vector field X
$\mathcal{L}_X$	Lie derivative with respect to the vector field X
∇	Affine connection on a manifold
Δ	Liouville vector field on a vector bundle $\pi \colon E \to B$

J_M	Vertical endomorphism
$J^1\pi$	The jet bundle of sections of π
$\mathcal{W}$	Extended jet–momentum bundle
$\mathcal{W}_r$	Restricted jet–momentum bundle
$\hat{\mathcal{C}}$	Coupling 1–form in $\mathcal{W}$
f_s, Y_s	Control vector fields on a manifold
(γ, u)	A curve on a manifold $M \times U$ or $Q \times U$
$\mathcal{F}$	Cost function
$\mathcal{S}[\gamma, u]$	Functional to be minimized
$(M, U, X, \mathcal{F}, I, x_a, x_b)$	Optimal control problem
$\widehat{M}$	Extended manifold, $\mathbb{R} \times M$
$\widehat{\gamma}$	Curve on the extended manifold $\widehat{M}$
$\widehat{X}$	Vector field X extended to the manifold $\widehat{M}$
(γ^*, u^*)	Solution to an optimal control problem
$(\widehat{M}, U, \widehat{X}, I, x_a, x_b)$	Extended optimal control problem
$\pi_1 = \{t_1, l_1, u_1\}$	Perturbation data for the control
$u[\pi_1^s]$	Perturbation of u specified by the data π_1
$v[\pi_1]$	Perturbation vector associated to the perturbation data π_1
K_t	Tangent perturbation cone at time t
$B(p, r)$	Open ball in the Euclidean space at center p with radius r
$\overline{B(p, r)}$	Closed ball in the Euclidean space at center p with radius r
H	(Pontryagin's) Hamiltonian function
$\widehat{\lambda}$	Extended momenta given by Pontryagin's Maximum Principle on M
$\pi_\pm = \{\tau, l_\tau, \delta\tau, u_\tau\}$	Data perturbation in the free optimal control problem
$K_t^\pm$	Time perturbation cone at time t
$N_i^{[0]}$	Constraint submanifold at step i for abnormality
$N_i^{[-1]}$	Constraint submanifold at step i for normality
$\mathscr{Y} = \{Y_1, \ldots, Y_k\}$	Family of control or input vector fields
$\Sigma = (Q, \nabla, \mathscr{Y}, U)$	Affine connection control system
$\Sigma_{\mathcal{F}}(x_a, x_b)$	Optimal control problem for an affine connection control system with endpoint conditions x_a, x_b in Q
Υ	A curve on the tangent bundle TQ
Z	Geodesic spray associated with a connection ∇
$\langle \cdot : \cdot \rangle$	Symmetric product of two vector fields

Λ	Momenta on T^*TQ given by the mechanical Pontryagin's Maximum Principle
B_x	Vector–valued quadratic form at $x \in Q$
$(\lambda B)_x$	Real quadratic form inherited from B_x for $\lambda \in \operatorname{ann} \mathscr{Y}_x$
$\Sigma_{\mathcal{D}} = (Q, g, F, \mathcal{D})$	Nonholonomic control system
$\Sigma_m = (\Sigma_{\mathcal{D}}, \mathcal{F})$	Nonholonomic optimal control problem for the system $\Sigma_{\mathcal{D}}$ with cost function $\mathcal{F}$
$\Sigma_k = (\Sigma_m, \mathcal{G})$	Kinematic optimal control problem for the kinematic control system associated to $\Sigma_{\mathcal{D}}$ with cost function $\mathcal{G}$
H_m	Pontryagin's Hamiltonian function for the problem Σ_m
H_k	Pontryagin's Hamiltonian function for the problem Σ_k
$\widehat{a}$	Extended momenta associated with the problem Σ_k or extended kinematic momenta
$\widehat{v}_m$	Extended perturbation vectors associated with Σ_m
$\widehat{v}_k$	Extended perturbation vectors associated with Σ_k
$\mathcal{W}^X$	Extended control–jet–momentum bundle
$\mathcal{W}_r^X$	Restricted control–jet–momentum bundle
$\mathcal{W}^{M_C}$	Extended control–jet–momentum bundle for implicit optimal control problems
$\mathcal{W}_r^{M_C}$	Restricted control–jet–momentum bundle for implicit optimal control problems

Chapter 1
Introduction

Optimal control theory is a young research area that appears in a wide variety of fields such as medicine, economics, traffic flow, engineering, astronomy. However, applications and understanding do not always come together. In order to gain insight, differential geometry has been used in control theory, giving rise to geometric control theory in the 70's [Agrachev 2002a, Agrachev and Sachkov 2004, Bloch 2003, Bonnard and Caillau 2006, Bonnard and Chyba 2003, Boscain and Piccoli 2004, Bressan and Piccoli 2007, Bullo and Lewis 2005a;b, Echeverría-Enríquez et al. 2003, Jurdjevic 1997, Langerock 2003a, Lewis 2006, Nijmeijer and van der Schaft 1990, Sussmann 1978; 1998; 1999; 2000, Sussmann and Jurdjevic 1972, Troutman 1996].

1.1 Historical remarks

If we look back in time, we will realise that optimal control problems have existed for a long time as claimed by Sussmann and Willems [1997]. There, the brachystochrone problem suggested by J. Bernoulli in the June 1696 issue of *Acta Eruditorum* is considered as the starting point of optimal control theory. The challenge posed to mathematicians by J. Bernoulli is the following:

> *If, in a vertical plane, two points A and B are given, then it is required to specify the orbit AMB of the movable point M, along which it, starting from A, and under the influence of its own weight, arrives at B in the shortest possible time. So that those who are keen of such matters will be tempted to solve this problem, is it good to know that it is not, as it may seem, purely speculative and without practical use. Rather it even appears, and this may be hard to believe, that it is very useful also for other branches of science than mechanics. In order to avoid a hasty conclusion, it should be remarked that the straight line is certainly the line of shortest distance between A and B, but it is not the one which is traveled in the shortest time. However, the curve AMB—which I shall divulge if by the end of this year nobody else has found it—is very well known among geometers.*

Thus, control theory studies the properties of a dynamical system with some degrees of freedom given by the controls; e.g., in the previous problem those are associated with the point M. When we want to find the trajectory of a dynamical system that minimizes a functional such as energy, time or distance, we are confronting a problem in optimal control. In general, to find a solution of these kinds of problems is not straightforward. A valuable tactic to deal with

optimal control problems is to restrict the candidate solutions through necessary conditions of optimality such as those given by Pontryagin's Maximum Principle [Pontryagin et al. 1962].

In some sense, optimal control theory is regarded as a generalization of the calculus of variations [Bullo and Lewis 2005b, Lewis 2006, Sussmann and Willems 1997]. A main difference between these two theories is that in the former the controls can take values in a closed control set and inequality constraints are accepted, whereas in the latter the control set is always open.

Pontryagin's Maximum Principle was introduced to the mathematical community in the International Congress of Mathematicians held in 1958 in Edinburgh, Scotland, by a group of Russian researchers working in Steklov Mathematical Institute [Pontryagin et al. 1962]. The Russian school focused on this research under a request by the military service. The Maximum Principle is considered one of the outstanding points in optimal control theory.

1.2   State–of–the–art

The classical approach to optimal control problems was from the point of view of the differential equations [Athans and Falb 1966, Lee and Markus 1967, Pontryagin et al. 1962] and of the functional analysis [Giaquinta and Hildebrandt 1996a;b, Zeidler 1985], but later the approach was from the differential geometry [Agrachev and Sachkov 2004, Bressan and Piccoli 2007, Jurdjevic 1997, Sussmann 1998]. However, the Maximum Principle admits other points of view such as stochastic control systems [Bensoussan 1984, Haussmann 1986] and discrete control systems [Chyba et al. 2008, Guibout and Bloch 2004, Hwang and Fan 1967]. Lately, the Skinner-Rusk formulation [Skinner and Rusk 1983] has been applied to study optimal control problem for non–autonomous control systems. As a result, the necessary conditions of Pontryagin's Maximum Principle have been obtained, as long as the differentiability with respect to controls is assumed [Barbero-Liñán et al. 2007].

A Hamiltonian formalism to optimal control problems is provided by the necessary conditions stated in Pontryagin's Maximum Principle. The solutions to the problem are in a manifold, but the Maximum Principle relates solutions to a lift to the cotangent bundle of that manifold. Thus, in order to find candidate optimal solutions, not only the controls but also the momenta must be chosen appropriately so that the necessary conditions in the Maximum Principle are fulfilled. These conditions are, in fact, first–order necessary conditions and they are not always enough to determine all the degrees of freedom in the problem. That is why sometimes it is necessary to use the high–order Maximum Principle [Bianchini 1998, Kawski 2003, Knobloch 1981, Krener 1977]. But, even when we succeed in finding the controls and the momenta in such a way that Hamilton's equations can be integrated to obtain a trajectory on the manifold, the controls and the momenta are not necessarily unique. In other words, different controls and different momenta can give the same trajectory on the manifold, although the necessary conditions in the Maximum Principle will be satisfied in different ways. The momenta and the control determine the kinds of trajectories, which can be abnormal, normal, strict abnormal, strict normal and singular. We point out that these different kinds of extremals do not provide a partition of the set of trajectories in the manifold, because it may happen that a same trajectory admits more than one momenta so that the trajectory is in two different categories.

For years, abnormal extremals were discarded because it was thought that they could not

be optimal [Hamenstädt 1990, Strichartz 1986]. The idea was that abnormal extremals were isolated curves and thus it was impossible to consider any variation of these curves. However, Montgomery [1994] proved that there exist abnormal minimizers by giving an example in sub-Riemannian geometry. Then Liu and Sussmann [1995] made an effort to characterize strictly abnormal minimizers in a general way for the length-minimizing problem in subRiemannian geometry if there are only two controls. They succeeded in giving a set with abnormal extremals that contains strict abnormal curves that are locally optimal for the considered control–linear system [Liu and Sussmann 1994a;b; 1995, Sussmann 1996]. Let us give a short review of some work that has been done concerning abnormal extremals.

Agrachev and Sarychev [1995b; 1996] study second–order necessary conditions for optimality since the first–order necessary conditions give little information about abnormal extremals. They state necessary and sufficient conditions for local optimality and rigidity (i.e., isolation) of abnormal extremals using the derivatives of the endpoint mapping, related to the reachable set. They also use the theory of the Morse index.

Agrachev and Sarychev [1995a; 1996] consider subRiemannian geometry for a distribution of rank 2. They arrive at the same definition of regular abnormal extremals given by Liu and Sussmann [1995], but they also focus on control–affine systems. They also prove that when the distribution is bracket–generating and the controls are constrained there exist local rigid abnormal minimizers, something that does not happen for unconstrained controls.

Agrachev and Zelenko [2007] characterize the affine line subbundle that gives the abnormal extremals for control–affine systems with one or two input vector fields. For two input vector fields, they concentrate on the study of manifolds of dimension 4 and 5.

Bonnard and Trélat [2001] characterize the abnormal directions looking at the subRiemannian sphere. In order to do this, they consider mainly the case of Martinet distributions.

Boscain and Piccoli [2002] study the time–optimal control problem for a control–affine system with one input vector field. It turns out that the abnormal extremals are concatenations of determined arcs coming from switching the control. They establish a classification of all the possibilities.

Langerock [2001; 2003a;b;c] characterizes the abnormal extremals geometrically by constructing a suitable connection.

What makes abnormal extremals more special is that the abnormality does not depend on the cost function. Hence, the abnormal extremals can be determined exclusively using the geometry of the control system. Thus abnormality and controllability must be closely related. In fact, in order to have abnormal minimizers, the system cannot be controllable. Moreover, the set of the trajectories given by the control system determines the reachable set, independently of the cost function. That is why it is thought that the study of the reachability and/or the controllability [Jurdjevic 1997, Nijmeijer and van der Schaft 1990, Sussmann 1978; 1987] could help to characterize abnormal extremals through the geometry of the reachable set [Bullo and Lewis 2005b, Langerock 2003a]. In control theory, controllability is still one of the properties under active research [Agrachev 1999, Aguilar and Lewis 2008, Basto-Gonçalves 1998, Bullo and Lewis 2005c, Cortés and Martínez 2003, Tyner 2007] and the same happens with abnormality in optimal control theory as already shown.

In contrast to the previous paragraph, the cost function is essential to prove that abnormal extremals are strict abnormal minimizers. That is why the existence or non–existence of abnormal minimizers is only known for specific control problems, mainly time–optimal control problems and optimal control problems with control–quadratic cost functions for control–linear and control–affine systems [Agrachev and Sarychev 1995a; 1999, Agrachev and Zelenko 2007, Bonnard and Chyba 2003, Chitour et al. 2006; 2008, Trélat 2001, Zelenko 1999].

In general, the necessary conditions given by the Maximum Principle are useful for determining the control that will give us a possible optimal trajectory, except for abnormal and singular extremals [Kupka 1987, Pelletier 1999, Zelenko 1999]. In such cases, it is necessary to consider high–order necessary conditions that will help us to determine the control, as studied by Krener [1977].

When a problem is difficult to deal with, we restrict to particular cases in order to figure out a first idea of possible solutions. That is why we restrict Pontryagin's Maximum Principle to optimal control problems for mechanical systems described by affine connections [Bullo and Lewis 2005b]. Singular extremals have already been studied in [Chyba and Haberkorn 2005, Chyba et al. 2003]. In fact, the singular extremals are also abnormal for time–optimal problems.

The main aim of this dissertation is to give a detailed geometric study of how to characterize abnormal extremals in nonlinear control theory and also for particular cases such as mechanical control systems and control–affine systems. We believe that a better geometric understanding of abnormality will give insights into strict abnormality because every strict abnormal optimal curve is also an abnormal optimal curve.

In order to achieve our aim, we have gone through the entire proof of Pontryagin's Maximum Principle translating it into a geometric framework, but preserving the outline of the original proof. All details have been carefully proved, making us to go into the details of concepts such as time–dependent variational equations and their properties, perturbation vectors and the separation conditions given by hyperplanes. Afterwards, we focus on some geometric approaches to the abnormality, establishing connections with the controllability.

Then we propose a method to characterize all the different kinds of extremals for any optimal control problem for any control system. In order to do that, the presymplectic theory is used to adapt the so–called presymplectic constraint algorithm by Gotay and Nester [1979], Gotay et al. [1978].

The next step is the study of the abnormality for affine connection control systems that model mechanical systems. After applying the adapted presymplectic constraint algorithm, we consider particular optimal control problems in order to give more information about the different extremals. Then, we focus on some examples with small dimension where geometric elements, as for instance, the symmetric product and the vector–valued quadratic forms, arise in the reasoning about abnormality. Some ideas between these elements and the abnormality are given.

We consider another approach to study the abnormality for the mechanical case that consists of taking advantage of the results known in subRiemannian geometry. The control system in the problem of finding the shortest paths in subRiemannian geometry is a kinematic control system. Thus, we look into the nonholonomic control systems and their equivalence with the kinematic

control system [Bloch 2003, Bullo and Lewis 2005a;c, Muñoz-Lecanda and Yániz-Fernández 2008] in order to find some connection between the optimal control problems associated with both control systems.

Finally, we put aside our interest in abnormality, and we focus on the Skinner–Rusk formalism to give a unified approach to the non–autonomous optimal control problems.

1.3   Contributions and scheme of the thesis

Here let us point out the contributions in the area of differential geometry and optimal control theory provided by this dissertation. We also give a brief description of the contents of every chapter.

Chapter 2

This chapter is a review of the main elements of differential geometry used in this dissertation. A special importance is given to the study of the time–dependent variational equations, the different definitions of a connection on a fiber bundle and the Skinner–Rusk formalism for non–autonomous systems.

In spite of being a review, the study of the time–dependent variational equations given in §2.2.2 gives a clear picture of the flows of the complete lift and of the cotangent lift of a time–dependent vector field via Propositions 2.2.3, 2.2.4, 2.2.6 and Corollary 2.2.5. These results although known, to our knowledge, have not appeared in the literature.

Chapter 3

In this chapter we give the background in control theory necessary for the understanding of the subsequent chapters. We focus mainly on the sufficient conditions for controllability.

Chapter 4

Our main contribution is the complete geometric version of the proof of Pontryagin's Maximum Principle in §4.2 and §4.4 in a symplectic framework, with all the details about the different perturbation vectors in §4.1.3 and §4.3.2. The complete proof we give of Proposition 4.1.12, although known, to our knowledge, there is not a self–contained proof of it in the literature.

In §4.5, we give a necessary condition for abnormality, valid for any optimal control problem and related with controllability, Proposition 4.5.2. The proof of Pontryagin's Maximum Principle suggests that all the perturbation vectors generate a linear approximation of the reachable set in some sense. That sense will become clear in Proposition 4.5.3, which proves a result often assumed as true in the literature.

Finally, we are able to give a picture of the separation condition for a particular example in §4.6.2 to show the differences when there exist momenta associated with the trajectory such that it is both abnormal and normal. We will see that the separation condition is another key point in the proof of Pontryagin's Maximum Principle and must be well–understood geometrically to understand abnormality.

Chapter 5

The Maximum Principle is treated from the presymplectic viewpoint. We have two different versions of Hamilton's equations where the presymplectic constraint algorithm in the sense of Gotay and Nester [1979], Gotay et al. [1978] can be applied.

Apart from the new results in §5.2.1 and §5.2.2, Proposition 5.2.9 summarizes how to use the algorithm to characterize the different kinds of extremals when the domain of definition of the curves is known. If the domain is not given, the adaptation of the presymplectic constraint algorithm is described in §5.3.

To highlight the generality of the geometric process described, the usual examples in the literature are revisited in §5.4: geodesics in Riemannian and in subRiemannian geometry and control–affine systems.

In §5.5, we give for first time to our knowledge, a strict abnormal extremal for an optimal control problem for a mechanical control system.

Chapter 6

In this chapter we focus on the mechanical control systems, called affine connection control systems, which are defined in §6.1. The controllability and the accessibility for these mechanical control systems is described in §6.2. Some sufficient conditions for accessibility are obtained from Proposition 6.2.7, although known, never stated clearly in the literature.

After studying the concepts of control theory related to the affine connection control systems, we move to optimal control theory in §6.3 stating the optimal control problems for affine connection control systems. In §6.4 we review an intrinsic version of Pontryagin's Maximum Principle given in [Bullo and Lewis 2005b]. Hence, a complete general study of Pontryagin's Maximum Principle is found in this dissertation. Moreover, this review will be useful for establishing a comparison with the approach considered in Chapter 7 to describe Pontryagin's Maximum Principle for nonholonomic optimal control problems.

Then, in §6.5 and §6.6 we consider the presymplectic approach to mechanical optimal control problems. The geometric method in Chapter 5 is used to obtain results about the abnormal minimizers for particular cases in §6.7 such as problems with a control–quadratic cost function, Propositions 6.7.1 and 6.7.2, and the time–optimal control problem, Proposition 6.7.4.

Moreover, in §6.8 particular examples for different values of the rank of the distribution spanned by the input vector fields are studied. All the information provided by the constraint algorithm is translated into the language of vector–valued quadratic forms in such a way it is possible to state Conjecture 6.8.5 about necessary conditions for the existence of abnormal optimal curves.

Chapter 7

Under suitable assumptions, there exists an equivalence between trajectories of kinematic control systems and nonholonomic mechanical control systems [Bloch 2003, Bullo and Lewis 2005a;c, Muñoz-Lecanda and Yániz-Fernández 2008]. Here the equivalence or not between solutions to optimal control problems associated with both control systems is studied. As stated

in Proposition 7.1.6, the solution to the nonholonomic optimal control problem determines a solution to the kinematic optimal control problem for specific cost functions. The equivalence between the solutions is fulfilled for the time–optimal control problem, Proposition 7.1.10 and Remark 7.1.11.

In §7.2, we study the relationship between the different kinds of extremals for the non-holonomic case and the kinematic case. It turns out that the approach given by the mechanical version of Pontryagin's Maximum Principle here is more natural than the one considered in [Bullo and Lewis 2005b] and reviewed in §6.4 since the new momenta that show up has a particular meaning because of the corresponding extended system.

Finally, in §7.2.4, all the previously proven results give us a locally strict abnormal minimizer for a nonholonomic optimal control problem obtained from a locally strict abnormal minimizer for the corresponding kinematic control problem.

Chapter 8

Skinner and Rusk [1983] suggested the so–called unified Skinner–Rusk formalism, which is applied here to give an intrinsic formalism for the non–autonomous optimal control problems, including the implicit optimal control problems that appear in applications such as descriptor systems. The new main results appear in Theorem 8.1.6 and in Propostion 8.2.2.

A unified formalism is considered for first time in optimal control problems for the controlled Lagrangian mechanical systems §8.3.1 and for the descriptor systems §8.3.2.

This is the only chapter where we do not focus on the abnormality. The results are given just for normal trajectories because we are more interested in the new approach to optimal control theory for non–autonomous control systems given by the unified formalism inherited from Skinner–Rusk.

Chapter 9

To conclude, we review all the main contributions and point out the future research lines to work on.

Appendices A, B and C

Notions related with analysis, geometry and algebra are introduced here. They do not belong to differential geometry, although they are essential for the development of the dissertation. Some of the results are not clearly proved in the literature, as for instance, Proposition A.1.7, which is necessary in the proof of Pontryagin's Maximum Principle.

Chapter 2
Background and notation

A minimum knowledge in differential geometry is assumed in this work, as for instance some of the topics studied in [Abraham and Marsden 1978, Abraham et al. 1988, Conlon 1993, do Carmo 1992, Kobayashi and Nomizu 1996, Kolář et al. 1993, Lee 2003, Saunders 1989]. However, the main geometric elements and their notations are briefly reviewed here. Some of the sections in this chapter are explained in more detail than others, because they are important for this dissertation and also because some concepts are not always clearly presented in the literature, although in general they are assumed to be known.

In §2.1 we give a panorama of differential geometry used in this work, for the purpose of fixing notation. Then, in §2.2 we focus on time–dependent vector fields. For instance, a vector field depending on parameters defined in §2.2.1 can be understood as a time–dependent vector field, if the parameters are functions of time. In Chapter 4 vector fields depending on parameters define a control system and the variations of the system obtained by modifying the parameters are considered. These variations depend on the time–dependent variational equations, which are explained in §2.2.2 in detail because we do not know any reference where these equations are studied.

Symplectic geometry is reviewed in §2.3 because the approach to optimal control problems considered in Chapter 4 is symplectic. On the other hand, in Chapter 5 and §6.5 the approach to optimal control problems is presymplectic since we adapt the presymplectic constraint algorithm developed by Gotay and Nester [1979], Gotay et al. [1978] and reviewed in §2.3.2. That adaptation, explained in general in Chapter 5, is useful for characterizing the different kinds of solutions to optimal control problems.

The notion of Ehresmann connections and the splittings associated to them in §2.4 are important to study the control mechanical systems in optimal control theory in Chapter 6. See [León and Rodrigues 1989] for more details on connections.

The general unified formalism for non–autonomous systems [Barbero-Liñán et al. 2008, Cortés et al. 2002b] due to Skinner and Rusk [1983] is introduced in §2.5, because in Chapter 8 the unified formalism for non–autonomous control systems is described. Some notions of the geometry of the jet bundles and the forced Euler–Lagrange equations must be included in §2.6 to construct one of the examples in Chapter 8.

2.1   Manifolds and tensor fields

Here we present the usual definitions and notations in differential geometry, for more details see [Abraham and Marsden 1978, Abraham et al. 1988, Conlon 1993, Kobayashi and Nomizu

1996, Lee 2003]. In the entire work, unless otherwise stated, M is a manifold of dimension m that is real, second countable and $\mathcal{C}^\infty$.

The *tangent bundle* of M is denoted by TM. The canonical tangent projection assigns to each tangent vector its base point, $\tau_M \colon TM \to M$. For each base point x in M, the tangent space $T_x M$ at x is a $\mathbb{R}$–vector space. Thus, we can consider the dual space $T_x^* M$, which is called the *cotangent space* at x of M and is the set of $\mathbb{R}$–linear mappings from $T_x M$ to $\mathbb{R}$. The union of all the cotangent spaces for every x in M is called the *cotangent bundle* of M, denoted by $T^* M$. Elements in $T_x^* M$ are called *covectors* or *momenta* at the point $x \in M$. The canonical cotangent projection assigns to each covector its base point, $\pi_M \colon T^* M \to M$.

Given two manifolds M and N, we may consider a smooth mapping $f \colon M \to N$. The set of all these mappings is denoted by $\mathcal{C}^\infty(M, N)$. The tangent map of f is a mapping between the tangent bundles, $Tf \colon TM \to TN$. When $N = \mathbb{R}$, we have the set of smooth real–valued functions denoted by $\mathcal{C}^\infty(M)$.

A *vector field* X on M is a smooth mapping $X \colon M \to TM$ such that $\tau_M \circ X = \mathrm{Id}_M$. In other words, it assigns to each $x \in M$ the tangent vector $X(x) \in T_x M$. The set of all vector fields on M is denoted by $\mathfrak{X}(M)$. An *integral curve of a vector field* is a curve γ on M satisfying $\dot{\gamma}(t) = X(\gamma(t))$. Given an initial condition x_0, there always exists a unique integral curve $\phi_{x_0} \colon I \to M$ of X with that initial condition because of the results about existence and uniqueness of solutions for differential equations [Coddington and Levinson 1955]. The flow of X is a mapping $\phi \colon I \times M \to M$, such that $\phi(t, x_0) = \phi_{x_0}(t)$ and for every $t \in I$, $\phi_t \colon M \to M$ is a diffeomorphism on M given by $\phi_t(x) = \phi(t, x)$. Observe that $\phi_0(x) = x$ for every $x \in M$ and $\phi_{s+r} = \phi_s \circ \phi_r$ for $s, r \in I$.

In fact, the flow $\phi \colon I \times M \to M$ is only defined for the so–called *complete vector fields*. Otherwise, for every $x \in M$ there exists $\epsilon > 0$, a neighbourhood U_x of x and a mapping $\phi \colon (-\epsilon, \epsilon) \times U_x \to M$ with the same properties as the flow just defined [Abraham et al. 1988]. In the sequel, for simplicity it is assumed to have complete vector fields. If not, everything must be understood locally.

Similar to the definition of vector fields, a *1–form* ω on M is a mapping $\omega \colon M \to T^* M$ such that $\pi_M \circ \omega = \mathrm{Id}_M$. In other words, it assigns to each point x in M a covector $\omega(x) \in T_x^* M$. The set of all the 1–forms is denoted by $\Omega^1(M)$.

In fact, the vector fields and the 1–forms are particular cases of tensor fields on M. Given $r, s \in \mathbb{N} \cup \{0\}$, an *$r$–contravariant and s–covariant tensor field* T on M is a $\mathcal{C}^\infty$–section of $T_s^r(M) = (TM \otimes \overset{r}{\ldots} \otimes TM) \otimes (T^* M \otimes \overset{s}{\ldots} \otimes T^* M)$; that is, it associates to each point $x \in M$ an $\mathbb{R}$–multilinear mapping:

$$T(x) \colon \overbrace{T_x^* M \times \ldots \times T_x^* M}^{r \text{ times}} \times \overbrace{T_x M \times \ldots \times T_x M}^{s \text{ times}} \longrightarrow \mathbb{R}.$$

That geometric element is also called an (r, s)–tensor field on M. Thus a vector field is a $(1, 0)$–tensor field and a 1–form is a $(0, 1)$–tensor field. The set of all the tensor fields on M is denoted by $\mathcal{T}(M)$. The skew–symmetric s–covariant tensor fields are called *s–forms*. The set of all the s–forms is denoted by $\Omega^s(M)$. By convention $\Omega^0(M) = \mathcal{C}^\infty(M)$, then $\Omega(M) = \oplus_{s=0,\, s\in\mathbb{N}}^\infty \Omega^s(M)$ is the exterior algebra M with the addition and the exterior product

as operations. The *exterior derivative* is denoted by the mapping $\mathrm{d}\colon \Omega^s(M) \to \Omega^{s+1}(M)$. An s–form β is *closed* if $\mathrm{d}\beta = 0$ and it is *exact* if there exists $\eta \in \Omega^{s-1}(M)$ such that $\beta = \mathrm{d}\eta$. Given a vector field X and a s–form β, the *inner or interior product of X and β* is a $(s-1)$–form denoted by $i(X)\beta$ or $i_X\beta$. For a vector field X, $\mathcal{L}_X\colon T^r_s(M) \to T^r_s(M)$ is the *Lie derivative operator with respect to X*. The Lie derivative of a vector field Y with respect to X is exactly the Lie bracket of vector fields, $\mathcal{L}_X Y = [X, Y]$. The set $\mathfrak{X}(M)$ is a real Lie algebra with this Lie bracket.

Let $f\colon M \to N$ be a smooth mapping and $\omega \in \Omega^s(N)$. The *pullback $f^*\omega$ of ω by f* is given by $f^*\omega(x)(v_1, \ldots, v_s) = \omega(f(x))(T_x f(v_1), \ldots, T_x f(v_s))$, where $v_i \in T_x M$. Observe that the pullback defines a mapping $f^*\colon \Omega^s(N) \to \Omega^s(M)$. If f is a diffeomorphism, the pushforward f_* is defined by $f_* = \left(f^{-1}\right)_* = (f^*)^{-1}$. For a vector field X on M, the pushforward is the vector field f_*X on N given by $T_{f^{-1}(y)}f(X(f^{-1}(y)))$ for $y \in N$.

Let us refresh some basic notions about fiber bundles. For smooth manifolds M and B, a *differentiable fiber bundle* is an onto submersion $\pi\colon M \to B$ such that it is locally trivial. That is, there exists a manifold F such that, for every $b \in B$, there exists a neighbourhood W of b and a diffeomorphism $\varphi\colon \pi^{-1}(W) \to W \times F$ such that $\pi \circ \varphi^{-1}_{|W} = \mathrm{Id}_W$, and moreover, given two fiber bundle charts (W_1, ψ_1) and (W_2, ψ_2) adapted to π, the mapping $\psi_1 \circ \psi_2^{-1}$, called *transition function*, is a diffeomorphism. Here F is called the *typical fiber*. A *section of a fiber bundle* is a mapping $X\colon B \to M$ such that $\pi \circ X = \mathrm{Id}_B$. The set of sections is denoted by $\Gamma(\pi)$ or $\Gamma(B)$ if there is no doubt about the fiber bundle used. The *vertical subbundle $V(\pi)$ of π* is the subbundle of TM given by the kernel of $T\pi$. The set of *vertical vector fields with respect to π* are the vector fields on M taking values in $V(\pi)$. This set is denoted by $\mathfrak{X}^V(M, \pi)$.

A particular class of fiber bundles are the *vector bundles* whose fibers have the structure of vector space and the transition functions are isomorphisms. The set of sections of a vector bundle is a $\mathcal{C}^\infty(B)$–module.

Distributions and codistributions

Let M be a manifold. A *smooth regular distribution D on M* is a subbundle of TM of fiber k–dimensional. In other words, for every $x \in M$, D_x is a vector subspace of $T_x M$. The *rank of D at $x \in M$* is the dimension of the subspace D_x. A similar element can be defined as a subbundle of T^*M and is called a *codistribution*.

A distribution is called *involutive* if $[X, Y] \in \Gamma(D)$ for every $X, Y \in \Gamma(D)$. An *integrable submanifold of D* is a submanifold N such that $T_x N \subset D_x$ for every $x \in N$. A distribution D is *completely integrable* if, for every $x \in M$, there exists a maximal integrable submanifold N of D such that $T_x N = D_x$. Frobenius theorem guarantees that involutivity and integrability are equivalent as long as the distribution is regular.

Given a distribution D on M, the *annihilator of D* is a codistribution on M given by

$$\mathrm{ann}\, D_x = D_x^0 = D_x^\perp = \{\alpha \in T_x^*M \mid \alpha(v) = \langle \alpha, v \rangle = 0, \ \forall \, v \in D_x\} \qquad (2.1.1)$$

for every $x \in M$.

A foliation $\mathcal{F}$ of a smooth manifold M is a disjoint collection of immersed submanifolds on

M whose union is equal to M. Each of the submanifolds in $\mathcal{F}$ is called a *leaf* of the foliation. For an integrable distribution, the set of maximal connected integrable submanifolds defines a foliation.

Riemannian geometry

A *pseudo–Riemannian metric* on a manifold M is a symmetric non–degenerate section of $T_2^0(M)$. If it is positive definite on every fiber, we have a *Riemannian metric*. A *Riemannian manifold* is a pair (M, g) such that M is a smooth manifold and g is a Riemannian metric on M. A pseudo–Riemannian metric defines the musical isomorphisms of $\mathcal{C}^\infty(M)$–modules:

$$g^\sharp \colon \Omega^1(M) \to \mathfrak{X}(M), \quad g^\flat \colon \mathfrak{X}(M) \to \Omega^1(M).$$

The mapping $g^\flat$ is defined by $g^\flat(X) = i_X g \colon \mathfrak{X}(M) \to \mathbb{R}$, $Y \mapsto g(X, Y)$, and $g^\sharp$ is its inverse.

A Riemannian manifold has associated an affine connection; that is, a mapping

$$\begin{aligned} \nabla \colon \mathfrak{X}(M) \times \mathfrak{X}(M) &\longrightarrow \mathfrak{X}(M) \\ (X, Y) &\longmapsto \nabla(X, Y) = \nabla_X Y, \end{aligned}$$

satisfying the following properties:

1. it is $\mathbb{R}$–linear on X and on Y;

2. $\nabla_{fX} Y = f \nabla_X Y$;

3. $\nabla_X fY = f \nabla_X Y + (\mathcal{L}_X f) Y$;

for every $f \in \mathcal{C}^\infty(M)$.

The mapping $\nabla_X Y$ is called the *covariant derivative of Y with respect to X*. Given local coordinates (x^i) on M, the *Christoffel symbols for the affine connection* in these coordinates are given by

$$\nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^r} = \Gamma^i_{jr} \frac{\partial}{\partial x^i}.$$

From the properties of the affine connection, we have

$$\nabla_X Y = \left(X^j \frac{\partial Y^i}{\partial x^j} + \Gamma^i_{jr} X^j Y^r \right) \frac{\partial}{\partial x^i},$$

where $X = X^i \partial / \partial x^i$ and $Y = Y^i \partial / \partial x^i$.

Given $X \in \mathfrak{X}(M)$, the mapping $\nabla_X \colon \mathcal{T}(M) \to \mathcal{T}(M)$ is the natural extension of the affine connection as a derivation of order 0 that commutes with the inner product or contractions, see [Conlon 1993, Lee 2003] for more details.

If (M, g) is a Riemannian manifold, the *Levi–Civita connection* is the unique affine connection on M such that $\nabla_X Y - \nabla_Y X = [X, Y]$ and $\nabla_X g = 0$. Then the Christoffel symbols of the Levi–Civita connection are given in terms of the components of the metric as follows:

$$\Gamma^i_{jr} = \frac{1}{2} g^{il} \left(\frac{\partial g_{lj}}{\partial x^r} + \frac{\partial g_{lr}}{\partial x^j} - \frac{\partial g_{jr}}{\partial x^l} \right).$$

A curve $\gamma \colon I \to M$ on a Riemannian manifold is called a *geodesic* if $\nabla_{\dot{\gamma}(t)}\dot{\gamma}(t) = 0$. The tangent curve to γ is an integral curve of the *geodesic spray* Z; that is, a second–order vector field on TM that links with particular linear Ehresmann connections in §6.4.1.1. Locally, the integral curves of Z satisfy

$$\ddot{x}^i + \Gamma^i_{jr}\dot{x}^j\dot{x}^r = 0.$$

2.2 Time–dependent vector fields

As we will see in §3.1, control systems are associated to a time–dependent vector field through a vector field along a projection defined in §2.2.1. For $I \subset \mathbb{R}$, a *differentiable time–dependent vector field* X is a mapping $X \colon I \times M \to TM$ such that each $(t, x) \in I \times M$ is assigned to a tangent vector $X(t, x)$ in $T_x M$. For every $(s, x) \in I \times M$, *the integral curve of X with initial condition* (s, x) is denoted by $\Phi^X_{(s,x)} \colon J_{(s,x)} \subset I \to M$ and satisfies

1. $\Phi^X_{(s,x)}(s) = x$ and

2. $\left.\dfrac{\mathrm{d}}{\mathrm{d}t}\right|_t \Phi^X_{(s,x)} = X\left(t, \Phi^X_{(s,x)}(t)\right)$, $t \in J_{(s,x)}$.

The domain of $\Phi^X_{(s,x)}$ is denoted by $J_{(s,x)} \subset I$ because the domain depends on the initial condition for the integral curves.

The *time–dependent flow* or *evolution operator of X* is the mapping

$$\Phi^X \colon \quad \begin{aligned} I \times I \times M &\longrightarrow M \\ (t, s, x) &\longmapsto \Phi^X(t, s, x) = \Phi^X_{(s,x)}(t) \end{aligned} \tag{2.2.2}$$

and Φ^X satisfies

1. $\Phi^X(s, s, x) = x$ and

2. $\left.\dfrac{\mathrm{d}}{\mathrm{d}t}\right|_t \left(\Phi^X(t, s, x)\right) = X\left(t, \Phi^X(t, s, x)\right)$.

To be more precise, the evolution operator is only defined in a maximal open neighborhood of $\Delta_I \times M$, where Δ_I is the diagonal of $I \times I$, unless the completeness of the vector field is assumed.

To obtain the original vector field through the evolution operator, the expression in the above second assertion must be evaluated at $s = t$,

$$\left[\left.\dfrac{\mathrm{d}}{\mathrm{d}t}\right|_t \left(\Phi^X(t, s, x)\right)\right]\Bigg|_{s=t} = X(t, x).$$

There is a time–independent vector field on the manifold $I \times M$ associated to X and given by $\widetilde{X}(t, x) = (\partial/\partial t + X(t, x))_{(t,x)}$. For $(t, s, x) \in I \times I \times M$, the flow of $\widetilde{X}$ is $\Phi^{\widetilde{X}} \colon I \times I \times M \to$

$I \times M$ such that $\Phi^{\widetilde{X}}_{(s,x)}$ is the integral curve of $\widetilde{X}$ with initial condition (s,x) at time 0 and $\Phi^{\widetilde{X}}(t,(s,x)) = (s+t, \Phi^X(s+t,(s,x)))$. The theorems in differential equations about the existence and uniqueness of solutions guarantee the existence and uniqueness of the evolution operator Φ^X.

For $(t,s) \in I \times I$,

$$
\begin{aligned}
\Phi^X_{(t,s)} \colon M &\longrightarrow M \\
x &\longmapsto \Phi^X_{(t,s)}(x) = \Phi^X_{(s,x)}(t)
\end{aligned}
$$

is a diffeomorphism on M satisfying $\Phi^X_{(t,s)} = \Phi^X_{(t,r)} \circ \Phi^X_{(r,s)}$ for $r \in I$. For more details see [Kolář et al. 1993].

2.2.1   Vector fields along a projection

A natural way to understand control theory in differential geometry is by means of the notion of a vector field depending on parameters, see §3.1. Properties about how the integral curves of differential equations depending on parameters evolve are explained in [Coddington and Levinson 1955, Hairer 1999, Kolář et al. 1993] and used in §4.1.3 and in §4.3.2.

Let M be a differentiable manifold of dimension m and U be a set in $\mathbb{R}^k$. Consider the projection $\pi \colon M \times U \to M$.

Definition 2.2.1. *A **vector field** X on M **along the projection** π is a mapping $X \colon M \times U \to TM$ such that X is continuous on $M \times U$, continuously differentiable on M for every $u \in U$ and $\tau_M \circ X = \pi$, where $\tau_M \colon TM \to M$ is the canonical tangent projection.*

The set of vector fields along the projection π is denoted by $\mathfrak{X}(\pi)$. If (W, x^i) is a local chart at x in M, then a vector field X along the projection is given locally by $f^i \partial/\partial x^i$, where f^i are functions defined on $W \times U$.

Let $I = [a,b] \subset \mathbb{R}$ be a closed interval, $(\gamma, u) \colon I \to M \times U$ is an integral curve of X if $\dot{\gamma}(t) = X(\gamma(t), u(t))$. All these elements come together in Diagram (2.2.3).

$$(2.2.3)$$

In other words, X is a vector field depending on parameters in U. In this work, the parameters are called *controls* and are assumed to be measurable mappings $u \colon I \to U$ such that $\operatorname{Im} u$ is bounded. Given the parameter u, we have a time–dependent vector field on M,

$$
\begin{aligned}
X^{\{u\}} \colon I \times M &\longrightarrow TM \\
(t,x) &\longmapsto X^{\{u\}}(t,x) = X(x, u(t)).
\end{aligned}
\qquad (2.2.4)
$$

For an integral curve (γ, u) of X, it is said that γ is an integral curve of $X^{\{u\}}$, as shown in the following commutative diagram:

$$
\begin{array}{ccc}
I \times M & \xrightarrow{X^{\{u\}}} & TM \\
{\scriptstyle(\gamma,\,\mathrm{Id})}\Big\uparrow & {\scriptstyle\dot\gamma}\nearrow \quad \Big\uparrow {\scriptstyle X} & \\
I & \xrightarrow[(\gamma,u)]{} & M \times U
\end{array}
\tag{2.2.5}
$$

That is, $X^{\{u\}} \circ (\gamma, \mathrm{Id}) = \dot\gamma = X \circ (\gamma, u)$.

Remark 2.2.2. As the controls $u\colon I \to U$ are measurable and bounded, the vector fields $X^{\{u\}}$ are measurable on t, and for a fixed t, they are differentiable on M. Hence, the notion of Carathéodory vector fields must be considered [Cañizo–Rincón 2004, Coddington and Levinson 1955, Filippov 1988] from now on. Then, we only consider absolutely continuous curves $\gamma\colon I \to M$ to be *generalized integral curves* of the vector field $X^{\{u\}}$; that is, they only satisfy $\dot\gamma = X \circ (\gamma, u)$ at points where γ is derivable, which happens almost everywhere. The existence and uniqueness of these integral curves are guaranteed once the parameter is fixed because of the theorems of existence and uniqueness of differential equations depending on parameters. For more details about absolute continuity, see Appendix A and [Cañizo–Rincón 2004, Coddington and Levinson 1955, Varberg 1965].

2.2.2 Time–dependent variational equations

The variational equations give us an approach to how the integral curves of vector fields vary when the initial condition varies along a curve. These equations have a formulation on the tangent and the cotangent bundle. Here we are interested in studying the variational equations associated to time–dependent vector fields defined in §2.2, and in proving the relationship stated in §2.2.2.3 between the solutions of variational equations on the tangent bundle described in §2.2.2.1 and the ones on the cotangent bundle in §2.2.2.2. See [Kolář et al. 1993] for more details about the required concepts.

2.2.2.1 Complete lift

From the evolution operator of a time–dependent vector field on M in Equation (2.2.2), it is determined the evolution operator of a particular time–dependent vector field on TM.

Let $X_t\colon M \to TM$ be a vector field on M such that $X_t(x) = X(t, x)$ for every $t \in I$. The *complete or tangent lift of X_t to TM* is the time–dependent vector field X_t^T on TM satisfying

$$
X_t^T = \kappa_M \circ TX_t,
$$

where κ_M is the canonical involution of TTM; that is, a mapping $\kappa_M\colon TTM \to TTM$ such that $\kappa_M^2 = \mathrm{Id}$ and $\tau_{TM} \circ \kappa_M = T\tau_M$. See [Kolář et al. 1993] for more details in the definition. Moreover, observe that X_t is a vector field that makes Diagram (2.2.6). If $(x, v) \in TM$, then $TX_t(x, v) = (x, X_t(x), T_x X_t(v)) \in T_{(x, X_t(x))}(TM)$.

Let (W, x^i) be a local chart at x in M such that $X_t = X_t^i \, \partial/\partial x^i$ where $X_t^i(x) = X^i(t, x)$ and $X^i \in \mathcal{C}^\infty(I \times W)$. If (x^i, v^i) are the induced local coordinates in TM, then locally

$$X^T(t, x, v) = X_t^T(x, v) = X^i(t, x) \left. \frac{\partial}{\partial x^i} \right|_{(x,v)} + \frac{\partial X^i}{\partial x^j}(t, x) \, v^j \left. \frac{\partial}{\partial v^i} \right|_{(x,v)} .$$

The equations satisfied by the integral curves of X^T are called *variational equations for X*.

$$
\begin{array}{ccc}
TTM & \xrightarrow{\;\tau_{TM}\;} & TM \\
T\tau_M \Big\Updownarrow TX_t & \tau_M \;\;\;\; \Big\Updownarrow X_t & \\
TM & \xrightarrow{\;\;\tau_M\;\;} & M
\end{array}
\qquad (2.2.6)
$$

Proposition 2.2.3. *If X is a time–dependent vector field on M and Φ^X is the evolution operator of X, then the map $\Psi \colon I \times I \times TM \to TM$ defined by*

$$\Psi(t, s, (x, v)) = \left(\Phi^X(t, s, x), T_x \Phi^X_{(t,s)}(v) \right)$$

is the evolution operator of the complete lift X^T of X.

(*Proof*) We have to prove that

1. $\Psi(s, s, (x, v)) = (x, v)$;

2. $\left. \dfrac{\mathrm{d}}{\mathrm{d}t} \right|_t (\Psi(t, s, (x, v))) = X^T(t, \Psi(t, s, (x, v)))$.

The first item is proved easily,

$$\Psi(s, s, (x, v)) = \left(\Phi^X(s, s, x), T_x \left(\Phi^X_{(s,s)} \right)(v) \right) = (x, v),$$

because $\Phi^X_{(s,s)} = \mathrm{Id}$.

As for the second assertion, we use that $\Phi^X_{(t,s)} \colon M \to M$ is a $\mathcal{C}^\infty$ diffeomorphism satisfying

$$
\begin{aligned}
T_t \left(T_x \Phi^X_{(t,s)}(v) \right) 1 &= \left. \frac{\mathrm{d}}{\mathrm{d}t} \right|_t \left(T_x \Phi^X_{(t,s)}(v) \right) 1 = \left(T_x \left(\left. \frac{\mathrm{d}}{\mathrm{d}t} \right|_t \left(\Phi^X_{(t,s)} \right) 1 \right) \right)(v) \\
&= T_x \left(T_t \Phi^X_{(s,x)} 1 \right)(v),
\end{aligned}
$$

and we obtain

$$
\left. \frac{\mathrm{d}}{\mathrm{d}t} \right|_t (\Psi(t, s, x, v)) = \left(\left. \frac{\mathrm{d}}{\mathrm{d}t} \right|_t \left(\Phi^X(t, s, x) \right), \left. \frac{\mathrm{d}}{\mathrm{d}t} \right|_t \left(T_x \Phi^X_{(t,s)}(v) \right) \right)
$$

$$
= \left(X(t, \Phi^X(t, s, x)), \left(T_x \left(\left. \frac{\mathrm{d}}{\mathrm{d}t} \right|_t \left(\Phi^X_{(t,s)} \right) \right) \right)(v) \right)
$$
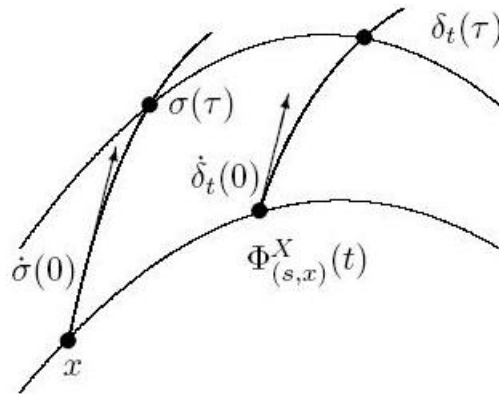
Figure 2.1: *Idea of the linear approximation of integral curves of a vector field.*

$$= \left(X(t, \Phi^X(t, s, x)), \left(T_x \left(X_t(\Phi^X(t, s, x)) \right) \right)(v) \right)$$

$$= \left(X(t, \Phi^X(t, s, x)), \left(T_{\Phi^X(t,s,x)} \left(X_t \right) \circ T_x(\Phi^X(t, s, x)) \right)(v) \right)$$

$$= \left(X(t, \Phi^X(t, s, x)), T_{\Phi^X(t,s,x)} \left(X_t \right) \left(T_x(\Phi^X_{(t,s)})(v) \right) \right) = X^T(t, \Psi(t, s, x, v)).$$

Hence, the evolution operator of X^T is the complete lift of the evolution operator of X. The integral curves of X^T are vector fields along the integral curves of X. ∎

About the geometric meaning of the complete lift

The integral curves of X^T must be understood as the linear approximation of the integral curves of X when the initial condition varies along a curve in M. This idea will appear again in Chapter 4.

Let us explain Figure 2.1. Given an integral curve of X with initial condition (s, x), we consider a curve σ starting at the point x of the integral curve. Every point of σ can be considered as the initial condition at time s for an integral curve of X. Thus the flow of X transports the curve σ at a different curve δ_t point by point. The resultant curve is related with the complete lift of X as the following results prove.

Proposition 2.2.4. *Let $X \colon I \times M \to TM$ be a time–dependent vector field with evolution operator Φ^X and $(s, x) \in I \times M$. For $\epsilon > 0$, let $\sigma \colon (-\epsilon, \epsilon) \subset \mathbb{R} \to M$ be a $\mathcal{C}^\infty$ curve such that $\sigma(0) = x = \Phi^X(s, s, x)$. For every $t \in I$, consider the curve*

$$\begin{aligned} \delta_t \colon (-\epsilon, \epsilon) &\longrightarrow M \\ \tau &\longmapsto \delta_t(\tau) = \Phi^X_{(s,\sigma(\tau))}(t). \end{aligned}$$

Then $\dot{\delta}_t(0) = T_x \Phi^X_{(t,s)}(\dot{\sigma}(0))$.

(*Proof*)

$$
\begin{aligned}
\dot{\delta}_t(0) &= (T_0 \delta_t(\tau)) \left. \frac{\mathrm{d}}{\mathrm{d}\tau} \right|_0 = \left(T_0 \left(\Phi^X_{(s,\sigma(\tau))}(t) \right) \right) \left. \frac{\mathrm{d}}{\mathrm{d}\tau} \right|_0 = \left(T_0 \left(\Phi^X_{(t,s)}(\sigma(\tau)) \right) \right) \left. \frac{\mathrm{d}}{\mathrm{d}\tau} \right|_0 \\
&= T_{\sigma(0)} \Phi^X_{(t,s)} \left(T_0(\sigma(\tau)) \left. \frac{\mathrm{d}}{\mathrm{d}\tau} \right|_0 \right) = T_{\sigma(0)} \Phi^X_{(t,s)} (\dot{\sigma}(0)) = T_x \Phi^X_{(t,s)} (\dot{\sigma}(0)) \, .
\end{aligned}
$$

∎

Observe that the curve δ_t satisfies the following properties

1. $\delta_s(\tau) = \sigma(\tau)$, and

2. $\delta_t(0) = \Phi^X_{(s,x)}(t)$.

Corollary 2.2.5. *Let X be a time–dependent vector field on M. For $x \in M$, $v \in T_x M$ and for $\epsilon > 0$, let $\sigma \colon (-\epsilon, \epsilon) \subset \mathbb{R} \to M$ be a $\mathcal{C}^\infty$ curve such that $\sigma(0) = x$ and $\dot{\sigma}(0) = v$. If δ_t is the curve defined in Proposition 2.2.4, then $\dot{\delta}_{(\cdot)}(\tau) \colon I \to TM$, $t \mapsto \dot{\delta}_t(\tau)$ is the integral curve of X^T with initial condition $(s, \dot{\sigma}(\tau))$.*

(*Proof*) The proof just comes from Propositions 2.2.3 and 2.2.4 and the definition of the curve δ_t. ∎

2.2.2.2   Cotangent lift

Given $(t, s) \in I \times I$, the evolution operator Φ^X defines the following diffeomorphism on TM

$$
T\Phi^X_{(t,s)} \colon TM \longrightarrow TM,
$$

which is a linear isomorphism on the fibers on TM. We consider the mapping

$$
\Lambda_{(t,s)} \colon T^* M \longrightarrow T^* M \tag{2.2.7}
$$

to be the inverse of the transpose of $T\Phi^X_{(t,s)}$ on every fiber of $T^* M$; that is, for $p \in T^*_x M$ and $v \in T_{\Phi^X_{(t,s)}(x)} M$,

$$
\left(\Lambda_{(t,s)}(x, p) \right) \left(\Phi^X_{(t,s)}, v \right) = \left\langle p, \left(T_x \Phi^X_{(t,s)} \right)^{-1} (v) \right\rangle = \left\langle p, T_{\Phi^X_{(t,s)}(x)} \left(\Phi^X_{(t,s)} \right)^{-1} (v) \right\rangle .
$$

The mapping $\Lambda \colon I \times I \times T^* M \to T^* M$, $\Lambda(t, s, (x, p)) = \Lambda_{(t,s)}(x, p)$, is the evolution operator of a vector field on $T^* M$ called the *cotangent lift of X* and denoted by X^{T^*}. The intrinsic expression of the flow of X^{T^*} is given by

$$
\Lambda(t, s, (x, p)) = \left(\Phi^X(t, s, x), \left({}^\tau \left(T_x \Phi^X_{(t,s)} \right) \right)^{-1} (p) \right),
$$

where $\left({}^\tau \left(T_x \Phi^X_{(t,s)} \right) \right)^{-1}$ is the explicit way to write $\Lambda_{(t,s)}$ in Equation (2.2.7).

In local coordinates (x, p) for T^*M,

$$X^{T^*}(t, x, p) = X_t^{T^*}(x, p) = X^i(t, x) \left. \frac{\partial}{\partial x^i} \right|_{(x,p)} - \frac{\partial X^j}{\partial x^i}(t, x) \, p_j \left. \frac{\partial}{\partial p_i} \right|_{(x,p)}.$$

The equations satisfied by the integral curves of the cotangent lift in the fibers are the *adjoint variational equations on the cotangent bundle for* X. In the literature, they are sometimes called *adjoint equations for* X.

2.2.2.3 A property for the complete and cotangent lift

The previous propositions allow us to determine an invariant function along integral curves of X.

Proposition 2.2.6. *Let* $X : I \times M \to TM$ *be a time–dependent vector field and let* $X^T : I \times TM \to TTM$ *and* $X^{T^*} : I \times T^*M \to TT^*M$ *be the complete lift and cotangent lift of* X, *respectively. If* $\gamma : I \to M$ *is an integral curve of* X *with initial condition* (s, x), $V : I \to TM$ *is the integral curve of* X^T *with initial condition* (s, v) *where* $v \in T_x M$, *and* $\Lambda : I \to T^*M$ *is the integral curve of* X^{T^*} *with initial condition* (s, p) *where* $p \in T_x^* M$, *then*

$$\langle \Lambda, V \rangle \colon I \quad \to \quad \mathbb{R}$$
$$t \quad \mapsto \quad \langle \Lambda(t), V(t) \rangle$$

is constant along γ.

(*Proof*) If Φ^X is the evolution operator of X, the evolution operators of X^T and X^{T^*} are

$$\Phi^{X^T}(t, s, (x, v)) = \left(\Phi^X(t, s, x), T_x \Phi^X_{(t,s)}(v) \right),$$
$$\Phi^{X^{T^*}}(t, s, (x, p)) = \left(\Phi^X(t, s, x), \left({}^\tau T_x \Phi^X_{(t,s)} \right)^{-1} (p) \right),$$

respectively, because of Proposition 2.2.3 and §2.2.2.2. Hence

$$\langle \Lambda(t), V(t) \rangle = \left\langle \left({}^\tau T_x \Phi^X_{(t,s)} \right)^{-1}(p), T_x \Phi^X_{(t,s)}(v) \right\rangle$$
$$= \left\langle {}^\tau \left(\left(T_x \Phi^X_{(t,s)} \right)^{-1} \right)(p), T_x \Phi^X_{(t,s)}(v) \right\rangle$$
$$= \left\langle p, \left(\left(T_x \Phi^X_{(t,s)} \right)^{-1} \circ \left(T_x \Phi^X_{(t,s)} \right) \right)(v) \right\rangle = \langle p, v \rangle = \text{constant}.$$

■

2.3 Symplectic geometry

Let M be a smooth manifold and $\Omega \in \Omega^2(M)$. Given $x \in M$, the kernel of Ω at x is defined by

$$\ker \Omega_x = \{ v \in T_x M \mid i_v \Omega = 0 \},$$

which is a subspace of the tangent space $T_x M$. It is said that Ω is *regular* if the dimension of $\ker \Omega$ does not depend on the point $x \in M$.

Under the assumption of regularity of Ω,

$$\ker \Omega = \bigcup_{x \in M} \ker \Omega_x$$

is a vector subbundle of the tangent bundle TM; that is, a regular distribution on M. The set of all the vector fields $X \in \mathfrak{X}(M)$ such that $X(x) \in \ker \Omega_x$ for all $x \in M$ is also denoted by $\ker \Omega$. The vector subbundle $\ker \Omega$ is involutive if and only if Ω is a closed form.

2.3.1 Symplectic manifolds

A *symplectic manifold* is a pair (M, Ω) where M is a m–dimensional manifold and Ω is a closed non–degenerate 2–form on M. Note that to have a symplectic manifold, the dimension of M must be even. For instance, the cotangent bundle has associated a symplectic structure.

The non–degeneracy of Ω guarantees that $\Omega^\flat \colon \mathfrak{X}(M) \to \Omega^1(M)$, defined by the inner product $\Omega^\flat(X) = i_X \Omega$, is a $\mathcal{C}^\infty(M)$–module isomorphism. The inverse of $\Omega^\flat$ is $\Omega^\sharp$. These two mappings are the so–called *canonical musical isomorphisms*. The *Hamiltonian vector field X_f on M associated with* $f \in \mathcal{C}^\infty(M)$ is $X_f = \Omega^\sharp(\mathrm{d}f)$, where $\mathrm{d} \colon \Omega^s(M) \to \Omega^{s+1}(M)$ is the exterior derivative.

A vector field X on M is *locally Hamiltonian* if Ω is invariant under the vector field; that is, $\mathcal{L}_X \Omega = 0$, where $\mathcal{L}$ is the Lie derivative with respect to X of any tensor field. The invariancy of Ω under Y implies that the 1–form $i_X \Omega$ is closed and, by Poincaré's Lemma, it is locally exact.

A *regular Hamiltonian system* is given by (M, Ω, α) where (M, Ω) is a symplectic manifold and α is a closed 1–form on M. Poincaré's Lemma guarantees that given $x \in M$ there exists an open neighbourhood W of x and $H \in \mathcal{C}^\infty(W)$ such that $\alpha_{|W} = \mathrm{d}H$. Then, (W, Ω, H) is called a *locally Hamiltonian system*. If α is an exact form, then $\alpha = \mathrm{d}H$ and H is called the *global Hamiltonian function*. Thus, a regular Hamiltonian system is given by (M, Ω, H) and is associated with Hamilton's equation $i_{X_H} \Omega = \mathrm{d}H$ that always has a unique solution under the hypothesis of symplecticity.

A natural way to define a Hamiltonian system on the cotangent bundle T^*M is by means of a vector field X on M. We take $H_X \colon T^*M \to \mathbb{R}$, $H_X(p_x) = \langle p, X(x) \rangle$ where $p \in T_x^*M$, as a Hamiltonian function to obtain the Hamiltonian system (T^*M, Ω, H_X), with Ω being the natural 2–form on T^*M. The corresponding Hamiltonian vector field is the cotangent lift of X defined in §2.2.2.2, as proved in [Barbero-Liñán and Muñoz Lecanda 2008b]. This proof follows analogously to the proof of Proposition 2.2.3.

2.3.2 Presymplectic constraint algorithm

The Dirac–Bergmann theory of constraints developed in the fifties for quantum field theory gave rise to the presymplectic constraint algorithm, which has been already adapted and used to

study singular optimal control problems [Delgado-Téllez and Ibort 2003] and to study optimal control problems with nonholonomic constraints [León et al. 2004]. In this dissertation, this constraint algorithm will be used in Chapters 5 and 6.

A *presymplectic form* on M is a closed and regular 2–form. A *presymplectic manifold* is a manifold M with a presymplectic form $\Omega \in \Omega^2(M)$. It is obvious that a symplectic manifold is presymplectic with $\ker \Omega = \{0\}$.

If (M, Ω) is a presymplectic manifold, some usual notions of symplectic manifolds explained in §2.3.1 also appear here. So if $H \in \mathcal{C}^\infty(M)$, we may consider the equation

$$i_X \Omega = \mathrm{d}H, \tag{2.3.8}$$

where the unknown, the vector field $X \in \mathfrak{X}(M)$, is called the Hamiltonian vector field associated with the Hamiltonian function H.

If $\ker \Omega \neq \{0\}$, the mapping $\Omega^\flat \colon TM \to T^*M$ given by $\Omega^\flat(v_x) = i_{v_x} \Omega$ is not onto. Thus Equation (2.3.8) does not always have a solution. It is indispensable to claim for $\mathrm{d}H \in \mathrm{Im}\,\Omega^\flat$. This condition can depend on the point $x \in M$ where we compute $\Omega_x^\flat$. With this in mind, we define a *presymplectic system* (M, Ω, H) as a presymplectic manifold (M, Ω) and a function $H \in \mathcal{C}^\infty(M)$.

Statement 2.3.1. *(Presymplectic problem) Given a presymplectic system* (M, Ω, H)*, find a pair* (N, X) *such that*

    (a) N *is a submanifold of* M,

    (b) $X \in \mathfrak{X}(M)$ *is tangent to* N *on* N, *and*

    (c) N *is maximal among all the submanifolds satisfying* (a) *and* (b).

The solution to this problem gives rise to the so–called presymplectic algorithm described as follows; see [Cariñena 1990, Gotay and Nester 1979] for more details. The condition (c) cannot be assured.

Step zero: Let $N_0 = \{x \in M \mid \exists\, v_x \in T_x M,\ i_{v_x} \Omega = \mathrm{d}_x H\}$, which is called the *primary constraint submanifold*.

Proposition 2.3.2. $N_0 = \{x \in M \mid (\mathcal{L}_Z H)_x = 0\,,\ Z \in \ker \Omega\}$.

(*Proof*) It is a straightforward consequence of the fact that if $\alpha_x \in T_x^* M$, we have $\alpha_x \in \mathrm{Im}\,\Omega_x^\flat$ if and only if $\ker \Omega_x \subset \ker \alpha_x$. ∎

On N_0 there exists a solution of the presymplectic equation (2.3.8), but the solution is not unique. Indeed, if X_0 is a solution, then $X_0 + \ker \Omega$ is the set of all the solutions. We may consider X_0 as a vector field defined on M because N_0 is closed. So X_0 defined on N_0 can be extended to M by using partitions of unity on M.

We assume that N_0 is a submanifold of M.

Take the pair $(N_0, X_0 + \ker \Omega)$, rewritten as (N_0, X^{N_0}) where X^{N_0} denotes the set of all the vector fields solving the problem on N_0.

Step one: Now let
$$N_1 = \{x \in N_0 \,|\, \exists\, X \in X^{N_0} \,,\; X(x) \in T_x N_0\},$$

providing a new pair (N_1, X^{N_1}) where X^{N_1} is the set of the vector fields solution and we assume again that N_1 is a submanifold. This step is usually called *stabilization step* or it is said that the *tangency condition* is imposed.

Observe that the vector fields X^{N_1} are tangent to N_0, but not necessarily to N_1. Hence, inductively, we arrive at (N_i, X^{N_i}) where we assume that N_i is a submanifold of M and we define
$$N_{i+1} = \{x \in N_i \,|\, \exists\, X \in X^{N_i} \,,\quad X(x) \in T_x N_i\}.$$

So we obtain the sequence

$$M \supseteq N_0 \supseteq N_1 \supseteq \ldots \supseteq N_i \supseteq N_{i+1} \supseteq \ldots .$$

Let

$$N_f = \bigcap_{i \geq 0} N_i, \quad X^{N_f} = \bigcap_{i \geq 0} X^{N_i}.$$

If N_f is a nontrivial submanifold of M, then (N_f, X^{N_f}) is the solution to the problem. If at one step $N_i = N_{i+1}$, the final submanifold is N_i. It could be a discrete set of points or even an empty set.

Observe that each step of the algorithm can reduce the set of points of M where there exist solutions; that is, $N_i \subseteq N_{i-1}$, and can also reduce the degrees of freedom of the set of vector fields solution, $X^{N_i} \subseteq X^{N_{i-1}}$.

2.4 Ehresmann connections

There are different ways to describe a connection. Some similar descriptions for a connection are given in [Echeverría-Enríquez et al., León and Rodrigues 1989, Saunders 1989]. In §2.4.2, we introduce an equivalent definition for a connection given by an almost product structure used in [León and Martín de Diego 1996] to study constrained dynamics. From §2.4.4 on, we focus on particular Ehresmann connections useful in Chapter 6.

2.4.1 Notion of connection

Let B be a differentiable manifold and $\pi : E \to B$ be a differentiable fiber bundle with typical fiber F. Let $\dim B = m$ and $\dim F = n$. We denote by $\Gamma(B, E)$ or $\Gamma(\pi)$ the set of global sections of π. So, if $\varphi \in \Gamma(\pi)$, then $\varphi : B \to E$ is a differentiable mapping and $\pi \circ \varphi = \mathrm{Id}_B$.

Let (W, x^μ) be a local chart at x in B for $\mu = 1, \ldots, m$ and y^A be a local coordinate system

in the fibers $\pi^{-1}(x)$ where $x \in W$ for $A = 1, \ldots, n$. The following proposition is proved in [Echeverría-Enríquez et al., Saunders 1989]

Proposition 2.4.1. *Let $\pi : E \to B$ be a fiber bundle. The following elements can be canonically constructed one from the other:*

1. *A π–semibasic 1–form ∇ on E with values in TE; that is,*

$$\nabla \in \Gamma\left(E, \pi^*(T^*B)\right) \otimes \Gamma\left(E, TE\right),$$

   such that $\nabla^*\alpha = \alpha$ for every π–semibasic form $\alpha \in \Omega^1(E)$.

2. *A subbundle $H(E)$ of TE such that*

$$TE = H(E) \oplus V(\pi). \tag{2.4.9}$$

3. *A (global) section of $\pi^1 \colon J^1E \to E$; that is, a mapping $\Psi \colon E \to J^1E$ such that $\pi^1 \circ \Psi = \mathrm{Id}_E$.*

Recall that a k–form Ω on E is π–*semibasic* if $i_Y\Omega = 0$ for every $Y \in \mathfrak{X}^V(E, \pi) = \Gamma(V(\pi)) = \ker T\pi$. In a local coordinate system (x^μ, y^A) adapted to π, the vertical bundle associated with π is spanned by $\left\{\frac{\partial}{\partial y^A}\right\}$ for $A = 1, \ldots, n$ and the most general local expression of a semibasic 1–form on E with values in TE is

$$\nabla = f_\mu \mathrm{d}x^\mu \otimes \left(g^\nu \frac{\partial}{\partial x^\nu} + h^A \frac{\partial}{\partial y^A}\right),$$

where $f_\mu, g^\nu, h^A \in \mathcal{C}^\infty(E)$. As ∇^* is the identity on semibasic forms, it follows that $\nabla^* dx^\mu = dx^\mu$, so the local expression of a connection form ∇ is

$$\nabla = \mathrm{d}x^\mu \otimes \left(\frac{\partial}{\partial x^\mu} + \Gamma_\mu^A \frac{\partial}{\partial y^A}\right), \tag{2.4.10}$$

where $\Gamma_\mu^A \in \mathcal{C}^\infty(E)$. For every $e \in E$, the horizontal subspace $H_e(E)$ of $T_e E$ associated with ∇ is locally spanned by

$$\left(\frac{\partial}{\partial x^\mu} + \Gamma_\mu^A \frac{\partial}{\partial y^A}\right)_e,$$

for $\mu = 1, \ldots, m$. Hence $H(E) = \mathrm{Im}\nabla$.

The splitting $(2.4.9)$ of TE determines the projectors

$$h_\nabla : TE \longrightarrow H(E), \quad v_\nabla : TE \longrightarrow V(\pi), \tag{2.4.11}$$

such that

$$h_\nabla^2 = h_\nabla, \quad v_\nabla^2 = v_\nabla, \quad h_\nabla v_\nabla = v_\nabla h_\nabla = 0.$$

A concept related to a connection is the curvature.

Definition 2.4.2. *The curvature of a connection* ∇ *is a* $(1, 2)$*–tensor field on* E *given by*

$$\mathcal{R}_\nabla \left(Z_1, Z_2 \right) : \ = \left(\mathrm{Id} - \nabla \right) \left(\left[\nabla \left(Z_1 \right), \nabla \left(Z_1 \right) \right] \right) = i \left(\left[\nabla \left(Z_1 \right), \nabla \left(Z_1 \right) \right] \right) \left(\mathrm{Id} - \nabla \right)$$

for every $Z_1,\ Z_2 \in \mathfrak{X}(E)$.

The curvature measures the lack of integrability of the horizontal subbundle. This subbundle is integrable if the curvature is zero. Then it is said that the connection is flat.

Using the local expression of the connection form ∇ in fibered coordinates (2.4.10), we have

$$\mathcal{R}_\nabla = \frac{1}{2} \left(\frac{\partial \Gamma_\eta^B}{\partial x^\mu} - \frac{\partial \Gamma_\mu^B}{\partial x^\eta} + \Gamma_\mu^A \frac{\partial \Gamma_\eta^B}{\partial y^A} - \Gamma_\eta^A \frac{\partial \Gamma_\mu^B}{\partial y^A} \right) \left(\mathrm{d}x^\mu \wedge \mathrm{d}x^\eta \right) \otimes \frac{\partial}{\partial y^B}.$$

2.4.2 Connection associated with an almost product structure on a fiber bundle

We give a notion of connection equivalent to the ones in Proposition 2.4.1. See [León and Rodrigues 1989, León and Martín de Diego 1996] for more details about the following geometric structure.

Definition 2.4.3. *Let* $\pi \colon E \ \to \ B$ *be a fiber bundle. An **almost product structure** Γ is a* $(1, 1)$*–tensor field on* E *such that* $\Gamma^2 = \mathrm{Id}$ *and* $\ker(\Gamma + \mathrm{Id}) = \mathrm{V}(\pi)$.

In local coordinates (x^μ, y^A) adapted to π,

$$\Gamma = \mathrm{d}x^\mu \otimes \left(\frac{\partial}{\partial x^\mu} + f_\mu^A \frac{\partial}{\partial y^A} \right) - \mathrm{d}y^A \otimes \frac{\partial}{\partial y^A}.$$

The almost product structure is not a projector, but its associated horizontal projector is $h_\Gamma = \frac{1}{2}(\mathrm{Id}+\Gamma)$. Then, the corresponding Nijenhuis bracket of vector fields on E is defined as follows

$$\begin{aligned}
\frac{1}{2} \left[h_\Gamma, h_\Gamma \right] \left(Z_1, Z_2 \right) \ &= \ \left[h_\Gamma \left(Z_1 \right), h_\Gamma \left(Z_2 \right) \right] - h_\Gamma \left(\left[h_\Gamma \left(Z_1 \right), Z_2 \right] + \left[Z_1, h_\Gamma \left(Z_2 \right) \right] \right) \\
&\quad + \ h_\Gamma^2 \left[Z_1, Z_2 \right]
\end{aligned}$$

for $Z_1,\ Z_2 \in \mathfrak{X}(E)$.

Definition 2.4.4. *The curvature* $\mathcal{R}_\Gamma$ *of the almost product structure* Γ *is a* $(1, 2)$*–tensor field on* E *such that*

$$\mathcal{R}_\Gamma = \frac{1}{2} \left[h_\Gamma, h_\Gamma \right],$$

where h_Γ *is the horizontal projector associated with* Γ *and* $[h_\Gamma, h_\Gamma]$ *is the Nijenhuis bracket.*

It can be checked that

$$\begin{aligned}
\mathcal{R}_\Gamma \left(h_\Gamma \left(Z_1 \right), h_\Gamma \left(Z_2 \right) \right) \ &= \ v_\Gamma \left(\left[h_\Gamma \left(Z_1 \right), h_\Gamma \left(Z_2 \right) \right] \right), \\
\mathcal{R}_\Gamma \left(h_\Gamma \left(Z_1 \right), v_\Gamma \left(Z_2 \right) \right) \ &= \ 0, \\
\mathcal{R}_\Gamma \left(v_\Gamma \left(Z_1 \right), v_\Gamma \left(Z_2 \right) \right) \ &= \ 0.
\end{aligned}$$

Thus, $\mathcal{R}_\Gamma (Z_1, Z_2) = \mathcal{R}_\Gamma (h_\Gamma (Z_1), h_\Gamma (Z_2))$.

Let us concentrate on the characterization of a connection as an almost product structure. Then, Proposition 2.4.1 about different descriptions of a connection ∇ on $\pi\colon E \to B$ can be completed with the following equivalence, similar result is proved in [León and Rodrigues 1989].

Proposition 2.4.5. *Let $\pi : E \to B$ be a fiber bundle. The following elements are equivalent:*

1. *A subbundle $H(E)$ of TE such that*

$$TE = H(E) \oplus V(\pi).$$

2. *An almost product structure such that $\Gamma^2 = \mathrm{Id}$ and $\ker(\Gamma + \mathrm{Id}) = V(\pi)$.*

(*Proof*) $1 \Rightarrow 2$. The existence of the projectors $(2.4.11)$ associated with the connection ∇ defines the following tensor field on E

$$\Gamma = h_\nabla - v_\nabla.$$

It can be easily proved that Γ is an almost product structure.

$2 \Rightarrow 1$. Conversely, given an almost product structure Γ, the connection form is given by the horizontal projector

$$h_\Gamma = \frac{1}{2} (\mathrm{Id} + \Gamma)$$

associated to Γ; that is,

$$\nabla = \frac{1}{2} (\mathrm{Id} + \Gamma). \tag{2.4.12}$$

It can be easily proved that effectively ∇ is a connection. ∎

In local coordinates (x^μ, y^A) adapted to π, an almost product structure from a connection form is given by

$$\Gamma = \mathrm{d}x^\mu \otimes \left(\frac{\partial}{\partial x^\mu} + 2\Gamma^A_\mu \frac{\partial}{\partial y^A} \right) - \mathrm{d}y^A \otimes \left(\frac{\partial}{\partial y^A} \right).$$

Remark 2.4.6. We have just proved that the connection form ∇ defined in Proposition 2.4.1 and the connection defined by the almost product structure Γ are equivalent. Thus it makes sense that the curvatures of both connections in Definitions 2.4.2 and 2.4.4 must be the same. The proof of this equality is just a computation.

To summarize, the tangent bundle TE admits four different and equivalent splittings:

$$TE = H(E) \oplus V(\pi) = \begin{cases} \mathrm{Im}\,\nabla \oplus \ker \nabla, \\ \ker(\Gamma - \mathrm{Id}) \oplus \ker(\Gamma + \mathrm{Id}), \\ \mathrm{Im}\, H_\Gamma \oplus \ker H_\Gamma, \\ \ker \mathrm{v}_\Gamma \oplus \mathrm{Im}\, \mathrm{v}_\Gamma. \end{cases}$$

For each $e \in E$, the tangent space $T_e E$ splits as $H_e(E) \oplus V_e(E)$. This gives an isomorphism restricting the tangent map of the fiber bundle π to $H_e(E)$, $(T_e \pi)_{|H_e(E)} \colon H_e(E) \to T_{\pi(e)} B$. The inverse of the isomorphism $(T_e \pi)_{|H_e(E)}$ defines

- the *horizontal lift of tangent vectors* $H^e_{\pi(e)} \colon T_{\pi(e)} B \to H_e(E) \subset T_e E$, locally given by

$$H^e_{\pi(e)} \left(\left. \frac{\partial}{\partial x^\mu} \right|_{\pi(e)} \right) = \left. \frac{\partial}{\partial x^\mu} \right|_e + \Gamma^A_\mu(e) \left. \frac{\partial}{\partial y^A} \right|_e \; ;$$

- the *horizontal lift of vector fields*, $\mathrm{h} \colon \mathfrak{X}(B) \to H(E) \subset TE$,

$$\mathrm{h}(X)(e) = \mathrm{h}^e_{\pi(e)}(X_{\pi(e)}) \tag{2.4.13}$$

for every $e \in E$.

2.4.3 Splitting of $\mathrm{T}^* \mathrm{E}$ according to an Ehresmann connection

For a connection ∇ on the fiber bundle $\pi \colon E \to B$ we have the splitting $TE = H(E) \oplus V(\pi)$. Then, there exists a splitting of $T^* E$,

$$T^* E = (H(E))^0 \oplus (V(\pi))^0 \,,$$

where $(H(E))^0$ and $(V(\pi))^0$ are the annihilators of the horizontal and the vertical subbundle of E with respect to π, respectively. In local coordinates (x^μ, y^A) adapted to π, a basis for the subbundles of $T^* E$ is

$$(H(E))^0 = \left\langle -\Gamma^A_\mu \mathrm{d} x^\mu + \mathrm{d} y^A \right\rangle_{A=1,\dots,n} , \quad (V(\pi))^0 = \left\langle \mathrm{d} x^\mu \right\rangle_{\mu=1,\dots,m} .$$

Observe that the dimension of $(H(E))^0$ is equal to the dimension of the vertical subbundle $V(\pi)$, whereas the dimension of $(V(\pi))^0$ is equal to the dimension of the horizontal subbundle $H(E)$ given by the Ehresmann connection.

2.4.4 Linear Ehresmann connection

Before introducing the linear Ehresmann connections, we need to introduce more geometric elements.

Definition 2.4.7. *Let $\pi \colon E \to B$ be a vector bundle and $X \in \mathfrak{X}(B)$. A **linear vector field** Y **over** X is a vector field on E that is π–projectable to X and $Y \colon E \to TE$ is a vector bundle map.*

Such vector fields make Diagram (2.4.14) commutative. In bundle coordinates (x, y) for E,

$$Y(x,y) = X^\mu(x) \left. \frac{\partial}{\partial x^\mu} \right|_{(x,y)} + Y^A_B(x) y^B \left. \frac{\partial}{\partial y^A} \right|_{(x,y)} .$$

$$
\begin{array}{ccc}
E & \xrightarrow{\;Y\;} & TE \\
{\scriptstyle\pi}\Big\downarrow & & \Big\downarrow{\scriptstyle T\pi} \\
B & \xrightarrow{\;X\;} & TB
\end{array}
\tag{2.4.14}
$$

Let us consider the *Liouville vector field* Δ *on* E. Consider the 1–parameter group of dilations $\phi_t \colon E \to E$ such that $y_x \mapsto e^t y_x$ for $x \in B$, then the Liouville vector field is the infinitesimal generator of this 1–parameter group; i.e.,

$$
\begin{aligned}
\Delta \colon E &\longrightarrow TE \\
y_x &\longmapsto \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} (\phi_t(y_x)).
\end{aligned}
\tag{2.4.15}
$$

This vector field is also known as the vector of field of the dilations, since the flow is a homothety over the fibers with a positive ratio. Locally,

$$
\Delta = y^A \frac{\partial}{\partial y^A}.
$$

Now we can give other characterizations of the linear vector fields:

- a vector field Y is linear over X if and only if it is π–projectable and $\mathcal{L}_\Delta Y = 0$;

- a vector field Y is linear over X if and only if $\Phi_t^Y \colon E_b \to E_{\Phi_t^X(b)}$ is an isomorphism of the fibers.

For instance the complete lift of a vector field X defined in §2.2.3 is a linear vector field over X.

Definition 2.4.8. *An Ehresmann connection ∇ on a vector bundle $\pi \colon E \to B$ is **linear** if the connection form is invariant under the Liouville vector field; that is,*

$$
\mathcal{L}_\Delta \nabla = 0.
$$

In local coordinates adapted to $\pi \colon E \to B$, this condition implies that

$$
\Gamma_\mu^A = y^B \frac{\partial \Gamma_\mu^A}{\partial y^B}.
$$

Then, by Euler's Theorem for the homogenous functions, $\Gamma_\mu^A(x, y) = \Gamma_{\mu B}^A(x) y^B$. In other words, locally, the invariance of the connection under Liouville vector field is equivalent to the fact that Γ_μ^A are homogenous functions of degree 1 on the fiber.

There are many different ways of characterizing a linear Ehresmann connection as reviewed in [Echeverría-Enríquez et al.]. For instance:

- a connection ∇ is linear if and only if for every $f \in \mathcal{C}^\infty(B)$ and for every section

$\varphi\colon B \to E$ of π, the connection induces a covariant derivative given by

$$\widetilde{\nabla}(f\,\varphi) = \mathrm{d}f \otimes \varphi + f\,\nabla\varphi;$$

- a connection ∇ is linear if and only if the π–projectable horizontal vector fields are linear.

2.4.5 Dual of a linear Ehresmann connection

The definition of the dual of a linear vector field is useful for defining the dual of a linear Ehresmann connection.

If $\pi\colon E \to B$ is a vector bundle, consider the dual vector bundle $\pi^*\colon E^* \to B$.

Definition 2.4.9. *Let $X \in \mathfrak{X}(B)$, Y be a linear vector field on E over X. The **dual of** Y is the vector field $Y^*\colon E^* \to TE^*$ such that*

$$\phi_t^{Y^*} = \left(\phi_{-t}^{Y}\right)^*,$$

where ϕ^Y is the flow of Y.

In other words, Y^* is a vector field, π^*–projectable onto X; that is, the following diagram is commutative.

$$
\begin{array}{ccc}
E^* & \xrightarrow{\ Y^*\ } & TE^* \\
{\scriptstyle \pi^*}\Big\downarrow & & \Big\downarrow{\scriptstyle T\pi^*} \\
B & \xrightarrow{\ X\ } & TB
\end{array}
$$

For instance, the cotangent lift X^{T^*} in §2.2.2.2 is the dual of the linear vector field X^T. In bundle coordinates (x,α) for E^*, if $Y = X^\mu(x)\dfrac{\partial}{\partial x^\mu} + Y_B^A(x)y^B\dfrac{\partial}{\partial y^A}$, then

$$Y^* = X^\mu(x)\frac{\partial}{\partial x^\mu} - Y_A^B(x)\alpha_B\frac{\partial}{\partial \alpha_A}.$$

From this local expression, it is clear that Y^* is a linear vector field over X on the vector bundle π^*. Other possible characterizations of the dual of a linear vector field are described in [Kolář et al. 1993].

If ∇ is a linear Ehresmann connection on π, the horizontal lift of vector fields on B is denoted by $H\colon \mathfrak{X}(B) \to H(E) \subset TE$ as described in (2.4.13). Locally,

$$H(X)(x,y) = H_x^{(x,y)}\left(f^\mu\left.\frac{\partial}{\partial x^\mu}\right|_x\right) = f^\mu(x)\left(\left.\frac{\partial}{\partial x^\mu}\right|_{(x,y)} + \Gamma_{\mu B}^A(x)y^B\left.\frac{\partial}{\partial y^A}\right|_{(x,y)}\right),$$

for $(x,y) \in E$.

Lemma 2.4.10. *(Dual of a linear connection, [Kolář et al. 1993]) If ∇ is a linear connection on $\pi\colon E \to B$, then there exists a unique linear connection ∇^* on the dual bundle $\pi^*\colon E^* \to B$*

such that for any vector field X on B, the horizontal lift of X to E^* via ∇^* is the dual of the horizontal lift of X to E via ∇.

Observe that the linearity of ∇ is necessary to guarantee that the horizontal lift of a vector field on B is linear.

In bundle coordinates (x, α) for E^*, the horizontal lift $H^* \colon \mathfrak{X}(B) \to H(E^*) \subset TE^*$—defined analogously to the horizontal lift to E in (2.4.13)—is given by

$$
\begin{aligned}
\mathrm{h}^*(X)(x, \alpha) \ &= \ (\mathrm{h}^*)_x^{(x,\alpha)} \left(f^\mu \left. \frac{\partial}{\partial x^\mu} \right|_x \right) \\
&= \ f^\mu(x) \left(\left. \frac{\partial}{\partial x^\mu} \right|_{(x,\alpha)} - \Gamma_{\mu B}^A(x) \alpha_A \left. \frac{\partial}{\partial \alpha_B} \right|_{(x,\alpha)} \right)
\end{aligned}
$$

for $f^\mu \in \mathcal{C}^\infty(B)$. The connection coefficients of ∇^* are

$$
-\Gamma_{\mu B}^A(x) \alpha_A;
$$

that is, they are linear with respect to α. Equivalently, the dual of a linear connection is a linear connection on π^*.

Analogously with connections on π, ∇^* determines a splitting of the tangent space $T_e E^*$ for every $e \in E^*$,
$$
T_e E^* = H_e(E^*) \oplus V_e(\pi^*) \simeq T_{\pi^*(e)} B \oplus E_{\pi^*(e)}^*,
$$
since $(T_e \pi^*)_{|H_e(E^*)} \colon H_e(E^*) \to T_{\pi^*(e)} B$ is an isomorphism and $V_e(\pi^*) \simeq E_{\pi^*(e)}^*$.

2.4.6 Induced Ehresmann connection on $\tau_{\mathrm{M}} \colon \mathrm{TM} \to \mathrm{M}$ associated with a second–order differential equation

Now we are interested in studying the different connections that can be considered when the fiber bundle is the vector bundle given by the tangent projection $\tau_M \colon TM \to M$, for more details see [León and Rodrigues 1989, Crampin 1981]. Throughout this section (x^i, v^i), $i = 1, \ldots, m$, are the natural coordinates for TM.

Definition 2.4.11. *A vector field S on TM satisfies the **second–order** condition if*

$$
T\tau_M \circ S = \mathrm{Id}_{TM}.
$$

A vector field satisfying the second–order condition is also called *semi–spray*. Observe that S is a semi–spray if and only if $\mathrm{J}_M(S) = \Delta$, where Δ is the Liouville vector field in (2.4.15) and $\mathrm{J}_M \colon TTM \to TTM$ is the *vertical endomorphism*. Remember that J_M is a $(1, 1)$–tensor field on TM such that

$$
\begin{aligned}
\mathrm{J}_M \colon TTM \ &\longrightarrow \ TTM \\
w_{v_x} \ &\longmapsto \ \left. \frac{\mathrm{d}}{\mathrm{d}t} \right|_{t=0} (v_x + t T\tau_M(w_{v_x})),
\end{aligned} \tag{2.4.16}
$$

with local expression

$$J_M = \mathrm{d}x^j \otimes \frac{\partial}{\partial v^j}.$$

Thus, a semi–spray is given by

$$S(x,v) = v^i \left.\frac{\partial}{\partial x^i}\right|_{(x,v)} + S^i(x,v) \left.\frac{\partial}{\partial v^i}\right|_{(x,v)}.$$

Proposition 2.4.12. *If S is a vector field on TM satisfying the second–order condition, then the $(1,1)$–tensor field $\Gamma = -\mathcal{L}_S J_M$ on TM is an almost product structure.*

(*Proof*) It must be proved that the tensor $\Gamma = -\mathcal{L}_S J_M$ satisfies

$$\Gamma^2 = \mathrm{Id}, \quad \ker(\Gamma + \mathrm{Id}) = V(\tau_M).$$

The local expression for $-\mathcal{L}_S J_M$ is

$$-\mathcal{L}_S J_M \;=\; -\mathrm{d}v^j \otimes \frac{\partial}{\partial v^j} + \mathrm{d}x^j \otimes \left(\frac{\partial}{\partial x^j} + \frac{\partial S^i}{\partial v^j} \frac{\partial}{\partial v^i} \right).$$

As

$$-\mathcal{L}_S J_M + \mathrm{Id} = \mathrm{d}x^j \otimes \left(2\frac{\partial}{\partial x^j} + \frac{\partial S^i}{\partial v^j} \frac{\partial}{\partial v^i} \right),$$

$\ker\left(-\mathcal{L}_S J_M + \mathrm{Id}\right) = \left\langle \frac{\partial}{\partial v^i} \right\rangle = V(\tau_M)$. It is proved easily that $\left(-\mathcal{L}_S J_M\right)^2 = \mathrm{Id}$. Thus $-\mathcal{L}_S J_M$ is an almost product structure. ∎

Due to Proposition 2.4.5, we have a connection associated with a semi–spray. This connection defined by the almost product structure $-\mathcal{L}_S J_M$ is denoted by Γ^S_{TM}, where the subindex indicates the vector bundle which the connection is defined on. To recover the connection form ∇^S_{TM} associated with this almost product structure, we use the relation (2.4.12) given in § 2.4.2:

$$\nabla^S_{TM} = \frac{1}{2}(\mathrm{Id} + \Gamma^S_{TM}). \tag{2.4.17}$$

Locally,

$$\Gamma^S_{TM} = \mathrm{d}x^j \otimes \left(\frac{\partial}{\partial x^j} + \frac{1}{2}\frac{\partial S^i}{\partial v^j} \frac{\partial}{\partial v^i} \right).$$

Hence the connection coefficients are

$$\Gamma^i_j(x,v) = \frac{1}{2}\frac{\partial S^i}{\partial v^j}(x,v)$$

and the horizontal subspace at $v_x \in TM$ is expressed as

$$H_{v_x}(TM) = \left\langle \frac{\partial}{\partial x^i} + \frac{1}{2}\frac{\partial S^j}{\partial v^i} \frac{\partial}{\partial v^j} \right\rangle_{v_x}.$$

Another way to define a connection on τ_M is explained in [Crampin 1981] and it consists of

considering the following horizontal subbundle

$$H(TM) = \left\langle \frac{1}{2} \left(\left[\left(\frac{\partial}{\partial x^i} \right)^V, S \right] + \left(\frac{\partial}{\partial x^i} \right)^T \right) \right\rangle, \tag{2.4.18}$$

where $[\cdot\,,\,\cdot]$ is the Lie bracket of vector fields on TM and $(\cdot)^V$ is the vertical lift of vector fields on M. Then, the horizontal lift $\mathrm{h}_x^{v_x} \colon T_x M \to H_{v_x}(TM)$ is given by

$$\mathrm{h}_x^{v_x}(X_x) = \left(\frac{1}{2} \left([X^V, S] + X^T \right) \right)_{v_x}.$$

It can be proved that the horizontal subbundle defined by (2.4.17) and the one in (2.4.18) are the same.

As in §2.4, the connection form ∇_{TM}^S gives a splitting of the tangent space $T_{v_x} TM$ for every $v_x \in TM$,

$$T_{v_x} TM = H_{v_x}(TM) \oplus V_{v_x}(\tau_M).$$

As $T_{v_x} \tau_{M|H_{v_x}(TM)} \colon H_{v_x}(TM) \to T_x M$ is an isomorphism and $V_{v_x}(\tau_M) \simeq T_x M$,

$$T_{v_x} TM = H_{v_x}(TM) \oplus V_{v_x}(\tau_M) \simeq T_x M \oplus T_x M.$$

2.5 Skinner–Rusk unified formalism for non–autonomous systems

We introduce the Skinner–Rusk unified formalism in order to give an alternative approach to study optimal control problems for the non–autonomous control systems in Chapter 8.

This formalism is a particular case of the unified formalism for field theories developed in [Echeverría-Enríquez et al. 2004, León et al. 2003]. See [Cortés et al. 2002b] for an alternative but equivalent approach, and [Gràcia and Martín 2005] for an extension of this formalism to more general time–dependent singular differential equations.

In the jet bundle description of non–autonomous dynamical systems, the configuration bundle is $\pi \colon E \longrightarrow \mathbb{R}$, where E is a $(n+1)$–dimensional differentiable manifold endowed with local coordinates (t, x^i), and $\mathbb{R}$ has as a global coordinate t. The jet bundle of local sections of π, $J^1\pi$, is the *velocity phase space* of the system, with natural coordinates (t, x^i, v^i), adapted to the bundle $\pi \colon E \longrightarrow \mathbb{R}$, and natural projections

$$\pi^1 \colon J^1\pi \longrightarrow E, \quad \bar{\pi}^1 \colon J^1\pi \longrightarrow \mathbb{R}.$$

A Lagrangian density $\mathcal{L} \in \Omega^1(J^1\pi)$ is a $\bar{\pi}^1$–semibasic 1–form on $J^1\pi$, and it is usually written as $\mathcal{L} = L\,\mathrm{d}t$, where $L \in \mathcal{C}^\infty(J^1\pi)$ is the *Lagrangian function* determined by $\mathcal{L}$. In this section and in Chapter 8, we denote by $\mathrm{d}t$ the volume form in $\mathbb{R}$, and its pullbacks to all the corresponding manifolds.

The canonical structure of the bundle $J^1\pi$ allows us to define the *Poincaré–Cartan forms*

$\Theta_{\mathcal{L}}$ and $\Omega_{\mathcal{L}}$ associated with the Lagrangian density $\mathcal{L}$ and then the Euler–Lagrange equations are written intrinsically, see for instance[Echeverría-Enríquez et al. 1991, Saunders 1989].

Furthermore, we have the *extended momentum phase space* T^*E, and the *restricted momentum phase space* which is defined by $J^1\pi^* = T^*E/\pi^*T^*\mathbb{R}$. Local coordinates in these manifolds are (t, x^i, p, p_i) and (t, x^i, p_i), respectively. Then, the following natural projections are

$$\tau^1\colon J^1\pi^* \longrightarrow E, \ \bar{\tau}^1 = \pi \circ \tau^1\colon J^1\pi^* \longrightarrow \mathbb{R}, \ \mu\colon T^*E \longrightarrow J^1\pi^*, \ p\colon T^*E \longrightarrow \mathbb{R}.$$

Let $\Theta \in \Omega^1(T^*E)$ and $\Omega = -\mathrm{d}\Theta \in \Omega^2(T^*E)$ be the canonical forms of T^*E whose local expressions are

$$\Theta = p_i\mathrm{d}x^i + p\mathrm{d}t, \quad \Omega = \mathrm{d}x^i \wedge \mathrm{d}p_i + \mathrm{d}t \wedge \mathrm{d}p.$$

Hamilton's equations can be written intrinsically from these canonical structures; see, for instance, [Echeverría-Enríquez et al. 1991, Kuwabara 1984, Mangiarotti and Sardanashvily 1998, Rañada 1992, Struckmeier 2005].

Now we introduce the geometric framework for the unified Skinner–Rusk formalism for non–autonomous systems. We define the *extended jet–momentum bundle* $\mathcal{W}$ and the *restricted jet–momentum bundle* $\mathcal{W}_r$,

$$\mathcal{W} = J^1\pi \times_E T^*E, \quad \mathcal{W}_r = J^1\pi \times_E J^1\pi^*,$$

with natural coordinates (t, x^i, v^i, p, p_i) and (t, x^i, v^i, p_i), respectively. We have the natural submersions

$$\rho_1\colon \mathcal{W} \longrightarrow J^1\pi, \ \rho_2\colon \mathcal{W} \longrightarrow T^*E, \ \rho_E\colon \mathcal{W} \longrightarrow E, \ \rho_{\mathbb{R}}\colon \mathcal{W} \longrightarrow \mathbb{R}, \qquad (2.5.19)$$

$$\rho_1^r\colon \mathcal{W}_r \longrightarrow J^1\pi, \ \rho_2^r\colon \mathcal{W}_r \longrightarrow J^1\pi^*, \ \rho_E^r\colon \mathcal{W}_r \longrightarrow E, \ \rho_{\mathbb{R}}^r\colon \mathcal{W}_r \longrightarrow \mathbb{R}.$$

Note that $\pi^1 \circ \rho_1 = \tau^1 \circ \mu \circ \rho_2 = \rho_E$. In addition, for $\bar{y} \in J^1\pi$ and $\mathbf{p} \in T^*E$, there is also the natural projection

$$\begin{aligned} \mu_{\mathcal{W}} \ &:\ & \mathcal{W} \ &\longrightarrow \ & \mathcal{W}_r \\ & & (\bar{y}, \mathbf{p}) \ &\mapsto \ & (\bar{y}, [\mathbf{p}]), \end{aligned}$$

where $[\mathbf{p}] = \mu(\mathbf{p}) \in J^1\pi^*$. The elements in T^*E are in bold in order not to confuse it with the momenta p for the time. The bundle $\mathcal{W}$ is endowed with the following canonical structures.

Definition 2.5.1. *1. The **coupling** 1–**form** in $\mathcal{W}$ is the $\rho_{\mathbb{R}}$–semibasic 1–form $\hat{\mathcal{C}} \in \Omega^1(\mathcal{W})$ defined as follows: for every $w = (j^1\phi(t), \alpha) \in \mathcal{W}$—that is, $\alpha \in T^*_{\rho_E(w)}E$ and $V \in T_w\mathcal{W}$—then*

$$\hat{\mathcal{C}}(V) = \alpha(T_w(\phi \circ \rho_{\mathbb{R}})V).$$

*2. The **canonical** 1–**form** $\Theta_{\mathcal{W}} \in \Omega^1(\mathcal{W})$ is the ρ_E–semibasic form defined by $\Theta_{\mathcal{W}} = \rho_2^*\Theta$.*

*The **canonical** 2–**form** is $\Omega_{\mathcal{W}} = -\mathrm{d}\Theta_{\mathcal{W}} = \rho_2^*\Omega \in \Omega^2(\mathcal{W})$.*

As $\hat{\mathcal{C}}$ is a $\rho_{\mathbb{R}}$–semibasic form, there is $\hat{C} \in \mathcal{C}^\infty(\mathcal{W})$ such that $\hat{\mathcal{C}} = \hat{C}\mathrm{d}t$. Note also that

$\Omega_{\mathcal{W}}$ is degenerate, its kernel being the ρ_2–vertical vectors. Then $(\mathcal{W}, \Omega_{\mathcal{W}})$ is a presymplectic manifold.

The local expressions for $\Theta_{\mathcal{W}}$, $\Omega_{\mathcal{W}}$ and $\hat{\mathcal{C}}$ are

$$\Theta_{\mathcal{W}} = p_i \mathrm{d}x^i + p\mathrm{d}t, \quad \Omega_{\mathcal{W}} = -\mathrm{d}p_i \wedge \mathrm{d}x^i - \mathrm{d}p \wedge \mathrm{d}t, \quad \hat{\mathcal{C}} = (p + p_i v^i)\mathrm{d}t.$$

Given a Lagrangian density $\mathcal{L} \in \Omega^1(J^1\pi)$, we denote $\hat{\mathcal{L}} = \rho_1^*\mathcal{L} \in \Omega^1(\mathcal{W})$, and we can write $\hat{\mathcal{L}} = \hat{L}\mathrm{d}t$, with $\hat{L} = \rho_1^*L \in \mathcal{C}^\infty(\mathcal{W})$. We define a *Hamiltonian submanifold*

$$\mathcal{W}_0 = \{w \in \mathcal{W} \mid \hat{\mathcal{L}}(w) = \hat{\mathcal{C}}(w)\}.$$

So, $\mathcal{W}_0$ is the submanifold of $\mathcal{W}$ defined by the regular constraint function $\hat{C} - \hat{L} = 0$. Observe that this function is globally defined in $\mathcal{W}$, using the dynamical data and the geometry. In local coordinates this constraint function is

$$p + p_i v^i - \hat{L}(t, x^j, v^j) = 0, \tag{2.5.20}$$

where p is the momenta corresponding with the time. The meaning of this function will be clear when we apply this formalism to optimal control problems, see §8.1.1. The natural imbedding is $\jmath_0 \colon \mathcal{W}_0 \hookrightarrow \mathcal{W}$, and we have the projections (submersions), see Diagram (2.5.21):

$$\rho_1^0 \colon \mathcal{W}_0 \longrightarrow J^1\pi, \; \rho_2^0 \colon \mathcal{W}_0 \longrightarrow T^*E, \; \rho_E^0 \colon \mathcal{W}_0 \longrightarrow E, \; \rho_{\mathbb{R}}^0 \colon \mathcal{W}_0 \longrightarrow \mathbb{R},$$

which are the restrictions to $\mathcal{W}_0$ of the projections (2.5.19), and

$$\hat{\rho}_2^0 = \mu \circ \rho_2^0 \colon \mathcal{W}_0 \longrightarrow J^1\pi^*.$$

Local coordinates in $\mathcal{W}_0$ are (t, x^i, v^i, p_i), and we have

$$\rho_1^0(t, x^i, v^i, p_i) = (t, x^i, v^i), \qquad \jmath_0(t, x^i, v^i, p_i) = (t, x^i, v^i, L - p_i v^i, p_i),$$
$$\hat{\rho}_2^0(t, x^i, v^i, p_i) = (t, x^i, p_i), \qquad \rho_2^0(t, x^i, v^i, p_i) = (t, x^i, L - p_i v^i, p_i).$$

Proposition 2.5.2. $\mathcal{W}_0$ *is a 1–codimensional* $\mu_{\mathcal{W}}$*–transversal submanifold of* $\mathcal{W}$*, which is diffeomorphic to* $\mathcal{W}_r$.

(*Proof*) For every $(\bar{y}, \mathbf{p}) \in \mathcal{W}_0$, we have $L(\bar{y}) \equiv \hat{L}(\bar{y}, \mathbf{p}) = \hat{C}(\bar{y}, \mathbf{p})$, and

$$(\mu_{\mathcal{W}} \circ \jmath_0)(\bar{y}, \mathbf{p}) = \mu_{\mathcal{W}}(\bar{y}, \mathbf{p}) = (\bar{y}, \mu(\mathbf{p})).$$

First, $\mu_{\mathcal{W}} \circ \jmath_0$ is injective: if $(\bar{y}_1, \mathbf{p}_1), (\bar{y}_2, \mathbf{p}_2) \in \mathcal{W}_0$, then we have

$$(\mu_{\mathcal{W}} \circ \jmath_0)(\bar{y}_1, \mathbf{p}_1) = (\mu_{\mathcal{W}} \circ \jmath_0)(\bar{y}_2, \mathbf{p}_2) \Rightarrow (\bar{y}_1, \mu(\mathbf{p}_1)) = (\bar{y}_2, \mu(\mathbf{p}_2))$$

$$\Rightarrow \bar{y}_1 = \bar{y}_2, \; \mu(\mathbf{p}_1) = \mu(\mathbf{p}_2),$$

hence $L(\bar{y}_1) = L(\bar{y}_2) = \hat{C}(\bar{y}_1, \mathbf{p}_1) = \hat{C}(\bar{y}_2, \mathbf{p}_2)$. In a local chart, the third equality gives

$$p(\mathbf{p}_1) + p_i(\mathbf{p}_1)v^i(\bar{y}_1) = p(\mathbf{p}_2) + p_i(\mathbf{p}_2)v^i(\bar{y}_2)$$

but $\mu(\mathbf{p}_1) = \mu(\mathbf{p}_2)$ implies that

$$p_i(\mathbf{p}_1) = p_i([\mathbf{p}_1]) = p_i([\mathbf{p}_2]) = p_i(\mathbf{p}_2)$$

therefore $p(\mathbf{p}_1) = p(\mathbf{p}_2)$ and hence $\mathbf{p}_1 = \mathbf{p}_2$.

Second, $\mu_\mathcal{W} \circ \jmath_0$ is onto, then, if $(\bar{y}, [\mathbf{p}]) \in \mathcal{W}_r$, there exists $(\bar{y}, \mathbf{q}) \in \jmath_0(\mathcal{W}_0)$ such that $[\mathbf{q}] = [\mathbf{p}]$. In fact, it suffices to take $[\mathbf{q}]$ such that, in a local chart of $J^1\pi \times_E T^*E = \mathcal{W}$

$$p_i(\mathbf{q}) = p_i([\mathbf{p}]) \text{ and } p(\mathbf{q}) = L(\bar{y}) - p_i([\mathbf{p}])v^i(\bar{y}).$$

Finally, since $\mathcal{W}_0$ is defined by the constraint function $\hat{C} - \hat{L}$ and, as $\ker \mu_{\mathcal{W}*} = \left\{ \dfrac{\partial}{\partial p} \right\}$ locally and $\dfrac{\partial}{\partial p}(\hat{C} - \hat{L}) = 1$, then $\mathcal{W}_0$ is $\mu_\mathcal{W}$–transversal.   ∎

As a consequence of this result, the submanifold $\mathcal{W}_0$ induces a section of the projection $\mu_\mathcal{W}$,

$$\hat{h}: \mathcal{W}_r \longrightarrow \mathcal{W}.$$

Locally, $\hat{h}$ is specified by giving the local *Hamiltonian function* $\hat{H} = -\hat{L} + p_i v^i$; that is, $\hat{h}(t, x^i, v^i, p_i) = (t, x^i, v^i, -\hat{H}, p_i)$. In this sense, $\hat{h}$ is said to be a *Hamiltonian section* of $\mu_\mathcal{W}$. The setting for this section is summarized in Diagram 2.5.21.

$$(2.5.21)$$

2.6   Particular background in jet bundles

Finally, we describe some geometric features about Tulczyjew's operators and contact systems so as to explain the Euler–Lagrange equations for forced systems. The notions in this section are necessary for studying the controlled Lagrangian systems in §8.3.1.

2.6.1   Jet bundles of order 1 and 2

Associated with every jet bundle $J^1\pi$, we have the *contact system* which is a subbundle $\mathcal{C}_\pi$ of $T^*J^1\pi$ whose fibres at every $j^1\phi(t) \in J^1\pi$ are defined by

$$
\begin{aligned}
\mathcal{C}_\pi(j^1\phi(t)) &= \Big\{ \alpha \in T^*_{j^1\phi(t)}(J^1\pi) \mid \\
&\quad \alpha = (T_{j^1\phi(t)}\pi^1 - T_{j^1\phi(t)}(\phi \circ \bar{\pi}^1))^*\beta, \ \beta \in V^*_{\phi(t)}(\pi) \Big\}.
\end{aligned}
$$

$$(2.6.22)$$

One may readily see that a local basis for the sections of this bundle is given by $\{\mathrm{d}x^i - v^i\mathrm{d}t\}$.

Now, denote by $J^2\pi$ the bundle of 2–jets of π. This jet bundle is equipped with natural coordinates (t, x^i, v^i, w^i) and canonical projections

$$\pi_1^2 \colon J^2\pi \longrightarrow J^1\pi, \quad \pi^2 \colon J^2\pi \longrightarrow E, \quad \bar\pi^2 \colon J^2\pi \longrightarrow \mathbb{R}\,.$$

Considering the bundle $J^1\bar\pi^1$, the canonical injection $\Upsilon \colon J^2\pi \longrightarrow J^1\bar\pi^1$ is given by

$$\Upsilon(j^2\phi(t)) = (j^1(j^1\phi))(t)\,. \tag{2.6.23}$$

Locally, $\Upsilon(t, x^i, v^i, w^i) = (t, x^i, v^i; v^i, w^i)$.

Thus, we have the Diagram (2.6.24) where $J^1\bar\pi^1 \simeq \mathbb{R} \times T(TQ)$ and the inclusion $\imath_1$ is locally given by $\imath_1(t, x, v, w) = (t, 1, x, v, v, w)$.

$$\tag{2.6.24}$$

Observe that $(\pi_1^2)^*T^*J^1\pi$ can be identified with a subbundle of $T^*J^2\pi$ by means of the natural injection $\hat\imath \colon (\pi_1^2)^*T^*J^1\pi \longrightarrow T^*J^2\pi$, defined as follows: for every $\hat p \in J^2\pi$, $\alpha \in T^*_{\pi_1^2(\hat p)}J^1\pi$, and $\mathrm{a} \in T_{\hat p}J^2\pi$,

$$(\hat\imath(\hat p, \alpha))(\mathrm{a}) = \alpha(T_{\hat p}\pi_1^2(\mathrm{a}))\,.$$

In the same way, we can identify $(\pi_1^2)^*\mathcal{C}_\pi$ as a subbundle of $(\pi_1^2)^*T^*J^1\pi$ by means of $\hat\imath$.

The set of sections of the bundles

$$T^*J^2\pi \longrightarrow J^2\pi, \ (\pi_1^2)^*T^*J^1\pi \longrightarrow J^2\pi, \text{ and } (\pi_1^2)^*\mathcal{C}_\pi \longrightarrow J^2\pi$$

have as local basis $(\mathrm{d}t, \mathrm{d}x^i, \mathrm{d}v^i, \mathrm{d}w^i)$, $(\mathrm{d}t, \mathrm{d}x^i, \mathrm{d}v^i)$, and $(\mathrm{d}x^i - v^i\mathrm{d}t)$, respectively.

Incidentally, $\Gamma(J^2\pi, (\pi_1^2)^*T^*J^1\pi) = \mathcal{C}^\infty(J^2\pi) \otimes_{\mathcal{C}^\infty(J^1\pi)} (\pi_1^2)^*\Omega^1(J^1\pi)$, which are the π_1^2–semibasic 1–forms in $J^2\pi$.

2.6.2   Tulczyjew's operators

Given a differentiable manifold Q and its tangent bundle $\tau_Q\colon TQ \longrightarrow Q$, we consider the following operators, introduced by Tulczyjew [1974]: first we have $i_T\colon \Omega^s(Q) \longrightarrow \Omega^{s-1}(TQ)$, which is defined as follows: for every $(\mathrm{p},v) \in TQ$, $\alpha \in \Omega^s(Q)$, and $X_1,\dots,X_{s-1} \in \mathfrak{X}(TQ)$,

$$(i_T\,\alpha)((\mathrm{p},v); X_1,\dots,X_{s-1}) = \alpha(\mathrm{p}; v, T_{(\mathrm{p},v)}\tau_Q((X_1)_{(\mathrm{p},v)}),\dots,T_{(\mathrm{p},v)}\tau_Q((X_{s-1})_{(\mathrm{p},v)})) \ .$$

Then, the so–called *total derivative* is a map $\mathrm{d}_T\colon \Omega^s(Q) \longrightarrow \Omega^s(TQ)$ defined by

$$\mathrm{d}_T = \mathrm{d} \circ i_T + i_T \circ \mathrm{d} \ .$$

For the case $s = 1$, using natural coordinates in TQ, we have the local expression

$$\mathrm{d}_T\alpha \equiv \mathrm{d}_T(A_j\mathrm{d}x^j) = A_j\mathrm{d}v^j + v^i\frac{\partial A_j}{\partial x^i}\mathrm{d}x^j \ .$$

2.6.3   Implicit Euler–Lagrange equations

Let $\mathcal{L} \in \Omega^1(J^1\pi)$ be a Lagrangian density and $L \in \mathcal{C}^\infty(J^1\pi)$ be its associated Lagrangian function. Observe that

$$\mathrm{d}_T\Theta_\mathcal{L} \in \Omega^1(TJ^1\pi), \ \imath_1^*\mathrm{d}_T\Theta_\mathcal{L} \in \Omega^1(J^2\pi), \ (\pi_1^2)^*\mathrm{d}L \in \Omega^1(J^2\pi).$$

Then, a simple calculation in coordinates shows that $\imath_1^*\mathrm{d}_T\Theta_\mathcal{L} - (\pi_1^2)^*\mathrm{d}L$ is a section of the bundle $\hat{\imath}((\pi_1^2)^*\mathcal{C}_\pi) \longrightarrow J^2\pi$.

The Euler–Lagrange equations for this Lagrangian are a system of second–order differential equations on Q; that is, in implicit form, a submanifold D of $J^2\pi$ implicitly determined by:

$$D = \{\hat{p} \in J^2\pi \mid (\imath_1^*\mathrm{d}_T\Theta_\mathcal{L} - (\pi_1^2)^*\mathrm{d}L)(\hat{p}) = 0\} = \{\hat{p} \in J^2\pi \mid \mathcal{E}_\mathcal{L}(\hat{p}) = 0\} = \mathcal{E}_\mathcal{L}^{-1}(0) \ ,$$

where $\mathcal{E}_\mathcal{L} = \imath_1^*\mathrm{d}_T\Theta_\mathcal{L} - (\pi_1^2)^*\mathrm{d}L$. Then, a section $\phi\colon \mathbb{R}\longrightarrow\mathbb{R}\times Q$ is a solution to the Lagrangian system if, and only if, $\operatorname{Im} j^2\phi \subset \mathcal{E}_\mathcal{L}^{-1}(0)$. In fact, working locally

$$
\begin{aligned}
\mathrm{d}_T\Theta_\mathcal{L} &= \frac{\partial L}{\partial v^r}\mathrm{d}v^r - \left(\frac{\partial L}{\partial v^j}v^j - L\right)\mathrm{d}\dot{\imath} + \left(\dot{\imath}\frac{\partial^2 L}{\partial t\partial v^r} + v^i\frac{\partial^2 L}{\partial x^i\partial v^r}\right.\\
&\quad + \left. w^i\frac{\partial^2 L}{\partial v^i\partial v^r}\right)\mathrm{d}x^r - \left[\dot{\imath}\left(v^j\dot{\imath}\frac{\partial^2 L}{\partial t\partial v^j} - \frac{\partial L}{\partial t}\right)\right.\\
&\quad + \left. v^i\left(v^j\frac{\partial^2 L}{\partial x^i\partial v^j} - \frac{\partial L}{\partial x^i}\right) + w^i\left(\frac{\partial L}{\partial v^i} + v^j\frac{\partial^2 L}{\partial v^i\partial v^j} - \frac{\partial L}{\partial v^i}\right)\right]\mathrm{d}t;\\
\imath_1^*\mathrm{d}_T\Theta_\mathcal{L} &= \frac{\partial L}{\partial v^r}\mathrm{d}v^r + \left(\frac{\partial^2 L}{\partial t\partial v^r} + v^i\frac{\partial^2 L}{\partial x^i\partial v^r} + w^i\frac{\partial^2 L}{\partial v^i\partial v^r}\right)\mathrm{d}x^r\\
&\quad - \left[v^j\frac{\partial^2 L}{\partial t\partial v^j} - \frac{\partial L}{\partial t} + v^i\left(v^j\frac{\partial^2 L}{\partial x^i\partial v^j} - \frac{\partial L}{\partial x^i}\right) + w^iv^j\frac{\partial^2 L}{\partial v^i\partial v^j}\right]\mathrm{d}t;\\
(\pi_1^2)^*\mathrm{d}L &= \frac{\partial L}{\partial t}\mathrm{d}t + \frac{\partial L}{\partial x^r}\mathrm{d}x^r + \frac{\partial L}{\partial v^r}\mathrm{d}v^r \ ;
\end{aligned}
$$

and then

$$\imath_1^* \mathrm{d}_T \Theta_{\mathcal{L}} - (\pi_1^2)^* \mathrm{d}L = \left(\frac{\partial^2 L}{\partial v^i \partial v^r} w^i + \frac{\partial^2 L}{\partial x^i \partial v^r} v^i + \frac{\partial^2 L}{\partial t \partial v^r} - \frac{\partial L}{\partial x^r} \right) (\mathrm{d}x^r - v^r \mathrm{d}t)$$

$$= \left[\frac{d}{dt} \left(\frac{\partial L}{\partial v^r} \right) - \frac{\partial L}{\partial x^r} \right] (\mathrm{d}x^r - v^r \mathrm{d}t) .$$

Now, suppose that there are external forces operating on the Lagrangian system $(J^1\pi, \mathcal{L})$. A force depending on velocities is a section $F \colon J^1\pi \longrightarrow \mathcal{C}_\pi$, see (2.6.22). As above, the corresponding Euler–Lagrange equations are a system of second–order differential equations on Q, given in implicit form by the submanifold D_F of $J^2\pi$ determined by:

$$D_F = \{\hat{p} \in J^2\pi \mid (\imath_1^* \mathrm{d}_T \Theta_{\mathcal{L}} - (\pi_1^2)^* \mathrm{d}L)(\hat{p}) = (F \circ \pi_1^2)(\hat{p})\}$$

$$= \{\hat{p} \in J^2\pi \mid \mathcal{E}_{\mathcal{L}}(\hat{p}) = (F \circ \pi_1^2)(\hat{p})\} .$$

A section $\phi \colon \mathbb{R} \longrightarrow \mathbb{R} \times Q$ is a solution to the Lagrangian system if, and only if,

$$\mathcal{E}_{\mathcal{L}}(j^2\phi) = (\pi_1^2)^*[(F \circ \pi_1^2)(j^2\phi)] = (\pi_1^2)^* F(j^1\phi) . \tag{2.6.25}$$

In natural coordinates we have

$$\left[\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial v^r} \right) - \frac{\partial L}{\partial x^r} \right] (\mathrm{d}x^r - v^r \mathrm{d}t) = F_j(\mathrm{d}x^j - v^j \mathrm{d}t) ;$$

that is,

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial v^j} \right) - \frac{\partial L}{\partial x^j} = F_j .$$

Chapter 3
Background in control theory

$\mathbf{A}$fter a review in Chapter 2 of all the main tools in differential geometry required for the subsequent chapters, we give a brief review of control theory before getting to the main core of this work.

In control theory, there are interesting properties that have attracted the attention of many researchers [Agrachev 1999, Agrachev and Sachkov 2004, Aguilar and Lewis 2008, Basto-Gonçalves 1998, Bianchini and Stefani 1993, Bloch 2003, Bullo and Lewis 2005a, Cortés Monforte 2002, Cortés and Martínez 2003, Jurdjevic 1997, Nijmeijer and van der Schaft 1990, Ostrowski and Burdick 1997, Sussmann 1978; 1983; 1987, Sussmann and Jurdjevic 1972]. Some of these properties are related with the following questions: is the interior of the reachable set empty?, what is the set of points reached by the control system from an initial point?, is there any trajectory of the control system that connects two given points? In control terminology, these questions refer to the accessibility and the controllability; the former is better characterized than the latter, that still has a lot to do. We review the notions and properties about accessibility and controllability in §3.2 and in §3.3 that will used at some point of this work for particular control systems defined in §3.1.

3.1 Control systems

Whenever we have a dynamical system where we can chose some parameters in order to compute trajectories satisfying different properties, we deal with a control system. This can be defined formally as follows, using Definition 2.2.1.

Definition 3.1.1. *Let M be an m–dimensional manifold and U be a subset of $\mathbb{R}^k$. A **control system** on M is a vector field X along the projection $\pi\colon M \times U \to M$. A **trajectory** or an **integral curve of the control system** X is a curve $(\gamma, u)\colon I \subset \mathbb{R} \to M \times U$ such that γ is absolutely continuous, u is measurable and bounded, and $\dot{\gamma}(t) = X(\gamma(t), u(t))$ a.e. for $t \in I$.*

The set of control systems is denoted by $\mathfrak{X}(\pi)$; that is, the set of vector fields along the projection. The curve $u\colon I \to U$ is called the *control*. Some particular cases of control systems are:

- Control–affine systems with the dynamics given by

$$X(x, u) = f_0(x) + u^s f_s(x),$$

where f_0, f_s are vector fields on M. Usually f_0 is called the *drift vector field* and f_s are called the *control* or *input vector fields*.

- Control–linear systems with the dynamics given by

$$X(x, u) = u^s f_s(x),$$

where f_s are vector fields on M called the *control vector fields*. In fact, a control–linear system is a particular control–affine system without drift.

Some examples of control–affine systems are mechanical control systems modeling a wide range of rigid bodies such as the robotic leg, the snakeboard, the rolling disk and so on; [Bloch 2003, Bullo and Lewis 2005a, Chyba et al. 2003, Ostrowski and Burdick 1997]. These systems are called affine connection control systems and considered in Chapter 6. Examples of control–linear systems appear in Riemannian and subRiemannian geometry [Agrachev and Sarychev 1995a, Bonnard and Trélat 2001, Jean 2003, Langerock 2001, Liu and Sussmann 1995, Montgomery 1994, Strichartz 1986], as described in §5.4.1 and §5.4.2, respectively.

3.2 Accessibility and controllability

In a control system $X \in \mathfrak{X}(\pi)$, we are usually interested in the set of points that can be reached from a initial point through trajectories $(\gamma, u) \colon I \subset \mathbb{R} \to M \times U$ of the control system, where $I = [a, b]$. In this regard, the following definition is essential.

Definition 3.2.1. *Let M be a manifold, U be a set in $\mathbb{R}^k$ and X be a vector field along the projection $\pi \colon M \times U \to M$. The **reachable set from** $x_0 \in M$ **at time** $T \in I$ is the set of points described by*

$$\mathcal{R}(x_0, T) = \{ x \in M \quad | \quad \text{there exists } (\gamma, u) \colon [a, b] \to M \times U \text{ such that}$$
$$\dot{\gamma}(t) = X(\gamma(t), u(t)), \ \gamma(a) = x_0, \ \gamma(T) = x \}.$$

Another definition given by the reachable points is the *reachable set from a point $x_0 \in M$ up to time T*; that is,

$$\mathcal{R}(x_0, \leq T) = \bigcup_{a \leq t \leq T} \mathcal{R}(x_0, t).$$

Now it is possible to introduce the notion of accessibility.

Definition 3.2.2. *Let $\pi \colon M \times U \to M$ be a projection and $X \in \mathfrak{X}(\pi)$. The system defined by X is **accessible from** $x_0 \in M$ if there exits a $T > a$ so that the interior of $\mathcal{R}(x_0, \leq t)$ is nonempty for all $t \in (a, T]$. If the system is accessible from every $x_0 \in M$, the system is **accessible**.*

Remark 3.2.3. It is usual to assume as initial time 0, but in order to get used to the notation for the domain of definition of the curves in this dissertation we use here a as initial time.

From now on in this chapter, we restrict to the control–affine systems

$$X(x, u) = f_0(x) + u^s f_s(x), \tag{3.2.1}$$

where $f_0, f_s \in \mathfrak{X}(M)$, to define the following elements and their associated properties. See [Bullo and Lewis 2005a, Jurdjevic 1997, Nijmeijer and van der Schaft 1990] for more details.

Definition 3.2.4. *The **accessibility algebra** C of the system (3.2.1) is the smallest Lie subalgebra of $\mathfrak{X}(M)$ that contains $\{f_0, f_1, \ldots, f_k\} \subset \mathfrak{X}(M)$. The **accessibility distribution** $\mathcal{C}$ is the distribution on M that is $\mathbb{R}$–spanned by the accessibility algebra.*

For a description of the accessibility distribution for a general control system $X \in \mathfrak{X}(\pi)$ see [Langerock 2003a; 2008].

Proposition 3.2.5. *[Nijmeijer and van der Schaft 1990, Theorem 3.9] For the system* (3.2.1)*, if the accessibility distribution $\mathcal{C}$ has maximum rank, equal to the dimension of the manifold, at $x_0 \in M$, then for any $T > a$, the reachable set $\mathcal{R}(x_0, \leq T)$ contains a nonempty open subset of M.*

Proposition 3.2.6. *[Nijmeijer and van der Schaft 1990, Corollary 3.13] If the system* (3.2.1) *is accessible, then the rank of the accessibility distribution $\mathcal{C}$ at $x \in M$ is maximum for any x on an open dense subset of M.*

If the vector fields are analytic instead of $\mathcal{C}^\infty$, then the condition is also necessary [Bullo and Lewis 2005a].

As studied in [Tyner 2007], the control–affine system (3.2.1) admits a linearization along a trajectory $(\gamma, u) \colon [a, b] \to M \times U$. Let X_γ be the vector field whose integral curve is γ, the linearization of the system is given by the control–affine system on TM

$$X_{TM}(v_x, w) = X_\gamma^T(v_x) + w^s f_s^V(x),$$

where f_s^V denotes the vertical lift to TM of the vector field f_s on M and w are the controls taking values in U. The *reachable set for the linearized system from v_{x_0}* is defined by

$$\begin{aligned}
\mathcal{R}_{X_{TM}}(v_{x_0}, T) = \{ v_x \in TM \quad | \quad & \text{there exists } (\Upsilon, w) \colon [a, b] \to TM \times U \text{ such that} \\
& \dot{\Upsilon}(t) = X_{TM}(\Upsilon(t), w(t)), \\
& \Upsilon(a) = v_{x_0}, \ \Upsilon(T) = v_x \}.
\end{aligned} \tag{3.2.2}$$

Let us next focus on the controllability, carefully studied in [Bullo and Lewis 2005a, Lewis 2001, Sussmann and Jurdjevic 1972, Nijmeijer and van der Schaft 1990].

Definition 3.2.7. *Let $(\gamma, u) \colon [a, b] \to M \times U$ be a trajectory of the control system (3.2.1). Let $t_1 \in [a, b]$ such that $\gamma(t_1) = x$. The system (3.2.1) is:*

*(a) **controllable at** x **along** (γ, u) if $\gamma(t)$ is in the interior of $\mathcal{R}(x, t)$ for each $t > t_1$;*

(b) **linearly controllable at** x **along** γ *if* $\mathcal{R}_{X_{TM}}(0_x, t) = T_{\gamma(t)}M$ *for each* $t > t_1$;

(c) **small–time locally controllable (STLC) from** x *if there exists* $T > 0$ *so that* x *is in the interior of* $\mathcal{R}(x, \leq t)$ *for each* $t \in [t_1, T]$.

For a better understanding of the linear controllability see [Tyner 2007]. Roughly speaking, the idea is that the reachable set of the linearized system is useful to obtain a linearization of the reachable set in the sense used in Proposition 4.1.12. In this Proposition, the manifold and the tangent space can be identified locally with the same Euclidean space $\mathbb{R}^m$ through the local charts. This identification becomes clearer in Chapter 4, in particular in Proposition 4.1.12, and in §4.5.2.

In order to point out the differences between accessibility and controllability, let us show Figure 3.1 where $\mathcal{R}$ stands for the reachable set from x_0. In the first two frames we consider the reachable set up to time T and in the third one the reachable set at time T. As it is observed,
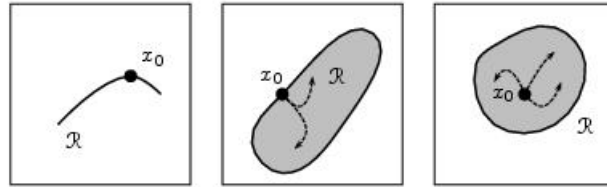
Figure 3.1: *No accessible, accessible and controllable, respectively.*

the difference between accessibility and controllability at a point $x_0 \in M$ depends on the topological description of the point with respect to the reachable set; that is, if the point is in the interior or not of the reachable set. Thus it is clear to see that controllability implies accessibility, but not always in the other way round as the figure depicts. For instance, the second picture is accessible, but not controllable. However, the third one is both accessible and controllable.

3.3   Sufficient conditions for small–time locally controllability

Accessibility is a notion that is well–studied, but this is not the case with controllability. There are different results about zeroth–order, first–order and second–order sufficient conditions for STLC related with the Lie brackets of the vector fields involved in the system [Bianchini and Stefani 1993, Lewis and Murray 1997, Sussmann 1978; 1983; 1987]. There are also some algebraic sufficient conditions for STLC under research by Aguilar and Lewis [2008]. Here we just recall one of those results, one that will be useful in Chapter 6.

We are only interested in the sufficient condition for STLC stated in Theorem 3.3.2 because in Chapters 4 and 6 that result connects with particular solutions to an optimal control problem.

Let $\mathbf{X} = \{X_0, X_1, \ldots, X_l\}$ be a set of indeterminates, the free Lie algebra $\mathrm{L}(\mathbf{X})$ is the

$\mathbb{R}$–vector space generated by the indeterminates and their formal brackets, that satisfy the relations of skew–symmetry and the Jacobi identity. Thus, the product associated with this algebra satisfies the properties of the Lie bracket used in the Lie algebras of vector fields on a manifold. See [Lewis and Murray 1997] and references therein for more details about finitely generated free Lie algebras and for the proofs of the results stated here.

Proposition 3.3.1. *Every element of* $\mathrm{L}(\mathbf{X})$ *is a finite linear combination of brackets of the form*

$$[X_l, [X_{l-1}, [\ldots, [X_2, X_1], \ldots]]]$$

where $X_i \in \mathbf{X}$ *for every* $i = 1, \ldots, l$.

Let $\mathrm{Br}(\mathbf{X})$ be the set of brackets in $\mathrm{L}(\mathbf{X})$; that is, the Lie monomials in $\mathrm{L}(\mathbf{X})$. Thus, $\mathrm{Br}(\mathbf{X})$ generates $\mathrm{L}(\mathbf{X})$ as a $\mathbb{R}$–vector space. For every element B in $\mathrm{Br}(\mathbf{X})$, it makes sense to count the number of times that X_s appears in B. That number is denoted by $|B|_s$. The sum $\sum_{s=0}^{l} |B|_s$ is called the *degree of B*. If $|B|_s$ is an even number for every $s = 1, \ldots, l$ and $|B|_0$ is an odd number, then the bracket $B \in \mathrm{Br}(\mathbf{X})$ is a *bad bracket*. Otherwise, $B \in \mathrm{Br}(\mathbf{X})$ is a *good bracket*.

Given the family of vector fields $\mathcal{V} = \{f_0, f_1, \ldots, f_l\}$ on M, we define a bijection ϕ between $\mathbf{X}$ and $\mathcal{V}$ as follows: $\phi(X_s) = f_s$ for $s = 0, \ldots, l$. Then, by the universal property of a free Lie algebra, we have the following Lie algebra homomorphism

$$\mathrm{Ev}(\phi) \colon \mathrm{L}(\mathbf{X}) \longrightarrow \mathfrak{X}(M),$$

such that each occurrence of X_s in a bracket in $\mathrm{L}(\mathbf{X})$ is replaced by f_s according to the bijection ϕ. Thus, the smallest Lie subalgebra of $\mathfrak{X}(M)$ which contains $\mathcal{V}$ is exactly the image of $\mathrm{L}(\mathbf{X})$ under the homomorphism $\mathrm{Ev}(\phi)$. Observe that the image of $\mathrm{L}(\mathbf{X})$ is the accessibility algebra of the system given by $\mathcal{V}$, see Definition 3.2.4.

For every $x \in M$, $\mathrm{Ev}_x(\phi) \colon \mathrm{L}(\mathbf{X}) \to T_x M$ evaluates the vector field $\mathrm{Ev}(\phi)(X)$ at x. It is said that $\mathcal{V}$ is satisfies the *Lie algebra rank condition (LARC) at $x \in M$* if $\mathrm{Ev}_x(\phi)(\mathrm{L}(\mathbf{X})) = T_x M$.

Let S_l be the permutation group of l elements. For $\pi \in S_l$ and $B \in \mathrm{Br}(\mathbf{X})$, $\overline{\pi}(B)$ is the bracket obtained by fixing X_0 and sending X_s to $X_{\pi(s)}$ for $s = 1, \ldots, l$. Define the bracket

$$\beta(B) = \sum_{\pi \in S_l} \overline{\pi}(B).$$

Theorem 3.3.2. *([Sussmann 1987]) Let* $\mathbf{X}$ *be the free Lie algebra generated by the set of indeterminates* $\{X_0, X_1, \ldots, X_l\}$ *and the control–affine system* $X(x, u) = f_0(x) + u^s f_s(x)$.

Consider the bijection $\phi : \mathbf{X} \to \{f_0, f_1, \ldots, f_l\}$ *which sends* X_s *to* f_s *for* $s = 0, \ldots, l$.

Suppose that the given control system is such that every bad bracket $B \in \mathrm{Br}(\mathbf{X})$ *satisfies*

that

$$\mathrm{Ev}_x(\phi)(\beta(B)) = \sum_{i=1}^{r} \mu^i \mathrm{Ev}_x(\phi)(C_i)$$

where C_i are good brackets in $\mathrm{Br}(\mathbf{X})$ of lower degree than B and $\mu^i \in \mathbb{R}$. If the control–affine system satisfies the LARC at x, then the system is STLC at x.

Thus, when the bad brackets do not satisfy the property in Theorem 3.3.2, it is said that they are obstructions to the controllability of the system. This result will be useful in Chapter 6 to discuss when a tentative solution to an optimal control problem can be found.

Chapter 4
Geometric Pontryagin's Maximum Principle

In 1958 the International Congress of Mathematicians was held in Edinburgh, Scotland, where, for the first time, L. S. Pontryagin talked publicly about the Maximum Principle. This Principle was developed by a research group on automatic control created by Pontryagin in the 1950's. He was engaged in applied mathematics by his friend A. Andronov because scientists in the Steklov Mathematical Institute were asked to carry out applied research, especially in the field of aircraft dynamics.

At the same time, in the regular seminars on automatic control in the Institute of Automatics and Telemechanics, A. Feldbaum introduced Pontryagin and his collegues to the time–optimization problem. This allowed them to study how to find the best way of piloting an aircraft in order to defeat a zenith fire point in the shortest time as a time–optimization problem.

Since the equations for modelling aircraft dynamics are nonlinear and since the control of the rear end of the aircraft runs over a bounded subset, it was necessary to reformulate the calculus of variations known at that time. Taking into account ideas suggested by McShane [1939], Pontryagin and his collaborators managed to state and prove the Maximum Principle, which was published in Russian in 1961 and translated into English [Pontryagin et al. 1962] the following year. See [Boltyanski et al. 1999] for more historical remarks.

Pontryagin's Maximum Principle is considered as an outstanding achievement of the optimal control theory. It has been used in a wide range of applications, such as medicine, traffic flow, robotics, economy, etc. Nevertheless, it is worth remarking that the Maximum Principle does not give sufficient conditions to compute an optimal trajectory; it only provides necessary conditions. Thus only candidates to be optimal trajectories are found. To determine if they are optimal or not, other results related to the existence of solutions for these problems are needed. See [Agrachev and Sachkov 2004, Athans and Falb 1966, Filippov 1962, Lee and Markus 1967] for more details.

Having in mind the definition of a vector field along a projection given in Chapter 2, we give two different statements of Pontryagin's Maximum Principle. In §4.1, §4.2, it is studied the optimal control problem with both the time interval and the endpoints given. If the final time is not given and the endpoints are not fixed but they must be in specific submanifolds, then the problem is studied in §4.3, §4.4. These four sections have been written in an analogous way. First of all, two different but equivalent statements of the optimal control problems are

given. The so–called extended system is the useful one for the proofs in §4.2, §4.4 because the functional to be minimized is included as a new coordinate of the system. In §4.1.4 and §4.3.3 the associated Hamiltonian problem, that leads to the statements of Maximum Principle and their proofs in §4.2 and §4.4, is explained.

One part of the proof of Pontryagin's Maximum Principle consists of perturbing the given optimal curve, therefore we introduce in §4.1.3 and §4.3.2 how this curve can be perturbed depending on the known data. Above all, it is important the complete proof of Proposition 4.1.12, although known, to our knowledge, there is not a self–contained proof of it in the literature.

Our purpose is to give an intrinsic proof of the Maximum Principle, but at some point it will be necessary the use of local results and coordinate expressions. For the understanding and the proof of the Maximum Principle, it is essential to refresh all the elements defined in Chapter 2.1 and Appendices A and B, as well the notations and properties related with them. These above–mentioned elements are mainly vector fields along projections, Lebesgue times, convex cones and separating hyperplanes.

In §4.5, we focus on obtaining necessary conditions for the existence of abnormal optimal curves. These conditions are related with the concepts of accessibility and controllability introduced in Chapter 3. The linear controllability of the system defined in §3.2 is used to characterize abnormality, what justifies the careful study in §4.5.2 about the connection between the reachable set and the tangent perturbation cone. The way to think about this relation is that the tangent perturbation cone contains, in some sense, the vectors tangent to perturbation curves.

Finally, in §4.6 some examples are given to illustrate how the Maximum Principle is used to solve optimal control problems and also to highlight the geometric part in the proof.

4.1 Pontryagin's Maximum Principle for fixed time and fixed endpoints

First we study Pontryagin's Maximum Principle with fixed time and fixed endpoints, that corresponds with the most simplified version of the necessary conditions for optimality.

4.1.1 Statement of optimal control problem and notation

Let M be a differentiable manifold of dimension m and $U \subset \mathbb{R}^k$ a subset. Let us consider the trivial Euclidean bundle $\pi\colon M \times U \to M$.

Let X be a vector field along the projection $\pi\colon M \times U \to M$ as in Definition 2.2.1. If (V, x^i) is a local chart at x in M, the local expression of the vector field is $X = f^i \partial/\partial x^i$ where f^i are functions defined on $V \times U$. Given a measurable and bounded control $u\colon I \to U$, this vector field can be rewritten as $X^{\{u\}}$ according to Equation (2.2.4).

Let $I \subset \mathbb{R}$ be an interval and $(\gamma, u)\colon I \to M \times U$ a curve where γ is an absolutely

continuous curve. Given $\mathcal{F}\colon M \times U \to \mathbb{R}$, let us consider the functional

$$\mathcal{S}[\gamma, u] = \int_I \mathcal{F}(\gamma, u)\,\mathrm{d}t$$

defined on curves (γ, u) with a compact interval as domain. The function $\mathcal{F}\colon M \times U \to \mathbb{R}$ is continuous on $M \times U$ and continuously differentiable with respect to M on $M \times U$.

Statement 4.1.1. *(Optimal Control Problem, OCP) Given the elements $M, U, X, \mathcal{F}, I = [a, b]$ and the endpoint conditions $x_a, x_b \in M$, consider the following problem.*

Find (γ^*, u^*) such that

(1) $\gamma^*(a) = x_a$, $\gamma^*(b) = x_b$ (endpoint conditions),

(2) γ^* is an integral curve of $X^{\{u^*\}}$: $\dot{\gamma}^*(t) = X(\gamma^*(t), u^*(t))$, for a.e. $t \in I$, and

(3) $\mathcal{S}[\gamma^*, u^*]$ is minimum over all curves (γ, u) satisfying (1) and (2), (minimal condition).

The tuple $(M, U, X, \mathcal{F}, I, x_a, x_b)$ denotes the *optimal control problem*. The function $\mathcal{F}$ is called the *cost function* of the problem. The mappings $u\colon I \to U$ are called *controls*.

Comments:

1. The curves considered in the previous statement satisfy the same properties as the generalized integral curves of Carathéodory vector fields along a projection described in §2.2.1. That is, γ is absolutely continuous and the controls u are measurable and bounded.

2. Locally, condition (2) is equivalent to the fact that the curve (γ^*, u^*) satisfies the differential equation $\dot{x}^i = f^i$.

4.1.2 The extended problem

Taking into account the elements defining the optimal control problem and their properties, we state an equivalent problem.

Given the OCP $(M, U, X, \mathcal{F}, I, x_a, x_b)$, let us consider the extended manifold $\widehat{M} = \mathbb{R} \times M$ and the trivial Euclidean bundle $\widehat{\pi}\colon \widehat{M} \times U \to \widehat{M}$.

Let $\widehat{X}$ be the following vector field along the projection $\widehat{\pi}\colon \widehat{M} \times U \to \widehat{M}$:

$$\widehat{X}(x^0, x, u) = \mathcal{F}(x, u)\partial/\partial x^0|_{(x^0, x, u)} + X(x, u),$$

where x^0 is the natural coordinate on $\mathbb{R}$. According to Equation (2.2.4), this vector field can be rewritten as $\widehat{X}^{\{u\}}$.

Given a curve $(\widehat{\gamma}, u) = ((x^0 \circ \widehat{\gamma}, \gamma), u)\colon I \to \widehat{M} \times U$ such that $\widehat{\gamma}$ is absolutely continuous and u is measurable and bounded, the previous elements come together in the following

diagram:

$$
\begin{array}{ccc}
 & & T\widehat{M} \\
 & \overset{\widehat{X}}{\nearrow} & \downarrow \tau_{\widehat{M}} \\
\widehat{M} \times U & \overset{\widehat{\pi}}{\longrightarrow} & \widehat{M} \\
(\widehat{\gamma}, u) \uparrow & \overset{\widehat{\gamma}}{\nearrow} & \downarrow \pi_2 \\
I & \overset{\gamma}{\longrightarrow} & M
\end{array}
$$

where π_2 is the projection of $\widehat{M}$ onto M.

Statement 4.1.2. *(Extended Optimal Control Problem, $\widehat{\mathrm{OCP}}$) Given the above–mentioned OCP $(M, U, X, \mathcal{F}, I, x_a, x_b)$, $\widehat{M}$ and $\widehat{X}$, consider the following problem.*

Find $(\widehat{\gamma}^*, u^*)$ such that

(1) $\widehat{\gamma}^*(a) = (0, x_a)$, $\gamma^*(b) = x_b$ (endpoint conditions),

(2) $\widehat{\gamma}^*$ is an integral curve of $\widehat{X}^{\{u^*\}}$: $\dot{\widehat{\gamma}}^*(t) = \widehat{X}(\widehat{\gamma}^*(t), u^*(t))$, for a.e. $t \in I$, and

(3) $\gamma^{*^0}(b)$ is minimum over all curves $(\widehat{\gamma}, u)$ satisfying (1) and (2), (minimal condition).

The tuple $(\widehat{M}, U, \widehat{X}, I, x_a, x_b)$ denotes the *extended optimal control problem.*

Comments:

1. The functional $\gamma^{*^0}(b)$ to be minimized in the $\widehat{\mathrm{OCP}}$ is equal to the functional defined in the OCP. That is to say, we have

$$
\widehat{\mathcal{S}}[\widehat{\gamma}, u] = \gamma^0(b) = \int_a^b \mathcal{F}(\gamma, u)\mathrm{d}t = \mathcal{S}[\gamma, u]
$$

   for curves $(\widehat{\gamma}, u)$.

2. Locally, the condition (2) is equivalent to the fact that the curve $(\widehat{\gamma}^*, u^*)$ satisfies the differential equations $\dot{x}^0 = \mathcal{F}$, $\dot{x}^i = f^i$.

The elements in the problem $(\widehat{M}, U, \widehat{X}, I, x_a, x_b)$ satisfy properties analogous to the ones fulfilled by the elements in the problem $(M, U, X, \mathcal{F}, I, x_a, x_b)$, but for different spaces; see §2.1, §4.1.1 for more details about the properties.

4.1.3 Perturbation and associated cones

The following constructions can be defined for any vector field depending on parameters—see §2.2.1—in particular, for those vector fields defining a control system. In order not to make the notation harder, we will construct everything on M, but the same can be done on $\widehat{M}$ or on any other convenient manifold, as for instance the tangent bundle TQ for the mechanical case; see Chapter 7.

4.1.3.1 Elementary perturbation vectors: class I

Now we study how integral curves of the time–dependent vector field $X^{\{u\}} \colon I \times M \to TM$, introduced in §2.2.2, change when the control u is perturbed in a small interval.

In the sequel, a measurable and bounded control $u \colon I = [a, b] \subset \mathbb{R} \to U$ and an absolutely continuous integral curve $\gamma \colon I \to M$ of $X^{\{u\}}$ are given. Let $\pi_1 = \{t_1, l_1, u_1\}$, where t_1 is a Lebesgue time in (a, b) always for the $X \circ (\gamma, u)$—i.e., it satisfies Equation (A.2.3)—$l_1 \in \mathbb{R}^+$, $u_1 \in U$. From now on, to simplify, t_1 is called just a Lebesgue time. For every $s \in \mathbb{R}^+$ small enough such that $a < t_1 - l_1 s$, consider $u[\pi_1^s] \colon I \to U$ defined by

$$u[\pi_1^s](t) = \begin{cases} u_1, & t \in [t_1 - l_1 s, t_1], \\ u(t), & \text{elsewhere.} \end{cases}$$

Definition 4.1.3. *The function $u[\pi_1^s]$ is called an **elementary perturbation of** u **specified by the data** $\pi_1 = \{t_1, l_1, u_1\}$. It is also called a **needle–like variation**.*

Associated to $u[\pi_1^s]$, consider the mapping $\gamma[\pi_1^s] \colon I \to M$, the generalized integral curve of $X^{\{u[\pi_1^s]\}}$ with initial condition $(a, \gamma(a))$. Figure 4.1 shows the situation obtained as a result of perturbing the control.
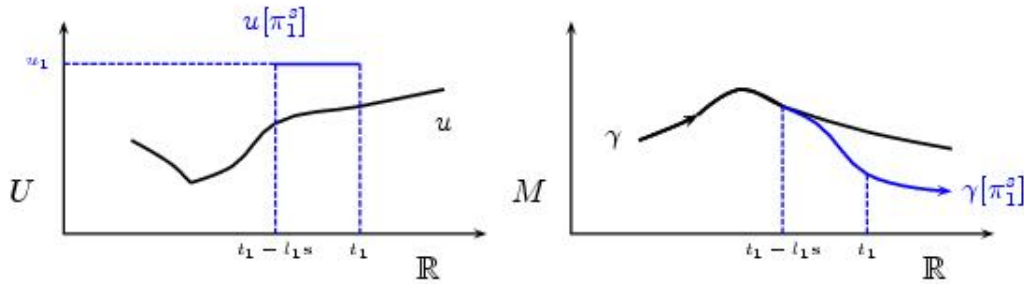
Figure 4.1: *Elementary perturbation of u specified by the data π_1 and the integral curve on M of $X^{\{u[\pi_1^s]\}}$.*

Given $\epsilon > 0$, define the map

$$\begin{aligned} \varphi_{\pi_1} \colon I \times [0, \epsilon] &\longrightarrow M \\ (\,t\,,\,s\,) &\longmapsto \varphi_{\pi_1}(t, s) = \gamma[\pi_1^s](t). \end{aligned}$$

For every $t \in I$, $\varphi_{\pi_1}^t \colon [0, \epsilon] \to M$ is given by $\varphi_{\pi_1}^t(s) = \varphi_{\pi_1}(t, s)$.

As the controls are assumed to be measurable and bounded, it makes sense to define the distance between two controls $u, \overline{u} \colon I \to U$ as follows

$$d(u, \overline{u}) = \int_I \|u(t) - \overline{u}(t)\| \; \mathrm{d}t,$$

where $\|\cdot\|$ is the usual norm in $\mathbb{R}^k$. Here, a bounded control means that there exists a compact

set in U that contains Im u. The control $u[\pi_1^s]$ depends continuously on the parameters s and $\pi_1 = \{t_1, l_1, u_1\}$; that is, given $\epsilon > 0$ there exists $\delta > 0$ such that if $|t_1 - t_2| < \delta$, $|l_1 - l_2| < \delta$, $\|u_1 - u_2\| < \delta$, $|s_1 - s_2| < \delta$, then $d(u[\pi_1^{s_1}], u[\pi_2^{s_2}]) < \epsilon$.

Hence the curve $\varphi_{\pi_1}^t$ depends continuously on s and $\pi_1 = \{t_1, l_1, u_1\}$, then it converges uniformly to γ as s tends to 0. See [Cañizo–Rincón 2004, Coddington and Levinson 1955] for more details of the differential equations depending continuously on parameters.

Let us prove that for a Lebesgue time t_1 the curve $\varphi_{\pi_1}^{t_1}$ has a tangent vector at $s = 0$.

Proposition 4.1.4. *Let t_1 be a Lebesgue time. If $u[\pi_1^s]$ is an elementary perturbation of u specified by the data $\pi_1 = \{t_1, l_1, u_1\}$, then the curve $\varphi_{\pi_1}^{t_1} : [0, \epsilon] \to M$, $\varphi_{\pi_1}^{t_1}(s) = \gamma[\pi_1^s](t_1)$, is differentiable at $s = 0$ and its tangent vector is $[X(\gamma(t_1), u_1) - X(\gamma(t_1), u(t_1))]\, l_1$.*

(*Proof*) It is enough to prove that for every differentiable function $g \colon M \to \mathbb{R}$, there exists

$$A = \lim_{s \to 0} \frac{g(\varphi_{\pi_1}^{t_1}(s)) - g(\varphi_{\pi_1}^{t_1}(0))}{s}.$$

As this is a derivation on the functions defined on a neighbourhood of $\gamma(t_1)$, it is enough to prove the proposition for the coordinate functions x^i of a local chart at $\gamma(t_1)$. Thus take $g = x^i$:

$$
\begin{aligned}
A &= \lim_{s \to 0} \frac{(x^i \circ \varphi_{\pi_1}^{t_1})(s) - (x^i \circ \varphi_{\pi_1}^{t_1})(0)}{s} = \lim_{s \to 0} \frac{(x^i \circ \gamma[\pi_1^s])(t_1) - (x^i \circ \gamma)(t_1)}{s} \\
&= \lim_{s \to 0} \frac{\gamma^i[\pi_1^s](t_1) - \gamma^i(t_1)}{s}.
\end{aligned}
$$

As γ is an absolutely continuous integral curve of $X^{\{u\}}$, $\dot{\gamma}(t) = X(\gamma(t), u(t))$ at every Lebesgue time $t \in I$. Then, integrating,

$$\gamma^i(t_1) - \gamma^i(a) = \int_a^{t_1} f^i(\gamma(t), u(t))\mathrm{d}t$$

and similarly for $\gamma[\pi_1^s]$ and $u[\pi_1^s]$. As Figure 4.1 shows, $\gamma[\pi_1^s](t) = \gamma(t)$ and $u[\pi_1^s](t) = u(t)$ for $t \in [a, t_1 - l_1 s)$. Then

$$
\begin{aligned}
A &= \lim_{s \to 0} \frac{\int_a^{t_1} f^i(\gamma[\pi_1^s](t), u[\pi_1^s](t))\mathrm{d}t - \int_a^{t_1} f^i(\gamma(t), u(t))\mathrm{d}t}{s} \\
&= \lim_{s \to 0} \frac{\int_{t_1 - l_1 s}^{t_1} f^i(\gamma[\pi_1^s](t), u_1)\mathrm{d}t - \int_{t_1 - l_1 s}^{t_1} f^i(\gamma(t), u(t))\mathrm{d}t}{s}.
\end{aligned}
$$

As t_1 is a Lebesgue time, we use Equation (A.2.3):

$$\int_{t-h}^t X(\gamma(s), u(s))\mathrm{d}s = h X(\gamma(t), u(t)) + o(h)$$

in such a way that

$$A = \lim_{s \to 0} \frac{f^i(\gamma[\pi_1^s](t_1), u_1)l_1 s - f^i(\gamma(t_1), u(t_1))l_1 s + o(s)}{s}$$

$$= \lim_{s \to 0} [f^i(\gamma[\pi_1^s](t_1), u_1) - f^i(\gamma(t_1), u(t_1))] \, l_1.$$

As f^i is continuous on M, we have

$$
\begin{aligned}
A &= [f^i(\lim_{s \to 0} \gamma[\pi_1^s](t_1), u_1) - f^i(\gamma(t_1), u(t_1))] \, l_1 = [f^i(\gamma(t_1), u_1) - f^i(\gamma(t_1), u(t_1))] \, l_1 \\
&= [(X(\gamma(t_1), u_1) - X(\gamma(t_1), u(t_1))) \, l_1] \, (x^i).
\end{aligned}
$$

$\blacksquare$

Definition 4.1.5. *The tangent vector* $v[\pi_1] = l_1 \, [X(\gamma(t_1), u_1) - X(\gamma(t_1), u(t_1))] \in T_{\gamma(t_1)} M$ *is the **elementary perturbation vector associated to the perturbation data** $\pi_1 = \{t_1, l_1, u_1\}$. It is also called a **perturbation vector of class I**.*

Comments:

(a) The previous proof shows the importance of defining perturbations only at Lebesgue times, otherwise the elementary perturbation vectors may not exist.

(b) Observe that if we change $\pi_1 = \{t_1, l_1, u_1\}$ for $\pi_2 = \{t_1, l_2, u_1\}$, then it is satisfied that $v[\pi_1] = (l_1/l_2) \, v[\pi_2]$. If $v[\pi_1]$ is a perturbation vector of class I and $\lambda \in \mathbb{R}^+$, then $\lambda \, v[\pi_1]$ is also a perturbation vector of class I with perturbation data $\{t_1, \lambda \, l_1, u_1\}$.

(c) We write $\mathcal{L}(w)g$ for the derivative of the function $g \in \mathcal{C}^\infty(M)$ in the direction given by the vector $w \in T_x M$. Due to Proposition 4.1.4, for every differentiable function $g \colon M \to \mathbb{R}$ we have

$$\frac{g(\varphi_{\pi_1}^{t_1}(s)) - g(\gamma(t_1)) - s \, \mathcal{L}(v[\pi_1])g}{s} \xrightarrow[s \to 0]{} 0.$$

Hence

$$g\left(\varphi_{\pi_1}^{t_1}(s)\right) = g\left(\gamma(t_1)\right) + s \, \mathcal{L}(v[\pi_1])g + o(s).$$

If (x^i) are coordinates of a local chart at $\gamma(t_1)$,

$$x^i\left(\varphi_{\pi_1}^{t_1}(s)\right) = x^i\left(\gamma(t_1)\right) + s \, v[\pi_1]^i + o(s).$$

That is,

$$\left(\varphi_{\pi_1}^{t_1}\right)^i(s) = \gamma^i(t_1) + s \, v[\pi_1]^i + o(s).$$

Now, if we identify the open set of the local chart and the tangent space to M at $\gamma(t_1)$ with the same space $\mathbb{R}^m$, we write the following linear approximation

$$\varphi_{\pi_1}^{t_1}(s) = \gamma(t_1) + s \, v[\pi_1] + o(s). \tag{4.1.1}$$

The initial condition for the velocity given by the elementary perturbation vector evolves along the reference trajectory γ through the integral curves of the complete lift $\left(X^T\right)^{\{u\}}$ of $X^{\{u\}}$, as explained in §2.2.2.1. Note that $\varphi_{\pi_1}^t(s) = \Phi_{(t,t_1)}^{X^{\{u\}}}\left(\varphi_{\pi_1}^{t_1}(s)\right)$ for $t \geq t_1$ because of the definition of φ_{π_1} and $u[\pi_1^s]$.

Proposition 4.1.6. *Let $V[\pi_1]\colon [t_1, b] \to TM$ be the integral curve of the complete lift $\left(X^T\right)^{\{u\}}$ of $X^{\{u\}}$ with initial condition $(t_1, v[\pi_1])$ where $v[\pi_1] \in T_{\gamma(t_1)}M$. For every Lebesgue time $t \in (t_1, b]$, $V[\pi_1](t)$ is the tangent vector to the curve $\varphi^t_{\pi_1}\colon [0, \epsilon] \to M$ at $s = 0$.*

(*Proof*) The proof follows from Proposition 2.2.3 and the definition of the curves considered. ∎

4.1.3.2   Perturbation vectors of class II

The control can be perturbed twice instead of only once, in fact it may be modified a finite number of times. If t_2 is a Lebesgue time greater than t_1 and we perturb the control with $\pi_1 = \{t_1, l_1, u_1\}$ and $\pi_2 = \{t_2, l_2, u_2\}$, then the perturbation data $\pi_{12} = \{(t_1, t_2), (l_1, l_2), (u_1, u_2)\}$ is obtained and is given by

$$u[\pi^s_{12}](t) = \begin{cases} u_1, & t \in [t_1 - l_1 s, t_1], \\ u_2, & t \in [t_2 - l_2 s, t_2], \\ u(t), & \text{elsewhere,} \end{cases}$$

for every $s \in \mathbb{R}^+$ small enough such that $[t_1 - l_1 s, t_1] \cap [t_2 - l_2 s, t_2] = \emptyset$. Then $\gamma[\pi^s_{12}]\colon I \longrightarrow M$ is the generalized integral curve of $X^{\{u[\pi^s_{12}]\}}$ with initial condition $(a, \gamma(a))$. Observe that $\gamma[\pi^0_{12}](t) = \gamma(t)$. Consider the curve $\varphi^{t_2}_{\pi_{12}}\colon [0, \epsilon] \to M$ given by $\varphi^{t_2}_{\pi_{12}}(s) = \gamma[\pi^s_{12}](t_2)$.

Proposition 4.1.7. *Let t_1, t_2 be Lebesgue times such that $t_1 < t_2$. If $u[\pi^s_{12}]$ is the perturbation of u specified by the perturbation data $\pi_{12} = \{(t_1, t_2), (l_1, l_2), (u_1, u_2)\}$, then the vector tangent to $\varphi^{t_2}_{\pi_{12}}\colon [0, \epsilon] \to M$ at $s = 0$ is $v[\pi_2] + V[\pi_1](t_2)$, where $V[\pi_1]\colon [t_1, b] \to TM$ is the generalized integral curve of $\left(X^T\right)^{\{u\}}$ with initial condition $(t_1, v[\pi_1])$ where $v[\pi_1] \in T_{\gamma(t_1)}M$.*

(*Proof*) Here we perturb the control first with π_1 along γ and we obtain $u[\pi^s_1]$. Then we perturb this last control with the other perturbation data, π_2, along $\gamma[\pi^s_1]$. Then the superindices of the tangent vectors denotes the curve along which the perturbation is made. As in the proof of Proposition 4.1.4,

$$\begin{aligned} A &= \lim_{s \to 0} \frac{(x^i \circ \varphi^{t_2}_{\pi_{12}})(s) - (x^i \circ \varphi^{t_2}_{\pi_{12}})(0)}{s} = \lim_{s \to 0} \frac{(x^i \circ \gamma[\pi^s_{12}])(t_2) - (x^i \circ \gamma)(t_2)}{s} \\ &= \lim_{s \to 0} \frac{\gamma^i[\pi^s_{12}](t_2) - \gamma^i(t_2)}{s} = \lim_{s \to 0} \left(\frac{\gamma^i[\pi^s_{12}](t_2) - \gamma^i[\pi^s_1](t_2)}{s} + \frac{\gamma^i[\pi^s_1](t_2) - \gamma^i(t_2)}{s} \right). \end{aligned}$$

We understand $\gamma[\pi^s_{12}]$ as the result of perturbing $\gamma[\pi^s_1]$ with π_2, and use the linear approximation in Equation (4.1.1) for $\gamma[\pi^s_{12}](t_2)$ and $\gamma[\pi^s_1](t_2)$ according to Proposition 4.1.4:

$$\varphi^{t_2}_{\pi_{12}}(s) = \gamma[\pi^s_{12}](t_2) = \gamma[\pi^s_1](t_2) + s\, v[\pi_2]^{\gamma[\pi^s_1]} + o(s),$$

$$\gamma[\pi^s_1](t_2) = \gamma(t_2) + s\, V[\pi_1]^\gamma(t_2) + o(s).$$

Then

$$A = \lim_{s \to 0} \left(\frac{s(v[\pi_2]^{\gamma[\pi^s_1]})^i}{s} + \frac{s(V[\pi_1]^\gamma)^i(t_2)}{s} \right) = \lim_{s \to 0} \left((v[\pi_2]^{\gamma[\pi^s_1]})^i + (V[\pi_1]^\gamma)^i(t_2) \right).$$

As $\gamma[\pi_1^s]$ depends on s and $\lim_{s\to 0}\gamma[\pi_1^s](t) = \gamma(t)$, $A = \mathcal{L}\left(v[\pi_2]^\gamma + V[\pi_1]^\gamma(t_2)\right)x^i$. ∎

Considering identifications similar to the ones used to write Equation (4.1.1), we have

$$\varphi_{\pi_{12}}^{t_2}(s) = \gamma(t_2) + sv[\pi_2] + sV[\pi_1](t_2) + o(s).$$

Now we define how the control changes when it is perturbed twice at the same time. If t_1 is a Lebesgue time, $\pi_1' = \{t_1, l_1', u_1'\}$ and $\pi_1'' = \{t_1, l_1'', u_1''\}$ are perturbation data, then $\pi_{11} = \{(t_1, t_1), (l_1', l_1''), (u_1', u_1'')\}$ is a perturbation data given by

$$u[\pi_{11}^s](t) = \begin{cases} u_1', & t \in [t_1 - (l_1' + l_1'')s, t_1 - l_1''s], \\ u_1'', & t \in [t_1 - l_1''s, t_1], \\ u(t), & \text{elsewhere,} \end{cases}$$

for every $s \in \mathbb{R}^+$ small enough such that $a < t_1 - (l_1' + l_1'')s$. Then $\gamma[\pi_{11}^s]\colon I \longrightarrow M$ is the generalized integral curve of $X^{\{u[\pi_{11}^s]\}}$ with initial condition $(a, \gamma(a))$. Observe that $\gamma[\pi_{11}^0](t) = \gamma(t)$. Consider the curve $\varphi_{\pi_{11}}^{t_1}\colon [0, \epsilon] \to M$, defined by $\varphi_{\pi_{11}}^{t_1}(s) = \gamma[\pi_{11}^s](t_1)$.

Proposition 4.1.8. *Let t_1 be a Lebesgue time. If $u[\pi_{11}^s]$ is the perturbation of the control u specified by the data $\pi_{11} = \{(t_1, t_1), (l_1', l_1''), (u_1', u_1'')\}$, then the vector tangent to $\varphi_{\pi_{11}}^{t_1}\colon [0, \epsilon] \to M$ at $s = 0$ is $v[\pi_1'] + v[\pi_1'']$, where $v[\pi_1']$ and $v[\pi_1'']$ are the perturbation vectors of class I associated to π_1' and π_1'' respectively.*

(*Proof*) As in the proof of Proposition 4.1.4,

$$A = \lim_{s\to 0}\frac{(x^i \circ \varphi_{\pi_{11}}^{t_1})(s) - (x^i \circ \varphi_{\pi_{11}}^{t_1})(0)}{s} = \lim_{s\to 0}\frac{\gamma^i[\pi_{11}^s](t_1) - \gamma^i(t_1)}{s}.$$

As γ is an absolutely continuous integral curve of $X^{\{u\}}$, $\dot\gamma(t) = X(\gamma(t), u(t))$ at every Lebesgue time $t \in I$. Then, integrating,

$$\gamma^i(t_1) - \gamma^i(a) = \int_a^{t_1} f^i(\gamma(t), u(t))\mathrm{d}t$$

and similarly for $\gamma[\pi_{11}^s]$ and $u[\pi_{11}^s]$. Observe that $\gamma[\pi_{11}^s](t) = \gamma(t)$ and $u[\pi_{11}^s](t) = u(t)$ for $t \in [a, t_1 - (l_1' + l_1'')s)$. Then

$$\begin{aligned} A &= \lim_{s\to 0}\frac{\int_a^{t_1} f^i(\gamma[\pi_{11}^s](t), u[\pi_{11}^s](t))\mathrm{d}t - \int_a^{t_1} f^i(\gamma(t), u(t))\mathrm{d}t}{s} \\ &= \lim_{s\to 0}\frac{\int_{t_1-(l_1'+l_1'')s}^{t_1} f^i(\gamma[\pi_{11}^s](t), u[\pi_{11}^s](t))\mathrm{d}t - \int_{t_1-(l_1'+l_1'')s}^{t_1} f^i(\gamma(t), u(t))\mathrm{d}t}{s} \\ &= \lim_{s\to 0}\left(\frac{\int_{t_1-(l_1'+l_1'')s}^{t_1-l_1''s}\left[f^i(\gamma[\pi_1'^s](t), u_1') - f^i(\gamma(t), u(t))\right]\mathrm{d}t}{s}\right. \\ &\quad + \left.\frac{\int_{t_1-l_1''s}^{t_1}\left[f^i(\gamma[\pi_{11}^s](t), u_1'') - f^i(\gamma(t), u(t))\right]\mathrm{d}t}{s}\right). \end{aligned}$$

As t_1 and $t_1 - l_1'' s$ are Lebesgue times such that $a < t_1 - (l_1' + l_1'')s$, Equation (A.2.3) is used. Now we have

$$
\begin{aligned}
A &= \lim_{s \to 0} \left(\frac{f^i(\gamma[\pi_1'^s](t_1 - l_1'' s), u_1')l_1' s - f^i(\gamma(t_1 - l_1'' s), u(t_1 - l_1'' s))l_1' s}{s} \right. \\
&\quad + \left. \frac{f^i(\gamma[\pi_{11}^s](t_1), u_1'')l_1'' s - f^i(\gamma(t_1), u(t_1))l_1'' s}{s} \right) \\
&= \lim_{s \to 0} \left(\left(f^i(\gamma[\pi_1'^s](t_1 - l_1'' s), u_1') - f^i(\gamma(t_1 - l_1'' s), u(t_1 - l_1'' s)) \right) l_1' \right. \\
&\quad + \left. \left(f^i(\gamma[\pi_{11}^s](t_1), u_1'') - f^i(\gamma(t_1), u(t_1)) \right) l_1'' \right).
\end{aligned}
$$

As f^i is continuous on $M \times U$ we have

$$
\begin{aligned}
A &= \left(f^i \left(\lim_{s \to 0} \gamma[\pi_1'^s](t_1 - l_1'' s), u_1' \right) - f^i \left(\lim_{s \to 0} \gamma(t_1 - l_1'' s), \lim_{s \to 0} u(t_1 - l_1'' s) \right) \right) l_1' \\
&\quad + \left(f^i \left(\lim_{s \to 0} \gamma[\pi_{11}^s](t_1), u_1'' \right) - f^i \left(\gamma(t_1), u(t_1) \right) \right) l_1'' \\
&= \left(f^i(\gamma(t_1), u_1') - f^i(\gamma(t_1), u(t_1)) \right) l_1' + \left(f^i(\gamma(t_1), u_1'') - f^i(\gamma(t_1), u(t_1)) \right) l_1'' \\
&= \mathcal{L} \left(v[\pi_1'] + v[\pi_1''] \right) (x^i).
\end{aligned}
$$

∎

Analogous to the linear approximation (4.1.1), we have

$$
\varphi_{\pi_{11}}^{t_1}(s) = \gamma(t_1) + sv[\pi_1'] + sv[\pi_1''] + o(s).
$$

If we perturb the control r times, $\pi = \{\pi_1, \ldots, \pi_r\}$, with $a < t_1 \leq \ldots \leq t_r < b$, then $\gamma[\pi^s](t)$ is the generalized integral curve of $X^{\{u[\pi^s]\}}$ with initial condition $(a, \gamma(a))$. Consider the curve $\varphi_\pi^t \colon [0, \epsilon] \to M$ for $t \in [t_r, b]$ given by $\varphi_\pi^t(s) = \gamma[\pi^s](t)$.

Corollary 4.1.9. *Let t be a Lebesgue time in $[t_r, b]$. If $u[\pi^s]$ is the perturbation of the control u specified by the data $\pi = \{\pi_1, \ldots, \pi_r\}$, then the vector tangent to the curve $\varphi_\pi^t \colon [0, \epsilon] \to M$ at $s = 0$ is $V[\pi_1](t) + \ldots + V[\pi_r](t)$, where $V[\pi_i] \colon [t_i, b] \to TM$ is the generalized integral curve of $\left(X^T \right)^{\{u\}}$ with initial condition $(t_i, v[\pi_i])$ where $v[\pi_i] \in T_{\gamma(t_i)}M$ for $i = 1, \ldots, r$.*

This corollary may be easily proved by induction using Propositions 4.1.4, 4.1.7, 4.1.8, where all the possibilities of combination of perturbation data have been studied. If w is the vector tangent to φ_π^t at $s = 0$, the perturbation data will be denoted by π_w. Bearing in mind the different combination of vectors in Definition B.1.2, we make the following definition:

Definition 4.1.10. *The conic non–negative combinations of perturbation vectors of class I and displacements by the flow of $X^{\{u\}}$ of perturbation vectors of class I are called **perturbation vectors of class II**.*

4.1.3.3   Perturbation cones

Considering all the elementary perturbation vectors, we define a closed convex cone at every time containing at least all displacements of these vectors. To transport all the elementary perturbation vectors, the pushforward of the flow of the vector field $X^{\{u\}}$ is used. See §2.2.2 for notation related with the evolution operator of time–dependent vector fields. Observe that the second comment after Definition 4.1.5 guarantees that the set of elementary perturbation vectors is a cone.

Definition 4.1.11. *For* $t \in (a,b] \subset \mathbb{R}$*, the **tangent perturbation cone** K_t is the smallest closed convex cone in $T_{\gamma(t)}M$ that contains all the displacements by the flow of $X^{\{u\}}$ of all the elementary perturbations vectors from all Lebesgue times τ smaller than t:*

$$K_t = \overline{\mathrm{conv}\left(\bigcup_{\substack{a < \tau \leq t \\ \tau \text{ is a Lebesgue time}}} (\Phi^{X^{\{u\}}}_{(t,\tau)})_*(\mathcal{V}_\tau) \right)},$$

where $\mathcal{V}_\tau$ denotes the set of elementary perturbation vectors at τ and $\mathrm{conv}(A)$ means the convex hull of the set A.

To prove the following statement, we use results in Appendices A and B; precisely Corollary A.3.2, and Propositions B.1.4 and B.1.5.

Proposition 4.1.12. *Let* $t \in (a,b]$*. If v is a nonzero vector in the interior of the tangent perturbation cone K_t, then there exists $\epsilon > 0$ such that for every $s \in (0, \epsilon)$ there exist $s' > 0$ and a perturbation of the control $u[\pi^s_{w_0}]$ such that $\gamma[\pi^s_{w_0}](t) = \gamma(t) + s'v$ where w_0 is the perturbation vector associated to the perturbation of the control.*

(*Proof*)   As v is interior to K_t, by Proposition B.1.5, item (d), v is in the interior of the cone

$$\mathcal{C} = \mathrm{conv}\left(\bigcup_{\substack{a < \tau \leq t \\ \tau \text{ is a Lebesgue time}}} \left(\Phi^{X^{\{u\}}}_{(t,\tau)} \right)_* \mathcal{V}_\tau \right),$$

where $\mathcal{V}_\tau$ is the cone of elementary perturbation vectors at time τ. Hence, v can be expressed as a convex finite combination of perturbation vectors of class I by Proposition B.1.4.

Let $(W, (x^i))$ be a local chart of M at $\gamma(t)$. We suppose that the image of the local chart and W are identified locally with an open subset of $\mathbb{R}^m$. Through the local chart we also identify $T_{\gamma(t)}M$ with $\mathbb{R}^m$. We consider the affine hyperplane Π orthogonal to v at the endpoint of the vector v and identify Π with $\mathbb{R}^{m-1}$.

A "closed" cone denotes a closed cone without the vertex. Observe that such a cone is not closed, that is why we use the inverted commas. We can choose a "closed" convex cone $\widetilde{\mathcal{C}}$ contained in the interior of $\mathcal{C}$ such that v lies in the interior of $\widetilde{\mathcal{C}}$ and $\langle w, v \rangle > 0$ for every $w \in \widetilde{\mathcal{C}}$. For example, we can consider a circular cone with axis v satisfying the two previous

conditions, as assumed from now on. Hence

$$\Pi \cap \widetilde{\mathcal{C}} = v + \overline{B(0, R)},$$

where $\overline{B(0, R)}$ is the closure of an open ball in the subspace orthogonal to v, denoted by $v^\perp$. For $r \in v^\perp$, we will write r instead of $0v + r$ as a vector in $\mathbb{R}^m$.

Let us construct a diffeomorphism from the cone $\widetilde{\mathcal{C}}$ to a cylinder of $\mathbb{R}^m$. If $w \in \widetilde{\mathcal{C}}$, the orthogonal decomposition of w induced by v and $v^\perp$ is

$$w = \frac{\langle w, v \rangle}{\|v\|} \frac{v}{\|v\|} + \left(w - \frac{\langle w, v \rangle}{\langle v, v \rangle} v \right) = \frac{\langle w, v \rangle}{\langle v, v \rangle} \left[v + \left(\frac{\langle v, v \rangle}{\langle w, v \rangle} w - v \right) \right].$$

Observe that $\frac{\langle v, v \rangle}{\langle w, v \rangle} w - v$ is a vector in $\overline{B(0, R)} \subset v^\perp$. Considering the "closed" cone $\widetilde{\mathcal{C}}$ without the vertex, we have the map

$$
\begin{aligned}
g \colon \widetilde{\mathcal{C}} &\longrightarrow \mathbb{R}^+ \times \overline{B(0, R)} \\
w &\longmapsto \left(\frac{\langle w, v \rangle}{\langle v, v \rangle}, \frac{\langle v, v \rangle}{\langle w, v \rangle} w - v \right) = (s, r),
\end{aligned}
$$

that is a $\mathcal{C}^\infty$ diffeomorphism with inverse given by

$$
\begin{aligned}
g^{-1} \colon \mathbb{R}^+ \times \overline{B(0, R)} &\longrightarrow \widetilde{\mathcal{C}} \\
(s, r) &\longmapsto s(v + r) = w.
\end{aligned}
$$

Note that g and g^{-1} can be extended to an open cone, without the vertex, containing $\widetilde{\mathcal{C}}$, so the condition that g is a diffeomorphism is clear.

If we truncate $\widetilde{\mathcal{C}}$ by the affine hyperplane Π, we obtain a bounded convex set $\widetilde{\mathcal{C}}_v$. The restriction of g to $\widetilde{\mathcal{C}}_v$ is $g_v \colon \widetilde{\mathcal{C}}_v \to (0, 1] \times \overline{B(0, R)}$, that is also a $\mathcal{C}^\infty$ diffeomorphism with inverse $g_v^{-1} \colon (0, 1] \times \overline{B(0, R)} \to \widetilde{\mathcal{C}}_v$.

If $r \in \overline{B(0, R)}$, then $w_0 = v + r$ is interior to $\mathcal{C}$. Hence, associated to w_0 we have a perturbation π_{w_0} of the control u. Let $\gamma[\pi_{w_0}^s] \colon I \to M$ be the generalized integral curve of $X^{\{u[\pi_{w_0}^s]\}}$ with initial condition $(a, \gamma(a))$ and consider the map

$$
\begin{aligned}
\Gamma \colon [0, 1] \times \overline{B(0, R)} &\longrightarrow M \\
(s, r) &\longmapsto \Gamma(s, r) = \gamma[\pi_{w_0}^s](t) \\
(0, r) &\longmapsto \Gamma(0, r) = \gamma(t),
\end{aligned}
\qquad (4.1.2)
$$

which is continuous because $\gamma[\pi_{w_0}^s](t)$ depends continuously on s and $\pi_{w_0}^s$ and

$$\lim_{(s, r) \to (0, r_0)} \Gamma(s, r) = \gamma(t) = \Gamma(0, r_0).$$

Hence, for every $\epsilon > 0$, there exist $\delta_1, \delta_2 > 0$ such that, if $|s| < \delta_1$ and $\|r\| < \delta_2$, then $\|\Gamma(s, r) - \Gamma(0, 0)\| = \|\gamma[\pi_{w_0}^s](t) - \gamma(t)\| < \epsilon$.

Taking $\epsilon > 0$ such that $B(\gamma(t), \epsilon)$ is contained in the open set W of the local chart at $\gamma(t)$,

there exist $\delta_1, \delta_2 > 0$ such that, if $|s| < \delta_1$ and $\|r\| < \delta_2$, then $\gamma[\pi^s_{w_0}](t) \in W$.

We consider now the map

$$
\begin{array}{rcl}
\Delta: & [0, \delta_1] \times \overline{B(0, \delta_2)} & \longrightarrow & T_{\gamma(t)}M \simeq \mathbb{R}^m \\
& (s, r) & \longmapsto & \Delta(s, r) = \gamma[\pi^s_{w_0}](t) - \gamma(t) \\
& (0, r) & \longmapsto & \Delta(0, r) = 0
\end{array}
$$

that is continuous because $\lim_{(s,r) \to (0, r_0)} \Delta(s, r) = 0 = \Delta(0, r_0)$. Note that we have identified W with $\mathbb{R}^m$ via the local chart. With this in mind and using Equation (4.1.1), we can write

$$
\gamma[\pi^s_{w_0}](t) - \gamma(t) = s(v + r) + o_r(s),
$$

where $o_r(s) \in \mathbb{R}^m$.

We are going to show that, taking (s, r) in an appropriate subset, $\Delta(s, r)$ lies in the interior of the cone $\widetilde{\mathcal{C}}$.

Next take a section of the cone through a plane containing v and w, and compute the distance from the endpoint of w to the boundary of the cone $\widetilde{\mathcal{C}}$. This is given by

$$
\frac{s\,(R - \|r\|)}{\sqrt{1 + \left(\frac{R}{\|v\|}\right)^2}}.
$$

This is the maximum value for the radius of an open ball centered at the endpoint of $s(v + r)$ to be contained in $\widetilde{\mathcal{C}}$.

Define the function

$$
\begin{array}{rcl}
\Theta: & [0, \delta_1] \times \overline{B(0, \delta_2)} & \longrightarrow & \mathbb{R}^m \\
& (s, r) & \longmapsto & \left(\gamma[\pi^s_{w_0}](t) - \gamma(t) - s(v + r)\right)/s \ = \ o_r(s)/s \\
& (0, r) & \longmapsto & 0
\end{array}
$$

which is continuous because $\lim_{(s,r) \to (0, r_0)} \Theta(s, r) = 0 = \Theta(0, r_0)$. Taking

$$
\epsilon = \frac{R - \delta_2}{\sqrt{1 + \left(\frac{R}{\|v\|}\right)^2}},
$$

there exist $\overline{\delta}_1, \overline{\delta}_2 > 0$ such that, if $|s| < \overline{\delta}_1$ and $\|r\| < \overline{\delta}_2$, then $\|\Theta(s, r)\| = \|o_r(s)/s\| < \epsilon$.

If $(s, r) \in (0, \overline{\delta}_1) \times \overline{B(0, \overline{\delta}_2)}$, then

$$
\|\Delta(s, r) - s(v + r)\| = \|s(v + r) + o_r(s) - s(v + r)\| = \|o_r(s)\| \leq s\epsilon < s\frac{R - \|r\|}{\sqrt{1 + \left(\frac{R}{\|v\|}\right)^2}}
$$

since $\|r\| \leq \overline{\delta}_2 < \delta_2 < R$. Thus we conclude that $\Delta(s, r) = s(v + r) + o_r(s)$ is in the interior

of the cone $\widetilde{\mathcal{C}}$ for every $(s,r) \in (0,\overline{\delta}_1) \times \overline{B(0,\overline{\delta}_2)}$.

Now, for $s \in (0,\overline{\delta}_1)$, we define the continuous mapping

$$
\begin{aligned}
G_s \colon \overline{B(0,\overline{\delta}_2)} &\longrightarrow \overline{B(0,R)} \subset \mathbb{R}^{m-1} \\
r &\longmapsto G_s(r) = (\pi_2 \circ g \circ \Delta)(s,r),
\end{aligned}
\tag{4.1.3}
$$

where $\pi_2 \colon \mathbb{R}^+ \times \overline{B(0,R)} \to \overline{B(0,R)}$, $\pi_2(s,r) = r$. Observe that for $r_0 \in \overline{B(0,\overline{\delta}_2)}$ we have

$$
\lim_{(s,r)\to(0,r_0)} G_s(r) = \lim_{(s,r)\to(0,r_0)} \left[\frac{\langle v,v\rangle}{s\langle v,v\rangle + \langle o(s),v\rangle}(s(v+r)+o(s)) - v \right] = r
$$

and

$$
(g \circ \Delta)(s,r) = g(\gamma[\pi_{w_0}^s](t) - \gamma(t)) = g(s(v+r) + o_r(s)) = (s',r').
\tag{4.1.4}
$$

Suppose that there exists $r \in \overline{B(0,R)}$ such that $G_s(r) = 0$. Then, applying g^{-1} to the above equation, we have

$$
\Delta(s,r) = \gamma[\pi_{w_0}^s](t) - \gamma(t) = g^{-1}(s',0) = s'v.
\tag{4.1.5}
$$

Hence, to conclude the proof we need to show that there exists r with $G_s(r) = 0$ for s small enough. To apply Corollary A.3.2, there must exist $r' \in B(0,\overline{\delta}_2)$ such that $\|G_s(r) - r\| < \|r - r'\|$ for every $r \in \partial\left(\overline{B(0,\overline{\delta}_2)}\right)$. We will show that the condition is fulfilled for $r' = 0$.

Consider the mapping

$$
\begin{aligned}
\mathcal{G} \colon [0,\overline{\delta}_1] \times \overline{B(0,\overline{\delta}_2)} &\longrightarrow \overline{B(0,R)} \subset \mathbb{R}^{m-1} \\
(s,r) &\longmapsto \mathcal{G}(s,r) = G_s(r) - r \\
(0,r) &\longmapsto \mathcal{G}(0,r) = 0.
\end{aligned}
$$

For $r_0 \in \overline{B(0,\overline{\delta}_2)}$, we have $\lim_{(s,r)\to(0,r_0)} \mathcal{G}(s,r) = \lim_{(s,r)\to(0,r_0)}(G_s(r) - r) = 0$. Thus $\mathcal{G}$ is continuous.

Given $r_0 \in \partial\overline{B(0,\overline{\delta}_2)}$, take $\epsilon = \overline{\delta}_2/2$. Then there exist $\delta_0(0,r_0)$, $\delta_1(0,r_0) > 0$ such that, if $|s| < \delta_0(0,r_0)$ and $\|r - r_0\| < \delta_1(0,r_0)$, then $\|\mathcal{G}(s,r) - \mathcal{G}(0,r_0)\| < \overline{\delta}_2/2$. Hence

$$
\left\{ B(r_0,\delta_1(0,r_0)) \,|\, r_0 \in \partial\overline{B(0,\overline{\delta}_2)} \right\}
$$

is an open covering of the boundary $\overline{B(0,\overline{\delta}_2)}$; i.e., $\partial\overline{B(0,\overline{\delta}_2)}$. As this is a compact set, there exists a finite subcovering,

$$
\{ B(r_1,\delta_1(0,r_1)), \dots, B(r_l,\delta(0,r_l)) \}.
$$

Take δ as the minimum of $\{\delta_0(0,r_1), \dots, \delta_0(0,r_l)\}$. Let us see that, for every $(s,r) \in [0,\delta] \times \partial\overline{B(0,\overline{\delta}_2)}$, $\|G_s(r) - r\| < \|r\|$. As r is in an open set of the finite subcovering,

$$
\|\mathcal{G}(s,r)\| = \|G_s(r) - r\| < \frac{\overline{\delta}_2}{2} < \overline{\delta}_2 = \|r\|.
$$

Hence, using Corollary A.3.2, for every $s \in (0, \delta)$ the set $G_s(\overline{B(0, \overline{\delta}_2)})$ covers the origin; that is, there exists $r \in \overline{B(0, \overline{\delta}_2)}$ such that

$$G_s(r) = (\pi_2 \circ g \circ \Delta)(s, r) = 0.$$

Then, because of the definition of the mapping G_s in Equation (4.1.3) and Equations (4.1.4) and (4.1.5), there exists $s' \in \mathbb{R}^+$ such that

$$\gamma[\pi_{w_0}^s](t) = \gamma(t) + s'v.$$

In other words, we have a trajectory coming from a perturbation of the control that meets the ray generated by v, as wanted. ∎

4.1.4 Pontryagin's Maximum Principle in the symplectic formalism for the optimal control problem

In this section, the OCP is transformed into a Hamiltonian problem that will allow us to state Pontryagin's Maximum Principle.

Given the OCP $(M, U, X, \mathcal{F}, I, x_a, x_b)$ and the $\widehat{\text{OCP}}$ $(\widehat{M}, U, \widehat{X}, I, x_a, x_b)$, let us consider the cotangent bundle $T^*\widehat{M}$ with its natural symplectic structure that will be denoted by Ω. If $(\widehat{x}, \widehat{p}) = (x^0, x, p_0, p) = (x^0, x^1, \ldots, x^m, p_0, p_1, \ldots, p_m)$ are local natural coordinates on $T^*\widehat{M}$, the form Ω has as its local expression $\Omega = \mathrm{d}x^0 \wedge \mathrm{d}p_0 + \mathrm{d}x^i \wedge \mathrm{d}p_i$.

For each $u \in U$, $H^u \colon T^*\widehat{M} \to \mathbb{R}$ is the Hamiltonian function defined by

$$H^u(\widehat{p}) = H(\widehat{p}, u) = \langle \widehat{p}, \widehat{X}(\widehat{x}, u) \rangle = p_0 \mathcal{F}(x, u) + \sum_{i=1}^{m} p_i f^i(x, u),$$

where $\widehat{p} \in T_{\widehat{x}}^* \widehat{M}$. The tuple $(T^*\widehat{M}, \Omega, H^u)$ is a Hamiltonian system. Using the notation in Equation (2.2.4), the associated Hamiltonian vector field $Y^{\{u\}}$ satisfies the equation

$$i_{Y^{\{u\}}} \Omega = \mathrm{d}H^u.$$

Thus we get a family of Hamiltonian systems parameterized by u, and given by $H \colon T^*\widehat{M} \times U \to \mathbb{R}$. The associated Hamiltonian vector field $Y \colon T^*\widehat{M} \times U \to T(T^*\widehat{M})$ is a vector field along the projection $\widehat{\pi}_1 \colon T^*\widehat{M} \times U \to T^*\widehat{M}$. Its local expression is

$$\begin{aligned} Y(\widehat{p}, u) &= \left(\mathcal{F}(x, u) \frac{\partial}{\partial x^0} + f^i(x, u) \frac{\partial}{\partial x^i} + 0 \frac{\partial}{\partial p_0} \right. \\ &+ \left. \left(-p_0 \frac{\partial \mathcal{F}}{\partial x^i}(x, u) - p_j \frac{\partial f^j}{\partial x^i}(x, u) \right) \frac{\partial}{\partial p_i} \right)_{(\widehat{x}, \widehat{p}, u)}. \end{aligned}$$

It should be noted that $Y = \widehat{X}^{T^*}$; that is, Y is the cotangent lift of $\widehat{X}$. See §2.2.2.2 for definition and properties of the cotangent lift.

Consider a curve $(\widehat{\lambda}, u)\colon I = [a, b] \to T^*\widehat{M} \times U$ such that it is absolutely continuous on $T^*\widehat{M}$, it is measurable and bounded on U, and $\widehat{\gamma} = \pi_{\widehat{M}} \circ \widehat{\lambda}$ with the natural projection $\pi_{\widehat{M}}\colon T^*\widehat{M} \to \widehat{M}$. The previous elements come together in the following diagram:

$$
\begin{array}{ccc}
 & & T(T^*\widehat{M}) \\
 & \overset{\widehat{X}^{T^*}}{\nearrow} & \downarrow \tau_{T^*\widehat{M}} \\
\mathbb{R} \xleftarrow{H} T^*\widehat{M} \times U \xrightarrow{\widehat{\pi}_1} & & T^*\widehat{M} \\
\uparrow (\widehat{\lambda}, u) \quad \overset{\widehat{\lambda}}{\nearrow} & & \downarrow \pi_{\widehat{M}} \\
I \xrightarrow{\widehat{\gamma}} & & \widehat{M} \\
 & \overset{\gamma}{\searrow} & \downarrow \pi_2 \\
 & & M
\end{array}
$$

Statement 4.1.13. *(**Hamiltonian Problem, HP**) Given the OCP $(M, U, X, \mathcal{F}, I, x_a, x_b)$, and the equivalent $\widehat{\mathrm{OCP}}$ $(\widehat{M}, U, \widehat{X}, I, x_a, x_b)$, consider the following problem.*

Find $(\widehat{\lambda}, u)$ such that

(1) $\widehat{\gamma}(a) = (0, x_a)$ and $\gamma(b) = x_b$, if $\widehat{\gamma} = \pi_{\widehat{M}} \circ \widehat{\lambda}$, $\gamma = \pi_2 \circ \widehat{\gamma}$;

(2) $\dot{\widehat{\lambda}}(t) = \widehat{X}^{T^*}(\widehat{\lambda}(t), u(t))$, for a.e. $t \in I$.

The tuple $(T^*\widehat{M}, U, \widehat{X}^{T^*}, I, x_a, x_b)$ denotes the *Hamiltonian problem* as it has just been defined and the elements satisfy the same properties as in §2.2.1.

Comments: The Hamiltonian problem satisfies analogous conditions to those satisfied by the OCP and the $\widehat{\mathrm{OCP}}$ defined in §4.1.1 and §4.1.2 respectively.

1. Given $(\widehat{\lambda}, u)$, the function $u\colon I \to U$ allows us to construct a time–dependent vector field on $T^*\widehat{M}$, $\left(\widehat{X}^{T^*}\right)^{\{u\}}\colon I \times T^*\widehat{M} \to T(T^*\widehat{M})$, defined by

$$
\left(\widehat{X}^{T^*}\right)^{\{u\}}(t, \widehat{x}, \widehat{p}) = \widehat{X}^{T^*}(\widehat{x}, \widehat{p}, u(t)).
$$

   Condition (2) shows that $\widehat{\lambda}$ is an integral curve of $\left(\widehat{X}^{T^*}\right)^{\{u\}}$.

2. Condition (2) is equivalent to the commutativity of Diagram (2.2.5) replacing M, $X^{\{u\}}$, γ by $T^*\widehat{M}$, $\left(\widehat{X}^{T^*}\right)^{\{u\}}$, $\widehat{\lambda}$, respectively.

3. The vector field $\left(\widehat{X}^{T^*}\right)^{\{u\}}$ is $\pi_{\widehat{M}}$–projectable and projects onto $\widehat{X}^{\{u\}}$. Thus if $\widehat{\lambda}$ is an integral curve of $\left(\widehat{X}^{T^*}\right)^{\{u\}}$, $\widehat{\gamma} = \pi_{\widehat{M}} \circ \widehat{\lambda}$ is an integral curve of $\widehat{X}^{\{u\}}$.

4. Locally, conditions (1) and (2) in Statement 4.1.13 are equivalent to the fact that the curve $(\widehat{\lambda}, u)$ satisfies the Hamilton equations of the system $(T^*\widehat{M}, \Omega, H^u)$,

$$
\begin{aligned}
\dot{x}^0 &= \frac{\partial H^u}{\partial p_0} = \mathcal{F}, \\
\dot{x}^i &= \frac{\partial H^u}{\partial p_i} = f^i, \\
\dot{p}_0 &= -\frac{\partial H^u}{\partial x^0} = 0 \Rightarrow p_0 = \text{ct}, \qquad (4.1.6) \\
\dot{p}_i &= -\frac{\partial H^u}{\partial x^i} = -p_0 \frac{\partial \mathcal{F}}{\partial x^i} - p_j \frac{\partial f^j}{\partial x^i}, \qquad (4.1.7)
\end{aligned}
$$

and satisfies the conditions $\widehat{\gamma}(a) = (0, x_a)$, $\gamma(b) = x_b$.

In the literature of optimal control, the system of differential equations given by Equations (4.1.6), (4.1.7) is called the *adjoint system*. In differential geometry, the adjoint system is the differential equations satisfied by the fiber coordinates of an integral curve of the cotangent lift of a vector field on M. See §2.2.2.2 for more details.

Note that there is no initial condition for $\widehat{p} = (p_0, p_1, \ldots, p_m)$, hence HP is not a Cauchy problem.

Comment: So far we have considered a fixed control $u \in U$. Therefore we have been working with a family of Hamiltonian systems on the manifold $\left(T^*\widehat{M}, \Omega\right)$ given by the Hamiltonians $\{H^u \mid u \in U\}$.

Given $u \colon I \to U$, we consider the Hamiltonian $H^{u(t)}$. The equation of the Hamiltonian vector field for the Hamiltonian system $\left(T^*\widehat{M}, \Omega, H^{u(t)}\right)$ is

$$
i_{Y^{\{u(t)\}}} \Omega = \mathrm{d}_{\widehat{M}} H^{u(t)},
$$

where $\mathrm{d}_{\widehat{M}}$ is the exterior differential on the manifold $T^*\widehat{M}$. Observe that we have studied the system defined by $(T^*\widehat{M}, \Omega, H^{u(t)})$ as an autonomous system by fixing the time t. The corresponding Hamiltonian vector field $Y^{\{u(t)\}}$ is a time–dependent vector field whose integral curves satisfy the equation

$$
\dot{\widehat{\lambda}}(t) = Y^{\{u(t)\}}(\widehat{\lambda}(t)), \quad t \in I. \qquad (4.1.8)
$$

Observe that $Y^{\{u(t)\}} = (\widehat{X}^{T^*})^{\{u(t)\}}$.

Now we are ready to state Pontryagin's Maximum Principle that provides the necessary conditions, which are in general not sufficient, to find solutions of the optimal control problem.

Theorem 4.1.14. *(Pontryagin's Maximum Principle, PMP) If $(\widehat{\gamma}^*, u^*) \colon I = [a, b] \to \widehat{M} \times U$ is a solution of the extended optimal control problem, Statement 4.1.2, such that $\widehat{\gamma}^*$ is absolutely continuous and u^* is measurable and bounded, then there exists $(\widehat{\lambda}^*, u^*) \colon I \to T^*\widehat{M} \times U$ along γ^* such that:*

1. *it is a solution of the Hamiltonian problem; that is, it satisfies Equation (4.1.8) and the initial conditions $\widehat{\gamma}^*(a) = (0, x_a)$ and $\gamma^*(b) = x_b$, if $\gamma^* = \pi_2 \circ \widehat{\gamma}^*$;*

2. $\widehat{\gamma}^* = \pi_{\widehat{M}} \circ \widehat{\lambda}^*$;

3.   (a) $H\left(\widehat{\lambda}^*(t), u^*(t)\right) = \sup_{u \in U} H(\widehat{\lambda}^*(t), u)$ *almost everywhere;*

    (b) $\sup_{u \in U} H(\widehat{\lambda}^*(t), u)$ *is constant everywhere;*

    (c) $\widehat{\lambda}^*(t) \neq 0 \in T^*_{\widehat{\gamma}^*(t)} \widehat{M}$ *for each* $t \in [a, b]$;

    (d) $\lambda_0^*(t)$ *is constant and* $\lambda_0^*(t) \leq 0$.

Comments:

1. There exists an abuse of notation between $\widehat{\lambda}(t) \in T^* \widehat{M}$ and $\widehat{\lambda}(t) \in T^*_{\widehat{\gamma}(t)} \widehat{M}$. We assume that the meaning of $\widehat{\lambda}$ in each situation will be clear from the context.

2. Condition (2) is immediately satisfied because $\widehat{\lambda}^*$ is a covector along $\widehat{\gamma}^*$.

3. Conditions (3a) and (3b) imply that the Hamiltonian function is constant along the optimal curve for almost every $t \in [a, b]$.

4. In item (3a), if U is a closed set, then the maximum of the Hamiltonian over the controls is considered instead of the the supremum over the controls. But in condition (3b) we can always consider the maximum, instead of the supremum, because item (3a) guarantees that the supremum of the Hamiltonian is reached in the optimal curve. Thus, the supremum is, in fact, a maximum. In general we will refer to (3a) as the condition of the maximization of the Hamiltonian over the controls.

5. Condition (3c) implies that $\lambda_0^*(t) \neq 0 \in T^*_{\gamma^{*0}(t)} \mathbb{R}$ or $\lambda^*(t) \neq 0 \in T^*_{\gamma^*(t)} M$ for each $t \in [a, b]$. Locally the condition (3c) states that for each $t \in [a, b]$ there exists a coordinate of $\widehat{\lambda}^*(t)$ nonzero, $(p_i \circ \widehat{\lambda}^*)(t) = \lambda_i^*(t) \neq 0$.

6. From the Hamilton's equations of the system $\left(T^* \widehat{M}, \Omega, H^{u(t)}\right)$, it is concluded that λ_0 is constant along the integral curves of $(\widehat{X}^{T^*})^{\{u(t)\}}$, since $\dot{p}_0 = 0$—see Equation (4.1.6). Hence the first result in (3d) is immediate for every integral curve of $(\widehat{X}^{T^*})^{\{u(t)\}}$. As λ_0^* is constant, $\widehat{\lambda}^*$ may be normalized without loss of generality. Thus it is assumed that either $\lambda_0^* = 0$ or $\lambda_0^* = -1$ because of the second result in (3d).

7. Pontryagin's Maximum Principle only guarantees that, given a solution of $\widehat{OCP}$, there exists a solution of HP. Hence, in principle, both problems are not equivalent.

    Observe that the Maximum Principle guarantees the existence of a covector along the optimal curve, but it does not say anything about the uniqueness of the covector. Indeed, this covector may not be unique as Figure 4.2 shows.

    Depending on the covector we associate with the optimal curves, different curves can be defined.

Definition 4.1.15. *A curve* $(\widehat{\gamma}, u) \colon [a, b] \to \widehat{M} \times U$ *for* $\widehat{OCP}$ *is:*
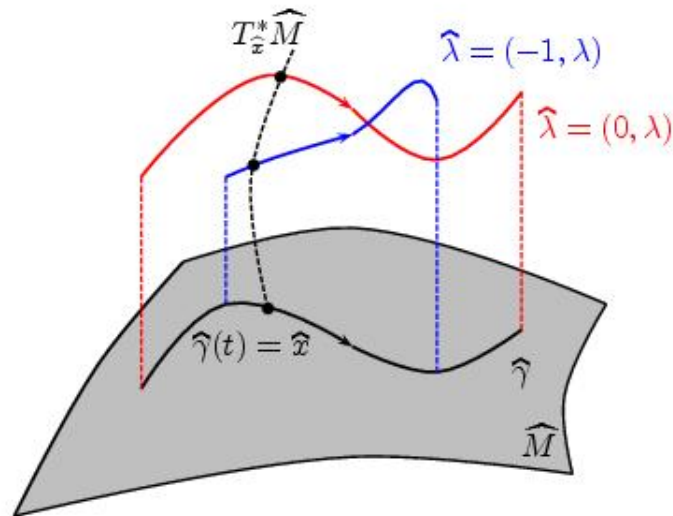
Figure 4.2: *An optimal curve with two different lifts to $T^*\widehat{M}$ so that it is abnormal and normal at the same time.*

1. an **extremal** if there exist $\widehat{\lambda}\colon [a,b] \to T^*\widehat{M}$ such that $\widehat{\gamma} = \pi_{T\widehat{M}} \circ \widehat{\lambda}$ and $(\widehat{\lambda}, u)$ satisfies the necessary conditions of PMP;

2. a **normal extremal** if it is an extremal with $\lambda_0 = -1$, then $\widehat{\lambda}$ is called a **normal lift or momenta**;

3. an **abnormal extremal** if it is an extremal with $\lambda_0 = 0$, then $\widehat{\lambda}$ is called an **abnormal lift or momenta**;

4. a **strictly abnormal extremal** if it is not a normal extremal, but it is abnormal;

5. a **strictly normal extremal** if it is not an abnormal extremal, but it is normal;

6. a **singular extremal** if it is an extremal and

$$H\big(\widehat{\lambda}(t), \widetilde{u}(t)\big) = \sup_{w \in U} H\big(\widehat{\lambda}(t), w\big)$$

almost everywhere for any $\widetilde{u}\colon I \to U$.

The curve $(\widehat{\lambda}, u)\colon [a,b] \to T^*\widehat{M} \times U$ is called **biextremal for OCP**.

In [Agrachev and Sachkov 2004, Anisi 2003, Kirschner et al. 1997, Loewen 2004, Sussmann and Tang 1991, Troutman 1996] there are some examples of optimal control problems whose solutions are searched using Pontryagin's Maximum Principle.

As proved in Proposition 6.5.2 and Remark 6.5.3, the abnormal and singular extremals are connected for some particular control systems. Usually in these systems the controls given abnormal and/or singular extremals cannot be determined using PMP in Theorem 4.1.14 and

the high–order Maximum Principle [Bianchini 1998, Kawski 2003, Knobloch 1981, Krener 1977] is necessary.

Observe that if $\widehat{\gamma}\colon I \to \widehat{M}$ is an integral curve of a vector field, there always exists a lift of $\widehat{\gamma}$ to a curve $\widehat{\lambda}\colon I \to T^*\widehat{M}$, given an initial condition for the cofibers. The curve $\widehat{\gamma}$ is an integral curve of the cotangent lift of the given vector field on $\widehat{M}$. Analogously, if the system is given by a vector field along the projection $\widehat{\pi}\colon \widehat{M} \times U \to \widehat{M}$.

Therefore, the items 1 and 2 in Theorem 4.1.14 do not provide any information related with the optimality. They only ask for the fulfilment of a final condition in the integral curve. The accomplishment of this depends on the accessibility of the problem, see Definition 3.2.2 and [Bullo and Lewis 2005a, Nijmeijer and van der Schaft 1990].

The real contribution of PMP is the third item related with the optimality through the maximization of the Hamiltonian, that will be only satisfied if the initial conditions for the fibers in $T^*\widehat{M}$ are chosen suitably. This is the key element of the proof of Pontryagin's Maximum Principle. In other words, we can always find a cotangent lift of an integral curve such that conditions 1 and 2 are satisfied under the assumption of accessibility, but it is not guaranteed the fulfilment of the conditions in assertion 3 in Theorem 4.1.14. That is why the initial condition for the fibers in $T^*\widehat{M}$ must be chosen conveniently as shows the proof of Theorem 4.1.14, see §4.2.

If we write the Hamiltonian function for the abnormal and the normal case, the difference is that the cost function does not play any role in the Hamiltonian for abnormal extremals. That is why it is said the abnormality only depends on the geometry of the control system. But to determine the optimality of the abnormal extremals, the cost function is essential. In fact, for the same control system different optimal control problems can be stated depending for each cost function and it might happen that the abnormal extremals are minimizers only for some of them problems.

As for strict abnormality, it must be proved that the abnormal extremals are not normal. In other words, there does not exist any covector along the given extremal that satisfies Hamilton's equations for normality. The cost function is also essential for proving that an extremal only admits an abnormal lift.

4.2   Proof of Pontryagin's Maximum Principle for fixed time and fixed endpoints

To prove Pontryagin's Maximum Principle it is necessary to use analytic results about absolute continuity and lower semicontinuity for real functions, and properties of convex cones. For the details see Appendices A and B and references therein. The reader is referred to §4.1.3 for results on perturbations of a reference trajectory in a control system.

In the literature on optimal control, the proof of the Maximum Principle has been discussed taking into account varying hypotheses [Agrachev and Sachkov 2004, Athans and Falb 1966,

Bonnard and Caillau 2006, Bressan and Piccoli 2007, Jurdjevic 1997, Sussmann 1998; 2000; 2005]. Most authors believe and justify that the origin of this Principle is the calculus of variations; see [Zeidler 1985] for instance.

(*Proof*) (*Theorem 4.1.14: Pontryagin's Maximum Principle, PMP*)
1. As $(\widehat{\gamma}^*, u^*)$ is a solution of $\widehat{\text{OCP}}$, if τ is in $[a, b]$, for every initial condition $\widehat{\lambda}_\tau$ in $T^*_{\widehat{\gamma}^*(\tau)}\widehat{M}$, we have a solution of HP, $\widehat{\lambda} \colon [a, b] \to T^*\widehat{M}$ along $\widehat{\gamma}^*$, satisfying that initial condition. The covector $\widehat{\lambda}_\tau$ must be chosen so that the remaining conditions of the PMP are satisfied.

According to §4.1.3, we construct the tangent perturbation cone $\widehat{K}_b$ in $T_{\widehat{\gamma}^*(b)}\widehat{M}$ that contains all tangent vectors associated with perturbations of the trajectory $\widehat{\gamma}^*$ corresponding to variations of u^*; see Definition 4.1.11.

Let us consider the vector $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)} \in T_{\widehat{\gamma}^*(b)}\widehat{M}$—the zero in bold emphasises that $\mathbf{0}$ is a vector in $T_{\gamma^*(b)}M$—which has the following properties:

1. the variation of $x^0(t) = \int_a^t F(\gamma^*(s), u^*(s))\mathrm{d}s$ along $(-1, \mathbf{0})$ is negative;

2. it is not interior to $\widehat{K}_b$.

   Let us prove the second assertion. Take a local chart at $\widehat{\gamma}^*(b)$ and work on the image of the local chart, in $\mathbb{R}^{m+1}$, without changing the notation.
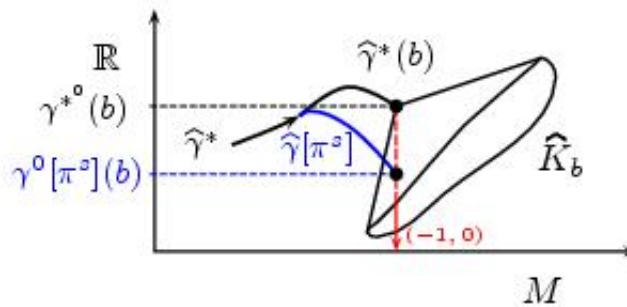
Figure 4.3: *Situation if $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)}$ is interior to $\widehat{K}_b$.*

If $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)}$ was interior to $\widehat{K}_b$, by Proposition 4.1.12 there would exist a positive number ϵ such that, for every $s \in (0, \epsilon)$, there would exist a positive number s', close to s, and a perturbation of the control $u[\pi^s]$ such that

$$\widehat{\gamma}[\pi^s](b) = (\gamma^0[\pi^s](b), \gamma[\pi^s](b)) = \widehat{\gamma}^*(b) + s'(-1, \mathbf{0}).$$

For this perturbed trajectory we have

$$\gamma^0[\pi^s](b) < \gamma^{*0}(b) \quad \text{and} \quad \gamma[\pi^s](b) = \gamma^*(b).$$

Hence there would be a trajectory, $\widehat{\gamma}[\pi^s]$, from $\gamma^*(a)$ to $\gamma^*(b)$ with less cost than $\widehat{\gamma}^*$, as pictured in Figure 4.3. Hence $\widehat{\gamma}^*$ would not be optimal. In other words, $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)}$

is the direction of decreasing of the functional to be minimized in the extended optimal control problem.

The second property implies that $\widehat{K}_b$ cannot be equal to $T_{\widehat{\gamma}^*(b)}\widehat{M}$. As $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)}$ is not interior to $\widehat{K}_b$, there exists a separating hyperplane of $\widehat{K}_b$ and $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)}$ by Proposition B.2.9; that is, there exists a nonzero covector determining a separating hyperplane. Let $\widehat{\lambda}_b \in T^*_{\widehat{\gamma}^*(b)}\widehat{M}$ be nonzero such that $\ker \widehat{\lambda}_b$ is a separating hyperplane satisfying

$$\begin{aligned} \langle \widehat{\lambda}_b, (-1, \mathbf{0}) \rangle &\geq 0, \\ \langle \widehat{\lambda}_b, \widehat{v}_b \rangle &\leq 0 \quad \forall \widehat{v}_b \in \widehat{K}_b. \end{aligned} \tag{4.2.9}$$

Observe that if $\widehat{\lambda}_b = 0 \in T^*_{\widehat{\gamma}^*(b)}\widehat{M}$, $\ker \widehat{\lambda}_b$ does not determine a hyperplane, but the whole space $T^*_{\widehat{\gamma}^*(b)}\widehat{M}$.

Given the initial condition $\widehat{\lambda}_b \in T^*_{\widehat{\gamma}^*(b)}\widehat{M}$, there exists only one integral curve $\widehat{\lambda}^*$ of $(\widehat{X}^{T^*})^{\{u^*\}}$ such that $\widehat{\lambda}^*(b) = \widehat{\lambda}_b$. Hence $(\widehat{\lambda}^*, u^*)$ is a solution of HP.

2. Obviously, by construction, $\widehat{\gamma}^* = \pi_{\widehat{M}} \circ \widehat{\lambda}^*$.

Now we prove that $\widehat{\lambda}^*$, the solution of HP, satisfies the remaining conditions of the PMP.

(3a) $H\left(\widehat{\lambda}^*(t), u^*(t)\right) = \sup_{u \in U} H(\widehat{\lambda}^*(t), u)$ almost everywhere.

We are going to prove the statement for every Lebesgue time, hence it will be true almost everywhere because of Definition A.2.4 and Remarks A.2.5, A.2.6. Suppose that there exists a control $\widetilde{u} \colon I \to U$ and a Lebesgue time t_1 such that u^* does not give the supremum of the Hamiltonian at t_1; that is,

$$H(\widehat{\lambda}^*(t_1), \widetilde{u}(t_1)) > H(\widehat{\lambda}^*(t_1), u^*(t_1)).$$

As $H(\widehat{p}, u) = \langle \widehat{p}, \widehat{X}(\widehat{x}, u) \rangle$,

$$\langle \widehat{\lambda}^*(t_1), \widehat{X}(\widehat{\gamma}^*(t_1), \widetilde{u}(t_1)) - \widehat{X}(\widehat{\gamma}^*(t_1), u^*(t_1)) \rangle > 0;$$

that is, $\langle \widehat{\lambda}^*(t_1), \widehat{v}[\pi_1] \rangle > 0$ where

$$\widehat{v}[\pi_1] = \widehat{X}(\widehat{\gamma}^*(t_1), \widetilde{u}(t_1)) - \widehat{X}(\widehat{\gamma}^*(t_1), u^*(t_1)) \in \widehat{K}_{t_1} \subset T_{\widehat{\gamma}^*(t_1)}\widehat{M}$$

is the elementary perturbation vector associated with the perturbation data $\pi_1 = \{t_1, 1, \widetilde{u}(t_1)\}$ by Proposition 4.1.4.

Let $\widehat{V}[\pi_1] \colon [t_1, b] \to T\widehat{M}$ be the integral curve of $(\widehat{X}^T)^{\{u^*\}}$ with $(t_1, \widehat{v}[\pi_1])$ as initial condition. For $\widehat{\lambda}^*$, solution of HP, the continuous function $\langle \widehat{\lambda}^*, \widehat{V}[\pi_1] \rangle \colon [t_1, b] \to \mathbb{R}$ is constant everywhere by Proposition 2.2.6. Hence $\langle \widehat{\lambda}^*(t_1), \widehat{v}[\pi_1] \rangle > 0$ implies that $\langle \widehat{\lambda}_b, \widehat{V}[\pi_1](b) \rangle > 0$, which is a contradiction with $\langle \widehat{\lambda}_b, \widehat{v}_b \rangle \leq 0$ for every $\widehat{v}_b \in \widehat{K}_b$ in (4.2.9), since $\widehat{V}[\pi_1](b) \in \widehat{K}_b$.

Therefore,

$$H(\widehat{\lambda}^*(t), u^*(t)) = \sup_{u \in U} H(\widehat{\lambda}^*(t), u)$$

at every Lebesgue time on $[a, b]$, so almost everywhere.

(3b) $\sup_{u \in U} H(\widehat{\lambda}^*(t), u)$ is constant everywhere.

In fact, because of (3a) we know that the supremum is achieved along the optimal curve. Thus (3b) is equivalent to prove that $\max_{u \in U} H(\widehat{\lambda}^*(t), u)$ is constant everywhere. To simplify the notation we define the function

$$\begin{aligned} \mathcal{M} \circ \widehat{\lambda}^* \colon I &\longrightarrow \mathbb{R} \\ t &\longmapsto \mathcal{M}(\widehat{\lambda}^*(t)) = \max_{u \in U} H(\widehat{\lambda}^*(t), u). \end{aligned}$$

In order to prove (3b), it is enough to see that $\mathcal{M}(\widehat{\lambda}^*(t))$ is constant everywhere.

First let us see that $\mathcal{M} \circ \widehat{\lambda}^*$ is lower semicontinuous on I. See Appendix A.1 for details of this property. As $\mathcal{M}(\widehat{\lambda}^*(t))$ is the maximum of the Hamiltonian function with respect to control, for every $\epsilon > 0$, there exists a control $u_{\mathcal{M}} \colon I \to U$ such that

$$H(\widehat{\lambda}^*(t), u_{\mathcal{M}}(t)) \geq \mathcal{M}(\widehat{\lambda}^*(t)) - \frac{\epsilon}{2} \tag{4.2.10}$$

everywhere.

For each constant control $\widetilde{u} \in U$, $H^{\widetilde{u}} \circ \widehat{\lambda}^* = H(\widehat{\lambda}^*, \widetilde{u}) \colon I \to \mathbb{R}$ is continuous on I. Hence for every $t_0 \in I$ and $\epsilon > 0$, there exists $\delta > 0$ such that $\mid t - t_0 \mid < \delta$, we have

$$\mid H^{\widetilde{u}}(\widehat{\lambda}^*(t)) - H^{\widetilde{u}}(\widehat{\lambda}^*(t_0)) \mid < \frac{\epsilon}{2}.$$

If $\widetilde{u} = u_{\mathcal{M}}(t_0)$, then using the continuity of $H^{\widetilde{u}} \circ \widehat{\lambda}^*$

$$\begin{aligned} \mathcal{M}(\widehat{\lambda}^*(t)) &= \max_{u \in U} H(\widehat{\lambda}^*(t), u) \geq H(\widehat{\lambda}^*(t), u_{\mathcal{M}}(t_0)) \\ &\geq H(\widehat{\lambda}^*(t_0), u_{\mathcal{M}}(t_0)) - \frac{\epsilon}{2} \geq \mathcal{M}(\widehat{\lambda}^*(t_0)) - \epsilon. \end{aligned}$$

The last inequality is true by evaluating Equation (4.2.10) at t_0. Hence $\mathcal{M} \circ \widehat{\lambda}^*$ is lower semicontinuous at every $t_0 \in I$; that is, $\mathcal{M} \circ \widehat{\lambda}^*$ is lower semicontinuous on I.

The control u^* is bounded, that means $\operatorname{Im} u^*$ is contained in a compact set $D \subset U$. Let us define a new function

$$\begin{aligned} \mathcal{M}_D \colon T^*\widehat{M} &\longrightarrow \mathbb{R} \\ \beta &\longmapsto \mathcal{M}_D(\beta) = \max_{\widetilde{u} \in D} H(\beta, \widetilde{u}). \end{aligned}$$

As $H(\beta, \cdot) \colon D \to \mathbb{R}, \widetilde{u} \mapsto H(\beta, \widetilde{u})$ is continuous by hypothesis and D is a compact, for every

$\beta \in T^* \widehat{M}$ there exists a control $\widetilde{w}_\beta$ that gives us the maximum of $H(\beta, \widetilde{u})$

$$\mathcal{M}_D(\beta) = \max_{\widetilde{u} \in D} H(\beta, \widetilde{u}) = H(\beta, \widetilde{w}_\beta). \tag{4.2.11}$$

Hence $\mathcal{M}_D$ is well–defined on $T^* \widehat{M}$. The following sketch explains in a compact way the necessary steps to prove that $\mathcal{M} \circ \lambda^*$ is constant everywhere. In this sketch, the circled figures in bold refer to statements which are going to be proved in the next paragraphs and a.c. stands for absolutely continuous and a.e. for almost everywhere.

$$
\begin{array}{l}
H^{\widetilde{u}} \in \mathcal{C}^1(T^* \widehat{M}) \\
\qquad \Downarrow \textcircled{1} \\
H^{\widetilde{u}} \text{ is locally Lipschitz } \forall\, \widetilde{u} \in D \\
\qquad \Downarrow \textcircled{2} \\
\mathcal{M}_D \text{ is locally Lipschitz on Im } \widehat{\lambda}^* \\[4pt]
\widehat{\lambda}^* \text{ is a.c.}
\end{array}
$$

$\mathcal{M} \circ \widehat{\lambda}^*$ is lower semicontinuous on I

$\Rightarrow \textcircled{3}\ \mathcal{M}_D \circ \widehat{\lambda}^*$ is a.c. $\Rightarrow \mathcal{M}_D \circ \widehat{\lambda}^*$ is continuous

$\textcircled{4}\ \mathcal{M}_D(\widehat{\lambda}^*(t)) \le \mathcal{M}(\widehat{\lambda}^*(t)),\ \forall\, t \in [a, b]$
$\textcircled{5}\ \mathcal{M}_D(\widehat{\lambda}^*(t)) = \mathcal{M}(\widehat{\lambda}^*(t))$ a.e.

$\mathcal{M}_D \circ \widehat{\lambda}^*$ is a.c.
$\textcircled{7}\ \mathcal{M}_D \circ \widehat{\lambda}^*$ has zero derivative

$\Rightarrow \textcircled{6}$

$\Rightarrow \textcircled{8}$

$$
\begin{array}{l}
\textcircled{6} \Rightarrow (A.15)\ \mathcal{M}_D(\widehat{\lambda}^*(t)) = \mathcal{M}(\widehat{\lambda}^*(t))\ \forall\, t \in [a,b] \\
\textcircled{8} \Rightarrow (A.16)\ \mathcal{M}_D(\widehat{\lambda}^*(t)) \text{ is constant } \forall\, t \in [a,b]
\end{array}
$$
$\Rightarrow \textcircled{9}\ \ \mathcal{M}(\widehat{\lambda}^*(t))$ is constant $\forall\, t \in [a,b]$.

$\textcircled{1}\ H^{\widetilde{u}} \in \mathcal{C}^1(T^* \widehat{M}) \Rightarrow H^{\widetilde{u}}$ is locally Lipschitz $\forall\, \widetilde{u} \in D$.

The Lipschitzian property applies to functions defined on a metric space. As the property we want to prove is local, we define the distance on a local chart as is explained in Appendix A.1. For every $\beta \in T^* \widehat{M}$, let (V_β, ϕ) be a local chart centered at β such that $\phi(\beta) = 0$ and $\phi(V_\beta) = B$, where B is an open ball centered at $0 \in \mathbb{R}^{2m+2}$. If β_1 and β_2 are in V_β, define $d_\phi(\beta_1, \beta_2) = d(\phi(\beta_1), \phi(\beta_2))$ where d is the Euclidean distance in $\mathbb{R}^{2m+2}$.

For every β in $T^* \widehat{M}$, we get an open neighbourhood V_β using the local chart (V_β, ϕ). As $H^{\widetilde{u}}$ is $\mathcal{C}^1(T^* \widehat{M})$ and $\widetilde{u}$ lies in the compact set D, by the Mean Value Theorem for every β in $T^* \widehat{M}$ there exists an open neighbourhood V_β such that $|H^{\widetilde{u}}(\beta_1) - H^{\widetilde{u}}(\beta_2)| < K_\beta d_\phi(\beta_1, \beta_2)$ where K_β does not depend on the control $\widetilde{u}$. Thus $H^{\widetilde{u}}$ is locally Lipschitz on $T^* \widehat{M}$. Moreover, the Lipschitz constant and the open neighbourhood V_β can be chosen so as to not depend on the control since $\widetilde{u}$ is in a compact set.

$\textcircled{2}\ H^{\widetilde{u}}$ is locally Lipschitz $\forall\, \widetilde{u} \in D \Rightarrow \mathcal{M}_D$ is locally Lipschitz on Im $\widehat{\lambda}^*$.

Let β be in Im $\widehat{\lambda}^*$, there exists an open convex neighbourhood V_β such that

$$|H^{\widetilde{u}}(\beta_1) - H^{\widetilde{u}}(\beta_2)| < K_\beta d(\beta_1, \beta_2)$$

for every $\widetilde{u}$ in D and β_1, β_2 in V_β. If $\widetilde{w}_1$, $\widetilde{w}_2$ are the controls in D maximizing $H(\beta_1, \widetilde{u})$ and

$H(\beta_2, \widetilde{u})$, respectively, then

$$H(\beta_1, \widetilde{w}_2) \leq H(\beta_1, \widetilde{w}_1),$$

$$H(\beta_2, \widetilde{w}_1) \leq H(\beta_2, \widetilde{w}_2).$$

Moreover, $H^{\widetilde{w}_1}$ and $H^{\widetilde{w}_2}$ are Lipschitz on V_β since the Lipschitz constant and the neighbourhood is independent of the control. Then using the last inequalities

$$
\begin{aligned}
-K_\beta d(\beta_1, \beta_2) &\leq H^{\widetilde{w}_2}(\beta_1) - H^{\widetilde{w}_2}(\beta_2) \leq H^{\widetilde{w}_1}(\beta_1) - H^{\widetilde{w}_2}(\beta_2) \\
&\leq H^{\widetilde{w}_1}(\beta_1) - H^{\widetilde{w}_1}(\beta_2) \leq K_\beta d(\beta_1, \beta_2).
\end{aligned}
$$

Observe that by Equation (4.2.11), $H^{\widetilde{w}_1}(\beta_1) - H^{\widetilde{w}_2}(\beta_2) = \mathcal{M}_D(\beta_1) - \mathcal{M}_D(\beta_2)$. Hence

$$|\mathcal{M}_D(\beta_1) - \mathcal{M}_D(\beta_2)| \leq K_\beta d(\beta_1, \beta_2), \quad \forall\, \beta_1, \beta_2 \in V_\beta, \tag{4.2.12}$$

that is, $\mathcal{M}_D$ is locally Lipschitz on $\operatorname{Im}\widehat{\lambda}^*$. As $\widehat{\lambda}^*$ is absolutely continuous, $\operatorname{Im}\widehat{\lambda}^*$ is compact. Thus we may choose a Lipschitz constant independent of the point β. Hence

$$|\mathcal{M}_D(\beta_1) - \mathcal{M}_D(\beta_2)| \leq K d(\beta_1, \beta_2), \quad \forall\, \beta_1, \beta_2 \in V_\beta.$$

③ $\mathcal{M}_D$ is locally Lipschitz on $\operatorname{Im}\widehat{\lambda}^*$ and $\widehat{\lambda}^*$ is absolutely continuous $\Rightarrow \mathcal{M}_D \circ \widehat{\lambda}^* \colon I \to \mathbb{R}$ is absolutely continuous $\Rightarrow \mathcal{M}_D \circ \widehat{\lambda}^* \colon I \to \mathbb{R}$ is continuous.

For every $t \in I$, let us consider the neighbourhood $V_{\widehat{\lambda}^*(t)}$ where Equation (4.2.12) is satisfied. As $\operatorname{Im}\widehat{\lambda}^*$ is a compact set,

- there exists a finite open subcovering $V_{\widehat{\lambda}^*(t_1)}, \ldots, V_{\widehat{\lambda}^*(t_r)}$ of $\{V_{\widehat{\lambda}^*(t)} \,;\, t \in I\}$, and

- there exists a Lebesgue number l of the subcovering, that is, for every two points in an open ball of diameter l there exists an open set of the finite subcovering containing both points.

For the Lebesgue number l, by the uniform continuity of $\widehat{\lambda}^*$, there exists a $\delta_l > 0$ such that for each t_1, t_2 in I with $|t_2 - t_1| < \delta_l$, then $d(\widehat{\lambda}^*(t_2), \widehat{\lambda}^*(t_1)) < l$. Thus there exists an open set of the finite subcovering containing $\widehat{\lambda}^*(t_1)$ and $\widehat{\lambda}^*(t_2)$.

On the other hand, taken $\epsilon > 0$ the absolutely continuity of $\widehat{\lambda}^*$ determines a $\delta_\epsilon > 0$.

To prove the absolute continuity of $\mathcal{M}_D \circ \widehat{\lambda}^*$, take $\delta = \min\{\delta_l, \delta_\epsilon\}$. Then, for every finite number of nonoverlapping subintervals (t_{i_1}, t_{i_2}) of I and with $\sum_{i=1}^{s} |t_{i_2} - t_{i_1}| < \delta$,

$$\sum_{i=1}^{s} \left| \mathcal{M}_D(\widehat{\lambda}^*(t_{i_2})) - \mathcal{M}_D(\widehat{\lambda}^*(t_{i_1})) \right| \leq \sum_{i=1}^{s} K d(\widehat{\lambda}^*(t_{i_2}), \widehat{\lambda}^*(t_{i_1})) \leq K\epsilon.$$

In the first step we use that $\delta < \delta_l$ to guarantee that $\widehat{\lambda}^*(t_{i_2})$ and $\widehat{\lambda}^*(t_{i_1})$ are contained in the same open set of the finite subcovering of $\operatorname{Im}\widehat{\lambda}^*$. That allows us to use the property of being locally Lipschitzian. Secondly, we use that $\delta < \delta_\epsilon$ to apply the absolute continuity of $\widehat{\lambda}^*$.

As $\mathcal{M}_D \circ \widehat{\lambda}^*$ is absolutely continuous on I, $\mathcal{M}_D \circ \widehat{\lambda}^*$ is continuous on I.

④ $\mathcal{M}_D(\widehat{\lambda}^*(t)) \leq \mathcal{M}(\widehat{\lambda}^*(t))$ everywhere.

Observe that

$$\mathcal{M}_D(\widehat{\lambda}^*(t)) = \max_{u \in D} H(\widehat{\lambda}^*(t), u) \leq \max_{u \in U} H(\widehat{\lambda}^*(t), u) = \mathcal{M}(\widehat{\lambda}^*(t)),$$

for each $t \in I$.

⑤ $\mathcal{M}(\widehat{\lambda}^*(t)) = \mathcal{M}_D(\widehat{\lambda}^*(t))$ almost everywhere.

For each $t \in I$ there exists a control $w(t)$ maximizing $H(\widehat{\lambda}^*(t), u)$ over the controls in D because of condition (3a),

$$\mathcal{M}_D(\widehat{\lambda}^*(t)) = \max_{u \in D} H(\widehat{\lambda}^*(t), u) = H(\widehat{\lambda}^*(t), w(t)).$$

As $u^*(t) \in D$ for each $t \in I$,

$$\max_{u \in D} H(\widehat{\lambda}^*(t), u) = \max_{u \in U} H(\widehat{\lambda}^*(t), u) = \mathcal{M}(\widehat{\lambda}^*(t)) = H(\widehat{\lambda}^*(t), u^*(t))$$

almost everywhere by (3a). Thus $\mathcal{M}(\widehat{\lambda}^*(t)) = \mathcal{M}_D(\widehat{\lambda}^*(t))$ a.e.

⑥ Applying Proposition A.1.7, we have $\mathcal{M}_D(\widehat{\lambda}^*(t)) = \mathcal{M}(\widehat{\lambda}^*(t))$ everywhere on I, because $\mathcal{M}_D \circ \widehat{\lambda}^*$ is continuous on I, $\mathcal{M} \circ \widehat{\lambda}^*$ is lower semicontinuous, $\mathcal{M}_D(\widehat{\lambda}^*(t)) \leq \mathcal{M}(\widehat{\lambda}^*(t))$ everywhere and $\mathcal{M}_D(\widehat{\lambda}^*(t)) = \mathcal{M}(\widehat{\lambda}^*(t))$ almost everywhere.

⑦ $\mathcal{M}_D \circ \widehat{\lambda}^*$ has zero derivative.

As $\mathcal{M}_D \circ \widehat{\lambda}^*$ is absolutely continuous on I, by Proposition A.1.4 it has a derivative almost everywhere. As the intersection of two sets of full measure is not empty—see Appendix A.1—there exists a $t_0 \in I$ such that $\mathcal{M}_D \circ \widehat{\lambda}^*$ is derivable at t_0 and $\mathcal{M}_D(\widehat{\lambda}^*(t_0)) = H(\widehat{\lambda}^*(t_0), u^*(t_0))$. For each $t \neq t_0$, by the definition of $\mathcal{M}_D$, we have

$$\mathcal{M}_D(\widehat{\lambda}^*(t)) = \max_{u \in D} H(\widehat{\lambda}^*(t), u) \geq H(\widehat{\lambda}^*(t), u^*(t_0))$$

because $u^*(t_0) \in D$. Thus

$$\mathcal{M}_D(\widehat{\lambda}^*(t)) - \mathcal{M}_D(\widehat{\lambda}^*(t_0)) \geq H(\widehat{\lambda}^*(t), u^*(t_0)) - H(\widehat{\lambda}^*(t_0), u^*(t_0)).$$

If $t - t_0 > 0$,

$$\frac{\mathcal{M}_D(\widehat{\lambda}^*(t)) - \mathcal{M}_D(\widehat{\lambda}^*(t_0))}{t - t_0} \geq \frac{H(\widehat{\lambda}^*(t), u^*(t_0)) - H(\widehat{\lambda}^*(t_0), u^*(t_0))}{t - t_0}.$$

Let us compute the right derivative of $\mathcal{M}_D \circ \widehat{\lambda}^*$ at t_0

$$\left. \frac{\mathrm{d}(\mathcal{M}_D \circ \widehat{\lambda}^*)}{\mathrm{d}t} \right|_{t=t_0^+} = \lim_{t \to t_0^+} \frac{\mathcal{M}_D(\widehat{\lambda}^*(t)) - \mathcal{M}_D(\widehat{\lambda}^*(t_0))}{t - t_0}$$

$$\geq \lim_{t \to t_0^+} \frac{H^{u^*(t_0)}\left(\widehat{\lambda}^*(t)\right) - H^{u^*(t_0)}\left(\widehat{\lambda}^*(t_0)\right)}{t - t_0} = \mathcal{L}_{\widehat{X}^{T^*\{u^*(t_0)\}}_{\widehat{\lambda}^*(t_0)}} H^{u^*(t_0)} = 0,$$

since $i_{\widehat{X}^{T^*\{u^*(t_0)\}}_{\widehat{\lambda}^*(t_0)}} \Omega = i\left(\widehat{X}^{T^*\{u^*(t_0)\}}_{\widehat{\lambda}^*(t_0)}\right) \Omega = \left(dH^{u^*(t_0)}\right)_{\widehat{\lambda}^*(t_0)}.$

Similarly, if $t - t_0 < 0$,

$$\left. \frac{\mathrm{d}(\mathcal{M}_D \circ \widehat{\lambda}^*)}{\mathrm{d}t} \right|_{t=t_0^-} \leq 0.$$

Hence the derivative of $\mathcal{M}_D \circ \widehat{\lambda}^*$ is zero almost everywhere.

⑧ Applying Theorem A.1.5, $\mathcal{M}_D \circ \widehat{\lambda}^*$ is constant everywhere, because $\mathcal{M}_D \circ \widehat{\lambda}^*$ is absolutely continuous.

⑨ As $\mathcal{M}_D(\widehat{\lambda}^*(t))$ and $\mathcal{M}(\widehat{\lambda}^*(t))$ coincide everywhere, $\mathcal{M} \circ \widehat{\lambda}^*$ is constant everywhere on I.

(3c) $\widehat{\lambda}^*(t) \neq 0 \in T^*_{\widehat{\gamma}^*(t)}\widehat{M}$ for each $t \in [a, b]$.

Let us suppose that there exists $\tau \in [a, b]$ such that $\widehat{\lambda}^*(\tau) = 0 \in T^*_{\widehat{\gamma}^*(\tau)}\widehat{M}$. As $\widehat{\lambda}^*$ is a generalized integral curve of $(\widehat{X}^{T^*})^{\{u^*\}}$, a linear vector field over $\widehat{X}$, we have $\widehat{\lambda}^*(t) = 0$ for each $t \in [a, b]$. As there exists at least a time such that $\widehat{\lambda}^*(\tau) \neq 0$, we arrive at a contradiction. Hence $\widehat{\lambda}^*(t) \neq 0$ for each $t \in [a, b]$.

(3d) $\lambda_0^*(t)$ is constant, $\lambda_0^*(t) \leq 0$.

From the equations satisfied by the generalized integral curves of $(\widehat{X}^{T^*})^{\{u^*\}}$, we have p_0 is constant. It was seen that $\langle \widehat{\lambda}_b, (-1, \mathbf{0}) \rangle \geq 0$ is equivalent to $(p_0 \circ \widehat{\lambda}^*)(b) = \lambda_0(b) \leq 0$. Hence $\lambda_0 \leq 0$ for each $t \in [a, b]$. ∎

Comment: As $\widehat{\lambda}_b$ is determined up to multiply by a positive real number, we may assume that $\lambda_0 \in \{-1, 0\}$.

The way in which perturbations have been used in this proof gives some clue concerning the fact that the tangent perturbation cone is understood as an approximation of the reachable set defined in §3.2. A precise meaning of this approximation is explained in §4.5.2 and §4.6.2.

The covector in the proof has been chosen such that

$$\langle \widehat{\lambda}_b, (-1, \mathbf{0}) \rangle \geq 0,$$
$$\langle \widehat{\lambda}_b, \widehat{v}_b \rangle \leq 0 \quad \forall \widehat{v}_b \in \widehat{K}_b.$$

In the abnormal case $\lambda_0 = 0$ and the first inequality is satisfied with equality. Thus the vector $(-1, \mathbf{0})$ is contained in the separating hyperplane, whereas for normal extremals $(-1, \mathbf{0})$ is not contained in the separating hyperplane. These two situations are shown in Figure 4.4. It would be interesting to determine geometrically what else must happen in order to have abnormal minimizers.
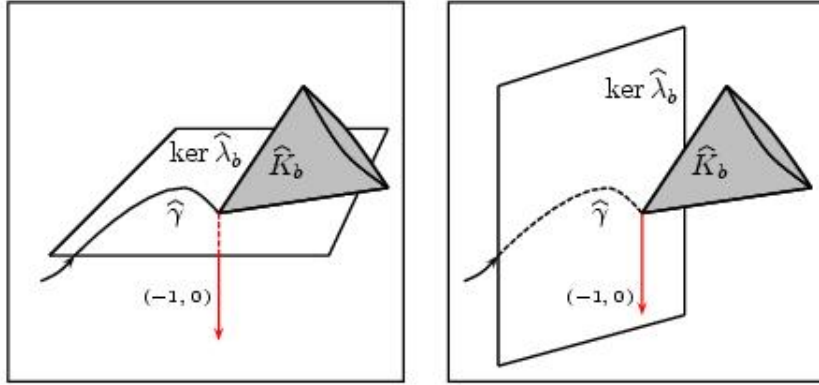
Figure 4.4: *Separation condition for an extremal being normal and abnormal, respectively.*

4.3 Pontryagin's Maximum Principle for nonfixed time and non-fixed endpoints

Now that Pontryagin's Maximum Principle has been proved for time and endpoints fixed, let us state the different problems related to Pontryagin's Maximum Principle with nonfixed time and nonfixed endpoints.

4.3.1 Statement of the optimal control problem with time and endpoints non-fixed

We consider the elements M, U, X, $\mathcal{F}$, $\mathcal{S}$ and π_2 with the same properties as in §2.2.1, §4.1.1. Let S_a and S_f be submanifolds of M.

Statement 4.3.1. (Free Optimal Control Problem, FOCP) *Given the elements M, U, X, $\mathcal{F}$, and the disjoint submanifolds of M, S_a and S_f, consider the following problem.*

Find $b \in \mathbb{R}$ and $(\gamma^*, u^*)\colon [a, b] \to M \times U$ such that

(1) $\gamma^*(a) \in S_a$, $\gamma^*(b) \in S_f$ (endpoint conditions),

(2) γ^* is an integral curve of $X^{\{u^*\}}$: $\dot{\gamma}^* = X^{\{u^*\}} \circ (\gamma^*, \mathrm{id})$, and

(3) $\mathcal{S}[\gamma^*, u^*] = \int_a^b F(\gamma^*(t), u^*(t))\mathrm{d}t$ is minimum over all curves (γ, u) satisfying (1) and (2) (minimal condition).

The tuple $(M, U, X, \mathcal{F}, S_a, S_f)$ denotes the *free optimal control problem.*

Statement 4.3.2. (Extended Free Optimal Control Problem, $\widehat{\mathrm{FOCP}}$) *Given the elements $\widehat{M}$ and $\widehat{X}$ defined in §4.1.2 and the FOCP, $(M, U, X, \mathcal{F}, S_a, S_f)$, consider the following problem.*

Find $b \in \mathbb{R}$ and $(\widehat{\gamma}^*, u^*)\colon [a, b] \to \widehat{M} \times U$, with $\gamma^* = \pi_2 \circ \widehat{\gamma}^*$, such that

(1) $\widehat{\gamma}^*(a) \in \{0\} \times S_a$, $\gamma^*(b) \in S_f$ (endpoint conditions),

(2) $\widehat{\gamma}^*$ is an integral curve of $\widehat{X}^{\{u^*\}}$: $\dot{\widehat{\gamma}}^* = \widehat{X}^{\{u^*\}} \circ (\widehat{\gamma}^*, \mathrm{id})$, and

(3) $\gamma^{0^*}(b)$ is minimum over all curves $(\widehat{\gamma}, u)$ satisfying (1) and (2) (minimal condition).

The tuple $(\widehat{M}, U, \widehat{X}, S_a, S_f)$ denotes the *extended free optimal control problem.*

4.3.2 Perturbation of the time and the endpoints

In this case of nonfixed time and nonfixed endpoint optimal control problems, we not only modify the control as explained in §4.1.3, but also modify the final time and the endpoint conditions. As was mentioned in §4.1.3, the following constructions obtained from perturbing the final time and the endpoint conditions are also general for any vector field depending on parameters.

4.3.2.1 Time perturbation vectors and associated cones

We study how to perturb the interval of definition of the control taking advantage of the fact that the final time is another unknown for the free optimal control problems.

Let X be a vector field on M along the projection $\pi\colon M \times U \to M$, $I \subset \mathbb{R}$ be a closed interval and $(\gamma, u)\colon I = [a, b] \to M \times U$ a curve such that γ is an integral curve of $X^{\{u\}}$.

Let $\pi_\pm = \{\tau, l_\tau, \delta\tau, u_\tau\}$, where τ is a Lebesgue time in (a, b) for $X \circ (\gamma, u)$, $l_\tau \in \mathbb{R}^+ \cup \{0\}$, $\delta\tau \in \mathbb{R}$, $u_\tau \in U$. For every $s \in \mathbb{R}^+$ small enough that $a < \tau - (l_\tau - \delta\tau)s$, consider $u[\pi_\pm^s]\colon [a, b + s\delta\tau] \to U$ defined by

$$
u[\pi_\pm^s](t) = \begin{cases} u(t), & t \in [a, \tau - (l_\tau - \delta\tau)s], \\ u_\tau, & t \in (\tau - (l_\tau - \delta\tau)s, \tau + s\delta\tau], \\ u(t), & t \in (\tau + s\delta\tau, b + s\delta\tau], \end{cases}
$$

if $\delta\tau < 0$, and by

$$
u[\pi_\pm^s](t) = \begin{cases} u(t), & t \in [a, \tau - (l_\tau - \delta\tau)s], \\ u_\tau, & t \in (\tau - (l_\tau - \delta\tau)s, \tau + s\delta\tau], \\ u(t - s\delta\tau), & t \in (\tau + s\delta\tau, b + s\delta\tau], \end{cases}
$$

if $\delta\tau \geq 0$.

Definition 4.3.3. *The function $u[\pi_\pm^s]$ is called a **perturbation of** u **specified by the data** $\pi_\pm = \{\tau, l_\tau, \delta\tau, u_\tau\}$.*

Associated to $u[\pi_\pm^s]$ we consider the mapping $\gamma[\pi_\pm^s]\colon [a, b + s\delta\tau] \to M$, the generalized integral curve of $X^{\{u[\pi_\pm^s]\}}$ with initial condition $(a, \gamma(a))$.

Given $\epsilon > 0$, define

$$
\begin{aligned} \varphi_{\pi_\pm}\colon [\tau, b] \times [0, \epsilon] &\longrightarrow M \\ (t, s) &\longmapsto \varphi_{\pi_\pm}(t, s) = \gamma[\pi_\pm^s](t + s\delta\tau). \end{aligned}
$$

For every $t \in [\tau, b]$, $\varphi_{\pi_\pm}^t : [0, \epsilon] \to M$ is given by $\varphi_{\pi_\pm}^t(s) = \varphi_{\pi_\pm}(t, s)$.

As explained in §4.1.3, the control $u[\pi_\pm^s]$ depends continuously on the parameters s and $\pi_\pm = \{\tau, l_\tau, \delta\tau, u_\tau\}$. Hence the curve $\varphi_{\pi_\pm}^t$ depends continuously on s and $\pi_\pm = \{\tau, l_\tau, \delta\tau, u_\tau\}$, and so it converges uniformly to γ as s tends to 0. See [Cañizo–Rincón 2004, Coddington and Levinson 1955] for more details of the differential equations depending continuously on parameters.

Let us prove that the curve $\varphi_{\pi_\pm}^\tau$ has a tangent vector at $s = 0$; cf. Proposition 4.1.4.

Proposition 4.3.4. *Let τ be a Lebesgue time. If $u[\pi_\pm^s]$ is the perturbation of the control u specified by the data $\pi_\pm = \{\tau, l_\tau, \delta\tau, u_\tau\}$ such that $\tau + s\delta\tau$ is a Lebesgue time, then the curve $\varphi_{\pi_\pm}^\tau : [0, \epsilon] \to M$ is differentiable at $s = 0$ and its tangent vector is $X(\gamma(\tau), u(\tau)) \delta\tau + [X(\gamma(\tau), u_\tau) - X(\gamma(\tau), u(\tau))] l_\tau$.*

(*Proof*) As in the proof of Proposition 4.1.4, we compute the limit

$$A = \lim_{s \to 0} \frac{(x^i \circ \varphi_{\pi_\pm}^\tau)(s) - (x^i \circ \varphi_{\pi_\pm}^\tau)(0)}{s} = \lim_{s \to 0} \frac{\gamma^i[\pi_\pm^s](\tau + s\delta\tau) - \gamma^i(\tau)}{s}.$$

As γ is an absolutely continuous integral curve of $X^{\{u\}}$, $\dot\gamma(t) = X(\gamma(t), u(t))$ at every Lebesgue time. Then, integrating,

$$\gamma^i(\tau) - \gamma^i(a) = \int_a^\tau f^i(\gamma(t), u(t)) \mathrm{d}t$$

and similarly for $\gamma[\pi_\pm^s]$ and $u[\pi_\pm^s]$. Observe that $\gamma[\pi_\pm^s](t) = \gamma(t)$ and $u[\pi_\pm^s](t) = u(t)$ for $t \in [a, \tau - (l_\tau - \delta\tau)s]$.

Here, we should consider three different possibilities:

- if $0 \le \delta\tau \le l_\tau$, then $\tau - (l_\tau - \delta\tau)s < \tau < \tau + s\delta\tau$;

- if $\delta\tau < 0$, then $\tau - (l_\tau - \delta\tau)s < \tau + s\delta\tau < \tau$;

- if $0 < l_\tau < \delta\tau$, then $\tau < \tau - (l_\tau - \delta\tau)s < \tau + s\delta\tau$.

We prove the proposition for the first case and the other cases follow analogously.

$$
\begin{aligned}
A &= \lim_{s \to 0} \frac{\int_a^{\tau + s\delta\tau} f^i(\gamma[\pi_\pm^s](t), u[\pi_\pm^s](t)) \mathrm{d}t - \int_a^\tau f^i(\gamma(t), u(t)) \mathrm{d}t}{s} \\
&= \lim_{s \to 0} \frac{\int_{\tau - (l_\tau - \delta\tau)s}^{\tau + s\delta\tau} f^i(\gamma[\pi_\pm^s](t), u_\tau) \mathrm{d}t - \int_{\tau - (l_\tau - \delta\tau)s}^\tau f^i(\gamma(t), u(t)) \mathrm{d}t}{s}.
\end{aligned}
$$

As $\tau + s\delta\tau$ is a Lebesgue time, using Equation (A.2.3) we have

$$
\begin{aligned}
A &= \lim_{s \to 0} \frac{f^i(\gamma[\pi_\pm^s](\tau + s\delta\tau), u_\tau) l_\tau s - f^i(\gamma(\tau), u(\tau))(l_\tau - \delta\tau)s + o(s)}{s} \\
&= \lim_{s \to 0} f^i(\gamma[\pi_\pm^s](\tau + s\delta\tau), u_\tau) l_\tau - f^i(\gamma(\tau), u(\tau))(l_\tau - \delta\tau).
\end{aligned}
$$

As f^i is continuous on M, we have

$$
\begin{aligned}
A &= [f^i(\gamma(\tau), u_\tau) - f^i(\gamma(\tau), u(\tau))] \, l_\tau + f^i(\gamma(\tau), u(\tau)) \, \delta\tau \\
&= \mathcal{L}([X(\gamma(\tau), u(\tau)) \, \delta\tau + (X(\gamma(\tau), u_\tau) - X(\gamma(\tau), u(\tau))) \, l_\tau])(x^i).
\end{aligned}
$$

∎

Definition 4.3.5. *The tangent vector*

$$
v[\pi_\pm] = X(\gamma(\tau), u(\tau)) \, \delta\tau + [X(\gamma(\tau), u_\tau) - X(\gamma(\tau), u(\tau))] \, l_\tau
$$

*is the **perturbation vector associated to the perturbation data** $\pi_\pm = \{\tau, l_\tau, \delta\tau, u_\tau\}$.*

If we disturb the control r times at r different Lebesgue times as in §4.1.3 and also the domain of the curve (γ, u) as just described, we have the perturbation data $\pi = \{\pi_1, \ldots, \pi_r, \pi_\pm\}$, with $a < t_1 \leq \ldots \leq t_r \leq \tau < b$. From the generalized integral curve $\gamma[\pi^s]$ of $X^{\{u[\pi^s]\}}$ with initial condition $(a, \gamma(a))$, we construct the curve $\varphi_\pi^t \colon [0, \epsilon] \to M$ for $t \in [\tau, b]$ given by $\varphi_\pi^t(s) = \gamma[\pi^s](t + s\delta\tau)$.

Corollary 4.3.6. *Let t be a Lebesgue time in $[\tau, b]$. If the data $\pi = \{\pi_1, \ldots, \pi_r, \pi_\pm\}$ defines a perturbation of the control u, then the vector tangent to the curve $\varphi_\pi^t \colon [0, \epsilon] \to M$ at $s = 0$ is $X(\gamma(t), u(t)) \, \delta\tau + V[\pi_1](t) + \ldots + V[\pi_r](t)$, where $V[\pi_i] \colon [t_i, b] \to TM$ is the generalized integral curve of $\left(X^T\right)^{\{u\}}$ with initial condition $(t_i, v[\pi_i])$ where $v[\pi_i] \in T_{\gamma(t_i)}M$ for $i = 1, \ldots, r$.*

(Proof) This corollary is proved taking into account Propositions 4.1.6, 4.1.7, 4.3.4, Corollary 4.1.9 and §2.2.2.1. ∎

Now, at a Lebesgue time $t \in (a, b)$, we construct a new cone such that it contains the tangent perturbation cone K_t in Definition 4.1.11 and $\pm X(\gamma(t), u(t))$.

Definition 4.3.7. *The **time perturbation cone** $K_t^\pm$ at every Lebesgue time t is the smallest closed cone in $T_{\gamma(t)}M$ containing K_t and $\pm X(\gamma(t), u(t))$.*

$$
K_t^\pm = \overline{\mathrm{conv}\left(\{\pm\mu X(\gamma(t), u(t)) \mid \mu \in \mathbb{R}\} \bigcup \left(\bigcup_{\substack{a < \tau \leq t \\ \tau \text{ is a Lebesgue time}}} \left(\Phi_{(t,\tau)}^{X^{\{u\}}}\right)_* \mathcal{V}_\tau \right) \right)},
$$

where $\mathcal{V}_\tau$ denotes the set of elementary perturbation vectors at τ, see Definition 4.1.11.

Enlarging the cone K_τ to $K_\tau^\pm$ allows us to introduce time variations.

Proposition 4.3.8. *If t_2 is a Lebesgue time greater than t_1, then*

$$
\left(\Phi_{(t_2,t_1)}^{X^{\{u\}}}\right)_* K_{t_1}^\pm \subset K_{t_2}^\pm.
$$

(*Proof*) We have

$$K_{t_1}^{\pm} = \text{conv}\left(\overline{\{\pm\mu X(\gamma(t_1), u(t_1)) \mid \mu \in \mathbb{R}\} \bigcup \left(\bigcup_{\substack{a < \tau \le t_1 \\ \tau \text{ is a Lebesgue time}}} \left(\Phi_{(t_1,\tau)}^{X^{\{u\}}}\right)_* \mathcal{V}_\tau\right)}\right).$$

Just for simplicity we use $\mathcal{C}_{t_1}^{\pm}$ to denote

$$\text{conv}\left(\{\pm\mu X(\gamma(t_1), u(t_1)) \mid \mu \in \mathbb{R}\} \bigcup \left(\bigcup_{\substack{a < \tau \le t_1 \\ \tau \text{ is a Lebesgue time}}} \left(\Phi_{(t_1,\tau)}^{X^{\{u\}}}\right)_* \mathcal{V}_\tau\right)\right).$$

1. The set $\mathcal{C}_{t_1}^{\pm}$ being convex, if v is interior to $K_{t_1}^{\pm}$, then v is interior to $\mathcal{C}_{t_1}^{\pm}$ by Proposition B.1.5, item (d). Hence, by Proposition B.1.4

$$v = l_0 X(\gamma(t_1), u(t_1)) + \sum_{i=1}^{r} l_i V[\pi_i](t_1),$$

   where every $V[\pi_i](t_1)$ is the transported elementary perturbation vector $v[\pi_i]$ from t_i to t_1 by the flow of $X^{\{u\}}$, $l_i \in [0, 1]$ for $i = 0, \ldots, r$ and $\sum_{i=0}^{r} l_i = 1$. By definition of the cone and the linearity of the flow, $\left(\Phi_{(t_2,t_1)}^{X^{\{u\}}}\right)_* v$ is in $K_{t_2}^{\pm}$, since

$$\left(\Phi_{(t_2,t_1)}^{X^{\{u\}}}\right)_* (X(\gamma(t_1), u(t_1))) = X(\gamma(t_2), u(t_2)),$$

   because both sides of the equality are the unique solutions of the variational equation along γ associated with $X^{\{u\}}$ with initial condition $(t_1, X(\gamma(t_1), u(t_1)))$. See §2.2.2.1 for more details.

2. If v is in the boundary of $K_{t_1}^{\pm}$, then there exists a sequence of vectors $(v_j)_{j \in \mathbb{N}}$ in the interior of $K_{t_1}^{\pm}$ such that

$$\lim_{j \to \infty} v_j = v.$$

   Due to the continuity of the flow

$$\lim_{j \to \infty} \left(\Phi_{(t_2,t_1)}^{X^{\{u\}}}\right)_* v_j = \left(\Phi_{(t_2,t_1)}^{X^{\{u\}}}\right)_* v.$$

   All the elements of the convergent sequence are in the closed cone $K_{t_2}^{\pm}$, hence the limit $\left(\Phi_{(t_2,t_1)}^{X^{\{u\}}}\right)_* v$ is also in $K_{t_2}^{\pm}$.

   If the interior of $K_{t_1}^{\pm}$ is empty, we consider the relative topology briefly defined in Appendix B.2 and the reasoning follows as before.

$$\blacksquare$$

For the time perturbation cone $K_\tau^{\pm}$ and the corresponding perturbation vectors, it can be proved properties analogous to the ones stated in Propositions 4.1.4, 4.1.6, 4.1.7 and 4.1.8.

Proposition 4.3.9. *Let $t \in (a, b)$ be a Lebesgue time. If v is a nonzero vector interior to $K_t^{\pm}$, then there exists $\epsilon > 0$ such that for every $s \in (0, \epsilon)$ there are $s' > 0$ and a perturbation of the control $u[\pi_{w_0}^s]$ such that $\gamma[\pi_{w_0}^s](t + s\delta t) = \gamma(t) + s'v$ where w_0 is the perturbation vector associated to the perturbation of the control.*

(*Proof*) The proof follows the same line as the proof of Proposition 4.1.12, but now the tangent space to M at $\gamma(t + s\delta t)$ is also identified with $\mathbb{R}^m$ through the local chart of M at $\gamma(t)$.

We use the same functions as in the proof of Proposition 4.1.12, but having in mind that $\Gamma(s, r) = \gamma[\pi_{w_0}^s](t)$ in Equation (4.1.2) is replaced by $\Gamma(s, r) = \gamma[\pi_{w_0}^s](t + s\delta t)$. ∎

4.3.2.2 Perturbing the endpoint conditions

Now we consider that the endpoint conditions for the integral curves of $X^{\{u\}}$ varies on submanifolds of M. Let S_a be a submanifold of M and $\gamma(a)$ in S_a; consider the integral curve $\gamma \colon I \to M$ of $X^{\{u\}}$ with initial condition $(a, \gamma(a))$.

We consider the curve $\gamma[\pi_{\pm}^s]$ obtained from a time perturbation of the control u associated with a vector in the time perturbation cone. The initial condition is disturbed along a curve $\delta \colon [0, \epsilon] \to S_a$ with initial tangent vector v_a in $T_{\gamma(a)}S_a$ and $\delta(0) = \gamma(a)$. Taking into account §2.2.2.1, §4.1.3.1 and considering that $T_{\gamma(a)}S_a$ and an open set at $\delta(a)$ are identified with $\mathbb{R}^m$, the integral curve $\gamma_{\delta(s)}[\pi_{\pm}^s] \colon I \to M$ of $X^{\{u[\pi_{\pm}^s]\}}$ with initial condition $(a, \delta(s))$ can be written as

$$\gamma_{\delta(s)}[\pi_{\pm}^s](t) = \gamma(t) + s \left(\Phi_{(t,a)}^{X^{\{u\}}} \right)_* v_a + sv[\pi_{\pm}](t) + o(s).$$

We define a cone that includes the time perturbation vectors, the elementary perturbation vectors and the vectors coming from changing the initial condition on S_a along different curves $\delta \colon [0, \epsilon] \to S_a$ through $\gamma(a)$ and contained in S_a.

Definition 4.3.10. *Let t be a Lebesgue time for the reference control u. The cone $\mathcal{K}_t$ is the smallest closed and convex cone in $T_{\gamma(t)}M$ containing the time perturbation cone at time t and the transported of the tangent space to S_a from a to t through the flow of $X^{\{u\}}$.*

$$\mathcal{K}_t = \overline{\mathrm{conv}(K_t^{\pm} \bigcup \left(\Phi_{(t,a)}^{X^{\{u\}}} \right)_* (T_{\gamma(a)}S_a))}$$

Proposition 4.3.11. *Let t be a Lebesgue time in (a, b) and $S \subset M$ be a submanifold with boundary. Suppose that $\gamma(t)$ is on the boundary of S. Let T be the half–plane tangent to S at $\gamma(t)$. If $\mathcal{K}_t$ and T are not separated, then there exists a perturbation of the control $u[\pi_{w_0}^s]$, where w_0 is the corresponding perturbation vector, and $x_a \in S_a$ such that the integral curve $\gamma_{x_a}[\pi_{w_0}^s]$ of $X^{\{u[\pi_{w_0}^s]\}}$ with initial condition (a, x_a) meets S at a point in the relative interior of S.*

(*Proof*) As $\mathcal{K}_t$ and T are not separated, by Proposition B.2.9 there no exists any hyperplane containing both and there is a vector v in the relative interior of both $\mathcal{K}_t$ and T. By Corollary

B.2.10, if $\mathcal{K}_t$ and T are not separated,

$$T_{\gamma(t)}M = \mathcal{K}_t - T.$$

See Appendix B for the notation and properties. If V is an open set of a local chart at $\gamma(t)$, we identify V with $\mathbb{R}^m$ and also the tangent space at $\gamma(t)$, $T_{\gamma(t)}M$, in the same sense as in Equation (4.1.1). Let us consider an orthonormal basis in $T_{\gamma(t)}M$, $\{e_1, \ldots, e_m\}$. If we take $e_0 = -(e_1 + \ldots + e_m)$, the vector $0 \in T_{\gamma(t)}M$ is expressed as an affine combination of $e_0, e_1, \ldots, e_m$:

$$0 = \frac{1}{m+1}e_0 + \ldots + \frac{1}{m+1}e_m. \tag{4.3.13}$$

Each w in $T_{\gamma(t)}M$ is written uniquely as

$$w = a^1 e_1 + \ldots + a^m e_m$$

and as an affine combination—see Definition B.1.2—of $e_0, e_1, \ldots, e_m$:

$$w = \sum_{i=0}^{m} b^i(w)e_i = re_0 + \sum_{i=0}^{m}(r + a^i)e_i \quad \text{with} \quad r = \frac{1 - \sum_{i=1}^{m} a^i}{m+1}.$$

Hence, we define the continuous mapping

$$\begin{aligned} \mathcal{G}: T_{\gamma(t)}M &\longrightarrow \mathbb{R}^{m+1} \\ w &\longmapsto (b^0(w), b^1(w), \ldots, b^m(w)) \end{aligned}$$

As $b^i(0) > 0$ for every $i = 0, \ldots, m$ because of Equation (4.3.13), there exists an open ball $B(0, r)$ centered at 0 with radius r such that for every $w \in B(0, r)$, $b^i(w) > 0$ for $i = 0, \ldots, m$. Now we consider the restriction of $\mathcal{G}$ to the closed ball $\overline{B(0, r)}$, $\mathcal{G}_{|\overline{B(0,r)}} : \overline{B(0, r)} \to [0, 1]^{m+1}$. Choose vectors $e_i^{\mathcal{K}} \in \mathcal{K}_t$ and $e_i^T \in T$ such that

$$e_i = e_i^{\mathcal{K}} - e_i^T.$$

As v lies in the relative interior of both convex sets, $e_i^{\mathcal{K}} + v$ is in the relative interior of $\mathcal{K}_t$ and $e_i^T + v$ is in the relative interior of T. Thus, if we rename the vectors, we always have $e_i = e_i^{\mathcal{K}} - e_i^T$, where each vector is in the relative interior to the corresponding set. For any $w \in \overline{B(0, r)}$,

$$w = \sum_{i=0}^{m} b^i(w)e_i = \sum_{i=0}^{m} b^i(w)\left(e_i^{\mathcal{K}} - e_i^T\right).$$

Then we can define

$$\begin{aligned} F_1: \overline{B(0, r)} &\longrightarrow \mathcal{K}_t \\ w &\longmapsto F_1(w) = \sum_{i=0}^{m} b^i(w)e_i^{\mathcal{K}}, \end{aligned}$$

$$\begin{aligned} F_2: \overline{B(0, r)} &\longrightarrow T \\ w &\longmapsto F_2(w) = \sum_{i=0}^{m} b^i(w)e_i^T, \end{aligned}$$

and let us consider the mapping

$$
\begin{aligned}
G \colon \mathbb{R} \times \overline{B(0,r)} &\longrightarrow \mathbb{R}^m \\
(s,w) &\longmapsto \frac{\gamma[\pi^s_{F_1(w)}](t) - \gamma[\pi^s_{F_2(w)}](t)}{s},
\end{aligned}
$$

where $\gamma[\pi^s_{F_1(w)}]$ is the perturbation curve associated to $\pi^s_{F_1(w)}$ and $\gamma[\pi^s_{F_2(w)}](t) = \gamma(t) + sF_2(w)$ is the straight line through $\gamma(t)$ with tangent vector $F_2(w)$. As the perturbation vectors are in the relative interior of the convex cones, we use Proposition 4.1.12 and the linear approximation (4.1.1) in such a way that $G(s,w) = F_1(w) - F_2(w) + o(1)$. Hence

$$
\lim_{s \to 0} G(s,w) = F_1(w) - F_2(w) = w.
$$

Hence, for any positive number ϵ, there exists $s_0 > 0$ such that if $s < s_0$ then $\|G(s,w) - w\| < \epsilon$. Take $\epsilon < r$, then

$$
\|G(s,w) - w\| < \epsilon < r = \|w\|
$$

for every w in the boundary of $\overline{B(0,r)}$. Thus the map $G_s \colon \overline{B(0,r)} \to \mathbb{R}^m$, $G_s(w) = G(s,w)$, satisfies the hypotheses of Corollary A.3.2 for the point 0 in $B(0,r)$. Hence, the point 0 is in the image of $\overline{B(0,r)}$ through G_s and there exists w such that $G_s(w) = 0$, that is,

$$
\gamma[\pi^s_{F_1(w)}](t) = \gamma[\pi^s_{F_2(w)}](t).
$$

Therefore, there exists a perturbation of the control, given by $u[\pi^s_{F_1(w)}]$ such that the associated trajectory meets S in an interior point since $F_2(w)$ lies in the relative interior of T. ∎

4.3.3 Pontryagin's Maximum Principle with time and endpoints nonfixed

Bearing in mind the symplectic formalism introduced in §2.3.1 and §4.1.4, we define the corresponding Hamiltonian Problem when the time and the endpoints are nonfixed.

Statement 4.3.12. (*Free Hamiltonian Problem, FHP*) *Given FOCP* $(M, U, X, \mathcal{F}, S_a, S_f)$, *and the equivalent* $\widehat{\mathrm{FOCP}}$ $(\widehat{M}, U, \widehat{X}, S_a, S_f)$, *consider the following problem.*

Find $b \in \mathbb{R}$ *and* $(\widehat{\lambda}, u) \colon [a,b] \to T^*\widehat{M} \times U$, *with* $\widehat{\gamma} = \pi_{\widehat{M}} \circ \widehat{\lambda}$ *and* $\gamma = \pi_2 \circ \widehat{\gamma}$, *such that*

(1) $\widehat{\gamma}(a) \in \{0\} \times S_a$, $\gamma(b) \in S_f$, *and*

(2) $\dot{\widehat{\lambda}}(t) = \widehat{X}^{T^*}(\widehat{\lambda}(t), u(t))$ *for a.e.* $t \in I$.

The tuple $(T^*\widehat{M}, U, \widehat{X}^{T^*}, S_a, S_f)$ denotes the *free Hamiltonian problem.*

Comments:

1. The minimum of the interval of definition of the curves is a, but the maximum is not fixed.

2. The curves γ, $\widehat{\gamma}$ and $\widehat{\lambda}$ are assumed to be absolutely continuous. So they are generalized integral curves of $X^{\{u\}}$, $\widehat{X}^{\{u\}}$ and $\left(\widehat{X}^{T^*}\right)^{\{u\}}$, respectively, in the sense defined in §2.2.1.

Now, we are ready to state the Free Pontryagin's Maximum Principle that provides the necessary conditions, but in general not sufficient, for finding solutions of the free optimal control problem.

Theorem 4.3.13. *(Free Pontryagin's Maximum Principle, FPMP)*
If $(\widehat{\gamma}^*, u^*)\colon [a, b] \to \widehat{M} \times U$ is a solution of the extended free optimal control problem, State-ment 4.3.2, then there exists $(\widehat{\lambda}^*, u^*)\colon [a, b] \to T^*\widehat{M} \times U$ such that:

1. *it is a solution of the associated free Hamiltonian problem in Statement 4.3.12;*

2. *$\widehat{\gamma}^* = \pi_{\widehat{M}} \circ \widehat{\lambda}^*$;*

3. *(a) $H\left(\widehat{\lambda}^*(t), u^*(t)\right) = \sup_{u \in U} H(\widehat{\lambda}^*(t), u)$ almost everywhere;*

    (b) $\sup_{u \in U} H(\widehat{\lambda}^*(t), u) = 0$ everywhere;

    (c) $\widehat{\lambda}^*(t) \neq 0 \in T^*_{\widehat{\gamma}^*(t)}\widehat{M}$ for each $t \in [a, b]$;

    (d) $\lambda_0^*(t)$ is constant, $\lambda_0^*(t) \leq 0$;

    (e) transversality conditions: $\lambda^*(a) \in \operatorname{ann} T_{\gamma^*(a)}S_a$ and $\lambda^*(b) \in \operatorname{ann} T_{\gamma^*(b)}S_f$.

Observe that once we have the optimal solution of the $\widehat{\text{FOCP}}$, the final time and the end-point conditions are known and fixed. We might think of just applying Theorem 4.1.14 in order to prove Theorem 4.3.13. However, this is not possible because the freedom to chose the final time and the endpoint conditions, only restricted to submanifolds, in Statement 4.3.2 is used in the proof of FPMP to consider variations of the optimal curve that are slightly different from the variations used in the case of fixed time, see §4.1.3 and §4.3.2 to compare them.

Apart from the transversality conditions, the main difference between FPMP and PMP is the fact that the domain of the curves in the optimal control problems is unknown. This introduces a new necessary condition: the maximum of the Hamiltonian must be zero, not just constant—see comment 4 after Theorem 4.1.14 to understand why the maximum can be used instead of the supremum. Then, from (3a) and (3b) it may be concluded that the Hamiltonian is zero almost everywhere. For instance, in the time optimal problems the Hamiltonian along extremals must be zero.

There are different statements of Pontryagin's Maximum Principle. In §4.1.4 we have con-sidered the statement of PMP for a fixed–time problem without transversality conditions to simplify the proof, although it may be stated the PMP for the fixed–time problem with vari-able endpoints where the transversality conditions appear. There also exists the PMP for the FOCP with the degenerate case that the submanifolds are only a point, then the Theorem is the following one.

Theorem 4.3.14. *(Free Pontryagin's Maximum Principle without variable endpoints)*
If $(\widehat{\gamma}^*, u^*) \colon [a, b] \to \widehat{M} \times U$ is a solution of the extended free optimal control problem, Statement 4.3.2, with endpoint submanifolds $S_a = \{x_a\}$ and $S_f = \{x_f\}$, then there exists $(\widehat{\lambda}^*, u^*) \colon [a, b] \to T^*\widehat{M} \times U$ such that:

1. *it is a solution of the associated free Hamiltonian problem in Statement 4.3.12;*

2. $\widehat{\gamma}^* = \pi_{\widehat{M}} \circ \widehat{\lambda}^*$;

3. (a) $H\left(\widehat{\lambda}^*(t), u^*(t)\right) = \sup_{u \in U} H\left(\widehat{\lambda}^*(t), u\right)$ *almost everywhere;*

   (b) $\sup_{u \in U} H\left(\widehat{\lambda}^*(t), u\right) = 0$ *everywhere;*

   (c) $\widehat{\lambda}^*(t) \neq 0 \in T^*_{\widehat{\gamma}^*(t)}\widehat{M}$ *for each $t \in [a, b]$;*

   (d) $\lambda_0^*(t)$ *is constant,* $\lambda_0^*(t) \leq 0$;

The only difference with Theorem 4.3.13 is that the transversality conditions do not appear.

4.4 Proof of Pontryagin's Maximum Principle for nonfixed time and nonfixed endpoints

In the proof of Theorem 4.3.13 we use notions about perturbations of the trajectories of a system introduced in §4.3.2. Observe that they are slightly different from the perturbations in §4.1.3 used to prove Theorem 4.1.14.

(*Proof*) (*Theorem 4.3.13: Free Pontryagin's Maximum Principle, FPMP*)

Given a solution of the $\widehat{\text{FOCP}}$, we only need an appropriate initial condition in the fibers of $\pi_{\widehat{M}} \colon T^*\widehat{M} \to \widehat{M}$ to find a solution of the FHP. Note that this initial condition is not given in the hypotheses of the Free Pontryagin's Maximum Principle. It is not possible to use Theorem 4.1.14 directly because the perturbation cones are not the same. Indeed, we need to consider changes in the interval of definition of the curves. These changes imply the inclusion of $\pm \widehat{X}(\widehat{\gamma}^*(t_1), u^*(t_1))$ in the perturbation cone at time t_1. All the times considered in this proof are Lebesgue times for the vector field along the optimal curve.

By Proposition 4.3.8, for $t_2 > t_1$,

$$\left(\Phi^{\widehat{X}^{\{u^*\}}}_{(t_2, t_1)}\right)_* \widehat{K}^{\pm}_{t_1} \subset \widehat{K}^{\pm}_{t_2}.$$

Let us consider the limit cone as follows

$$\widehat{K}^{\pm}_b = \overline{\bigcup_{\substack{a < \tau \leq b \\ \tau \text{ is a Lebesgue time}}} \left(\Phi^{\widehat{X}^{\{u\}}}_{(b, \tau)}\right)_* \widehat{K}^{\pm}_{\tau}}.$$

Observe that it is a closed cone and it is convex because it is the union of an increasing family of convex cones. Let us show that $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)}$ is not interior to $\widehat{K}_b^{\pm}$. Indeed, suppose that $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)}$ is interior to the limit cone, then it will be interior to

$$\bigcup_{\substack{a < \tau \leq b \\ \tau \text{ is a Lebesgue time}}} \left(\Phi_{(b,\tau)}^{\widehat{X}^{\{u\}}} \right)_* \widehat{K}_\tau^{\pm}$$

by Proposition B.1.5, item (d). As we have an increasing family of cones, there exists a time τ such that $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)}$ is interior to $\left(\Phi_{(b,\tau)}^{\widehat{X}^{\{u\}}} \right)_* \widehat{K}_\tau^{\pm}$. Let us see that this is not possible.

If $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)}$ is interior to $(\Phi_{(b,\tau)}^{\widehat{X}^{\{u\}}})_* \widehat{K}_\tau^{\pm}$, then, by Proposition 4.3.9, there exists $\epsilon > 0$ such that, for every $s \in (0, \epsilon)$, there exist $s' > 0$ and a perturbation of the control $u[\pi_{w_0}^s]$ such that

$$\widehat{\gamma}[\pi_{w_0}^s](b + s\delta\tau) = \widehat{\gamma}^*(b) + s'(-1, \mathbf{0}).$$

Hence

$$\gamma^0[\pi_{w_0}^s](b + s\delta\tau) < \gamma^{*0}(b) \quad \text{and} \quad \gamma[\pi_{w_0}^s](b + s\delta\tau) = \gamma^*(b).$$

That is, the trajectory $\gamma[\pi_{w_0}^s]$ arrives at the same endpoint as γ^* but with less cost. Then $\widehat{\gamma}^*$ cannot be optimal as assumed. Thus $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)}$ is not interior to $\widehat{K}_b^{\pm}$.

As $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)}$ is not in the interior of $\widehat{K}_b^{\pm}$, by Proposition B.2.9 there exists a covector $\widehat{\lambda}_b \in T_{\widehat{\gamma}^*(b)}^* \widehat{M}$ such that

$$\begin{aligned} \langle \widehat{\lambda}_b, (-1, \mathbf{0}) \rangle &\geq 0, \\ \langle \widehat{\lambda}_b, \widehat{v}_b \rangle &\leq 0 \quad \forall \, \widehat{v}_b \in \widehat{K}_b^{\pm}. \end{aligned}$$

The initial condition for the covector not only must satisfy the previous inequalities, but also the transversality conditions. In order to prove this, it is necessary to have the separability of two new cones.

(3e) Hence, the initial condition in the fibers of $T^* \widehat{M}$ may be chosen satisfying the transversality conditions. We consider the manifold with boundary given by

$$M_f = \{(x^0, x) \mid x \in S_f, \ x^0 \leq \gamma^{*0}(b)\}.$$

The set of tangent vectors to M_f at $\widehat{\gamma}^*(b)$ is the convex set whose generators are $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)}$ and $T_f = \{0\} \times T_{\gamma^*(b)} S_f$.

Given $\tau \in [a, b]$, consider the following closed convex sets

$$\mathcal{K}_\tau = \overline{\text{conv}(\widehat{K}_\tau^{\pm} \bigcup (\Phi_{(\tau,a)}^{\widehat{X}^{\{u^*\}}})_*(T_a))}, \qquad \text{where } T_a = \{0\} \times T_{\gamma^*(a)} S_a,$$

$$\mathcal{J}_\tau = \overline{\text{conv}((-1, \mathbf{0})_{\widehat{\gamma}^*(\tau)} \bigcup (\Phi_{(b,\tau)}^{\widehat{X}^{\{u^*\}}})_*^{-1}(T_f))}, \qquad \text{where } T_f = \{0\} \times T_{\gamma^*(b)} S_f,$$

and the manifold M_τ obtained transporting M_f from b to τ using the flow of $\widehat{X}^{\{u^*\}}$. Observe

that $\mathcal{J}_\tau$ is the closure of the set of tangent vectors to M_τ at the point $\widehat{\gamma}^*(\tau)$. We are going to show that the cones $\mathcal{K}_b$ and $\mathcal{J}_b$ are separated, using Proposition 4.3.11.

Observe that $\mathcal{J}_b$ is a half–plane tangent to M_f and $\widehat{\gamma}^*(b)$ is on the boundary of M_f by construction. Hence, if $\mathcal{K}_b$ and $\mathcal{J}_b$ were not separated, by Proposition 4.3.11 there would exist a perturbation of the control $u[\pi_{w_0}^s]$ and $x_a \in S_a$ such that the integral curve $\gamma_{x_a}[\pi_{w_0}^s]$ with initial condition (a, x_a) meets M_f at a point in the relative interior of M_f. Hence we have found a trajectory with less cost than the optimal one because of the definition of M_f. But this is not possible because of the optimality of $\widehat{\gamma}^*$. Thus $\mathcal{K}_b$ and $\mathcal{J}_b$ are separated. So, by Proposition B.2.9, there exists a covector $\widehat{\lambda}_b \in T^*_{\widehat{\gamma}^*(b)}\widehat{M}$ such that

$$\langle \widehat{\lambda}_b, \widehat{v}_b \rangle \leq 0 \quad \forall\, \widehat{v}_b \in \mathcal{K}_b, \tag{4.4.14}$$

$$\langle \widehat{\lambda}_b, \widehat{w}_b \rangle \geq 0 \quad \forall\, \widehat{w}_b \in \mathcal{J}_b. \tag{4.4.15}$$

This covector separates the vector $(-1, \mathbf{0})_{\widehat{\gamma}^*(b)} \in \mathcal{J}_b$ and the cone $\widehat{K}_b^\pm \subset \mathcal{K}_b^\pm$. Let $\widehat{\lambda}^*$ be the integral curve of $(\widehat{X}^{T^*})^{\{u^*\}}$ with initial condition $(b, \widehat{\lambda}_b)$.

As T_f is contained in $\mathcal{J}_b$, we have $\langle \widehat{\lambda}_b, \widehat{v} \rangle \geq 0$ for every $\widehat{v} \in T_f$. As T_f is a vector space, if $\widehat{v} \in T_f$, then $-\widehat{v} \in T_f$. Hence, we have

$$\langle \widehat{\lambda}_b, \widehat{v} \rangle = 0 \quad \text{for every} \ \widehat{v} \in T_f.$$

That is,

$$\langle \widehat{\lambda}_b, (0, v) \rangle = 0 \quad \text{for every} \ v \in T_{\gamma^*(b)}S_f.$$

This is equivalent to $\langle \lambda_b, v \rangle = 0$ for every $v \in T_{\gamma^*(b)}S_f$; that is, $\lambda_b = \lambda^*(b) \in T^*_{\gamma^*(b)}M$ is in the annihilator of $T_{\gamma^*(b)}S_f$ as wanted.

For every $\widehat{w}_b \in \mathcal{J}_b \subset T_{\widehat{\gamma}^*(b)}\widehat{M}$, if $\widehat{V} \colon I \to TM$ is the integral curve of $(\widehat{X}^T)^{\{u^*\}}$ with initial condition $(b, \widehat{w}_b)$ where $\widehat{w}_b \in T_{\widehat{\gamma}(b)}\widehat{M}$, then by Proposition 2.2.6 the continuous natural pairing function $\langle \widehat{\lambda}^*, \widehat{V} \rangle \colon I \to \mathbb{R}$ is constant everywhere and $\langle \widehat{\lambda}^*(a), \widehat{V}(a) \rangle \geq 0$ by Equation (4.4.15). As $(\Phi_{(b,a)}^{\widehat{X}^{\{u^*\}}})^{-1}_*(\mathcal{J}_b) = \mathcal{J}_a$ by the continuity and the linearity of the flow, the transversality condition at a is proved analogously as the transversality condition at b proved above.

Since $(\widehat{\gamma}^*, u^*)$ is a solution of the $\widehat{\mathrm{FOCP}}$, it is also a solution of $\widehat{\mathrm{OCP}}$ with time and endpoints fixed given by the curve. Hence, we can apply Pontryagin's Maximum Principle Theorem 4.1.14, for time and endpoints fixed. If the curve $(\widehat{\gamma}^*, u^*)$ is a solution of $\widehat{\mathrm{OCP}}$ with $I = [a, b]$ and endpoints $\widehat{\gamma}^*(a)$ and $\widehat{\gamma}^*(b)$, $(\widehat{\lambda}^*, u^*) \colon [a, b] \to T^*\widehat{M} \times U$ is a solution of the HP, such that $\widehat{\gamma}^* = \pi_{\widehat{M}} \circ \widehat{\lambda}^*$, and moreover $\widehat{\lambda}^*$ satisfies

(3a) $H\left(\widehat{\lambda}^*(t), u^*(t)\right) = \sup_{u \in U} H\left(\widehat{\lambda}^*(t), u\right)$ almost everywhere,

(3b) $\sup_{u \in U} H\left(\widehat{\lambda}^*(t), u\right)$ is constant everywhere,

(3c) $\widehat{\lambda}^*(t) \neq 0 \in T^*_{\widehat{\gamma}^*(t)}\widehat{M}$ for every $t \in [a, b]$, and

(3d) $\lambda_0^*(t)$ is constant, $\lambda_0^*(t) \le 0$.

Observe that it only remains to prove **(3b)** of the Free Pontryagin's Maximum Principle, since **(3a)**, **(3c)** and **(3d)** are the same in both Theorems 4.1.14, 4.3.13.

(3b) Due to (3a) we already know that the maximum of the Hamiltonian is constant everywhere along $(\widehat{\lambda}^*, u^*)$. Now, let us prove that it can be taken to be zero everywhere.

Take $\widehat{v}_b = \pm\widehat{X}(\widehat{\gamma}^*(b), u^*(b)) \in \widehat{K}_b^\pm \subset T_{\widehat{\gamma}^*(b)}M$. If $\widehat{V} \colon I \to T\widehat{M}$ is the integral curve of $(\widehat{X}^T)^{\{u^*\}}$ with initial condition $(b, \widehat{v}_b)$, then the continuous function $\langle\widehat{\lambda}^*, \widehat{V}\rangle \colon I \to \mathbb{R}$ is constant everywhere by Proposition 2.2.6. Thus,

$$\langle\widehat{\lambda}^*(t), \widehat{V}(t)\rangle = \langle\widehat{\lambda}^*(t), \pm\widehat{X}(\widehat{\gamma}^*(t), u^*(t))\rangle \le 0 \quad \text{for every} \ \ t \in I$$

by Equation (4.4.14), and this implies that

$$\langle\widehat{\lambda}^*(t), \widehat{X}(\widehat{\gamma}^*(t), u^*(t))\rangle = 0.$$

As $\langle\widehat{\lambda}^*(t), \widehat{X}(\widehat{\gamma}^*(t), u^*(t))\rangle = H(\widehat{\lambda}^*(t), u^*(t))$, the Hamiltonian function is zero everywhere and the maximum of the Hamiltonian function is zero everywhere by (3b) in Theorem 4.1.14.∎

Observe that the initial condition for the covector in this proof has been chosen such that the tangent spaces to the initial and final submanifolds are contained in the separating hyperplane defined by the covector. In this statement of the Maximum Principle the initial condition for the covector must satisfy more conditions (namely the transversality conditions) than in Theorem 4.1.14.

4.5   Abnormality, controllability and optimality

In control theory, the reachable sets are useful to determine the accessibility and the controllability of the systems as briefly commented in §3.2. In optimal control, the reachable set has great importance for distinguishing the abnormal optimal curves from the normal ones [Bullo and Lewis 2005b, Langerock 2003a]—see Definition 4.1.15 for the different optimal curves.

We are going to establish relationships between abnormality, controllability and optimality; always with the purpose of characterizing abnormality.

4.5.1   Generalization of the notion of linear controllable

The notion of linear controllability appears in Definition 3.2.7 for control–affine systems. We are going to extend this notion to any general control system having in mind the elementary perturbations of the controls described in §4.1.3.1 and the tangent perturbation cone in Definition 4.1.11.

Definition 4.5.1. *Let* $\pi \colon M \times U \to M$ *be a projection and* $(\gamma, u) \colon [a, b] \to M \times U$ *be a reference trajectory of a control system* $X \in \mathfrak{X}(\pi)$. *Let* $t_1 \in [a, b]$ *such that* $\gamma(t_1) = x$. *The*

*system is **linear controllable at** t_1 **along** γ if the tangent perturbation cone K_t is the entire tangent space $T_{\gamma(t)}M$ for each $t > t_1$.*

As stated in [Tyner 2007, Theorem 2.7], there exists an equivalent way to characterize the reachable set (3.2.2) of the linearized system of a control–affine system along a trajectory (γ, u); that is,

$$\mathcal{R}_{TM}(0_x, t) = \text{span}_{\mathbb{R}} \left\{ \bigcup_{\substack{a < \tau \leq t \\ v_\tau \in \{w^s f_s(\gamma(\tau)) \,|\, w \in U\}}} \Phi_{(\tau,t)}^{X_\gamma^T}(v_\tau) \right\}.$$

From here it is clear that Definition 4.5.1 coincides with the one of linear controllability in Definition 3.2.7 for control–affine systems.

Proposition 4.5.2. *Let $(\gamma, u)\colon I \to M \times U$ be an optimal curve of OCP in Statement 4.1.1.*

1. *If the optimal curve is abnormal, then the system is not linear controllable at $\gamma(t)$ for every time $t \in I$.*

2. *If the control system is linear controllable at $t_1 \in I$ along γ, then it is small–time locally controllable from $\gamma(t_1)$.*

(Proof)

1. If an optimal curve is abnormal, according to Definition 4.1.15 there exists a lift λ to the cotangent bundle satisfying the conditions in Theorem 4.1.14 with $p_0 = 0$. If so, the separation condition in Equation (4.2.9) is

$$\langle \lambda(t), K_t \rangle \leq 0.$$

   Assume that the system is linear controllable, then $K_t = T_{\gamma(t)}M$ and the inequality is satisfied only if λ is the zero momenta. But, because of Theorem 4.1.14, (p_0, λ) cannot vanish simultaneously.

   As a result of this contradiction, we have that if an optimal curve is abnormal, then the system is not linear controllable at $\gamma(t)$ for $t \in I$.

2. If the system is linear controllable at $\gamma(t)$, then $K_{\gamma(t)} = T_{\gamma(t)}M$. Thus, by Proposition 4.1.12, $\gamma(t)$ is in the interior of the reachable set; that is, the control system is small–time locally controllable from $\gamma(t)$.

$\blacksquare$

This proposition provides a necessary condition for abnormality: not to be linear controllable. A sufficient condition for not being linear controllable is not to be small–time locally controllable. At this point, the abnormality can be characterized using the results about sufficient conditions for small–time locally controllability in the literature of control systems, see for instance Theorem 3.3.2 and [Basto-Gonçalves 1998, Lewis and Murray 1997, Sussmann 1978; 1987].

4.5.2   The tangent perturbation cone as an approximation of the reachable set

The concept of linear controllable introduced in Definition 3.2.7, generalized in Definition 4.5.1 and used in Proposition 4.5.2 justifies the effort to understand the linear approximation of the reachable set given by the tangent perturbation cone. Moreover, a key point in the proof of Pontryagin's Maximum Principle that comes from Proposition 4.1.12 also depends on the understanding of that linear approximation of the reachable set in a neighborhood of a point in the optimal curve. This interpretation of the tangent perturbation cone has been studied by Agrachev [2002b], Agrachev and Sachkov [2004], but we will study it in a great and clear detail in this section.

In the sequel, we explain why this interpretation of the tangent perturbation cone is feasible. Remember from §2.2 that a time–dependent vector field on M has associated the evolution operator $\Phi^X \colon I \times I \times M \to M$, $(t, s, x) \mapsto \Phi^X(t, s, x)$ as defined in Equation (2.2.2).

Proposition 4.5.3. *Let X, Y be time–dependent vector fields on M, then there exists a time–dependent vector field Z such that*

$$\Phi^{X+Y}_{(t,s)}(x) = \left(\Phi^X_{(t,s)} \circ \Phi^Z_{(t,s)} \right)(x)$$

and $Z = \left(\Phi^X_{(t,s)} \right)^* Y = \left(\Phi^X_{(t,s)*} \right)^{-1} Y.$

(*Proof*) For any initial time s, we define the diffeomorphism $\tilde{\Phi}^X_s \colon I \times M \to I \times M$, $(t, x) \to (t, \Phi^X_s(t, x)) = (t, \Phi^X(t, s, x))$ such that $\tilde{\Phi}^X_s(s, x) = (s, x)$. We look for a time–dependent vector field Z on M such that

$$\tilde{\Phi}^{X+Y}_s(t, x) = (\tilde{\Phi}^X_s \circ \tilde{\Phi}^Z_s)(t, x) \, . \tag{4.5.16}$$

This expression has been assumed true in [Agrachev and Sachkov 2004, Bullo and Lewis 2000] for $s = 0$, but it has not been carefully proved. On the left–hand side of Equation (4.5.16) we have

$$\tilde{\Phi}^{X+Y}_s(t, x) = (t, \Phi^{X+Y}_s(t, x))$$

and the right–hand side is

$$(\tilde{\Phi}^X_s \circ \tilde{\Phi}^Z_s)(t, x) = \tilde{\Phi}^X_s(t, \Phi^Z_s(t, x)) = (t, \Phi^X_s(t, \Phi^Z_s(t, x))) \, .$$

Thus Equation (4.5.16) is satisfied if and only if

$$\Phi^{X+Y}_s(t, x) = \Phi^X_s(t, \Phi^Z_s(t, x)) \quad = \quad (\Phi^X_s \circ \tilde{\Phi}^Z_s)(t, x) \, , \tag{4.5.17}$$

or equivalently,

$$\Phi^{X+Y}_{(t,s)} \quad = \quad \Phi^X_{(t,s)} \circ \Phi^Z_{(t,s)} \, . \tag{4.5.18}$$

Let us differentiate with respect to t the left–hand side of Equation (4.5.17),

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{(s,x)}^{X+Y}(t) = (X+Y)(t, \Phi_{(s,x)}^{X+Y}(t)) = (X+Y)(t, \Phi_s^X(t, \Phi_s^Z(t,x))) . \qquad (4.5.19)$$

The differentiation with respect to time of the right–hand side of Equation (4.5.17), for f in $\mathcal{C}^\infty(M)$, is

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}(\Phi_s^X(t, \Phi_s^Z(t,x)))f &= \lim_{h\to 0} \frac{f((\Phi_s^X(t+h, \Phi_s^Z(t+h, x)))) - f((\Phi_s^X(t, \Phi_s^Z(t,x))))}{h} \\
&= \lim_{h\to 0} \left\{ \frac{(f \circ \Phi_{(t+h,s)}^X)(\Phi_{(t+h,s)}^Z(x)) - (f \circ \Phi_{(t+h,s)}^X)(\Phi_{(t,s)}^Z(x))}{h} \right. \\
&\quad + \left. \frac{(f \circ \Phi_s^X)(t+h, \Phi_s^Z(t,x)) - (f \circ \Phi_s^X)(t, \Phi_s^Z(t,x))}{h} \right\} \\
&= Z(t, \Phi_s^Z(t,x))(f \circ \Phi_{(t,s)}^X) + X(t, \Phi_s^X(t, \Phi_s^Z(t,x)))f \\
&= T_{\Phi_s^Z(t,x)}\Phi_{(t,s)}^X Z(t, \Phi_s^Z(t,x))f + X(t, \Phi_s^X(t, \Phi_s^Z(t,x)))f .
\end{aligned}
$$

Hence

$$\frac{\mathrm{d}}{\mathrm{d}t}(\Phi_s^X(t, \Phi_s^Z(t,x))) = T_{\Phi_s^Z(t,x)}\Phi_{(t,s)}^X Z(t, \Phi_s^Z(t,x)) + X(t, \Phi_s^X(t, \Phi_s^Z(t,x))) .$$

From Equation (4.5.19) we have

$$
\begin{aligned}
X(t, \Phi_s^X(t, \Phi_s^Z(t,x))) + Y(t, \Phi_s^X(t, \Phi_s^Z(t,x))) &= T_{\Phi_s^Z(t,x)}\Phi_{(t,s)}^X Z(t, \Phi_s^Z(t,x)) \\
&\quad + X(t, \Phi_s^X(t, \Phi_s^Z(t,x))) ,
\end{aligned}
$$

that is,

$$Y((\tilde{\Phi}_s^X \circ \tilde{\Phi}_s^Z)(t,x)) = T_{\Phi_s^Z(t,x)}\Phi_{(t,s)}^X Z(t, \Phi_s^Z(t,x)) .$$

Remember from §2.1 that the pushforward of a time–dependent vector field Z is another time–dependent vector field given by

$$(\Phi_{(t,s)*}^X Z)(t,x) = T_{(\Phi_{(t,s)}^X)^{-1}(x)}\Phi_{(t,s)}^X (Z(t, (\Phi_{(t,s)}^X)^{-1}(x))) .$$

Then

$$(Y \circ \tilde{\Phi}_s^X \circ \tilde{\Phi}_s^Z)(t,x) = (\Phi_{(t,s)*}^X Z)(t, \Phi_{(t,s)}^X(\Phi_s^Z(t,x)))$$

$$= (\Phi_{(t,s)*}^X Z)(\tilde{\Phi}_s^X(t, \Phi_s^Z(t,x))) = (\Phi_{(t,s)*}^X Z)(\tilde{\Phi}_s^X \circ \tilde{\Phi}_s^Z)(t,x) ,$$

or equivalently,

$$Y \circ \tilde{\Phi}_s^X \circ \tilde{\Phi}_s^Z = (\Phi_{(t,s)*}^X Z) \circ \tilde{\Phi}_s^X \circ \tilde{\Phi}_s^Z ,$$

that is, $Y = \Phi_{(t,s)*}^X Z$.

Hence $Z = (\Phi^X_{(t,s)*})^{-1}Y = (\Phi^X_{(t,s)})^*Y$. Now, going back to Equation (4.5.18) we have

$$\Phi^{X+Y}_{(t,s)}(x) = \left(\Phi^X_{(t,s)} \circ \Phi^{(\Phi^X_{(t,s)})^*Y}_{(t,s)}\right)(x) \,. \tag{4.5.20}$$

∎

Once we know how to express the flow of a sum of vector fields as a composition of flows of different vector fields, we are going to show that all the integral curves used to construct the reachable set in Definition 3.2.1 can be written as composition of flows associated with vector fields given by vectors in the tangent perturbation cone in Definition 4.1.11.

Each control system $X \in \mathfrak{X}(\pi)$ with the projection $\pi\colon M \times U \to M$ is a time–dependent vector field $X^{\{u\}}$ when the control is given. Consider the reference trajectory (γ, u) to be an integral curve of $X^{\{u\}}$ with initial condition x_0 at a. Take $\gamma(t_1)$ to be a reachable point from x_0 at time t_1. Let us consider another control $\tilde{u}\colon I \to U$ and the integral curve of $X^{\{\tilde{u}\}}$ with initial condition x_0 at a denoted by $\tilde{\gamma}$. Then $\tilde{\gamma}(t_1)$ is another reachable point from x_0 at time t_1.

Let us see how to reach the point $\tilde{\gamma}(t_1)$ using Equation (4.5.20),

$$
\begin{aligned}
\tilde{\gamma}(t_1) &= \Phi^{X^{\{\tilde{u}\}}}_{(t_1,a)}(x_0) = \Phi^{X^{\{u\}}+\left(X^{\{\tilde{u}\}}-X^{\{u\}}\right)}_{(t_1,a)}(x_0) \\
&= \left(\Phi^{X^{\{u\}}}_{(t_1,a)} \circ \Phi^{\left(\Phi^{X^{\{u\}}}_{(t_1,a)}\right)^*\left(X^{\{\tilde{u}\}}-X^{\{u\}}\right)}_{(t_1,a)}\right)(x_0) \\
&= \left(\Phi^{X^{\{u\}}}_{(t_1,a)} \circ \Phi^{\left(\Phi^{X^{\{u\}}}_{(t_1,a)}\right)^*\left(X^{\{\tilde{u}\}}-X^{\{u\}}\right)}_{(t_1,a)} \circ \left(\Phi^{X^{\{u\}}}_{(t_1,a)}\right)^{-1} \circ \Phi^{X^{\{u\}}}_{(t_1,a)}\right)(x_0) \\
&= \left(\Phi^{X^{\{u\}}}_{(t_1,a)} \circ \Phi^{\left(\Phi^{X^{\{u\}}}_{(t_1,a)}\right)^*\left(X^{\{\tilde{u}\}}-X^{\{u\}}\right)}_{(t_1,a)} \circ \left(\Phi^{X^{\{u\}}}_{(t_1,a)}\right)^{-1}\right)(\gamma(t_1)) \,.
\end{aligned}
\tag{4.5.21}
$$

Hence, from $\gamma(t_1)$ we can get every reachable point from x_0 at time t_1 through Equation (4.5.21) composing integral curves of the vector fields $X^{\{u\}}$ and $(\Phi^{X^{\{u\}}}_{(t_1,a)})^*(X^{\{\tilde{u}\}} - X^{\{u\}})\colon I \times M \to TM$, this latter with initial condition x_0 at a.

In fact this is true for any time τ in $[a, t_1]$, that is,

$$\tilde{\gamma}(\tau) = \left(\Phi^{X^{\{u\}}}_{(\tau,a)} \circ \Phi^{\left(\Phi^{X^{\{u\}}}_{(\tau,a)}\right)^*\left(X^{\{\tilde{u}\}}-X^{\{u\}}\right)}_{(\tau,a)} \circ \left(\Phi^{X^{\{u\}}}_{(\tau,a)}\right)^{-1}\right)(\gamma(\tau)) \,.$$

If we compose with the flow of $X^{\{u\}}$, we get a reachable point from x_0 at time t_1 because it is a concatenation of integral curves of the dynamical system,

$$
\begin{aligned}
\Phi^{X^{\{u\}}}_{(t_1,\tau)}(\tilde{\gamma}(\tau)) &= \left(\Phi^{X^{\{u\}}}_{(t_1,\tau)} \circ \Phi^{X^{\{u\}}}_{(\tau,a)} \circ \Phi^{\left(\Phi^{X^{\{u\}}}_{(\tau,a)}\right)^*\left(X^{\{\tilde{u}\}}-X^{\{u\}}\right)}_{(\tau,a)} \circ \left(\Phi^{X^{\{u\}}}_{(\tau,a)}\right)^{-1}\right)(\gamma(\tau)) \\
&= \left(\Phi^{X^{\{u\}}}_{(t_1,a)} \circ \Phi^{\left(\Phi^{X^{\{u\}}}_{(\tau,a)}\right)^*\left(X^{\{\tilde{u}\}}-X^{\{u\}}\right)}_{(\tau,a)} \circ \left(\Phi^{X^{\{u\}}}_{(t_1,a)}\right)^{-1}\right)(\gamma(t_1)) \,.
\end{aligned}
\tag{4.5.22}
$$

Hence, from $\gamma(t_1)$ we can also get reachable points from x_0 at time t_1 through composition of

integral curves of the vector fields $X^{\{u\}}$ and $\left(\Phi_{(\tau,a)}^{X^{\{u\}}}\right)^* \left(X^{\{\tilde{u}\}} - X^{\{u\}}\right)$, the latter with initial condition $\gamma(a)$ at time a.

On the other hand, the tangent perturbation cone at $\gamma(t_1)$ is given by the closure of the convex hull of all the tangent vectors $\left(\Phi_{(t_1,\tau)}^{X^{\{u\}}}\right)_* \left(X^{\{\tilde{u}\}}(\tau, \gamma(\tau)) - X^{\{u\}}(\tau, \gamma(\tau))\right)$ for every Lebesgue time τ in $[a, t_1]$. These vectors are related with the vector fields $X^{\{u\}}$ through Equations (4.5.21) and (4.5.22).

In this sense, we say that the tangent perturbation cone at $\gamma(t_1)$ is an approximation of the reachable set in a neighborhood of $\gamma(t_1)$.

4.6 Examples

Now some examples are given to make clear how to use Pontryagin's Maximum Principle as tool to find optimal solutions. First the Dubins car problem is studied in §4.6.1, see [Agrachev and Sachkov 2004, Anisi 2003, Sussmann and Tang 1991] for more details, in particular about the fact that not every extremal is optimal. Second, an example of a strict abnormal minimizer is given in [Bullo and Lewis 2005b] and it is used in §4.6.2 to show graphically what the separation conditions look like.

4.6.1 Dubins car

The time–optimal control problem posed here is considered from two different viewpoints: the variational calculus and Pontryagin's Maximum Principle.

Statement 4.6.1. *Let $M = \mathbb{R}^2 \times [-\pi/2, 3\pi/2)$ and $U = [-1, 1]$. Find the final time b and an integral curve $(\gamma, u) \colon I \to M \times U$ of*

$$
\left\{
\begin{array}{rcl}
\dot{x}^1 & = & \cos\theta, \\
\dot{x}^2 & = & \sin\theta, \\
\dot{\theta} & = & u,
\end{array}
\right.
$$

with local coordinates (x^1, x^2, θ) satisfying the endpoint conditions $\gamma(0) = (0, 0, \pi/2)$ and $\gamma(b) = (2, 2, 3\pi/2)$ and minimizing $\int_0^b dt$.

The problem consists of minimizing the time to go from an initial point to a final one with constant linear velocity (equal to 1). Observe that the control changes the angular velocity; that is, the direction of the car. It is impossible to change the direction of the car without moving it, since the linear velocity is always a nonzero constant. This is the simplest model for a car as appears in Figure 4.5.

Variational calculus: vakonomic

From a vakonomic viewpoint [Cortés et al. 2002a, Gràcia et al. 2003], the Lagrangian

$$
\begin{array}{rcl}
L \colon T(M \times U \times \mathbb{R}^3) & \longmapsto & \mathbb{R} \\
(x^1, x^2, \theta, u, \lambda_1, \lambda_2, \lambda_3, \dot{x}^1, \dot{x}^2, \dot{\theta}, \dot{u}, \dot{\lambda}_1, \dot{\lambda}_2, \dot{\lambda}_3) & \longrightarrow & 1 + \lambda_1(\dot{x}^1 - \cos\theta) \\
& & + \lambda_2\left(\dot{x}^2 - \sin\theta\right) + \lambda_3(\dot{\theta} - u)
\end{array}
$$

defines the Euler–Lagrange equations whose extremal curves are given by

$$\begin{cases} x^1(t) & = & (t - t_0)\cos\theta_0 + x_0^1, \\ x^2(t) & = & (t - t_0)\sin\theta_0 + x_0^2, \\ \theta(t) & = & \theta_0, \end{cases}$$

if we assume u to be constant. Thus the car moves along straight lines. Imposing the initial condition $\gamma(0) = (0, 0, \pi/2)$ we have

$$\begin{cases} x^1(t) & = & t\ \cos(\pi/2) = 0, \\ x^2(t) & = & t\ \sin(\pi/2) = t, \\ \theta(t) & = & \pi/2. \end{cases}$$

It is impossible to fulfill the endpoint conditions along this line. The only option would be concatenate different straight lines, but θ would not be a continuous function neither an absolutely continuous function, as the extremals are assumed to be in the optimal control problems considered along this chapter.

The conclusion is that calculus of variations do not show the solutions for this problem since the control set is closed. That is why Pontryagin's Maximum Principle is the suitable tool to solve this problem as explained in the following paragraph.
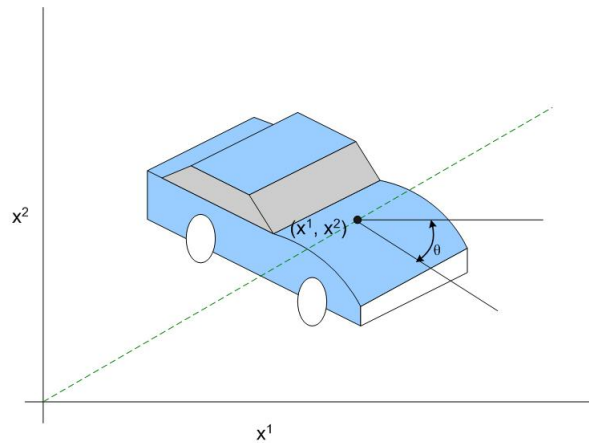
Figure 4.5: *Dubins car.*

Pontryagin's Maximum Principle

The result obtained to solve the problem is different when Pontryagin's Maximum Principle is used. Let us extend the system to $\widehat{M} = \mathbb{R}^3 \times [-\pi/2, 3\pi/2)$. For local coordinates, (x^0, x^1, x^2, θ), the equations of motion are

$$\begin{cases} \dot{x}^0 & = & 1, \\ \dot{x}^1 & = & \cos\theta, \\ \dot{x}^2 & = & \sin\theta, \\ \dot{\theta} & = & u. \end{cases}$$

We are looking for an integral curve $(\widehat{\gamma}, u) \colon I \to \widehat{M} \times U$ of

$$\widehat{X} = \frac{\partial}{\partial x^0} + \cos\theta \frac{\partial}{\partial x^1} + \sin\theta \frac{\partial}{\partial x^2} + u \frac{\partial}{\partial \theta}$$

satisfying the endpoint conditions $\widehat{\gamma}(0) = (0, 0, 0, \pi/2)$ and $\gamma(b) = (2, 2, 3\pi/2)$, minimizing $x^0 \circ \widehat{\gamma}$.

Pontryagin's Hamiltonian $H \colon T^*\widehat{M} \times U \to \mathbb{R}$ is

$$H(x^0, x^1, x^2, \theta, p_0, p_1, p_2, p_3, u) = p_0 + p_1 \cos\theta + p_2 \sin\theta + p_3 u.$$

An extremal $(\widehat{\gamma}, u) \colon [t_0, b_\gamma] \to \widehat{M} \times U$ must satisfy Hamilton's equations,

$$
\begin{aligned}
\dot{x}^0 &= 1, & \dot{p}_0 &= 0, \\
\dot{x}^1 &= \cos\theta, & \dot{p}_1 &= 0, \\
\dot{x}^2 &= \sin\theta, & \dot{p}_2 &= 0, \\
\dot{\theta} &= u, & \dot{p}_3 &= p_1 \sin\theta - p_2 \cos\theta,
\end{aligned}
$$

and maximize the Hamiltonian over the control, and be such that the Hamiltonian vanishes almost everywhere, Theorem 4.3.13. Due to the expression of the Hamiltonian and the control set, the control maximizing the Hamiltonian is

$$u = \begin{cases} 1, & \text{if } p_3 > 0, \\ \text{any value}, & \text{if } p_3 = 0, \\ -1, & \text{if } p_3 < 0. \end{cases}$$

When $p_3 = 0$, Hamilton's equations implies that θ must be constant and thus $u = 0$. Thus the control is always a constant function equal to -1, 0 or 1. Moreover, From the equations of motion, p_0, p_1 and p_2 are constant. That is why in the following paragraphs we do not write these functions.

Control u $= \pm 1$

The integral curve and the momenta are

$$
\begin{cases}
x^1(t) &= u \sin(u\,(t - t_0) + \theta_0) - u \sin\theta_0 + x_0^1, \\
x^2(t) &= -u \cos(u\,(t - t_0) + \theta_0) + u \cos\theta_0 + x_0^2, \\
\theta(t) &= u(t - t_0) + \theta_0, \\
p_3(t) &= p_1(-u \cos(u\,(t - t_0) + \theta_0) + u \cos\theta_0) - p_2\,(u \sin(u\,(t - t_0) + \theta_0) \\
&\quad - u \sin\theta_0) + p_3(t_0),
\end{cases}
$$

taking into account that $u = 1/u$ because $u = \pm 1$. As $u^2 = 1$, the Hamiltonian along these extremals is

$$
\begin{aligned}
H &= p_0 + p_1 \cos\theta + p_2 \sin\theta + [p_1(-u \cos\theta + u \cos\theta_0) - p_2(u \sin\theta - u \sin\theta_0) \\
&\quad + p_3(t_0)]\,u = p_0 + p_1 \cos\theta_0 + p_2 \sin\theta_0 + p_3(t_0)u = 0.
\end{aligned}
$$

The extremals can be

- abnormal and the associated momenta are $p_0 = 0$, $p_1 \cos \theta_0 + p_2 \sin \theta_0 + p_3(t_0)u = 0$;

- normal and the corresponding momenta are $p_0 = -1$, $p_1 \cos \theta_0 + p_2 \sin \theta_0 + p_3(t_0)u = 1$.

Observe that not all the initial conditions for the momenta can be chosen independently. The necessary conditions of Pontryagin's Maximum Principle can restrict the momenta, as happens in this case.

Control u=0

The integral curve for Hamilton's equations are

$$
\begin{cases}
x^1(t) &= (t - t_0) \cos \theta_0 + x_0^1, \\
x^2(t) &= (t - t_0) \sin \theta_0 + x_0^2, \\
\theta(t) &= \theta_0, \\
p_3(t) &= [p_1 \sin \theta_0 - p_2 \cos \theta_0](t - t_0) + p_3(t_0).
\end{cases}
$$

As $u = 0$, the Hamiltonian along the extremal is

$$
H = p_0 + p_1 \cos \theta_0 + p_2 \sin \theta_0 = 0.
$$

The extremals may be

- abnormal and the momenta is $p_0 = 0$, $p_1 = -\sin \theta_0$, $p_2 = \cos \theta_0$, $p_3(t) = -(t - t_0) + p_3(t_0)$;

- normal and then $p_0 = -1$, $p_1 = \cos \theta_0$, $p_2 = \sin \theta_0$, $p_3(t) = p_3(t_0)$.

Possible trajectories

The above assumptions show that the extremals on $\mathbb{R}^2$ are concatenations of arc of circles of radius 1 with center $(x_0^1 - u \sin \theta_0, x_0^2 + u \cos \theta_0)$ if $u = \pm 1$ and straight lines if $u = 0$. It must be easily proved that the endpoint conditions cannot be satisfied using only one of the trajectories. According to Agrachev and Sachkov [2004], the straight lines goes from an arc of circumference to another one tangent to both circumferences, so the trajectory is differentiable everywhere. This property is used to compute the trajectory in Figure 4.7. Moreover, the extremal curves do not have more than three different subtrajectories.

For the problem under study, there are three possible extremals satisfying all the necessary conditions and guaranteeing the absolutely continuity of the curves on $T^*\widehat{M}$.

In Figure 4.6 we have:

1. From $(0, 0, \pi/2)$ at time 0 to $(x^1(t_1), x^2(t_1), \theta(t_1))$ with control $u = -1$:

$$
\begin{aligned}
x^1(t) &= -\sin(-t + \pi/2) + 1, \\
x^2(t) &= \cos(-t + \pi/2), \\
\theta(t) &= -t + \pi/2, \\
p_3(t) &= p_1 \cos(-t + \pi/2) - p_2(-\sin(-t + \pi/2) + 1) + p_3(0),
\end{aligned}
$$

where p_1, p_2 are constants and $p_3(t) \leq 0$ for $t \in [0, t_1]$.

The abnormal lift must satisfy $p_0 = 0$ and $p_2 - p_3(0) = 0$.

The normal lift must satisfy $p_0 = -1$ and $p_2 - p_3(0) = 1$.

2. From $(x^1(t_1), x^2(t_1), \theta(t_1))$ to $(x^1(t_2), x^2(t_2), \theta(t_2))$ following a trajectory with control $u = 0$:

$$
\begin{aligned}
x^1(t) &= (t - t_1)\cos(-t_1 + \pi/2) - \sin(-t_1 + \pi/2) + 1, \\
x^2(t) &= (t - t_1)\sin(-t_1 + \pi/2) + \cos(-t_1 + \pi/2), \\
\theta(t) &= -t_1 + \pi/2, \\
p_3(t) &= (p_1 \sin(-t_1 + \pi/2) - p_2 \cos(-t_1 + \pi/2))(t - t_1) + p_3(t_1),
\end{aligned}
$$

where p_1, p_2 are constants and $p_3(t) = 0$ for $t \in [t_1, t_2]$.

The abnormal lift must satisfy $p_0 = 0$, $p_1 = -\sin(-t_1 + \pi/2)$, $p_2 = \cos(-t_1 + \pi/2)$. But then $p_3(t) = -t + t_1 + p_3(t_1)$ and it does not vanish for $t \in [t_1, t_2]$. Thus the abnormal lift is not possible.

The normal lift must satisfy $p_0 = -1$, $p_1 = \cos(-t_1 + \pi/2)$, $p_2 = \sin(-t_1 + \pi/2)$. Then $p_3(t) = p_3(t_1) = 0$. By the absolute continuity of the momenta curve, the normal lift on $[0, t_1]$ must be

$$
\begin{aligned}
p_1(t) &= \cos(-t_1 + \pi/2), \\
p_2(t) &= \sin(-t_1 + \pi/2), \\
p_3(t) &= \cos(t - t_1) - 1.
\end{aligned}
$$

Observe that $p_3(t) \leq 0$ for any $t \in \mathbb{R}$. It must be a change of the control at t_1 because $p_3(t_1) = 0$.

3. From $(x^1(t_2), x^2(t_2), \theta(t_2))$ to $(2, 2, 3\pi/2)$ at time b following a trajectory with control $u = 1$:

$$
\begin{aligned}
x^1(t) &= \sin(t - t_2 - t_1 + \pi/2) - 2\sin(-t_1 + \pi/2) + (t_2 - t_1)\cos(-t_1 + \pi/2) \\
&\quad + 1, \\
x^2(t) &= -\cos(t - t_2 - t_1 + \pi/2) + 2\cos(-t_1 + \pi/2) \\
&\quad + (t_2 - t_1)\sin(-t_1 + \pi/2), \\
\theta(t) &= t - t_2 - t_1 + \pi/2, \\
p_3(t) &= -\cos(-t + t_2) + 1,
\end{aligned}
$$

where $p_3(t) \geq 0$ for $t \in [t_2, b]$ and taking the constant values for p_1 and p_2 obtained on $[t_1, t_2]$.

The trajectory must satisfy the final condition. Then,

$$
3\pi/2 = \theta(b) = b - t_2 - t_1 + \pi/2 \Rightarrow b = t_1 + t_2 + \pi.
$$

As $x^1(b) = x^2(b) = 2$, we have a system of two equations to find t_1 and t_2. As the dependence on t_2 is affine, it is easy to prove that

$$
t_2 = t_1 + 2\cos(-t_1 + \pi/2) + 2\sin(-t_1 + \pi/2).
$$

A solution of the system is given by $t_1 = \pi/2$ and $t_2 = 2 + \pi/2$. So the final time is $2 + 2\pi$.

The abnormal lift must satisfy $p_0 = 0$, $p_3(t_2) = 1$, assuming that we connect with the normal lift on $[t_1, t_2]$ in such a way that we get an absolutely continuous momenta.

The normal lift must satisfy $p_0 = -1$, $p_3(t_2) = 0$, assuming that the momenta curve is absolutely continuous on $[0, b]$.

Thus, the extremal is normal on $[0, 2+2\pi]$ and also abnormal on $[0, \pi/2] \cup [2+\pi/2, 2+2\pi]$.
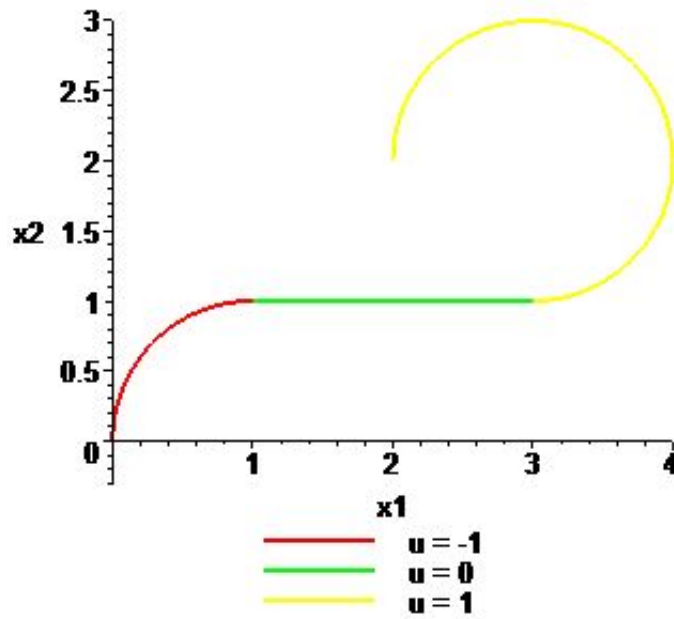
Figure 4.6: *The value of the functional for this trajectory is $2 + 2\pi$ seconds.*

In Figure 4.7 we have:

1. From $(0, 0, \pi/2)$ at time 0 to $(x^1(t_1), x^2(t_1), \theta(t_1))$ with control $u = 1$:

$$\begin{aligned}
x^1(t) &= \sin(t + \pi/2) - 1, \\
x^2(t) &= -\cos(t + \pi/2), \\
\theta(t) &= t + \pi/2, \\
p_3(t) &= -p_1 \cos(t + \pi/2) - p_2(\sin(t + \pi/2) - 1) + p_3(0),
\end{aligned}$$

where p_1, p_2 are constants and $p_3(t) \geq 0$ for $t \in [0, t_1]$.

The abnormal lift must satisfy $p_0 = 0$ and $p_2 + p_3(0) = 0$.

The normal lift must satisfy $p_0 = -1$ and $p_2 + p_3(0) = 1$.

2. From $(x^1(t_1), x^2(t_1), \theta(t_1))$ to $(x^1(t_2), x^2(t_2), \theta(t_2))$ following a trajectory with control

$u = 0$:

$$
\begin{aligned}
x^1(t) &= (t - t_1)\cos(t_1 + \pi/2) + \sin(t_1 + \pi/2) - 1, \\
x^2(t) &= (t - t_1)\sin(t_1 + \pi/2) - \cos(t_1 + \pi/2), \\
\theta(t) &= t_1 + \pi/2, \\
p_3(t) &= (p_1 \sin(t_1 + \pi/2) - p_2 \cos(t_1 + \pi/2))(t - t_1) + p_3(t_1),
\end{aligned}
$$

where p_1, p_2 are constants and $p_3(t) = 0$ for $t \in [t_1, t_2]$.

The abnormal lift must satisfy $p_0 = 0$, $p_1 = -\sin(t_1 + \pi/2)$, $p_2 = \cos(t_1 + \pi/2)$. But then $p_3(t) = -t + t_1 + p_3(t_1)$ and it does not vanish for $t \in [t_1, t_2]$. Thus the abnormal lift is not possible.

The normal lift must satisfy $p_0 = -1$, $p_1 = \cos(t_1 + \pi/2)$, $p_2 = \sin(t_1 + \pi/2)$. Then $p_3(t) = p_3(t_1) = 0$. By the absolute continuity of the momenta curve, the normal lift on $[0, t_1]$ must be

$$
\begin{aligned}
p_1(t) &= \cos(t_1 + \pi/2), \\
p_2(t) &= \sin(t_1 + \pi/2), \\
p_3(t) &= -\cos(-t + t_1) + 1.
\end{aligned}
$$

Observe that $p_3(t) \geq 0$ for any $t \in \mathbb{R}$. A change of the control might happen at t_1 because $p_3(t_1) = 0$.

3. From $(x^1(t_2), x^2(t_2), \theta(t_2))$ to $(2, 2, 3\pi/2 + 2\pi)$ at time b with control $u = 1$ identifying $3\pi/2$ with $3\pi/2 + 2\pi$:

$$
\begin{aligned}
x^1(t) &= \sin(t - t_2 + t_1 + \pi/2) + (t_2 - t_1)\cos(t_1 + \pi/2) - 1, \\
x^2(t) &= -\cos(t - t_2 + t_1 + \pi/2) + (t_2 - t_1)\sin(t_1 + \pi/2), \\
\theta(t) &= t - t_2 + t_1 + \pi/2, \\
p_3(t) &= -\cos(-t + t_2) + 1,
\end{aligned}
$$

where $p_3(t) \geq 0$ for $t \in [t_2, b]$ and taking the constant values for p_1 and p_2 obtained on $[t_1, t_2]$.

The trajectory must satisfy the final condition. Then,

$$
3\pi/2 + 2\pi = \theta(b) = b - t_2 + t_1 + \pi/2 \Rightarrow b = t_2 - t_1 + 3\pi.
$$

As $x^1(b) = x^2(b) = 2$, we have a system of two equations to find t_1 and t_2. As the dependence on t_2 is affine, it is easy to prove that

$$
t_2 = t_1 + 4\cos(t_1 + \pi/2) + 2\sin(t_1 + \pi/2).
$$

Numerically, the approximate solution of the system that gives a smooth switch is $t_1 \simeq 5.15$ and $t_2 \simeq 9.62$. Then the approximate final time is 13.89.

The abnormal lift must satisfy $p_0 = 0$, $p_3(t_2) = -1$, assuming that we connect with the normal lift on $[t_1, t_2]$ in such a way that we get an absolutely continuous momenta.

The normal lift must satisfy $p_0 = -1$, $p_3(t_2) = 0$, assuming that the momenta curve is absolutely continuous on $[0, b]$.

Thus, the extremal is normal on $[0, 13.89]$ and also abnormal on $[0, 5.15] \cup [9.62, 13.89]$.
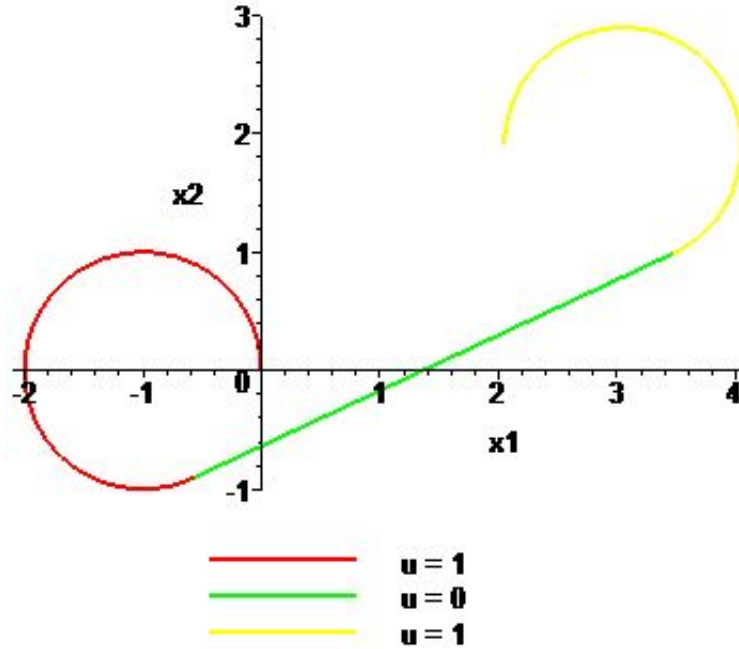
Figure 4.7: *The value of the functional for this trajectory is 13.89 seconds.*

In Figure 4.8 we have:

1. From $(0, 0, \pi/2)$ at time 0 to $(x^1(t_1), x^2(t_1), \theta(t_1))$ with a trajectory with control $u = 0$:

$$
\begin{aligned}
x^1(t) &= 0, \\
x^2(t) &= t, \\
\theta(t) &= \pi/2, \\
p_3(t) &= p_1 t + p_3(0),
\end{aligned}
$$

where p_1, p_2 is constant and $p_3(t) = 0$ for $t \in [0, t_1]$.

The abnormal lift must satisfy $p_0 = 0$, $p_1 = -1$ and $p_2 = 0$. But then $p_3(t) \neq 0$. Thus there is no abnormal lift.

The normal lift must satisfy $p_0 = -1$, $p_1 = 0$ and $p_2 = 1$. Then $p_3(t) \equiv 0$ for $t \in [0, t_1]$.

2. From $(0, x^2(t_1), \pi/2)$ at time t_1 to $(2, 2, -\pi/2)$ at time b with control $u = -1$ identifying the angle $3\pi/2$ with $-\pi/2$:

$$
\begin{aligned}
x^1(t) &= -\sin(-t + t_1 + \pi/2) + 1, \\
x^2(t) &= \cos(-t + t_1 + \pi/2) + t_1, \\
\theta(t) &= -t + t_1 + \pi/2, \\
p_3(t) &= p_1 \cos(-t + t_1 + \pi/2) - p_2(-\sin(-t + t_1 + \pi/2) + 1),
\end{aligned}
$$

where p_1 and p_2 are constants, $p_3(t_1) = 0$ and $p_3(t) \leq 0$ for $t \in [t_1, b]$.

The trajectory must satisfy the final condition. Then,

$$-\pi/2 = \theta(b) = -b + t_1 + \pi/2 \Rightarrow b = t_1 + \pi.$$

From $x^2(b) = 2$ we obtain $t_1 = 2$. Thus the point where the control switches is $(0, 2, \pi/2)$.

The abnormal lift must satisfy $p_0 = 0$, $p_2 - p_3(t_1) = 0$. As $p_3(t_1) = 0$, we have $p_2 = 0$. There exists an abnormal lift on $[2, 2 + \pi]$ as long as $p_1 \neq 0$.

The normal lift must satisfy $p_0 = -1$, $p_2 - p_3(t_1) = 1$. As $p_3(t_1) = 0$, we have $p_2 = 1$. By the assumption of absolute continuity of the momenta curve, $p_1 = 0$.

This third trajectory is normal on $[0, 2 + \pi]$ and abnormal only on $[2, 2 + \pi]$.
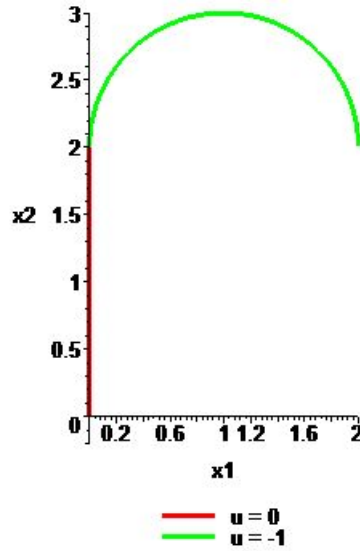
Figure 4.8: *The value of the functional for this trajectory is $2 + \pi$ seconds.*

Among the three trajectories studied, the optimal is the third because it gives us the smallest time to arrive from the given point to the other one. However, all of them satisfy the necessary conditions of the Pontryagin's Maximum Principle. Moreover, observe that all the trajectories are strict normal in the interval of time that $u = 0$ and are normal and abnormal when $u = \pm 1$.

4.6.2 A strict abnormal optimal solution

Consider now a linear oscillator on a manifold $M = \mathbb{R}^2$ with control set $U = [-1, 1] \subset \mathbb{R}$; see [Bullo and Lewis 2005b]. The control system is given by

$$\begin{cases} \dot{x}^1 = x^2, \\ \dot{x}^2 = -x^1 + u \end{cases} \tag{4.6.23}$$

and the final time, the integral curve and the control must be found such that the endpoint conditions given by $x(0) = (0,0)$ and $x(b) = (2,0)$ are satisfied and the time is minimized.

Considering the extended system, Pontryagin's Hamiltonian is

$$H(x^0, x^1, x^2, p_0, p_1, p_2, u) = p_0 + p_1 x^2 + p_2(-x^1 + u)$$

and Hamilton's equations are

$$
\begin{aligned}
\dot{x}^0 &= 1, & \dot{p}_0 &= 0, \\
\dot{x}^1 &= x^2, & \dot{p}_1 &= p_2, \\
\dot{x}^2 &= -x^1 + u, & \dot{p}_2 &= -p_1.
\end{aligned}
$$

Assuming the control to be constant, the momenta are

$$
\begin{aligned}
p_1(t) &= -A\cos t + B\sin t, \\
p_2(t) &= A\sin t + B\cos t.
\end{aligned}
$$

The maximum of the Hamiltonian is obtained when $u = \mathrm{sgn}(p_2)$. Observe that if $u = 0$ the fulfilment of the endpoint condition at time 0 by the integral curves of the control system (4.6.23) gives only constant curves. Hence, the options are

- **u = 1**: $(x^1(t), x^2(t)) = (-\cos t + 1, \sin t)$. The Hamiltonian along the extremal is $H = B + p_0 = 0$. The normal momenta is $(p_1(t), p_2(t)) = (-A\cos t + \sin t, A\sin t + \cos t)$ and the abnormal one is $(p_1(t), p_2(t)) = (-A\cos t, A\sin t)$.

- **u = −1**: $(x^1(t), x^2(t)) = (\cos t - 1, -\sin t)$. The Hamiltonian along the extremal is $H = -B + p_0 = 0$. The normal momenta is $(-A\cos t - \sin t, A\sin t - \cos t)$ and the abnormal one is $(-A\cos t, A\sin t)$.

Observe that both trajectories are normal and abnormal at the same time.

The extremals for this problem will be obtained as concatenation of integral curves with control equal to 1 and integral curves with control equal to -1, alternatively. A curve is followed until the momentum corresponding with x^2 vanishes since the sign of this momentum determines the value of the controls. The momenta must be absolutely continuous, therefore continuous on the domain. With this in mind, the reachable set from the origin up to time π is shown in Figures 4.9 and 4.10, assuming $A = 1$ for the integral curves starting at $(0,0)$. When the control switches, the initial condition for the new momenta is given such that the momenta are continuous.

In Figure 4.9 the normal lift of the integral curves of (4.6.23) has been considered. Only the pieces of the curves in magenta colour satisfy that $u = \mathrm{sgn}(p_2)$. Thus they are extremals for the optimal control problem associated to (4.6.23). All the curves in the figure are in the reachable set and have associated a curve in the cotangent bundle in a normal way, but not all of them are extremals. From Figure 4.9, it can be observed that there is not a normal extremal that goes from $(0,0)$ to $(2,0)$ in time π.
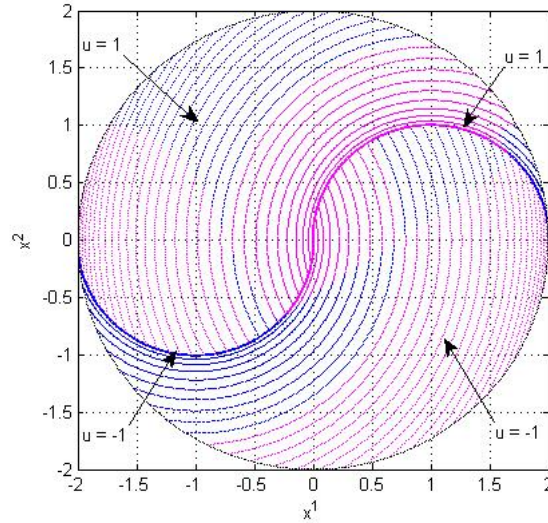
Figure 4.9: *Reachable set from $(0,0)$ up to π given by curves that have been associated with a normal lift*

In Figure 4.10, the curves in the reachable set have associated an abnormal lift. The pieces of curves in magenta colour are extremals because they satisfy $u = \text{sgn}(p_2)$. Observe that only the integral curves for $u = 1$ are abnormal extremals, the remaining curves are not extremals because $u \neq \text{sgn}(p_2)$. In particular, the integral curve from $(0,0)$ to $(2,0)$ is an abnormal extremal and it is the minimizer. In order to gain a better understanding of the lifts of the curves to the cotangent bundle, let us represent the separation condition for the optimal curve when it is considered as normal and as abnormal. In order to do this, we consider the set of admissible velocities given by the motion equations $\widehat{X}(x^1, x^2, u) = (1, x^2, -x^1 + u)$ with u taking constant values in $[-1, 1]$.

The tangent perturbation cone is constructed from the elementary perturbation vectors which are, by Definition 4.1.5,

$$\widehat{v}[\pi_1] = \left[\widehat{X}(\gamma(t_1), u_1) - \widehat{X}(\gamma(t_1), u(t_1)) \right] l_1 = (0, 0, u_1 - u)l_1,$$

where $l_1 \in \mathbb{R}^+$, $u_1 \in [-1, 1]$. It is necessary to consider the transport of these elementary perturbation vectors through the flow of the complete lift of $\widehat{X}$ and also the transport of the $\pm \widehat{X}(\gamma(t_1), 1)$ to construct the tangent perturbation cone in Definition 4.3.7; see §2.2.2.1. As the system is autonomous, we can always take the initial time to be zero. The result of the transport of the elementary perturbation vectors is

$$\widehat{V}[\pi_1] = (0, -D \sin t, u_1 - u)l_1, \quad D \in \mathbb{R}$$

If we follow the curve

$$\gamma(t) = (-\cos t + 1, \sin t)$$

from $(0,0)$ to $(1,1)$ in a normal way, then at time $t = 1$ the tangent perturbation cone and the
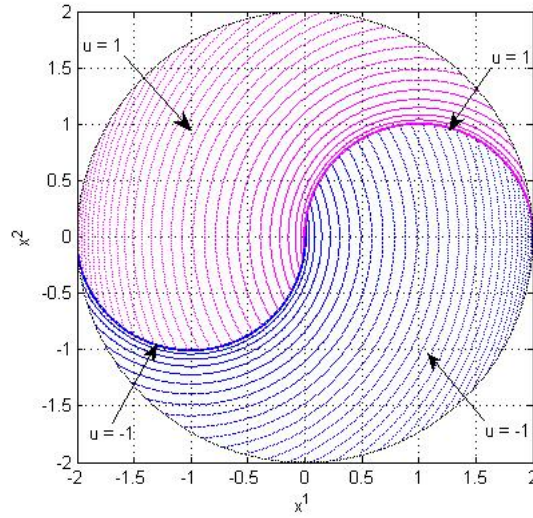
Figure 4.10: *Reachable set from $(0,0)$ up to time π given by curves that have been associated with an abnormal lift*

admissible velocities lie in the same half–space defined by the kernel of the following extended normal momenta for $A = 1$

$$\widehat{\lambda}(t) = (-1, -\cos t + \sin t, \sin t + \cos t).$$

That can be seen in Figure 4.11. However, the decreasing direction of the cost function in
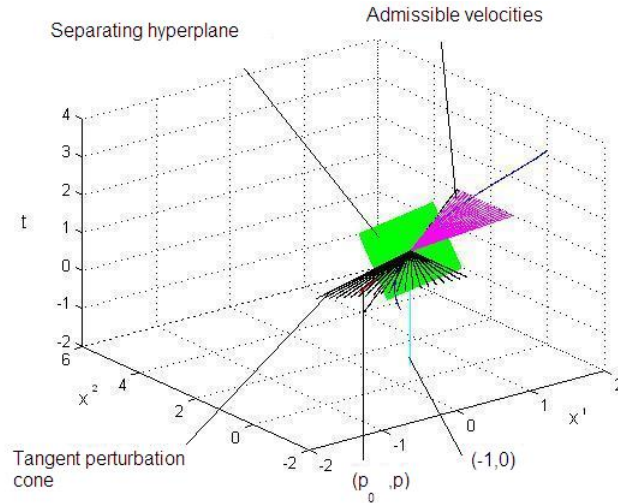
Figure 4.11: *Separation condition at $t = 1$ for the normal momenta.*

light blue colour in Figure 4.11 appears in the other half–space defined by the extended normal momentum. This direction is not contained in the separating hyperplane. To gather everything together we are identifying $\widehat{M}$, $T_{\widehat{\gamma}(1)}\widehat{M}$ and $T^*_{\widehat{\gamma}(1)}\widehat{M}$ with $\mathbb{R}^3$.

When the same trajectory is followed in an abnormal way using the abnormal extended momentum for $A = 1$

$$\widehat{\lambda}(t) = (0, -\cos t - \sin t, \sin t - \cos t),$$

the decreasing direction of the cost function is contained in the separating hyperplane as mentioned at the end of §4.2 and shown in Figure 4.12. Thus, the separation condition for $(-1, \mathbf{0})$ in (4.2.9) is satisfied with equality.
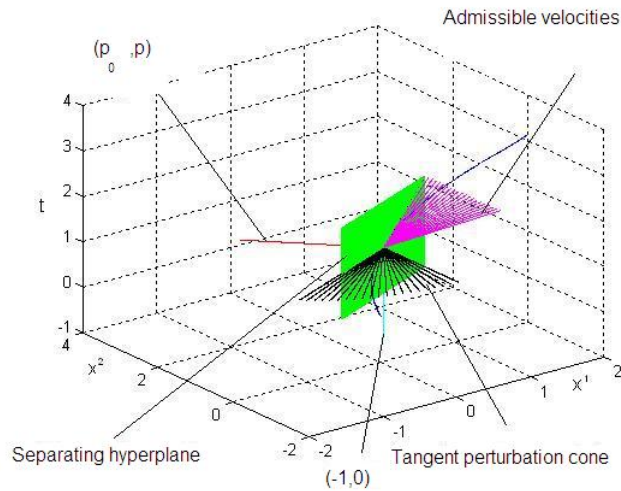
Figure 4.12: *Separation condition at $t = 1$ for the abnormal momenta.*

Now, if we evaluate the trajectories at time $t = 3$, then the normal momenta does not satisfy the separation condition (4.2.9) any more because the tangent perturbation cone is not contained in a half–space defined by the kernel of the extended momenta at $t = 3$, as shown in Figure 4.13. This implies the momenta p_2 has changed the sign and we should have been switched the control to $u = -1$ in order to follow a trajectory satisfying the necessary conditions of FPMP in Theorem 4.3.13.
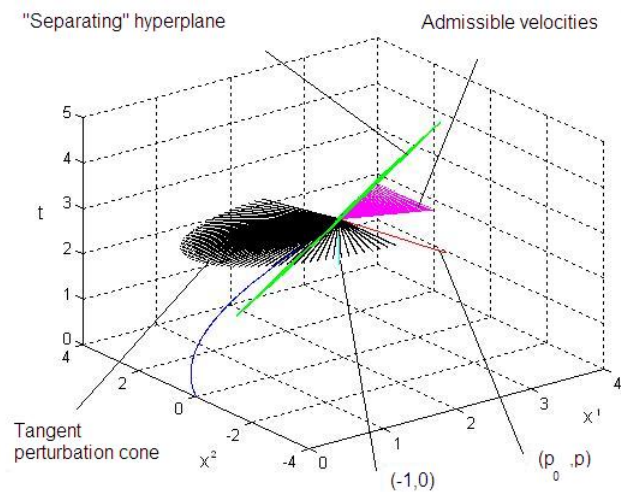
Figure 4.13: *The separation condition at $t = 3$ for the normal momenta is not satisfied.*

On the other hand, the curve followed in an abnormal way is still satisfying the separation condition at time $t = 3$ as appears in Figure 4.14.
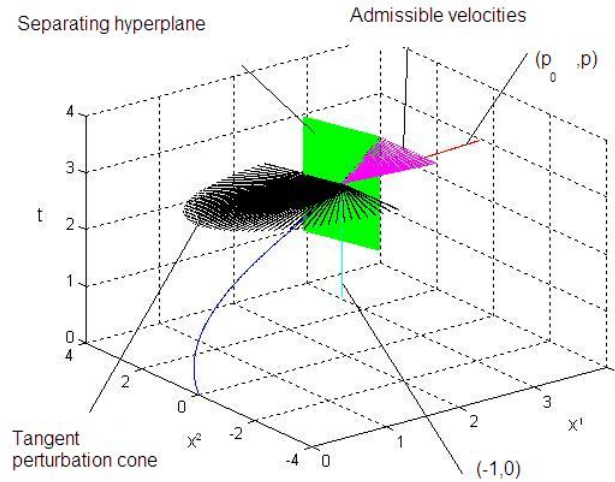
Figure 4.14: *Separation condition at $t = 3$ for the abnormal momenta.*

In conclusion, the time–optimal problem with the given endpoint conditions is solved exclusively with a strict abnormal trajectory. It will remain to prove there is no other trajectory that reaches $(2, 0)$ in less time. But the reachable set already shows that π is the minimum time for fulfilling the endpoint conditions.

To get a better understanding of the abnormal extremals, the reachable set in the extended system should be studied. If so, the abnormal extremals are in the boundary of the reachable set as mentioned in [Bullo and Lewis 2005b, Langerock 2003b] and shown in Figure 4.15. The colours are the same as in Figure 4.10 and they have the same meaning; that is, only the curves in magenta are extremals. The green curve is the strict abnormal minimizer for the problem. When we project onto $\mathbb{R}^2$, the Figure 4.10 is recovered.
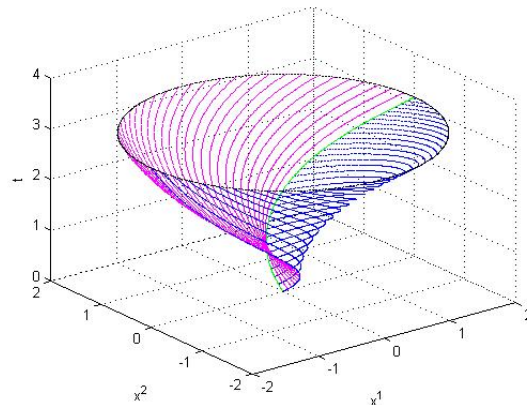
Figure 4.15: *Reachable set from $(0, 0, 0)$ up to time π of the extended control system of (4.6.23).*

Chapter 5
Constraint algorithm for the extremals

$\mathbf{A}$ geometric method is described to characterize the different kinds of extremals in optimal control theory. This comes from the use of a presymplectic constraint algorithm, in the sense given in [Cariñena 1990, Gotay and Nester 1979, Gotay et al. 1978, Gràcia and Pons 1992] and reviewed in §2.3.2, starting from the necessary conditions given by Pontryagin's Maximum Principle, Chapter 4. Apart from the design of this general algorithm useful for any optimal control problem, it is shown how to classify the set of extremals and, in particular, how to characterize the strict abnormality. The procedure will be used for mechanical systems in Chapter 6.

Despite the fact that the natural geometric framework for Pontryagin's Maximum Principle is the symplectic one [Barbero-Liñán and Muñoz Lecanda 2008b, Jurdjevic 1997, Lewis 2006, Sussmann 1998] (see also Chapter 4), for our purposes the presymplectic formalism [Barbero-Liñán and Muñoz Lecanda 2008a, Delgado-Téllez and Ibort 2003, Martínez 2004] will be more useful. We provide an implicit equation including some compatibility conditions that are necessary for the maximization of the Hamiltonian over the controls according to the classic Maximum Principle [Lee and Markus 1967, Pontryagin et al. 1962]. Hence, in the presymplectic framework, a weaker version of Pontryagin's Maximum Principle is stated, §5.1. Instead of the above classical necessary condition for optimality, we have an implicit differential equation that sets up a constraint algorithm, useful for determining where the normal extremals evolve and also for characterizing the abnormal extremals. We also obtain sufficient conditions to have both kinds of extremals. These conditions elucidate how to determine strict abnormality. This adaptation of the algorithm to the study of the extremals for the fixed time optimal control problems is mostly developed in §5.2, under the assumption that the control set is open and the differentiability with respect to the controls whenever is needed. After studying the fixed time problem, in §5.3 we explain how the algorithm works for the free optimal control problem given in Statement 4.3.1.

Finally, in §5.4 some results obtained by Agrachev and Sachkov [2004], Agrachev and Zelenko [2007], Liu and Sussmann [1995] are revisited using our procedure that provides a natural understanding of the dynamics of the extremals. In §5.5 a strict abnormal extremal for a control–affine system is given using the presymplectic constraint algorithm.

5.1   Presymplectic Pontryagin's Maximum Principle

As in Chapter 3, a *control system* is given by a vector field along a projection, see Definition 3.1.1. Let M be a smooth manifold, $\dim M = m$, U be an open set of $\mathbb{R}^k$ called the *control set* with $k \leq m$.

The problem to be solved is the fixed time optimal control problem, Statement 4.1.1, but a presymplectic viewpoint is considered, as for instance in [León et al. 2004, Delgado-Téllez and Ibort 2003, Echeverría-Enríquez et al. 2003, Martínez 2004]. In this approach, the main elements are the following.

- The *presymplectic manifold* $(T^*M \times U, \Omega)$, where Ω is the closed 2–form on $T^*M \times U$ given by the pullback through $\pi_1 \colon T^*M \times U \to T^*M$ of the canonical 2–form on T^*M.

- A *presymplectic Hamiltonian system* $(T^*M \times U, \Omega, H)$, where $H \colon T^*M \times U \to \mathbb{R}$ is the *Pontryagin's Hamiltonian function*:

$$H(\lambda, u) = \langle \lambda, X(x, u) \rangle + p_0 \mathcal{F}(x, u) = H_X(\lambda, u) + p_0 \mathcal{F}(x, u),$$

with $\lambda \in T_x^*M$, $p_0 \in \{-1, 0\}$ and the notation $H_X(\lambda, u) = \langle \lambda, X(x, u) \rangle$.

Observe that the kernel of Ω is precisely the π_1–vertical vector fields; that is, π_1–projectable vector fields $Z \in \mathfrak{X}(T^*M \times U)$ such that $(\pi_1)_* Z = 0$. Thus, Ω is degenerate. For details about the presymplectic formalism see [Cariñena 1990, Gotay and Nester 1979, Gotay et al. 1978, Gràcia and Pons 1992, Martínez 2004, Muñoz-Lecanda and Román-Roy 1992].

Theorem 5.1.1. *(Presymplectic Pontryagin's Maximum Principle) Let $U \subset \mathbb{R}^k$ be an open set. If $(\gamma, u) \colon [a, b] \to M \times U$ is a solution of the optimal control problem 4.1.1 with endpoint conditions x_a, x_b, then there exist $\lambda \colon [a, b] \to T^*M$ along γ and a constant $p_0 \in \{-1, 0\}$ such that:*

1. *(λ, u) is an integral curve of a Hamiltonian vector field X_H that satisfies*

$$i_{X_H} \Omega = \mathrm{d}H, \ \text{that is}, \ i_{(\dot{\lambda}(t), \dot{u}(t))} \Omega = \mathrm{d}H(\lambda(t), u(t)); \tag{5.1.1}$$

2. *(a) $H(\lambda(t), u(t))$ is constant everywhere in $t \in [a, b]$;*
   (b) $(p_0, \lambda(t)) \neq 0$ for each $t \in [a, b]$.

As Ω is degenerate, (5.1.1) does not necessarily have solution on the whole manifold $T^*M \times U$. As explained in §2.3.2, it may have a solution if we restrict the equation to the submanifold defined implicitly by

$$S = \{\beta \in T^*M \times U \mid i_v \, \mathrm{d}H = 0, \quad \text{for } v \in \ker \Omega_\beta\}. \tag{5.1.2}$$

Locally,

$$S = \left\{ \beta \in T^*M \times U \ \middle|\ \frac{\partial H}{\partial u^l}(\beta) = 0, \quad l = 1, \ldots, k \right\}.$$

Remark 5.1.2. Observe that S is defined implicitly by a necessary condition for the Hamiltonian to have an extremum over the controls as long as U is an open set. In the classical Pontryagin's Maximum Principle [Pontryagin et al. 1962], the Hamiltonian is equal to the maximum of the Hamiltonian over the controls. Therefore, Theorem 5.1.1 is a weaker version of the classical Maximum Principle.

The necessary conditions of Theorem 5.1.1 determine the same extremals as in Definition 4.1.15. Remember that Pontryagin's Maximum Principle lifts optimal solutions to the cotangent bundle. The uniqueness of the lifts is not guaranteed; that is, some extremals could be lifted in two different ways: normal and abnormal.

Remark 5.1.3. If the control set is not open, the presymplectic constraint algorithm can also be applied to characterize the different kinds of extremals following the process described in [López and Martínez 2000].

5.2   Characterization of extremals

Here we take advantage of the necessary conditions in Theorem 5.1.1 to determine where the different kinds of extremals in Definition 4.1.15 are contained. We are specially interested in strict abnormal extremals and abnormal extremals as a consequence of [Liu and Sussmann 1994b; 1995, Montgomery 1994], where it is proved the existence of strict abnormal minimizers for the problem of the shortest paths in subRiemannian geometry with two control vector fields. A meaningful and constructive procedure in the presymplectic formalism to find a solution to Statement 5.2.1 is the constraint algorithm [Cariñena 1990, Gotay and Nester 1979, Gotay et al. 1978, Gràcia and Pons 1992, Muñoz-Lecanda and Román-Roy 1992].

Statement 5.2.1. *Given a presymplectic system* $(T^*M \times U, \Omega, H)$, *find* (N, X) *such that*

(a) N *is a submanifold of* $T^*M \times U$,

(b) $X \in \mathfrak{X}(T^*M \times U)$ *is tangent to* N *and satisfies* $i_X \Omega = \mathrm{d}H$ *on* N,

(c) N *is maximal among all the submanifolds satisfying* (a) *and* (b).

As mentioned in §2.3.2, the presymplectic equation (5.1.1), $i_{X_H}\Omega = \mathrm{d}H$, has solution in the primary constraint submanifold

$$N_0 = \{(\lambda, u) \in T^*M \times U \mid \exists\, v_{(\lambda,u)} \in T_{(\lambda,u)}(T^*M \times U)\,, \quad i_{v_{(\lambda,u)}}\Omega = \mathrm{d}_{(\lambda,u)}H\},$$

or equivalently, $N_0 = \{(\lambda, u) \in T^*M \times U \mid (\mathcal{L}_Z H)_{(\lambda,u)} = 0\,, \ \forall\, Z \in \ker \Omega\}$, where $\mathcal{L}_Z$ is the Lie derivative with respect to Z. Observe that N_0 is exactly the submanifold S in (5.1.2). See [Gotay et al. 1978, Muñoz-Lecanda and Román-Roy 1992] for details on this equivalence.

Locally, the Hamiltonian vector field $X_H = A^i \partial/\partial x^i + B_j \partial/\partial p_j + C_l \partial/\partial u^l$ in the presymplectic equation (5.1.1) is given by $A^i = \partial H/\partial p_i$ and $B_j = -\partial H/\partial x^j$. Moreover, as $\partial/\partial u^l \in \ker \Omega$,

$$N_0 = \left\{(\lambda, u) \in T^*M \times U \,\middle|\, \frac{\partial H}{\partial u^l} = \lambda_j \frac{\partial X^j}{\partial u^l} + p_0 \frac{\partial \mathcal{F}}{\partial u^l} = 0\,, \quad l = 1, \dots, k\right\}. \tag{5.2.3}$$

The solution on N_0 is not necessarily unique. Indeed, if X_0 is a solution, then $X_0 + \ker \Omega$ is the set of all the solutions. We may consider X_0 as a vector field defined on $T^*M \times U$ because we assume that N_0 is a closed submanifold. If N_0 is not a submanifold we can use the adaptation

of presymplectic constraint algorithm described in [Krupková 1997]. Thus X_0 can be extended to $T^*M \times U$ by a construction using an open cover and partitions of unity [Lee 2003].

Take the pair $(N_0, X_0 + \ker \Omega)$, rewritten as (N_0, X^{N_0}). Observe that we are looking for an element in X^{N_0} tangent to N_0. Then,

$$N_1 = \{x \in N_0 \,|\, \exists\, X \in X^{N_0} \,,\, X(x) \in T_x N_0\}.$$

Locally,

$$N_1 = \left\{ (\lambda, u) \in N_0 \quad\middle|\quad 0 = X_H \left(\frac{\partial H}{\partial u^l} \right) = \frac{\partial H}{\partial p_i} \frac{\partial^2 H}{\partial x^i \partial u^l} - \frac{\partial H}{\partial x^j} \frac{\partial^2 H}{\partial p_j \partial u^l} \right.$$
$$\left. + C_r \frac{\partial^2 H}{\partial u^r \partial u^l}, \quad l = 1, \dots, k \right\}. \tag{5.2.4}$$

If the matrix $(\partial^2 H / \partial u^r \partial u^l)_{rl}$ is not invertible, the optimal control problem is called *singular* [Delgado-Téllez and Ibort 2003]; otherwise it is *regular*.

This step stabilizes the constraints in N_0 providing a new pair (N_1, X^{N_1}) where X^{N_1} is the set of the vector fields solution and tangent to N_0. Inductively, we arrive at (N_i, X^{N_i}) where we assume that N_i is a submanifold of $T^*M \times U$ and we define

$$N_{i+1} = \left\{ x \in N_i \,\middle|\, \exists\, X \in X^{N_i} \,,\, X(x) \in T_x N_i \right\},$$

obtaining the sequence

$$T^*M \times U \supseteq N_0 \supseteq N_1 \supseteq \dots \supseteq N_i \supseteq N_{i+1} \supseteq \dots$$

and the corresponding $X^{N_{i+1}}$. Observe that at each step the constraints should be independent to guarantee that the corresponding subsets in $T^*M \times U$ are submanifolds. Let

$$N_f = \bigcap_{i \geq 0} N_i, \quad X^{N_f} = \bigcap_{i \geq 0} X^{N_i},$$

if (N_f, X^{N_f}) is a nontrivial pair, it is the solution to the problem in Statement 5.2.1. If at one step $N_i = N_{i+1}$, the algorithm finishes with $N_f = N_i$.

Note that each step of the algorithm can reduce the set of points of $T^*M \times U$ where there exists solution, that is, $N_{i+1} \subsetneq N_i$, and can also reduce the degrees of freedom of the set of vector fields solution, $X^{N_{i+1}} \subsetneq X^{N_i}$. In terms of control systems, the desirable objectives are to restrict the problem to a smaller submanifold of $T^*M \times U$ and to determine the input controls. Observe that, generally, a step of the algorithm can provide us new constraints and the determination of some controls at the same time. Hence, either a unique vector field is found or the new constraints must be stabilized or the set must be split in such a way that the constraints define submanifolds. At the final step, we have either a unique or nonunique vector field and a submanifold that could be an empty or discrete set.

Remark 5.2.2. Observe that this procedure does not exclude any extremal, in contrast to the

method used in [Liu and Sussmann 1995] for the problem of the shortest paths in subRiemannian geometry. There, using a less geometric approach, the constant extremals are missed.

Now, let us focus again on optimal control problems where there are two distinct Hamiltonians depending on the value of the constant p_0. Thus, from Equation (5.2.3) it is deduced that the constraint algorithm must be run twice, once for each Hamiltonian, as is explained in §5.2.1 and §5.2.2.

5.2.1 Characterization of abnormality

First, we characterize a subset of $T^*M \times U$ where the abnormal biextremals evolve if they exist; see Definition 4.1.15. In this situation $p_0 = 0$ and the corresponding Pontryagin's Hamiltonian is $H^{[0]} = H_X$. Then the primary constraint submanifold (5.2.3) becomes

$$N_0^{[0]} = \left\{ (\lambda, u) \in T^*M \times U \,\middle|\, \lambda_j \frac{\partial X^j}{\partial u^l} = 0, \quad l = 1, \dots, k \right\}, \tag{5.2.5}$$

the submanifold (5.2.4) is

$$N_1^{[0]} = \left\{ (\lambda, u) \in N_0^{[0]} \,\middle|\, \lambda_j \left(X^i \frac{\partial^2 X^j}{\partial x^i \partial u^l} - \frac{\partial X^j}{\partial x^i} \frac{\partial X^i}{\partial u^l} + C_r \frac{\partial^2 X^j}{\partial u^r \partial u^l} \right) = 0, \quad l = 1, \dots, k \right\},$$

and the algorithm continues. Once we have the final constraint submanifold $N_f^{[0]}$ for abnormality, we have to delete the biextremals through the zero fiber because these biextremals do not satisfy the necessary condition (2b) of the presymplectic Pontryagin's Maximum Principle, Theorem 5.1.1. For the sake of simplicity and clarity, we denote this actual final constraint submanifold with the same symbol $N_f^{[0]}$.

Considering the natural projection of the cotangent bundle $\pi_M \colon T^*M \to M$ and the above–defined elements, we have the following result.

Proposition 5.2.3. *Let $N_f^{[0]}$ be the final constraint submanifold solution to the presymplectic problem in Statement 5.2.1 with the Hamiltonian $H^{[0]} = H_X$. If $N_f^{[0]} \neq \emptyset$, then there exists a curve in $(\lambda, u) \colon I \to N_f^{[0]}$ with $\lambda \neq 0$ such that*

$$(\gamma, u) = (\pi_M \times \mathrm{Id})(\lambda, u)$$

is an abnormal extremal in terms of the presymplectic Pontryagin's Maximum Principle, Theorem 5.1.1.

5.2.2 Characterization of normality

Analogous to § 5.2.1, for $p_0 = -1$, Pontryagin's Hamiltonian is $H^{[-1]} = H_X - \mathcal{F}$.

Then the primary constraint submanifold (5.2.3) becomes

$$N_0^{[-1]} = \left\{ (\lambda, u) \in T^*M \times U \,\middle|\, \lambda_j \frac{\partial X^j}{\partial u^l} - \frac{\partial \mathcal{F}}{\partial u^l} = 0, \quad l = 1, \dots, k \right\} \tag{5.2.6}$$

and the submanifold in (5.2.4) is

$$N_1^{[-1]} = \left\{ (\lambda, u) \in N_0^{[-1]} \;\middle|\; \lambda_j \left(X^i \frac{\partial^2 X^j}{\partial x^i \partial u^l} - \frac{\partial X^j}{\partial x^i} \frac{\partial X^i}{\partial u^l} \right) - X^i \frac{\partial^2 \mathcal{F}}{\partial x^i \partial u^l} \right.$$

$$\left. + C_r \left(\lambda_j \frac{\partial^2 X^j}{\partial u^r \partial u^l} - \frac{\partial^2 \mathcal{F}}{\partial u^r \partial u^l} \right) = 0, \quad l = 1, \ldots, k \right\}.$$

The determination of the controls for normal extremals depends on the given cost function. To better understand this process we refer the reader to the examples in §5.3, §5.4 and §5.5.

It can be observed that Hamilton's equations for $\dot{x}^i$ are the same for both Hamiltonian functions, for $p_0 = 0$ and $p_0 = -1$, since the cost function does not depend on the momenta. Hamilton's equations for $\dot{p}_i$ are equal for cost functions not depending on x^i. For instance, if the cost function is constant, as in the case of time–optimal control problems.

The final constraint submanifolds $N_f^{[0]}$ and $N_f^{[-1]}$ restrict the set of points where the biextremals of the optimal control problem in Statement 4.1.1 evolve. But, even in the case that Hamilton's equations are the same, $N_f^{[0]}$ and $N_f^{[-1]}$ could be different. Then the integral curves of the same vector field in $T^*M \times U$ along the same extremal in M may be different, depending on where the initial conditions for the momenta are taken. In other words, there may exist abnormal extremals being normal and vice versa. For a deeper study of the extremals, we need to project the biextremals onto the base manifold $M \times U$ using $\pi_M \times \mathrm{Id} \colon T^*M \times U \to M \times U$.

Summarizing all the above comments and keeping in mind the elements previously defined, we have the following propositions.

Proposition 5.2.4. *Let $(M, U, X, \mathcal{F}, I, x_a, x_b)$ be an optimal control problem in Statement 4.1.1 and $N_f^{[-1]}$ be the final constraint submanifold solution to the presymplectic problem in Statement 5.2.1 with the Hamiltonian $H^{[-1]} = H_X - \mathcal{F}$. If there exists a curve $(\lambda, u) \colon I \to N_f^{[-1]}$, then $(\gamma, u) = (\pi_M \times \mathrm{Id})(\lambda, u)$ is a normal extremal for the given OCP.*

Proposition 5.2.5. *Let $(M, U, X, \mathcal{F}, I, x_a, x_b)$ be an optimal control problem in Statement 4.1.1 and (γ, u) be an abnormal extremal for this OCP. If there exists a covector λ along γ such that $(\lambda, u) \in N_f^{[-1]}$, then (γ, u) is also a normal extremal for the given OCP.*

Let (γ, u) be a normal extremal for the given OCP. If there exists a covector λ along γ such that $(\lambda, u) \in N_f^{[0]}$, then (γ, u) is also an abnormal extremal for the OCP.

Proposition 5.2.6. *Let $(M, U, X, \mathcal{F}, I, x_a, x_b)$ be an optimal control problem in Statement 4.1.1, and $N_f^{[0]}$ and $N_f^{[-1]}$ be the final constraint submanifold solution to the presymplectic problem in Statement 5.2.1 with the Hamiltonian $H^{[0]} = H_X$ and $H^{[-1]} = H_X - \mathcal{F}$, respectively. If there exist curves $(\lambda^{[0]}, u^{[0]}) \colon I \to N_f^{[0]}$ with $\lambda^{[0]} \neq 0$ and $(\lambda^{[-1]}, u^{[-1]}) \colon I \to N_f^{[-1]}$ such that $\pi_M(\lambda^{[0]}) = \pi_M(\lambda^{[-1]}) = \gamma$, then γ is an abnormal extremal and also a normal extremal for the OCP.*

Remark 5.2.7. In this last proposition we do not consider the control as a part of the extremal,

because it may happen that different controls give the same extremals in M depending on the control system. So we project onto M the biextremals to compare them.

However, under some assumptions about the control system, such as control–affinity with independent control vector fields, different controls give different extremals. If this happens, we may project the biextremals onto $M \times U$ through $\pi_M \times \text{Id}$ to compare them.

Remark 5.2.8. The union of both final constraint submanifods do not cover exactly the set of extremals in Definition 4.1.15, because the condition (2a) in Theorem 5.1.1 is not included in the final constraint submanifold. See § 5.3 to get a better understanding.

5.2.3 Characterization of strict abnormality

The normal and abnormal extremals in Definition 4.1.15 do not constitute a disjoint partition of the set of extremals as Propositions 5.2.5 and 5.2.6 show. While in §5.2.1 we do not care about the cost function, in §5.2.2 the cost function is involved in the description of the extremal. To characterize strict abnormal extremals, the cost function is fundamental because these extremals are abnormal but not normal. The only way to guarantee that an extremal is not normal is to use the cost function.

As a consequence of the final constraint submanifolds obtained from the algorithm for abnormality and normality, strict abnormality can be studied. The adjective "strict" denotes that the extremal only admits one kind of lift to the cotangent bundle. To find strict abnormal extremals we have to project the final constraint submanifolds to M. In the intersection are the extremals admitting two different kinds of lifts: with $p_0 = 0$ and with $p_0 = -1$.

To summarize, all the biextremals in $N_f^{[0]}$ and $N_f^{[-1]}$ are projected to M via $\rho = \pi_M \circ \pi_1 \colon T^*M \times U \to M$ to be compared due to Remark 5.2.7, as shown in Figure 5.1.
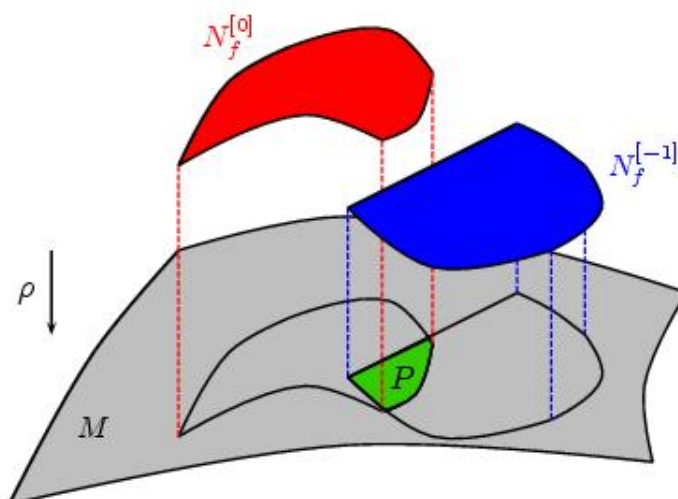
Figure 5.1: *Final situation in the constraint algorithm for the extremals.*

Proposition 5.2.9. *Let $(M, U, X, \mathcal{F}, I, x_a, x_b)$ be an optimal control problem in Statement 4.1.1, and $N_f^{[0]}$ and $N_f^{[-1]}$ be the final constraint submanifold solution to the presymplectic problem in Statement 5.2.1 with the Hamiltonian $H^{[0]} = H_X$ and $H^{[-1]} = H_X - \mathcal{F}$, respectively. Let $\rho = \pi_M \circ \pi_1 \colon T^*M \times U \to M$ be a projection with the projection $\pi_1 \colon T^*M \times U \to T^*M$ and $P = \rho\left(N_f^{[0]}\right) \cap \rho\left(N_f^{[-1]}\right)$.*

(i) If $P = \emptyset$ and $\rho\left(N_f^{[0]}\right) \neq \emptyset$, then all the abnormal extremals are strict.

(ii) If $P = \emptyset$ and $\rho\left(N_f^{[-1]}\right) \neq \emptyset$, then all the normal extremals are strict.

(iii) If $P \neq \emptyset$ and $\rho\left(N_f^{[0]}\right) = P$, then there are no strict abnormal extremals.

(iv) If $P \neq \emptyset$ and $\rho\left(N_f^{[0]}\right) \neq P$, then there are locally strict abnormal extremals.

(v) If $P \neq \emptyset$ and $\rho\left(N_f^{[0]}\right) = \rho\left(N_f^{[-1]}\right) = P$, then all the abnormal extremals are also normal and viceversa.

In item (iv), it is said that there are strict abnormal extremals, but only locally since the extremal could have pieces in P. So at some points the extremal can be locally normal.

5.3   Free optimal control problem

Now that the theory has been introduced, let us deal with the particular case of the free optimal control problem, Statement 4.3.1. In this case the interval of definition of the extremals is an unknown of the problem.

Pontryagin's Maximum Principle is the same as Theorem 5.1.1, but replacing (2a) by

$$(2\text{a}') \ H(\lambda(t), u(t)) \text{ is zero everywhere } t \in I.$$

Thus the presymplectic equation (5.1.1) must be restricted to the submanifold defined by the condition

$$H = H_X + p_0 \mathcal{F} = 0.$$

Hence, it must also be stabilized in the algorithm. Due to the properties of Hamiltonian systems [Abraham and Marsden 1978], the condition $H = 0$ is trivially stabilized. Thus its tangency condition does not add any new constraint to the submanifolds of the algorithm. The same happens with $H = \text{constant}$, but this is not a suitable constraint for a submanifold. This is why it is not included in the primary constraint submanifold for the fixed–time OCP in Statement 4.1.1 whose extremals have been studied in § 5.2. In contrast to Remark 5.2.8, the final constraint submanifolds we find here recover the entire set of extremals since all the necessary conditions of Theorem 5.1.1 are taking into account. The trivial stabilization of $H = 0$ makes it possible to run the algorithm putting aside that constraint. Then the same final constraint submanifolds for abnormality and normality as in § 5.2.1 and §5.2.2, respectively,

are obtained. Those submanifolds are renamed $N_{ff}^{[0]}$ and $N_{ff}^{[-1]}$ since the actual final constraint submanifolds are obtained by considering the vanishing of the Hamiltonian:

$$
\begin{aligned}
N_f^{[0]} &= N_{ff}^{[0]} \cap \{(\lambda, u) \in T^*M \times U \mid H_X = 0\}, \\
N_f^{[-1]} &= N_{ff}^{[-1]} \cap \{(\lambda, u) \in T^*M \times U \mid H_X - \mathcal{F} = 0\}.
\end{aligned}
$$

Due to condition (2b) in Theorem 5.1.1, the zero fiber must be deleted from $N_f^{[0]}$.

Proposition 5.3.1. *Let $(M, U, X, \mathcal{F}, x_a, x_b)$ be a free optimal control problem in Statement 4.3.1:*

1. *If $N_f^{[0]}$ has only zero covectors, there are no abnormal extremals.*

2. *If $N_f^{[0]}$ has nonzero covectors and $N_{ff}^{[0]} \subset \{(\lambda, u) \in T^*M \times U \mid H_X = 0\}$, then every abnormal extremal is strict and there are no normal extremals as long as $\mathcal{F}$ does not vanish along the extremal.*

(*Proof*) It is similar to the proof of Proposition 6.7.4.   ∎

5.4   Examples revisited

There are some classical optimal control problems where the classification of extremals has been described with different tools and approaches: geodesics in Riemannian geometry [Liu and Sussmann 1995], shortest–paths in subRiemannian geometry [Agrachev and Gauthier 2001, Liu and Sussmann 1995] and optimal control problems for control–affine systems [Agrachev and Sachkov 2004, Agrachev and Zelenko 2007, Trélat 2000; 2001]. All of them can be studied in a unified way by direct application of the method we have proposed in this chapter.

5.4.1   Geodesics in Riemannian geometry

Let M be an m–dimensional Riemannian manifold and assume that there exist m linear independent vector fields $\{Y_1, \ldots, Y_m\}$ on M. Consider the following control–linear system:

$$
X = u^1 Y_1 + \ldots + u^m Y_m.
$$

The problem of finding the geodesic curves in M can be addressed as an optimal control problem for the previous system with cost function $\mathcal{F}(x, u) = \|X\|$, where $\|\cdot\|$ is the Riemannian norm.

For abnormality, $p_0 = 0$, the primary constraint submanifold (5.2.5) is

$$
N_0^{[0]} = \{(\lambda, u) \in T^*M \times U \mid \langle \lambda, Y_l \rangle = 0, \quad l = 1, \ldots, m\}.
$$

Note that the controls do not appear in the primary constraint submanifold. As the number of controls coincides with the dimension of the state space, the annihilator of all the control

vector fields is the zero covector. But Theorem 5.1.1 says that $(p_0, \lambda(t)) \neq 0$. So in Riemannian geometry there are neither abnormal nor strict abnormal extremals, as stated in [Liu and Sussmann 1995].

For normality, $p_0 = -1$, the primary constraint submanifold (5.2.6) is

$$N_0^{[-1]} = \left\{ (\lambda, u) \in T^*M \times U \,\middle|\, \langle \lambda, Y_l \rangle - \frac{\partial \mathcal{F}}{\partial u^l} = 0, \quad l = 1, \ldots, m \right\}.$$

For instance, if the vector fields $\{Y_1, \ldots, Y_m\}$ are orthonormal, the cost function is

$$\mathcal{F}(x, u) = \frac{1}{2}((u^1)^2 + \ldots + (u^m)^2),$$

where $1/2$ is written by convention, and we have

$$N_0^{[-1]} = \{ (\lambda, u) \in T^*M \times U \mid \langle \lambda, Y_l \rangle - u^l = 0, \quad l = 1, \ldots, m \}.$$

Hence, all the controls are known and the Hamiltonian vector field X_H is uniquely determined. Then $N_f^{[-1]} = N_0^{[-1]}$ as explained in § 5.2. We are in case (ii) in Proposition 5.2.9.

The projections on M of the integral curves of X_H satisfy the well–known geodesic equations on M, as can be easily proved.

5.4.2 SubRiemannian geometry

As before, let M be an m–dimensional Riemannian manifold and $\{Y_1, \ldots, Y_k\}$ be linearly independent vector fields on M, but with $k < m$. Now the corresponding control–linear system is

$$X = u^1 Y_1 + \ldots + u^k Y_k. \tag{5.4.7}$$

We state a problem analogous to the one in the previous section: find the integral curve in M of the vector field (5.4.7) such that it minimizes the functional of the cost function $\mathcal{F}(x, u) = \|X\|$. This optimal control problem has as a solution the shortest paths—i.e., geodesics—in subRiemannian geometry.

For abnormality, $p_0 = 0$, the primary constraint submanifold (5.2.5) is

$$N_0^{[0]} = \{ (\lambda, u) \in T^*M \times U \mid \langle \lambda, Y_l \rangle = 0, \quad l = 1, \ldots, k \}$$

and the Hamiltonian vector field on $N_0^{[0]}$ is $X_{H^{[0]}} = \sum_{r=1}^{k} u^r X_{Y_r}$, where X_{Y_r} denotes $X_{H_{Y_r}}$.

The tangency condition is

$$
\begin{aligned}
0 = X_{H^{[0]}}(H_{Y_l}) &= \sum_{r=1}^{k} u^r X_{Y_r}(H_{Y_l}) = \sum_{r=1}^{k} u^r \mathrm{d}H_{Y_l}(X_{Y_r}) = -\sum_{r=1}^{k} u^r \{H_{Y_r}, H_{Y_l}\} \\
&= \sum_{r=1}^{k} u^r H_{[Y_r, Y_l]}
\end{aligned}
$$

for $l = 1, \ldots, k$. Here, the properties of the Poisson bracket, which is denoted by $\{\cdot, \cdot\}$, have been used. See [Abraham and Marsden 1978, pp. 192] for more details about the Poisson brackets. Then

$$N_1^{[0]} = \left\{ (\lambda, u) \in N_0^{[0]} \,\middle|\, \sum_{r=1}^{k} u^r H_{[Y_r, Y_l]} = 0, \quad l = 1, \ldots, k \right\}.$$

Liu and Sussmann [1995] gives a characterization of the abnormal extremals when there are only two control vector fields. In [Agrachev and Gauthier 2001] the abnormal extremals are studied more generally, without any assumption about the number of control vector fields.

For two input vector fields, the submanifold $N_1^{[0]}$ is defined implicitly by the constraints

$$\{u^1 H_{[Y_1, Y_2]} = 0, \ u^2 H_{[Y_2, Y_1]} = 0\}.$$

As both controls cannot be identically zero—otherwise there is no motion—then the only constraint is $H_{[Y_1, Y_2]} = 0$. Following the algorithm we obtain

$$N_2^{[0]} = \left\{ (\lambda, u) \in N_1^{[0]} \,\middle|\, u^1 H_{[Y_1, [Y_1, Y_2]]} + u^2 H_{[Y_2, [Y_1, Y_2]]} = 0 \right\}.$$

If we assume accessibility, then at least one of $H_{[Y_1, [Y_1, Y_2]]}$ and $H_{[Y_2, [Y_1, Y_2]]}$ must be nonzero. Hence, as we have a linear dependence between the controls, the motion is determined up to reparametrization.

To recap, the abnormal extremals, if they exist, are in

$$N_2^{[0]} - \left\{ (\lambda, u) \in N_2^{[0]} \,\middle|\, H_{[Y_1, [Y_1, Y_2]]} = 0, \ H_{[Y_2, [Y_1, Y_2]]} = 0 \right\}.$$

See [Liu and Sussmann 1995] for similar results with another approach.

For normality, $p_0 = -1$ and the primary constraint submanifold (5.2.6) is

$$N_0^{[-1]} = \left\{ (\lambda, u) \in T^*M \times U \,\middle|\, \langle \lambda, Y_l \rangle - \frac{\partial \mathcal{F}}{\partial u^l} = 0, \quad l = 1, \ldots, k \right\}.$$

As in § 5.4.1, if the vector fields $\{Y_1, \ldots, Y_k\}$ are orthonormal, the cost function is

$$\mathcal{F}(x, u) = \frac{1}{2}((u^1)^2 + \ldots + (u^k)^2)$$

and we have $N_0^{[-1]} = \{(\lambda, u) \in T^*M \times U \mid \langle \lambda, Y_l \rangle - u^k = 0, \quad l = 1, \ldots, k\}$. Thus, $N_f^{[-1]} = N_0^{[-1]}$.

Observe that the momenta associated with an abnormal extremal can be associated with a normal extremal if the controls are zero. As the system is control–linear, zero controls give constant curves. According to Proposition 5.2.9, we are either in case (iii) or (iv) or (v). In other words, it is guaranteed the existence of extremals being abnormal and also normal.

As the input vector fields are assumed to be linear independent, the curves in $T^*M \times U$ must be projected to M to be compared. The different biextremals—i.e., the curves in $T^*M \times U$—associated with an extremal are used to discuss in how many classes of Definition 4.1.15 the extremal can be.

5.4.3   Control–affine systems

Now we consider an m–dimensional manifold M and the control–affine system

$$X = Y + u^1 Y_1 + \ldots + u^k Y_k,$$

where $\{Y_1, \ldots, Y_k\}$ are linear independent vector fields and Y is the drift vector field. Let $\mathcal{F}$ be the cost function for an optimal control problem.

For abnormality, $p_0 = 0$, the primary constraint submanifold (5.2.5) is

$$N_0^{[0]} = \{(\lambda, u) \in T^*M \times U \mid \langle \lambda, Y_l \rangle = 0, \quad l = 1, \ldots, k\}$$

and the Hamiltonian vector field is $X_{H^{[0]}} = X_Y + \sum_{r=1}^{k} u^r X_{Y_r}$ on $N_0^{[0]}$, with the same notation as in §5.4.2.

The tangency condition is

$$0 = X_{H^{[0]}}(H_{Y_l}) = H_{[Y,Y_l]} + \sum_{r=1}^{k} u^r H_{[Y_j, Y_l]}, \quad l = 1, \ldots, k.$$

Then $N_1^{[0]} = \left\{ (\lambda, u) \in N_0^{[0]} \mid H_{[Y, Y_l]} + \sum_{r=1}^{k} u^r H_{[Y_r, Y_l]} = 0, \quad l = 1, \ldots, k \right\}$.

In [Agrachev and Sachkov 2004, Agrachev and Zelenko 2007, Trélat 2001] this situation is studied when there are at most two controls and in [Trélat 2000] more general results related to control–affine systems are given.

Depending of the rank of the matrices $A = (H_{[Y_r, Y_l]})$ and $B = (A \mid H_{[Y, Y_l]})$, we have the following situations, cf. [Agrachev and Sachkov 2004, Agrachev and Zelenko 2007, Trélat 2000; 2001]:

(i) The rank of A is maximum and then all the controls are determined. Hence, given the initial conditions for the momenta, the abnormal extremals are known.

(ii) The rank of A is not maximum and is equal to the rank of B. Then some controls are determined and others are free. There are no new constraints and the algorithm ends.

(iii) The rank of A is not maximum and different from the rank of B. Then some controls are determined and others are free. But there are also new constraints and the algorithm continues. At every step, a similar analysis to (i–iii) must be done to stabilize the new constraints.

For normality, $p_0 = -1$, and the primary constraint submanifold (5.2.6) is

$$N_0^{[-1]} = \left\{ (\lambda, u) \in T^*M \times U \;\middle|\; \langle \lambda, Y_l \rangle - \frac{\partial \mathcal{F}}{\partial u^l} = 0, \quad l = 1, \dots, k \right\}.$$

For instance, if the cost function is

$$\mathcal{F}(x, u) = \frac{1}{2}((u^1)^2 + \dots + (u^k)^2),$$

we have $N_0^{[-1]} = \{(\lambda, u) \in T^*M \times U \mid \langle \lambda, Y_l \rangle - u^k = 0, \quad l = 1, \dots, k\}$. Thus, $N_f^{[-1]} = N_0^{[-1]}$.

Suppose that the momenta that makes an extremal abnormal, it also makes the same extremal normal. Then all the controls for the extremal must be zero. In contrast with the example in §5.4.2, if the controls are zero, the curves are not necessarily constant because they are integral curves of the drift vector field of the control–affine system. According to Proposition 5.2.9, we are either in case (iii) or (iv) or (v). In other words, it is guaranteed the existence of extremals being abnormal and also normal.

5.5 A strict abnormal extremal in a control–affine system

Following the described method in §5.2, we find a strict abnormal extremal for a control–affine system on TQ, that, in fact, models an affine connection control mechanical system. See more details about these systems in §6.1 and [Bullo and Lewis 2005a]

Let $M = TQ = T\mathbb{R}^3$ (i.e., $Q = \mathbb{R}^3$), U be an open set in $\mathbb{R}^2$ containing the zero. In natural local coordinates (x, y, z, v_x, v_y, v_z) for TQ, the drift vector field of the control–system is

$$Z = v_x \frac{\partial}{\partial x} + v_y \frac{\partial}{\partial y} + v_z \frac{\partial}{\partial z} \,,$$

and the input vector fields are $Y_1 = \dfrac{\partial}{\partial v_x}$ and $Y_2 = (1 - x) \dfrac{\partial}{\partial v_y} + x^2 \dfrac{\partial}{\partial v_z}$. So the control system is given by $Z + u^1 Y_1 + u^2 Y_2$. The endpoint conditions in TQ are

$$v_a = (2, 0, 0, 0, v_y^0, 4(1 - v_y^0)), \; v_b = (2, 1, 0, 0, 2(1 - v_y^0), 4v_y^0 - 4)$$

with $v_y^0 \neq 1$. The cost function is $\mathcal{F} = \dfrac{(u^1)^2 + (u^2)^2}{2}$. Hence Pontryagin's Hamiltonian is

$$H(\lambda, u^1, u^2) = p_x v_x + p_y v_y + p_z v_z + u^1 q_x + u^2(1 - x)q_y + u^2 x^2 q_z + p_0 \frac{(u^1)^2 + (u^2)^2}{2},$$

with Hamilton's equations for abnormality and normality being

$$\begin{aligned}
\dot{x} &= v_x, & \dot{v}_x &= u^1, & \dot{p}_x &= q_y u^2 - 2q_z u^2 x, & \dot{q}_x &= -p_x, \\
\dot{y} &= v_y, & \dot{v}_y &= u^2(1 - x), & \dot{p}_y &= 0, & \dot{q}_y &= -p_y, \\
\dot{z} &= v_z, & \dot{v}_z &= u^2 x^2, & \dot{p}_z &= 0, & \dot{q}_z &= -p_z,
\end{aligned}$$

where p's and q's are the momenta for the states and the velocities, respectively, and the Hamiltonian vector field $X_H = \sum_{i \in \{x,y,z\}} (A^i \partial/\partial i + B^i \partial/\partial v_i + C_i \partial/\partial p_i + D_i \partial/\partial q_i) + E_1 \partial/\partial u^1 + E_2 \partial/\partial u^2$, where A^i, B^i, C_i, D_i are determined by Hamilton's equations.

The constraint algorithm for abnormality gives us

$$
\begin{aligned}
N_0^{[0]} &= \left\{ (\lambda, u) \in T^*TQ \times U \ \middle| \ \frac{\partial H^{[0]}}{\partial u^l}(\lambda, u) = H_{Y_l}(\lambda) = 0 \ , \text{ for } l = 1, 2 \right\} \\
&= \{ (\lambda, u) \in T^*TQ \times U \ | \ q_x = 0, \ \mathbf{q_y(1 - x) + q_z x^2 = 0} \} \\
N_1^{[0]} &= \{ (\lambda, u) \in N_0^{[0]} \ | \ H_{[Z, Y_l]}(\lambda) = 0 \ , \text{ for } l = 1, 2 \} \\
&= \{ (\lambda, u) \in N_0^{[0]} \ | \ p_x = 0, \ (-1 + x)p_y - x^2 p_z - v_x q_y + 2x v_x q_z = 0 \} \\
N_2^{[0]} &= \{ (\lambda, u) \in N_1^{[0]} \ | \ (H_{[Z, [Z, Y_l]] + \sum_{r=1}^2 u^r [Y_r, [Z, Y_l]]})(\lambda) = 0 \ , \text{ for } l = 1, 2 \} \\
&= \{ (\lambda, u) \in N_1^{[0]} \ | \ \mathbf{(-q_y + 2x q_z)u^2 = 0}, \\
&\qquad (-q_y + 2x q_z)u^1 = -(2p_y v_x - 4x v_x p_z + 2v_x^2 q_z) \} \ .
\end{aligned}
$$

In order to satisfy the endpoint conditions, to not have the zero covector, and to have a strict abnormal extremal, the subset defined by $x\,(x-1)\,q_z\,u^2 = 0$, coming from the above bold equations, must be deleted from the constraint submanifold. Then

$$
\begin{aligned}
N_2^{[0]} &= \{ (\lambda, u) \in T^*TQ \times U \backslash \{ x(x-1)q_z u^2 = 0 \} \ | \ q_x = 0, \ -q_y + 4q_z = 0, \ p_x = 0, \\
&\qquad p_y - 4p_z = 0, \ x = 2, \ v_x = 0 \} \\
N_3^{[0]} &= \{ (\lambda, u) \in N_2^{[0]} \ | \ v_x = 0, \ u^1 = 0 \} \\
N_4^{[0]} &= \{ (\lambda, u) \in N_2^{[0]} \ | \ u^1 = 0, \ E_1 = 0 \} = N_5^{[0]} = N_f^{[0]}.
\end{aligned}
$$

By restriction to the final constraint submanifold and integrating Hamilton's equations on $[0, 1]$, we have the abnormal lift

$$
\lambda(t) = (0, 4p_z^0, p_z^0, 0, -4p_z^0 t + 4q_z^0, -p_z^0 t + q_z^0)
$$

and

$$
\gamma(t) = \left(2, -u^2 \frac{t^2}{2} + v_y^0 t, 2u^2 t^2 + 4(1 - v_y^0)t, 0, -u^2 t + v_y^0, 4u^2 t + 4(1 - v_y^0) \right)
$$

with $u^2 = 2(v_y^0 - 1)$.

The constraint algorithm for normality gives us

$$
\begin{aligned}
N_0^{[-1]} &= \left\{ (\lambda, u) \in T^*TQ \times U \ \middle| \ \frac{\partial H^{[-1]}}{\partial u^l}(\lambda, u) = 0 \ , \text{ for } l = 1, 2 \right\} \\
&= \{ (\lambda, u) \in T^*TQ \times U \ | \ q_x - u^1 = 0, \\
&\qquad q_y(1 - x) + q_z x^2 - u^2 = 0 \} = N_1^{[-1]} = N_f^{[-1]}.
\end{aligned}
$$

If we substitute the curve γ in Hamilton's equations, we have $u^1 = 0$ and $u^2 = 2(v_y^0 - 1)$. Then, for the primary constraint submanifold, $q_x = 0$ and $u^2 = -q_y + 4q_z$. Due to Hamilton's equations, $p_x = 0$ and $0 = \dot{p}_x = -(u^2)^2$. This last equality is only possible if $v_y^0 = 1$, but that

was not the hypothesis. Thus there does not exist a lift with $p_0 = -1$ along γ; that is, γ is a strict abnormal extremal whenever $v_y^0 \neq 1$.

Chapter 6

Pontryagin's Maximum Principle for mechanical systems

After a general overview of optimal control theory providing new insights into this topic, we focus on a particular class of control systems: affine connection control systems. These systems, defined in §6.1, model a wide range of mechanical systems, such as the rolling disk, the snakeboard, the planar rigid body, the robotic leg; see for instance [Bloch 2003, Bullo and Lewis 2005a;b].

The mechanical control systems studied here are governed by second–order differential equations on a manifold Q, the so–called *configuration space* of the mechanical system. Thus they can be rewritten as first–order differential equations on TQ. Then we have a control–affine system on TQ. Notions related to the accessibility and the controllability of these mechanical systems are briefly reviewed in §6.2, but more details can be studied in [Bullo and Lewis 2005a, Cortés and Martínez 2003, Lewis and Murray 1997, Nijmeijer and van der Schaft 1990, Ostrowski and Burdick 1997, Sussmann 1987, Žefran et al. 1999].

In §6.3 we pose an optimal control problem for these mechanical systems. As these systems are reinterpreted as first–order differential equations on TQ, so the cost function for that problem may depend on the velocities and the endpoint conditions for the curves can be either on Q—i.e., without restrictions on the velocities—or on TQ. Thus, there are a few possible statements for the optimal control problem for mechanical systems.

The evolution of the previous chapters leads us to an intrinsic Pontryagin's Maximum Principle and to a weak Pontryagin's Maximum Principle in a presymplectic framework, both for the optimal control problems for affine connection control systems stated in §6.3.

According to the theory developed in §2.4.6, there exists a particular splitting of the tangent bundle TQ defined by the linear connection associated with a specific second–order vector field. Similar splittings for different tangent bundles are used to state intrinsically Pontryagin's Maximum Principle for the mechanical systems as explained in [Bullo and Lewis 2005b] and reviewed here in §6.4. The purpose of this review is to give a panorama, as complete as possible, of Pontryagin's Maximum Principle. Moreover, §6.4 is important to establish a comparison with some results in §7.2.1.

As described in Chapter 5, one of the contributions of this dissertation is the characterization of the different extremals in optimal control theory through the adaptation of a presymplectic constraint algorithm in the sense given by Cariñena [1990], Gotay and Nester [1979], Gotay

et al. [1978], Gràcia and Pons [1992], Muñoz-Lecanda and Román-Roy [1992]. To apply the process of characterizing extremals in optimal control problems for affine connection control systems, it is useful to use a weak or presymplectic mechanical Pontryagin's Maximum Principle, stated in §6.5 and similar to Theorem 5.1.1. Once the foundations are clear, the constraint submanifolds are computed for both abnormal and normal extremals in §6.6.

Some new results can be obtained in the study of abnormal extremals, in particular optimal control problems such as those with a control–quadratic cost function and the time–optimal control problem [Barbero-Liñán and Muñoz Lecanda 2008d], in §6.7.1 and §6.7.2, respectively. We mainly focus on the abnormal and strict abnormal extremals because of the interest in these as a result of some papers in mid 1990's [Liu and Sussmann 1995, Montgomery 1994], where the existence of strict abnormal minimizers is proved.

Furthermore, we consider some particular examples of optimal control problems with a given number of input vector fields and on a given configuration manifold, in order to better understand the constraint algorithm and to be mindful of the different situations that arise along the process, see §6.8. In the stop conditions to find the final constraint submanifolds, some geometric constructions such as symmetric products, vector–valued quadratic forms appear; see Appendix C and [Bullo and Lewis 2005a, Hirschorn and Lewis 2002]. There is still work to do in this direction, as shows Conjecture 6.8.5, to take advantage of the geometry in the problem as much as possible, especially to study abnormal extremals and necessary conditions for high–order maximum principle [Bianchini 1998, Kawski 2003, Knobloch 1981, Krener 1977].

6.1 Affine connection control systems

A *general simple mechanical system* is defined by $(Q, g, F, \mathcal{D}, \mathcal{Y}, U)$, where

- Q is a smooth n–dimensional manifold called the *configuration manifold*,

- g is a Riemannian metric on Q,

- F is a vector field along the projection τ_Q defining a vector force including all the external forces; e.g., the potential and the non–potential forces,

- $\mathcal{D}$ is the so–called *nonholonomic distribution* that restricts the set of velocities,

- $\mathcal{Y}$ is a set of k input vector fields on Q, and

- $U \subset \mathbb{R}^k$.

The Lagrange–d'Alembert principle [Bloch 2003] states that the solutions $\gamma \colon I \subset \mathbb{R} \to Q$ of this mechanical system are given by

$$\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) - u^s(t) Y_s(\gamma(t)) - F(\dot{\gamma}(t)) \in \mathcal{D}^{\perp}(\gamma(t)), \; \dot{\gamma}(t) \in \mathcal{D}(\gamma(t)),$$

where $\mathcal{D}^\perp$ is the g–orthogonal distribution to $\mathcal{D}$, ∇ is the Levi–Civita connection associated with g and the controls $u\colon I \to U$ are locally integrable; see [Bullo and Lewis 2005a; 2007, Lewis 1998] for more details. Thus, these general systems describe forced mechanical systems with nonholonomic constraints; that is, when the velocities at each point x in Q are restricted to be in a subspace $\mathcal{D}_x \subset T_xQ$. These general control systems will be studied in Chapter 7 beyond the scope of optimal control theory in the particular case of having the nonholonomic distribution equal to the distribution spanned by the input vector fields.

Here, we focus on the *affine connection control system (ACCS)* that is a mechanical system given by $\Sigma = (Q, \nabla, \mathscr{Y}, U)$ where ∇ is an affine connection on Q and $\mathscr{Y}$ is the family of input vector fields $\{Y_1, \ldots, Y_k\}$. Thus, there are neither external forces nor nonholonomic constraints ($\mathcal{D}_x = T_xQ$ for every $x \in Q$). An affine connection ∇ is considered in place of the connection associated with the Riemannian metric.

The dynamical equations of the control system are given by

$$\nabla_{\dot\gamma(t)}\dot\gamma(t) = u^l(t)Y_l(\gamma(t)), \tag{6.1.1}$$

where $\gamma\colon I \subset \mathbb{R} \to Q$ is absolutely continuous and the *controls* $u\colon I \to U \subset \mathbb{R}^k$ are locally integrable.

Equation $(6.1.1)$ can be rewritten as a first–order control–affine system on TQ,

$$\dot\Upsilon(t) = Z(\Upsilon(t)) + u^l(t)Y_l^V(\Upsilon(t)), \tag{6.1.2}$$

where $\Upsilon\colon I \to TQ$, Z is the geodesic spray associated to the affine connection on Q and is the drift vector field of the system, Y_l^V denotes the vertical lift of the vector field Y_l; see [Abraham and Marsden 1978]. The geodesic spray is the element of this control system that carries the information about the connection on Q. Observe that this control system can be understood as a control–linear system with $k+1$ input vector fields given by $\{Z, Y_1^V, \ldots, Y_k^V\}$ with the control set $\{1\} \times U$; that is, the control corresponding with the geodesic spray can only be equal to 1.

Remark 6.1.1. The linear independence of the input vector fields is assumed in the sequel to guarantee that different controls determine different curves in TQ satisfying Equation (6.1.2). This is also pointed out in Remark 5.2.7.

6.2 Accessibility and controllability for mechanical systems

Let us review the definitions and main results related with accessibility for ACCS [Bullo and Lewis 2005a;b, Cortés and Martínez 2003, Lewis and Murray 1997, Ostrowski and Burdick 1997].

Definition 6.2.1. *Let $\Sigma = (Q, \nabla, \mathscr{Y}, U)$ be an ACCS and $v_x \in T_xQ$.*

*1. The **reachable set from v_x at time T in Q** is*

$$R_{\Sigma,Q}(v_x, T) = \{\gamma(T) \mid (\gamma, u)\colon I \subset \mathbb{R} \to Q \times U \text{ satisfies } (6.1.1) , \ \dot\gamma(0) = v_x\}.$$

2. *The **reachable set from** v_x **at time** T **in** TQ is*

$$R_{\Sigma,TQ}(v_x, T) = \{ \Upsilon(T) \quad | \quad (\Upsilon, u) \colon I \subset \mathbb{R} \to TQ \times U \text{ satisfies (6.1.2)},$$
$$\Upsilon(0) = v_x \}.$$

3. *The **reachable set** $R_{\Sigma,Q}(v_x, \leq T)$ **from** v_x **up to** T **in** Q is*

$$R_{\Sigma,Q}(v_x, \leq T) = \bigcup_{0 \leq t \leq T} R_{\Sigma,Q}(v_x, t).$$

4. *The **reachable set** $R_{\Sigma,TQ}(v_x, \leq T)$ **from** v_x **up to** T **in** TQ is*

$$R_{\Sigma,TQ}(v_x, \leq T) = \bigcup_{0 \leq t \leq T} R_{\Sigma,TQ}(v_x, t).$$

Once the reachable sets are defined, we can introduce the notion of accessibility and controllability.

Definition 6.2.2. *Let $\Sigma = (Q, \nabla, \mathscr{Y}, U)$ be an ACCS and $v_x \in T_xQ$.*

1. *The system Σ is **accessible from** v_x if there exists $T > 0$ such that*

$$\operatorname{int} R_{\Sigma,TQ}(v_x, \leq t) \neq \emptyset$$

   for every $t \in (0, T]$.

2. *The system Σ is **configuration accessible from** v_x if there exists $T > 0$ such that $\operatorname{int} R_{\Sigma,Q}(v_x, \leq t) \neq \emptyset$ for every $t \in (0, T]$.*

3. *The system Σ is **small–time locally controllable from** v_x if there exists $T > 0$ such that $v_x \in \operatorname{int} R_{\Sigma,TQ}(v_x, \leq t)$ for every $t \in (0, T]$.*

4. *The system Σ is **small–time locally configuration controllable from** v_x if there exists $T > 0$ such that $x \in \operatorname{int} R_{\Sigma,Q}(v_x, \leq t)$ for every $t \in (0, T]$.*

Remark 6.2.3. Observe that if a system is accessible, then it is configuration accessible. Analogously, if the system is small–time locally controllable, then it is small–time locally configuration controllable. The converses are not necessarily true.

Remark 6.2.4. According to Ostrowski and Burdick [1997], the notion of small–time locally controllable only has sense if the initial velocity is assumed to be zero, otherwise we cannot guarantee that the trajectory stays in a neighbourhood of the initial condition. There exist results related with controllability for mechanical systems in the literature on control theory as long as the initial velocity is zero; see for instance [Lewis and Murray 1997].

Before proceeding, we need an assumption about the control set; for more details see [Bullo and Lewis 2005a;b;c]. In the following statement, $\operatorname{conv}(U)$ is the convex hull of the open control set (that is, the smallest convex set containing U), and $\operatorname{aff}(U)$ is the affine hull (that is, the smallest affine subspace of $\mathbb{R}^k$ containing U). See Appendix B for definitions of convexity and affinity.

Definition 6.2.5. *The control set $U \subset \mathbb{R}^k$ is **almost proper** if*

    1. $0 \in \mathrm{conv}(U)$, and

    2. $\mathrm{aff}(U) = \mathbb{R}^k$.

Using the notation in Chapter 3, this assumption on the control set guarantees that the span of the vector fields $\{f_0, f_1, \ldots, f_k\}$ defining a control–affine system is equal to the span of the vector fields $\{f_0 + \sum_{s=1}^{k} u^s f_s \mid u \in U\}$. This is useful to analyze the structure of the reachable set.

If the mechanical system is studied as a control–affine system on TQ; see Equation (6.1.2), then the notions of accessibility reviewed in §3.2 and described in [Nijmeijer and van der Schaft 1990, Sussmann and Jurdjevic 1972] can be used. Thus, according to Definition 3.2.4, the accessibility distribution of the system (6.1.2) is $\mathrm{Lie}^{\infty}\left(Z, Y_1^V, \ldots, Y_k^V\right)$. The system is accessible if

$$\mathrm{Lie}^{\infty}(Z, Y_1^V, \ldots, Y_k^V)_{v_x} = T_{v_x}TQ$$

because of Proposition 3.2.6.

At first, the mechanical system is defined on Q. Thus if the initial velocity is taken to be zero, there is a characterization of the accessibility in terms of constructions on Q without considering vector fields on the tangent bundle. In order to obtain results, it is necessary to study the *symmetric product* of two vector fields $X, Y \in \mathfrak{X}(Q)$ denoted by $\langle X : Y \rangle$ and defined as follows

$$\langle X : Y \rangle = \nabla_X Y + \nabla_Y X. \tag{6.2.3}$$

The geometric meaning of the symmetric product is the characterization of the geodesic invariance of a distribution; see [Bullo and Lewis 2005b] for more details.

Theorem 6.2.6. *([Bullo and Lewis 2005a, Cortés and Martínez 2003, Lewis and Murray 1997]) Let $\Sigma = (Q, \nabla, \mathscr{Y}, U)$ be an ACCS with the control set U being almost proper and $\mathrm{Sym}^{\infty}(\mathscr{Y})$ be the smallest distribution such that $\mathscr{Y} \subset \mathrm{Sym}^{\infty}(\mathscr{Y})$ and $\langle X : Y \rangle \in \Gamma\left(\mathrm{Sym}^{\infty}(\mathscr{Y})\right)$ for each $X, Y \in \Gamma\left(\mathrm{Sym}^{\infty}(\mathscr{Y})\right)$. If $\mathrm{Lie}^{\infty}(\mathrm{Sym}^{\infty}(\mathscr{Y}))_x = T_xQ$, then Σ is configuration accessible from 0_x. If $\mathrm{Sym}^{\infty}(\mathscr{Y})_x = T_xQ$, then Σ is accessible from 0_x.*

This result is related with the fact that

$$V_{0_x}(\tau_Q) = \mathrm{Sym}^{\infty}(\mathscr{Y})_x \text{ and } H_{0_x}(TQ) = \mathrm{Lie}^{\infty}\left(\mathrm{Sym}^{\infty}(\mathscr{Y})_x\right).$$

To try to generalize the results of accessibility for non–zero initial velocities, we compute some Lie brackets of the control vector fields and the drift vector field. In natural local coordinates (x, v) for TQ,

$$Z = v^i \frac{\partial}{\partial x^i} - \Gamma_{jl}^i v^j v^l \frac{\partial}{\partial v^i}, \quad Y^V = Y^i \frac{\partial}{\partial v^i}, \tag{6.2.4}$$

where $\Gamma_{jl}^i, Y^i \in \mathcal{C}^{\infty}(Q)$, and then we have

$$[Z, Y^V] = -Y^i \frac{\partial}{\partial x^i} + \left(v^i \frac{\partial Y^l}{\partial x^i} + Y^i v^j (\Gamma_{ij}^l + \Gamma_{ji}^l)\right) \frac{\partial}{\partial v^l}. \tag{6.2.5}$$

Proposition 6.2.7. *If Σ is an ACCS with the control vector fields $Y_1, \ldots, Y_k$ linearly independent, then the vector fields in the family $\{[Z, Y_r^V], Y_r^V\}_{r=1,\ldots,k}$ are linearly independent.*

(*Proof*) The proof of this result is due to the expression of the Lie bracket (6.2.5) and the hypothesis of linear independence of the control vector fields. The $(2n \times 2k)$–matrix given by the components of the family of vector fields $\{[Z, Y_1^V], \ldots, [Z, Y_k^V], Y_1^V, \ldots, Y_k^V\}$, according to Equation (6.2.5), is

$$\begin{pmatrix} (-Y_r^i)_{ir} & 0 \\ \left(v^j \frac{\partial Y_r^i}{\partial x^j} + Y_r^s v^j (\Gamma_{sj}^i + \Gamma_{js}^i) \right)_{ir} & (Y_r^i)_{ir} \end{pmatrix} \tag{6.2.6}$$

for $i = 1, \ldots, n$, $r = 1, \ldots, k$ and $0 \in M_{n \times k}$. Thus, we conclude that the matrix (6.2.6) has maximum rank. ∎

Remark 6.2.8. Observe that given a family of vector fields on Q and an affine connection on Q, if the involutive distribution containing this family spans the whole tangent space on Q, then the vertical lift of those vector fields and their Lie bracket with the geodesic spray span the tangent space of TQ.

As a result of the previous remark and Proposition 6.2.7 the following result is immediately proved.

Corollary 6.2.9. *Let $\Sigma = (Q, \nabla, \mathscr{Y}, U)$ be an ACCS with an almost proper control set U and $v_x \in TQ$. If $\mathrm{Lie}^\infty(\mathscr{Y})_x = T_xQ$, then Σ is accessible from v_x.*

Thus, Corollary 6.2.9 characterizes the accessibility of an affine connection control system even from a nonzero velocity. It would be interesting to give more specific results analogous in some sense to Theorem 6.2.6 as explained in Chapter 9.

6.3 Optimal control problem for affine connection control systems

As defined in §6.1, ACCS are given by $\Sigma = (Q, \nabla, \mathscr{Y}, U)$ where Q is a smooth n–dimensional manifold called the *configuration manifold*, ∇ is an affine connection on Q, $\mathscr{Y}$ is the set of input vector fields $\{Y_1, \ldots, Y_k\}$ and U is a set in $\mathbb{R}^k$. After studying these systems in control theory, let us associate them with an optimal control problem.

From Chapter 4, it is known that there are two different classes of optimal control problems: fixed time and nonfixed time, depending on whether the interval of definition of the curves is given or not. However, for mechanical systems there are more possible statements because the endpoint conditions can be in Q—that is, without restricting the velocities—or in TQ; see §6.1 for the reinterpretation of these systems as control–affine systems on TQ. Thus, it makes sense that the cost function depends on the velocities.

The assumptions for all the elements in the following problems are analogous to the assumptions considered in §2.2.1 and §4.1.1.

If $\mathcal{F}\colon TQ \times U \to \mathbb{R}$ is the cost function, we have the functional

$$\mathcal{S}[\Upsilon, u] = \int_I \mathcal{F}(\Upsilon(t), u(t))\, \mathrm{d}t$$

defined on curves $(\Upsilon, u)\colon I \to TQ \times U$ where $I = [a, b]$ is a compact interval in $\mathbb{R}$. Consider the following problems, where $\tau_Q\colon TQ \to Q$ is the natural projection of the tangent bundle:

Statement 6.3.1. *(Optimal control problem for ACCS without velocity endpoint conditions)*
Given $(Q, \nabla, \mathscr{Y}, U)$, $\mathcal{F}$, $I = [a, b]$, $x_a, x_b \in Q$. Find $(\gamma, u)\colon I \to Q \times U$ such that for a curve $\Upsilon\colon I \to TQ$ satisfying $\tau_Q \circ \Upsilon = \gamma$,

(1) $\gamma(a) = x_a$, $\gamma(b) = x_b$,

(2) $\dot{\Upsilon}(t) = Z(\Upsilon(t)) + u^s(t) Y_s^V(\Upsilon(t))$, and

(3) $\mathcal{S}[\Upsilon, u] = \int_I \mathcal{F}(\Upsilon(t), u(t))\mathrm{d}t$ is minimum over all curves on $TQ \times U$ satisfying (1) and (2).

Statement 6.3.2. *(Optimal control problem for ACCS with velocity endpoint conditions)*
Given $(Q, \nabla, \mathscr{Y}, U)$, $\mathcal{F}$, $I = [a, b]$, $v_a \in T_{x_a}Q$, $v_b \in T_{x_b}Q$. Find $(\gamma, u)\colon I \to Q \times U$ such that for a curve $\Upsilon\colon I \to TQ$ satisfying $\tau_Q \circ \Upsilon = \gamma$,

(1) $\Upsilon(a) = v_a$, $\Upsilon(b) = v_b$,

(2) $\dot{\Upsilon}(t) = Z(\Upsilon(t)) + u^s(t) Y_s^V(\Upsilon(t))$, and

(3) $\mathcal{S}[\Upsilon, u] = \int_I \mathcal{F}(\Upsilon(t), u(t))\mathrm{d}t$ is minimum over all curves on $TQ \times U$ satisfying (1) and (2).

The tuple $\Sigma_{\mathcal{F}} = (Q, \nabla, \mathscr{Y}, U, \mathcal{F}, I)$ denotes a fixed time optimal control problem. The notation $\Sigma_{\mathcal{F}}(x_a, x_b)$ is used when the velocity endpoint conditions are not restricted and $\Sigma_{\mathcal{F}}(v_a, v_b)$ when they are restricted.

Observe that the condition (2) in both previous Statements 6.3.1 and 6.3.2 implies that Υ is the natural lift of γ, that is, $\Upsilon = \dot{\gamma}$.

It is possible to define *free optimal control problems for ACCS with(out) velocity endpoint conditions*; analogous to the problem in Statement 4.3.1. Remember that in such problems the interval I is also an unknown and it must be computed.

6.4 Intrinsic mechanical Pontryagin's Maximum Principle

It has already been mentioned that an important resource in optimal control theory is Pontryagin's Maximum Principle, in a both geometric and computational sense. We are interested in studying the necessary conditions for optimality in mechanical control systems such as the ones defined in §6.1. The solution to the problem is on the manifold $TQ \times U$. Then, as we saw in Chapter 4, Pontryagin's Hamiltonian is a real–valued function on $T^*TQ \times U$. Particular

splittings of T^*TQ and other tangent and cotangent bundles are explained in §6.4.1 because they are necessary to state intrinsically the mechanical Pontryagin's Maximum Principle. In §6.4.2 an intrinsic version of mechanical Pontryagin's Maximum Principle is given, according to Bullo and Lewis [2005b].

The contents in this section are mainly a review of the intrinsic Pontryagin's Maximum Principle for the mechanical case so that this dissertation describes as complete as possible the geometric information related with Pontryagin's Maximum Principle.

6.4.1 Useful splittings

The technique of using splittings to the benefit of the geometry of the control system has been used previously in [Lewis and Murray 1997] to study the controllability of ACCS starting at zero velocity. Keeping in mind §2.4.4, it is possible to define the setting necessary for stating the mechanical Pontryagin's Maximum Principle.

6.4.1.1 Linear connection on $\tau_{\mathbf{Q}} \colon \mathbf{TQ} \to \mathbf{Q}$

The notion of an induced Ehresmann connection on $\tau_M \colon TM \to M$ has been defined in §2.4.4. This connection is associated with a semi–spray; that is, a vector field satisfying the second–order condition, Definition 2.4.11. The affine connection on Q defines a linear connection on τ_Q associated with the geodesic spray.

Definition 6.4.1. *A vector field* $S \colon TQ \to TTQ$ *is a* **spray** *if it satisfies the second–order condition and* $\mathcal{L}_\Delta S = S$, *where* $\mathcal{L}_\Delta$ *denotes the Lie derivative with respect to the Liouville vector field* Δ *in Equation (2.4.15).*

By Euler's Theorem for the homogeneous functions, in natural local coordinates (x, v) for TQ, a spray is given by

$$S(x,v) = v^i \left. \frac{\partial}{\partial x^i} \right|_{(x,v)} + S^i(x,v) \left. \frac{\partial}{\partial v^i} \right|_{(x,v)},$$

where S^i is a homogeneous function of degree 2 with respect to v:

$$S^i(x,v) = S^i_{jl}(x)v^j v^l.$$

When S^i_{jl} are the Christoffel symbols of the affine connection ∇, S is the geodesic spray Z, cf. Equation (6.2.4).

According to Propositions 2.4.5 and 2.4.12, the geodesic spray defines a linear connection ∇ on τ_Q as in Equation (2.4.12). This connection defines a splitting of the tangent bundle into horizontal subbundle and vertical subbundle:

$$T(TQ) = H(TQ) \oplus V(\tau_Q).$$

At each point $v_x \in TQ$, we have $T_{v_x}(TQ) = H_{v_x}(TQ) \oplus V_{v_x}(\tau_Q)$ and a basis for the subspaces are

$$H_{v_x}(TQ) \oplus V_{v_x}(\tau_Q) = \left\langle \frac{\partial}{\partial x^i} - \frac{1}{2}(\Gamma^j_{il} + \Gamma^j_{li})v^l \frac{\partial}{\partial v^j} \right\rangle_{v_x} \oplus \left\langle \frac{\partial}{\partial v^i} \right\rangle_{v_x}.$$

In order to simplify the forthcoming computations, the connection is assumed to be the Levi–Civita connection; that is, symmetric and compatible with the metric. The symmetry of the connection implies zero torsion, $\Gamma^j_{li} = \Gamma^j_{il}$. Then, the associated subspaces at $v_x \in TQ$ are spanned by

$$H_{v_x}(TQ) = \left\langle \frac{\partial}{\partial x^i} - \Gamma^l_{ij}v^j \frac{\partial}{\partial v^l} \right\rangle_{v_x}, \quad V_{v_x}(\tau_Q) = \left\langle \frac{\partial}{\partial v^i} \right\rangle_{v_x}.$$

6.4.1.2 Splitting of T^*TQ according to the linear connection on τ_Q

From the linear connection on $\tau_Q \colon TQ \to Q$ associated with the affine connection ∇ on Q, it is possible to give a splitting of T^*TQ as defined in §2.4.3. Recall that $T(TQ) = H(TQ) \oplus V(\tau_Q)$ so that

$$T^*(TQ) = (H(TQ))^* \oplus (V(\tau_Q))^*,$$

where $(H(TQ))^*$ and $(V(\tau_Q))^*$ are the dual subbundles of $H(TQ)$ and $V(\tau_Q)$, respectively.

For every $v_x \in TQ$, there exists the following isomorphism

$$T^*_{v_x}TQ = (H_{v_x}(TQ))^* \oplus (V_{v_x}(\tau_Q))^* \simeq T^*_x Q \oplus T^*_x Q.$$

Observe that the elements in $(H_{v_x}(TQ))^*$ annihilate the elements in $H_{v_x}(TQ)$ and the ones in $(V_{v_x}(\tau_Q))^*$ annihilate the elements in $V_{v_x}(\tau_Q)$. Local bases of the two subspaces of $T^*_{v_x}TQ$ are

$$(H_{v_x}(TQ))^* = \left\langle \mathrm{d}x^i \right\rangle_{v_x}, \quad (V_{v_x}(\tau_Q))^* = \left\langle \mathrm{d}v^i + \frac{1}{2}(\Gamma^i_{jl} + \Gamma^i_{lj})v^l \mathrm{d}x^j \right\rangle_{v_x},$$

for $i = 1, \ldots, n$, as a result of Equation (6.4.1.1).

6.4.1.3 Dual of a linear connection on $\tau_Q \colon TQ \to Q$ associated with an affine connection on Q

As described in §2.4.5, the dual of a linear connection on $\tau_Q \colon TQ \to Q$ is indeed a linear connection on $\pi_Q \colon T^*Q \to Q$.

Remember that a connection is defined once the horizontal subbundle is given; see Proposition 2.4.1. Due to Lemma 2.4.10, the horizontal subbundle of the linear connection in which we are interested is given by the dual of the horizontal lift of vector fields on Q to TQ via $\mathrm{h} \colon \mathfrak{X}(Q) \to H(TQ)$, see (2.4.13).

Hence, $T(T^*Q) = H(T^*Q) \oplus V(\pi_Q)$. Observe that for every $x \in Q$,

$$
\begin{array}{ccccc}
T^*_x Q & \overset{j_x}{\hookrightarrow} & T^*Q & \overset{\pi_Q}{\to} & Q \\
p & \mapsto & p_x & \mapsto & x
\end{array}
$$

where j_x is the inclusion map. As $\pi_Q \circ j_x$ is a constant function, we have $T_x^* Q \simeq V_{p_x}(\pi_Q)$. So it makes sense to define the vertical lift $(\mathrm{v}^*)_x^{p_x} : T_x^* Q \to V_{p_x}(\pi_Q)$ with local expression

$$(\mathrm{v}^*)_x^{p_x} (\mathrm{d}x_x^i) = \left. \frac{\partial}{\partial p_i} \right|_{p_x} .$$

Hence for every $p_x \in T^* Q$,

$$T_{p_x} T^* Q = H_{p_x}(T^* Q) \oplus V_{p_x}(\pi_Q) \simeq T_x Q \oplus T_x^* Q.$$

In local coordinates (x, p) adapted to $\pi_Q \colon T^* Q \to Q$, for every $p_x \in T^* Q$,

$$H_{p_x}(T^* Q) = \left\langle \frac{\partial}{\partial x^i} + \frac{1}{2} (\Gamma_{il}^j + \Gamma_{li}^j) p_j \frac{\partial}{\partial p_l} \right\rangle_{p_x}, \quad V_{p_x}(\pi_Q) = \left\langle \frac{\partial}{\partial p_i} \right\rangle_{p_x}, \quad \text{for } i = 1, \dots, n.$$

6.4.1.4   Linear connection on $\tau_{\mathbf{TQ}} \colon \mathbf{TTQ} \to \mathbf{TQ}$ associated with an affine connection on Q

The construction is similar to the one in §6.4.1.1, so we need a vector field on TQ that satisfies the second–order condition. A first try would be consider the complete lift Z^T of the geodesic spray associated with the affine connection ∇ on Q —see §2.2.2.1 for more details about this lift—but it does not satisfy the second–order condition as can be seen from the local expression.

Lemma 6.4.2. *Let Z be the geodesic spray associated with the affine connection ∇ on Q and $\kappa_Q \colon TTQ \to TTQ$ be the canonical involution (that is, $\kappa_Q \circ \kappa_Q = \mathrm{Id}_{TTQ}$ and $\tau_{TQ} \circ \kappa_Q = T\tau_Q$). The κ_Q–pushforward vector field $(\kappa_Q)_* Z^T$ of the complete lift of the geodesic spray satisfies the second–order condition on TTQ.*

(Proof) The vector field $(\kappa_Q)_* Z^T \in \mathfrak{X}(TTQ)$ satisfies the second–order condition on TTQ if, for every $W \in TTQ$,

$$T_W \tau_{TQ} \left[((\kappa_Q)_* Z^T)_W \right] = \mathrm{Id}(W).$$

Let $f \in \mathcal{C}^\infty(TQ)$. By the properties of the canonical involution, we have

$$T_W \tau_{TQ} \left[((\kappa_Q)_* Z^T)_W \right] f = ((\kappa_Q)_* Z^T)_W (f \circ \tau_{TQ})$$
$$= T_{\kappa_Q^{-1}(W)} \kappa_Q \left(Z^T_{\kappa_Q^{-1}(W)} \right) (f \circ \tau_{TQ}) = T_{\kappa_Q(W)} \kappa_Q \left(Z^T_{\kappa_Q(W)} \right) (f \circ \tau_{TQ})$$
$$= T_W \tau_{TQ} \left(\left[T_{\kappa_Q(W)} \kappa_Q \left(Z^T_{\kappa_Q(W)} \right) \right]_W \right) f = T_{\kappa_Q(W)} (\tau_{TQ} \circ \kappa_Q) \left(Z^T_{\kappa_Q(W)} \right) f$$
$$= T_{\kappa_Q(W)} (T\tau_Q) \left(Z^T_{\kappa_Q(W)} \right) f.$$

Then,

$$T_W \tau_{TQ} \left[((\kappa_Q)_* Z^T)_W \right] = T_{\kappa_Q(W)} (T\tau_Q) \left(Z^T_{\kappa_Q(W)} \right)$$
$$= \left((T\tau_Q \circ \kappa_Q)(W), D_{\kappa_Q(W)} \left(T\tau_Q Z^T_{\kappa_Q(W)} \right) \right) = W.$$

∎

According to Propositions 2.4.5 and 2.4.12 the vector field $(\kappa_Q)_* Z^T$ defines a linear connection on τ_{TQ} as in Equation (2.4.12) such that

$$T(TTQ) = H(TTQ) \oplus V(\tau_{TQ})$$

and, for every $W_{v_x} \in T_{v_x}TQ$, we have the isomorphism

$$T_{W_{v_x}}TTQ = H_{W_{v_x}}(TTQ) \oplus V_{W_{v_x}}(\tau_{TQ}) \simeq T_{v_x}TQ \oplus T_{v_x}TQ.$$

The linear connection on τ_Q then gives the isomorphism

$$T_{W_{v_x}}TTQ \simeq \overbrace{\overbrace{T_xQ}^{H_{v_x}(TQ)} \oplus \overbrace{T_xQ}^{V_{v_x}(\tau_Q)}}^{H_{W_{v_x}}(TTQ)} \oplus \overbrace{\overbrace{T_xQ}^{H_{v_x}(TQ)} \oplus \overbrace{T_xQ}^{V_{v_x}(\tau_Q)}}^{V_{W_{v_x}}(\tau_{TQ})} .$$

Thus, from vector at x on Q it is possible to define the lifts to vector in $T_{W_{v_x}}TTQ$ step by step using different connections; see [Bullo and Lewis 2005b] for a local expression of the basis of this last splitting.

6.4.1.5 Dual of a linear connection on $\tau_{\mathbf{TQ}} \colon \mathbf{TTQ} \to \mathbf{TQ}$ associated with an affine connection on $\mathbf{Q}$

As before, let us construct the dual of a linear connection as explained in Lemma 2.4.10. In other words, the linear connection on $\pi_{TQ} \colon T^*TQ \to TQ$ is defined by the horizontal subbundle given by the dual of the horizontal lift of vector fields on TQ to TTQ via $\mathrm{h}^T \colon \mathfrak{X}(TQ) \to \mathfrak{X}(TTQ)$ associated to the connection defined in §6.4.1.4 and defined analogously to (2.4.13).

Thus there exists a splitting $T(T^*TQ) = H(T^*TQ) \oplus V(\pi_{TQ})$ and, for every $\Lambda_{v_x} \in T^*TQ$,

$$T_{\Lambda_{v_x}}T^*TQ \simeq H_{\Lambda_{v_x}}(T^*TQ) \oplus V_{\Lambda_{v_x}}(\pi_{TQ}) \simeq T_{v_x}TQ \oplus T_{v_x}^*TQ.$$

The connection on $\tau_Q \colon TQ \to Q$ defined in §6.4.1.1 and the splitting of $T_{v_q}^*TQ$ in §6.4.1.2 allows us to write

$$T_{\Lambda_{v_x}}T^*TQ \simeq \overbrace{\overbrace{T_xQ}^{H_{v_x}(TQ)} \oplus \overbrace{T_xQ}^{V_{v_x}(\tau_Q)}}^{H_{\Lambda_{v_x}}(T^*TQ)} \oplus \overbrace{\overbrace{T_x^*Q}^{(H_{v_x}(TQ))^*} \oplus \overbrace{T_x^*Q}^{(V_{v_x}(\tau_Q))^*}}^{V_{\Lambda_{v_x}}(\pi_{TQ})} .$$

Thus tangent vectors in T_xQ can be naturally lifted to tangent vectors in $T_{\Lambda_{v_x}}T^*TQ$.

6.4.2 Intrinsic Pontryagin's Maximum Principle for affine connection control systems

We continue with the definition of all the geometric elements necessary for the suitable statement of Pontryagin's Maximum Principle in this chapter. We refer the reader to [Bullo and

Lewis 2005b] for a thorough study of this topic. Associated with the affine connection ∇ on Q defining the ACCS, we have the torsion tensor T; that is, the $(1,2)$–tensor field on Q defined by

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y],$$

and the curvature tensor R; that is, the $(1,3)$–tensor field on Q defined by

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]}Z,$$

for $X, Y, Z \in \mathfrak{X}(Q)$. Let us define the following tensor fields:

$$
\begin{aligned}
T^* &: \Omega^1(Q) \times \mathfrak{X}(Q) \to \Omega^1(Q), & T^*(\alpha, X)(Z) &= \langle \alpha, T(X,Z) \rangle, \\
R^* &: \Omega^1(Q) \times \mathfrak{X}(Q) \times \mathfrak{X}(Q) \to \Omega^1(Q), & R^*(\alpha, X, Y)(Z) &= \langle \alpha, R(X,Y)Z \rangle,
\end{aligned}
$$

for $Z \in \mathfrak{X}(Q)$. For any vector field $Y \in \mathfrak{X}(Q)$, we have

$$(\nabla Y)^* : \Omega^1(Q) \to \Omega^1(Q), \quad (\nabla Y)^*(\alpha)(X) = \langle \alpha, \nabla_X Y \rangle,$$

for $X \in \mathfrak{X}(Q)$.

Definition 6.4.3. *Given the affine connection control system $\Sigma = (Q, \nabla, \mathscr{Y}, U)$, a U–dependent $(0,r)$ tensor field on Q is a map $A \colon Q \times U \to T^0_r(Q)$ such that A is continuous and $x \mapsto A(x,u)$ is a smooth $(0,r)$–tensor field for every $u \in U \subset \mathbb{R}^k$. Analogously, a U–dependent $(r,0)$ tensor field on Q may be defined.*

Given $\mathcal{A} = (A, f)$ where A is a U–dependent $(0,r)$–tensor field on Q and $f \in \mathcal{C}^\infty(\mathbb{R})$, we define a *cost function for* $\Sigma = (Q, \nabla, \mathscr{Y}, U)$ *associated with* $\mathcal{A}$ as a function $\mathcal{F}_\mathcal{A} \colon TQ \times U \to \mathbb{R}$ such that

$$\mathcal{F}_\mathcal{A}(v_x, u) = f(A(x,u)(v_x, \overset{r}{\ldots}, v_x)). \tag{6.4.7}$$

For $v_1, \ldots, v_{r-1} \in T_x Q$, we define $\hat{A}(v_1, \ldots, v_{r-1}) \in T^*_x Q$ by

$$\left\langle \hat{A}(v_1, \ldots, v_{r-1}); v \right\rangle = A(v, v_1, \ldots, v_{r-1})$$

for $v \in T_x Q$. If B is a symmetric $(r,0)$–tensor field and $\alpha^1, \ldots, \alpha^{r-1} \in T^*_x Q$, then we define $\hat{B}(\alpha^1, \ldots, \alpha^{r-1}) \in T_x Q$ by

$$\left\langle \beta, \hat{B}(\alpha^1, \ldots, \alpha^{r-1}) \right\rangle = B\left(\beta, \alpha^1, \ldots, \alpha^{r-1}\right)$$

for $\beta \in T^*_x Q$.

Pontryagin's Hamiltonian for the ACCS Σ is given by $H \colon T^*(TQ) \times U \to \mathbb{R}$,

$$H(\Lambda_{v_x}, u) = \langle \Lambda_{v_x}, Z(v_x) + u^s Y^V_s(v_x) \rangle + p_0 \mathcal{F}_\mathcal{A}(v_x, u), \tag{6.4.8}$$

with $\Lambda_{v_x} \in T^*_{v_x} TQ$.

The splitting of $T^*_{v_x} TQ$ defined in §6.4.1.2 and the splitting of $T_{v_x} TQ$ in §6.4.1.1 enable us

to write the Hamiltonian as

$$H(\alpha_{v_x} \oplus \beta_{v_x}, u) = \langle \alpha_{v_x} \oplus \beta_{v_x}, v \oplus (u^s Y_s(x)) \rangle + p_0 \mathcal{F}_\mathcal{A}(v_x, u), \qquad (6.4.9)$$

where $\alpha_{v_x} \oplus \beta_{v_x} \in (H_{v_x}(TQ))^* \oplus (V_{v_x}(\tau_Q))^*$ with local expression

$$H(\alpha_{v_x} \oplus \beta_{v_x}, u) = \alpha_i v^i + \beta_i Y_s^i u^s + p_0 \mathcal{F}_\mathcal{A}(v_x, u).$$

Theorem 6.4.4. *(Pontryagin's Maximum Principle for Affine Connection Control Systems, [Bullo and Lewis 2005b]) Let $\Sigma = (Q, \nabla, \mathscr{Y}, U)$ be an affine connection control system with the cost function $\mathcal{F}_\mathcal{A}$ where $\mathcal{A} = (A, f)$. If $(\gamma, u) \colon I \to Q \times U$ is a solution of the fixed time optimal control problem in Statement 6.3.2 with velocity endpoint conditions $v_a \in T_{x_a}Q$, $v_b \in T_{x_b}Q$. Then there exists a smooth covector field $\lambda \colon I \to T^*Q$ along γ and a constant $p_0 \in \{-1, 0\}$ such that*

(i) for almost every $t \in I$ we have

$$\begin{aligned}
&\nabla^2_{\dot\gamma(t)}\lambda(t) + R^*(\lambda(t), \dot\gamma(t))\dot\gamma(t) - T^*(\nabla_{\dot\gamma(t)}\lambda(t), \dot\gamma(t)) \\
&= u^a(t)(\nabla Y_a)^*(\lambda(t)) - \lambda_0 f'(A(\dot\gamma(t), \ldots, \dot\gamma(t)))(\nabla A(\dot\gamma(t), \ldots, \dot\gamma(t)) \\
&\quad -r(\nabla_{\dot\gamma(t)}\hat{A})(\dot\gamma(t), \ldots, \dot\gamma(t)) - r(r-1)u(t)\hat{A}(Y_a(\gamma(t)), \dot\gamma(t), \ldots, \dot\gamma(t)) \\
&\quad +rT^*(\hat{A}(\dot\gamma(t), \ldots, \dot\gamma(t)), \dot\gamma(t))) \\
&\quad +r\lambda_0 f''(A(\dot\gamma(t), \ldots, \dot\gamma(t)))(\nabla A(\dot\gamma(t), \ldots, \dot\gamma(t); \dot\gamma(t)) \\
&\quad +ru^a(t)A(Y_a(\gamma(t)), \dot\gamma(t), \ldots, \dot\gamma(t)))\hat{A}(\dot\gamma(t), \ldots, \dot\gamma(t));
\end{aligned}$$

(ii) $H(\theta(t) \oplus \lambda(t), u(t)) = \sup_{w \in U} H(\theta(t) \oplus \lambda(t), w(t))$ where

$$\begin{aligned}
\theta(t) &= \tfrac{1}{2}T^*(\lambda(t), \dot\gamma(t)) - \nabla_{\dot\gamma(t)}\lambda(t) \\
&\quad - r\lambda_0 f'(A(\dot\gamma(t), \ldots, \dot\gamma(t)))\hat{A}(\dot\gamma(t), \ldots, \dot\gamma(t)),
\end{aligned} \qquad (6.4.10)$$

for $t \in I$;

(iii) either $p_0 = -1$ or $\theta(a) \oplus \lambda(a) \neq 0$;

(iv) $H(\theta(t) \oplus \lambda(t), u(t))$ is constant almost everywhere.

If (γ, u) is a solution of a free optimal control problem for ACCS with velocity endpoint conditions, $(i) - (iii)$ are satisfied and (iv) is replaced by $H(\theta(t) \oplus \lambda(t), u(t)) = 0$ almost everywhere.

If (γ, u) is a solution of the fixed time optimal control problem in Statement 6.3.1 without velocity endpoint conditions, $(i) - (iv)$ are satisfied and also $\lambda(a) = \lambda(b) = 0$.

If (γ, u) is a solution of a free optimal control problem for ACCS without velocity endpoint conditions, $(i) - (iii)$ are satisfied (iv) is replaced by $H(\theta(t) \oplus \lambda(t), u(t)) = 0$ almost everywhere, and it is also satisfied that $\lambda(a) = 0$ and $\lambda(b) = 0$.

The proof of this theorem consists of using Theorem 4.1.14 plus all the splittings explained in [Bullo and Lewis 2005b, Coombs 2000] and briefly reviewed in §6.4.1.

When the velocity endpoint conditions are not restricted and endpoint conditions on the configuration manifold are fixed, the transversality conditions in Theorem 4.3.13 are

$$\theta(a) \oplus \lambda(a) \in \operatorname{ann} T_{\dot\gamma(a)}(\{x_a\} \times T_{x_a}Q), \quad \theta(b) \oplus \lambda(b) \in \operatorname{ann} T_{\dot\gamma(b)}(\{x_b\} \times T_{x_b}Q).$$

These conditions imply that $\lambda(a) = \lambda(b) = 0$.

Remark 6.4.5. Observe that once the momenta associated to the vertical subbundle is known, the momenta of the horizontal subbundle is also determined independently of the controls because of Equation (6.4.10).

The necessary conditions in Theorem 6.4.4 define again different extremals, in the same way as in Definition 4.1.15.

6.5   Weak mechanical Pontryagin's Maximum Principle

The optimal control problem defined in §6.3 can be also solved through the method described in Chapter 5 as long as the control system is rewritten as a control–affine system on TQ—see Equation (6.1.2)—the control set U is open and all the elements are differentiable enough. The particular expressions of the drift vector field (the geodesic spray) and the control vector fields (being vertical lift of vector fields on Q) give a nice geometric description of the constraint submanifolds, particularly in the abnormal case. The geometry will be defined by vector–valued quadratic forms; see Appendix C and [Bullo and Lewis 2005a, Hirschorn and Lewis 2002] for more details.

As in §5.1, to state Pontryagin's Maximum Principle we use a presymplectic framework, in this case given by the presymplectic Hamiltonian system $(T^*(TQ) \times U, \Omega, H)$ where Ω is the pullback of the canonical 2–form on $T^*(TQ)$ through $\pi_1 \colon T^*(TQ) \times U \to T^*(TQ)$ and the Hamiltonian is given in Equation (6.4.8). For more details on the presymplectic formalism see §2.3.2 and [Cariñena 1990, Gotay and Nester 1979, Gotay et al. 1978, Gràcia and Pons 1992, Muñoz-Lecanda and Román-Roy 1992] and for the specific use in optimal control theory see Chapter 5 and [Barbero-Liñán and Muñoz Lecanda 2008a, Delgado-Téllez and Ibort 2003, Martínez 2004].

Theorem 6.5.1. *(Weak Pontryagin's Maximum Principle for ACCS)*
If $(\Upsilon, u) \colon [a, b] \to TQ \times U$ is a solution of the optimal control problem in Statement 6.3.1, $\Sigma_{\mathcal{F}}(x_a, x_b)$, then there exist $\Lambda \colon [a, b] \to T^*(TQ)$ along Υ, and $p_0 \in \{-1, 0\}$ such that:

1. *(Λ, u) is an integral curve of a Hamiltonian vector field X_H defined by*

$$i_{X_H}\Omega = \mathrm{d}H; \tag{6.5.11}$$

   that is, (Λ, u) satisfies $i_{\dot\Lambda(t)}\Omega = \mathrm{d}H(\Lambda(t), u(t))$;

2. *(a) $\sup_{\widetilde{u}\in U} H(\Lambda(t), \widetilde{u})$ is constant everywhere in $t \in [a, b]$;*

   (b) $(p_0, \Lambda(t)) \neq 0$ for each $t \in [a, b]$.

As Ω is a degenerate 2–form, the presymplectic equation (6.5.11) does not have solution in the entire manifold $T^*(TQ) \times U$, exactly the same as in §2.3.2 and §5.1. The equation has solutions if we restrict it to the primary submanifold defined by

$$N_0 = \{\beta \in T^*(TQ) \times U \mid i_v \mathrm{d}H = 0, \text{ for } v \in \ker \Omega_\beta\}\,.$$

Remark 5.1.2 explains that this submanifold is defined implicitly by the necessary condition for the maximization of the Hamiltonian over the controls along the optimal solution. Remember that the maximum can be used instead of the supremum as explained in comment 4 after Theorem 4.1.14.

The necessary conditions of Theorem 6.5.1 determine different kinds of extremals as in Definition 4.1.15. Among all the different extremals, there are abnormal and singular extremals. As the control system studied here is a control–affine system, the next result holds.

Proposition 6.5.2. *For an optimal control problem $Sigma_{\mathcal{F}}$ for ACCS, the abnormal extremals (Υ, u) on $TQ \times U$; i.e. the extremals satisfying the necessary conditions in Theorem 6.5.1 for $p_0 = 0$; are always singular extremals.*

(*Proof*) If (Υ, u) is an abnormal extremal, then the Hamiltonian to be considered is $H^{[0]} = H_Z + u^s H_{Y_s^V}$, and the presymplectic equation (6.5.11) has a solution when restricted to the points satisfying

$$\frac{\partial H^{[0]}}{\partial u^s} = H_{Y_s^V} = 0, \text{ for } s = 1, \ldots, k.$$

Thus the Hamiltonian becomes $H^{[0]} = H_Z$ and it does not depend on the controls. Neither, therefore, does the maximum. Then the abnormal extremals are singular, see Definition 4.1.15. ∎

Remark 6.5.3. In fact, an abnormal extremal is also singular provided that the system is control–affine, as in ACCS [Bullo and Lewis 2005b] or, in particular, for control–linear systems as in subRiemannian geometry [Liu and Sussmann 1995, Montgomery 1994]. As pointed out in §3.1, a control–linear system is a particular case of a control–affine system.

6.6 Constraint algorithm for extremals in optimal control problems for affine connection control systems

Through the techniques described in Chapter 5 and [Barbero-Liñán and Muñoz Lecanda 2008a] we determine where the dynamics of normal and abnormal extremals take place for the mechanical control systems considered in this chapter.

We have the presymplectic Hamiltonian system $(T^*(TQ) \times U, \Omega, H)$ with the Hamiltonian function $H = H_Z + u^s H_{Y_s^V} + p_0 \mathcal{F}$, where $H_Z(\Lambda) = \langle \Lambda, Z \rangle$ and similarly for $H_{Y_s^V}$.

The presymplectic equation (6.5.11), $i_{X_H} \Omega = \mathrm{d}H$, has a solution in the primary constraint

submanifold

$$N_0 = \left\{ (\Lambda, u) \in T^*(TQ) \times U \, \middle| \, \frac{\partial H}{\partial u^s} = H_{Y_s^V} + p_0 \frac{\partial \mathcal{F}}{\partial u^s} = 0, \quad s = 1, \dots, k \right\}. \quad (6.6.12)$$

The tangency condition of the Hamiltonian vector field X_H to N_0 on N_0 defines

$$N_1 = \left\{ (\Lambda, u) \in N_0 \, \middle| \, X_H \left(\frac{\partial H}{\partial u^s} \right) = 0, \quad s = 1, \dots, k \right\}. \quad (6.6.13)$$

Successive stabilization steps are done until a final constraint submanifold is found. All the propositions in §5.2 related with the different kinds of extremals are also valid here, included the characterization of the strict abnormality in Proposition 5.2.9. In this chapter more specific details about strict abnormality will be given in §6.7.1, §6.7, but before that we proceed with the general expression of the constraints for abnormality and normality in optimal control problems for ACCS.

6.6.1   Characterization of abnormality

First we characterize a subset of $T^*(TQ) \times U$ where the abnormal biextremals evolve when they exist. In this situation we take $p_0 = 0$ and the corresponding Pontryagin's Hamiltonian is

$$H^{[0]} = H_Z + u^s H_{Y_s^V}.$$

The primary constraint submanifold (6.6.12) becomes

$$N_0^{[0]} = \left\{ (\Lambda, u) \in T^*(TQ) \times U \, \middle| \, H_{Y_s^V} = 0, \quad s = 1, \dots, k \right\}. \quad (6.6.14)$$

Thus the abnormal biextremals must lie in the annihilator of the distribution spanned by $\mathscr{Y}^V$. Note that $N_0^{[0]}$ is a submanifold defined by the zeroes of k independent functions on T^*TQ if the k control vector fields are linear independent.

Remark 6.6.1. The Hamiltonian vector field on $N_0^{[0]}$ is $X_{H^{[0]}} = X_Z + u^s X_{Y_s^V}$, where X_Z denotes the Hamiltonian vector field associated to the Hamiltonian function H_Z, and analogously for $X_{Y_s^V}$.

The stabilization step consists of guaranteeing the tangency of the Hamiltonian vector field $X_{H^{[0]}}$ to $N_0^{[0]}$. Then the submanifold (6.6.13) becomes

$$N_1^{[0]} = \left\{ (\Lambda, u) \in N_0^{[0]} \, \middle| \, X_{H^{[0]}}(\Lambda, u) \in T_{(\Lambda, u)} N_0^{[0]} \right\}.$$

This tangency condition is

$$
\begin{aligned}
0 &= X_{H^{[0]}}(H_{Y_s^V}) = \mathrm{d}H_{Y_s^V}(X_{H^{[0]}}) = \mathrm{d}H_{Y_s^V}(X_Z + u^r X_{Y_r^V}) \\
&= -\{H_Z, H_{Y_s^V}\} - u^r \{H_{Y_r^V}, H_{Y_s^V}\} = H_{[Z, Y_s^V]} + u^r H_{[Y_r^V, Y_s^V]} \\
&= H_{[Z, Y_s^V]}, \quad (6.6.15)
\end{aligned}
$$

where properties of Poisson bracket explained in [Abraham and Marsden 1978, pp. 192] have

been used, and also the fact that Lie brackets of vertical lift of vector fields on Q are always zero. Then,

$$N_1^{[0]} = \left\{ (\Lambda, u) \in N_0^{[0]} \,\middle|\, H_{[Z, Y_s^V]} = 0, \; s = 1, \ldots, k \right\}. \tag{6.6.16}$$

This submanifold is defined as the zeroes of $2k$ independent functions by Proposition 6.2.7. That is, k new independent constraints have been added if the k control vector fields are linear independent.

The vector field $X_{H^{[0]}}$ restricted to $N_1^{[0]}$ is tangent to $N_0^{[0]}$, but we would like that $X_{H^{[0]}}$ is tangent to $N_1^{[0]}$. The next step of the constraint algorithm is similar to the previous one. We define

$$N_2^{[0]} = \left\{ (\Lambda, u) \in N_1^{[0]} \,\middle|\, X_{H^{[0]}}(\Lambda, u) \in T_{(\Lambda, u)} N_1^{[0]} \right\}.$$

The tangency condition is now

$$
\begin{aligned}
0 &= (X_Z + u^r X_{Y_r^V})(H_{[Z, Y_s^V]}) = \mathrm{d} H_{[Z, Y_s^V]}(X_Z + u^r X_{Y_r^V}) \\
&= H_{[Z, [Z, Y_s^V]]} + H_{[u^r Y_r^V, [Z, Y_s^V]]} = H_{[Z, [Z, Y_s^V]]} - ([Z, Y_s^V](u^r)) H_{Y_r^V} \\
&\quad + u^r H_{[Y_r^V, [Z, Y_s^V]]} = H_{[Z, [Z, Y_s^V]]} + u^r H_{[Y_r^V, [Z, Y_s^V]]},
\end{aligned}
$$

for $s = 1, \ldots, k$, using similar computations as in Equation (6.6.15) and keeping in mind that $(\Lambda, u) \in N_1^{[0]}$. Then,

$$N_2^{[0]} = \left\{ (\Lambda, u) \in N_1^{[0]} \,\middle|\, H_{[Z, [Z, Y_s^V]]} + u^r H_{[Y_r^V, [Z, Y_s^V]]} = 0, \; s = 1, \ldots, k \right\}.$$

The algorithm continues until a final constraint submanifold $N_f^{[0]}$ where the abnormal biextremals are is obtained. The vector field $X_{H^{[0]}}$ could be completely determined or not as remarked in Chapter 5 and [Barbero-Liñán and Muñoz Lecanda 2008a]. The biextremals in $N_f^{[0]}$ are not necessarily abnormal in the sense of Pontryagin's Maximum Principle: the zero fiber must be deleted in order not to contradict condition $(2b)$ in Theorem 6.5.1.

Remark 6.6.2. Observe that in $N_0^{[0]}$ and $N_1^{[0]}$ the controls do not appear, but they do in $N_2^{[0]}$. If we determine some controls in $N_2^{[0]}$ without having a new constraint to be stabilized, the algorithm stops. It could happen that the vector field is completely determined. So we avoid having to use Lie brackets of higher degree.

Example: For a system with two–input vector fields, the constraints in $N_2^{[0]}$ are

$$
\left.
\begin{aligned}
H_{[Z, [Z, Y_1^V]]} + u^1 H_{[Y_1^V, [Z, Y_1^V]]} + u^2 H_{[Y_2^V, [Z, Y_1^V]]} = 0, \\
H_{[Z, [Z, Y_2^V]]} + u^1 H_{[Y_1^V, [Z, Y_2^V]]} + u^2 H_{[Y_2^V, [Z, Y_2^V]]} = 0.
\end{aligned}
\right\} \tag{6.6.17}
$$

Here, it is useful to refresh the *symmetric product* of two vector fields $X, Y \in \mathfrak{X}(Q)$ denoted by $\langle X : Y \rangle$ and defined in Equation (6.2.3). Observe that

$$[Y_s^V, [Z, Y_l^V]] = \langle Y_s : Y_l \rangle^V.$$

This relation allows us to rewrite the matrix of coefficients of the linear–system (6.6.17) on the

controls as follows:

$$\begin{pmatrix} H_{\langle Y_1 : Y_1 \rangle^V} & H_{\langle Y_1 : Y_2 \rangle^V} \\ H_{\langle Y_1 : Y_2 \rangle^V} & H_{\langle Y_2 : Y_2 \rangle^V} \end{pmatrix}.$$

This matrix is symmetric since the symmetric product is symmetric. Observe that all the elements can be written in terms of the configuration manifold Q because all the Hamiltonians in this matrix are associated with vertical vector fields. Then, only the momenta in $\left(V_{\Upsilon(t)}(\tau_{TQ}) \right)^*$ take part, those momenta correspond with β in Equation (6.4.9). In other words, for every $x \in Q$ and $\mathscr{Y}_x = \{Y_1(x), Y_2(x)\}$ we define

$$\begin{aligned} B_{\mathscr{Y}_x}(\mathscr{Y}_x) \colon \mathscr{Y}_x \times \mathscr{Y}_x & \longrightarrow & T_x Q / \mathscr{Y}_x \\ (w_1, w_2) & \longmapsto & \pi_{\mathscr{Y}_x}(\langle W_1 : W_2 \rangle), \end{aligned} \tag{6.6.18}$$

where W_1 and W_2 are vector fields in $\mathscr{Y}$ extending $w_1, w_2 \in \mathscr{Y}_x$ and $\pi_{\mathscr{Y}_x} \colon T_x Q \to T_x Q / \mathscr{Y}_x$ is the natural projection. This mapping is a well–defined vector–valued quadratic form, because it does not depend on the extensions considered, as explained in Appendix C.

The matrix of the vector–valued quadratic form (6.6.18) is given by $n-2$ symmetric matrices $B^1, \ldots, B^{n-2}$ of dimension 2×2 such that $B^i_{rs}(x) = \langle \eta^i_x, \pi_{\mathscr{Y}_x}(\langle Y_r : Y_s \rangle) \rangle$, where η^i_x is a basis in $(T_x Q / \mathscr{Y}_x)^*$.

For any $\lambda \in (T_x Q / \mathscr{Y}_x)^* \simeq \operatorname{ann} \mathscr{Y}_x$, we have the real quadratic form

$$\begin{aligned} (\lambda B)_x = \lambda B_{\mathscr{Y}_x}(\mathscr{Y}_x) \colon \mathscr{Y}_x \times \mathscr{Y}_x & \longrightarrow & \mathbb{R} \\ (w_1, w_2) & \longmapsto & \langle \lambda, \langle W_1 : W_2 \rangle(x) \rangle = H_{\langle W_1 : W_2 \rangle}(\lambda), \end{aligned} \tag{6.6.19}$$

where $H_{\langle W_1 : W_2 \rangle} \colon \operatorname{ann} \mathscr{Y} \to \mathbb{R}$ and its associated matrix corresponds to the matrix of the system (6.6.17). Assuming the regularity of the matrix, the controls can be determined and there are no more constraints. Then, $N_2^{[0]} = N_f^{[0]}$. The regularity of the matrix can be given using the properties related with the definiteness of the vector–valued quadratic form (6.6.18) in Definition C.1.1. If it is definite or it is essentially indefinite and nonzero, then the matrix of coefficients is regular and the controls are

$$\begin{pmatrix} u^1 \\ u^2 \end{pmatrix} = \begin{pmatrix} H_{\langle Y_1 : Y_1 \rangle^V} & H_{\langle Y_1 : Y_2 \rangle^V} \\ H_{\langle Y_1 : Y_2 \rangle^V} & H_{\langle Y_2 : Y_2 \rangle^V} \end{pmatrix}^{-1} \begin{pmatrix} -H_{[Z,[Z,Y_1^V]]} \\ -H_{[Z,[Z,Y_2^V]]} \end{pmatrix} \tag{6.6.20}$$

The properties for vector–valued quadratic form in Lemma C.1.2 guarantee that the definiteness of these quadratic form implies some separation conditions that connect with the separation conditions for optimality that appear in the proof of Pontryagin's Maximum Principle in §4.2 and in §4.4. On the other hand, indefinite vector–valued quadratic forms are related with controllability in some sense [Hirschorn and Lewis 2002] and therefore abnormality cannot exist, as shown in §4.5.2.

Given the controls in (6.6.20) and an initial condition in $N_2^{[0]}$ for the fibers, we have a unique integral curve of the vector field $X_{H^{[0]}}$; that is, a solution of the Cauchy problem given by Hamilton's equations. So we have an abnormal extremal in this case.

Suppose now that the matrix of coefficients is singular on a strict submanifold $\widetilde{N}_1^{[0]}$ of $N_1^{[0]}$ with smaller dimension than $N_1^{[0]}$. Then we have to:

- search for abnormal extremals in the open submanifold $N = N_1^{[0]} \setminus \widetilde{N}_1^{[0]}$ where the matrix of coefficients is regular, and

- search for abnormal extremals in $\widetilde{N}_1^{[0]}$, once the constraints defining this submanifold are stabilized, we can also have new constraints from Equation (6.6.17) that must be stabilized.

6.6.2 Characterization of normality

For the normal case, $p_0 = -1$, Pontryagin's Hamiltonian is

$$H = H_Z + u^s H_{Y_s^V} - \mathcal{F}.$$

The primary constraint submanifold (6.6.12) becomes

$$N_0^{[-1]} = \left\{ (\Lambda, u) \in T^*(TQ) \times U \,\middle|\, H_{Y_s^V} - \frac{\partial \mathcal{F}}{\partial u^s} = 0, \ s = 1, \dots, k \right\}. \qquad (6.6.21)$$

Note the significant role that the cost function plays for normal extremals: the possibility for the controls to be determined essentially depends on the given cost function. To better understand this process we address the reader to Chapter 5, [Barbero-Liñán and Muñoz Lecanda 2008a] and to examples studied in §6.7.

6.7 Applications to optimal control problems

We apply the constraint algorithm described in §6.6, obtained from the theory developed in §5.2, to two particular optimal control problems. We are able to give new results about the abnormal extremals for these problems.

6.7.1 Some control–quadratic cost functions

A control–quadratic cost function is given by a quadratic form in $U \subset \mathbb{R}^k$

$$
\begin{aligned}
A \colon U &\longrightarrow \mathbb{R} \\
u &\longmapsto \frac{1}{2} u^t A u = \frac{1}{2} \sum_{r,s=1}^{k} a_{rs} u^r u^s
\end{aligned}
$$

where u is a k–vector, A is a $(k \times k)$–matrix and u^t denotes the transpose of the vector u. We assume that A is positive–definite, hence it is also regular. Then the cost function associated to A is

$$
\begin{aligned}
\mathcal{F} \colon TQ \times U &\longrightarrow \mathbb{R} \\
(v_x, u) &\longmapsto \frac{1}{2} u A u^t = \frac{1}{2} \sum_{r,s=1}^{k} a_{rs} u^r u^s
\end{aligned}
$$

that is a particular case of the cost functions defined in (6.4.7), given by a U–dependent $(0, 0)$ tensor field on Q.

The study for abnormality is the same as in §6.6.1 because the cost function does not play any role, so the primary constraint submanifold is (6.6.14). However, given the cost function, we can study more carefully the normal constraint submanifolds in §6.6.2.

Assume $p_0 = -1$. Pontryagin's Hamiltonian is $H = H_Z + u^s H_{Y_s^V} - 1/2 \sum_{r,s=1}^{k} a_{rs} u^r u^s$.

The primary constraint submanifold (6.6.21) is

$$N_0^{[-1]} = \left\{ (\Lambda, u) \in T^*(TQ) \times U \; \middle| \; H_{Y_s^V} - \sum_{r=1}^{k} a_{sr} u^r = 0, \; s = 1, \ldots, k \right\}. \qquad (6.7.22)$$

As A is regular, the controls are determined and the algorithm stops:

$$N_f^{[-1]} = N_0^{[-1]}.$$

Proposition 6.7.1. *Let $\Sigma_{\mathcal{F}}$ be an optimal control problem in Statement 6.3.1 or 6.3.2 with a cost function $\mathcal{F}$ given by a positive–definite quadratic form in $U \subset \mathbb{R}^k$. Let $(\Upsilon, u) \colon I \to TQ \times U$ be a nonconstant abnormal extremal for $\Sigma_{\mathcal{F}}$. If the controls u are zero, then the extremal is also normal; that is, the extremal is not strict abnormal.*

(*Proof*) If (Υ, u) is an abnormal extremal, there exists a nonzero momenta along Υ such that $(\Lambda, u) \in N_f^{[0]}$.

The extremal will admit a normal lift—see Definition 4.1.15—if there exists an initial momenta in the final normal constraint submanifold $N_f^{[-1]}$ given by (6.7.22). Observe that the abnormal lift vanishes the vertical lift of the controls vector fields. Thus the zero controls and the abnormal lift also satisfy the constraints in the normal primary constraint submanifold (6.7.22). Hence $(\Upsilon, 0)$ is a normal extremal. ∎

Proposition 6.7.2. *Let $\Sigma_{\mathcal{F}}$ be an optimal control problem in Statement 6.3.1 or 6.3.2 with a cost function $\mathcal{F}$ given by a positive–definite quadratic form in $U \subset \mathbb{R}^k$.*

1. *The curves $(\Upsilon, 0) \colon I \to TQ \times U$ are always normal extremals for $\Sigma_{\mathcal{F}}$. If they are abnormal extremals, then they are not strict abnormal.*

2. *The curves $(\Upsilon, 0) \colon I \to TQ \times U$ satisfying the endpoint conditions are minimizers of the functional $\mathcal{S}[\Upsilon, u] = \int_I \mathcal{F}(\Upsilon(t), u(t)) \, \mathrm{d}t$.*

(*Proof*) The first result follows from Proposition 6.7.1. The second one is immediate using the positive–definiteness of the quadratic form A defining the cost function. ∎

6.7.2 Time–optimal control problem

The time–optimal control problem is a free optimal control problem in Statement 4.3.1 with fixed endpoint conditions and with cost function $\mathcal{F} = 1$, that is a particular case of the cost functions defined in (6.4.7) with $f = 1$. Thus the constraint algorithm must be used as explained in §5.3. There is one unknown more, the final time, and there is also a new condition in Pontryagin's Maximum Principle that must be stabilized; that is, the Hamiltonian along the optimal curve is zero almost everywhere.

Statement 6.7.3. *(Time–optimal control problem)*
Given $\Sigma = (Q, \nabla, \mathscr{Y}, U)$ and $\mathcal{F} = 1$, find $I = [a, b] \subset \mathbb{R}$ and $(\gamma, u) \colon I \to Q \times U$ such that, given the endpoint conditions $x_a, x_b \in Q$, there exists $\Upsilon \colon I \to TQ$ along γ satisfying

(1) $\gamma(a) = x_a$, $\gamma(b) = x_b$,

(2) $\dot{\Upsilon}(t) = Z(\Upsilon(t)) + u^s(t) Y_s^V(\Upsilon(t))$, and

(3) $\mathcal{S}[\Upsilon, u] = \int_I \mathrm{d}t$ is minimized over all curves on $TQ \times U$ satisfying (1) and (2).

The abnormal and normal Hamiltonians are related by $H^{[0]} = H^{[-1]} + 1$. Then the abnormal and normal Hamiltonian vector fields are the same:

$$X_H = X_Z + u^s X_{Y_s^V}.$$

The presymplectic equation (6.5.11) must be restricted to the submanifold defined by the new condition

$$H = H_Z + u^s H_{Y_s^V} + p_0 = 0,$$

which also has to be stabilized in the algorithm. As pointed out in §5.3, the stabilization step, $0 = X_H(H)$, is immediately satisfied because of the properties of the Hamiltonian vector fields. Therefore, the tangency condition of $H = 0$ does not add any new constraint to the successive submanifolds of the algorithm and the condition $H = 0$ can be put aside until the end of the algorithm. Then the primary constraint submanifolds for abnormality (6.6.14) and normality (6.6.21) are the same:

$$N_0 = N_0^{[0]} = N_0^{[-1]} = \left\{ (\Lambda, u) \in T^*(TQ) \times U \,\middle|\, \frac{\partial H}{\partial u^s} = H_{Y_s^V} = 0, \ s = 1, \ldots, k \right\}.$$

The next constraint submanifold is

$$N_1 = \left\{ (\Lambda, u) \in T^*(TQ) \times U \,\middle|\, H_{Y_s^V} = 0, \ H_{[Z, Y_s^V]} = 0, \ s = 1, \ldots, k \right\},$$

and so on until getting the final submanifold N_f, when it exists.

Finally, the vanishing of the Hamiltonian must be considered to have the actual final constraint submanifolds

$$N_f^{[0]} = N_f \cap \left\{ (\Lambda, u) \in T^*(TQ) \times U \,\middle|\, H_Z + u^s H_{Y_s^V} = 0 \right\},$$

$$N_f^{[-1]} = N_f \cap \left\{ (\Lambda, u) \in T^*(TQ) \times U \mid H_Z + u^s H_{Y_s^V} = 1 \right\}.$$

Proposition 6.7.4. *Let* $\Sigma_1 = (Q, \nabla, \mathscr{Y}, U, 1)$ *be the time–optimal control problem in Statement 6.7.3 and let* $N_f^{[0]}$ *be its abnormal final constraint submanifold :*

1. *If* $N_f^{[0]}$ *only has zero covectors, then there are no abnormal extremals.*

2. *If* $N_f^{[0]}$ *has nonzero covectors and*

$$N_f \subset \left\{ (\Lambda, u) \in T^*(TQ) \times U \,\middle|\, (H_Z + u^s H_{Y_s^V}) = 0 \right\},$$

   then there are only strict abnormal extremals.

(*Proof*) First, if $N_f^{[0]}$ only has zero covectors, then there are no abnormal extremals for contradicting $(2b)$ in Theorem 6.5.1.

Secondly, if $N_f^{[0]}$ has nonzero covectors, then there exists $(\Lambda, u) \in N_f^{[0]}$ such that Λ is nonzero along an abnormal extremal (Υ, u) on $TQ \times U$; see Proposition 5.2.3. For $t \in I$, $(N_f^{[0]})_{(\Upsilon(t), u(t))}$ is a subspace in $T_{\Upsilon(t)}^*(TQ) \times \{u(t)\}$. Under the assumption of abnormality and the Hamiltonian being zero, we have $H_Z = 0$ along (Υ, u); that is, this constraint defines a subspace. If

$$(N_f)_{(\Upsilon(t), u(t))} \subseteq \left\{ (\widetilde{\Lambda}, \widetilde{u}) \in T_{\Upsilon(t)}^*(TQ) \times U \mid H_Z(\widetilde{\Lambda}) = 0 \right\},$$

then $(\Lambda(t), u(t)) \notin N_f^{[-1]}$ and $(\Upsilon(t), u(t))$ is a strict abnormal extremal for every $t \in I$. If the hypotheses are satisfied for all the extremals, then all the abnormal extremals are strict. ∎

6.8   Study of abnormality for particular cases

Let us consider different dimensions for the configuration manifold and the distribution of the input vector fields so as to illustrate the process explained in §6.6. Here we also explain how to reason with the vector–valued quadratic form to obtain conditions to determine the controls; see Appendix C and references therein for a brief introduction to these quadratic forms.

The abnormality in some other particular cases of control–affine systems have been studied in [Agrachev and Zelenko 2007, Chitour et al. 2006; 2008], but here we concentrate on the mechanical systems where the drift vector field and the control vector fields satisfy specific properties. These make them impossible to apply the results related with the Goh matrix used in [Chitour et al. 2006; 2008]. The Goh matrix is given by the Hamiltonian associated with the Lie brackets of the control vector fields. For the systems considered in this chapter the Goh matrix is $(H_{[Y_r^V, Y_s^V]})_{rs}$, which is identically zero on T^*TQ because $[Y_r^V, Y_s^V] = 0$. Then we are forced to go to the next stabilization step to determine the extremals. The matrix of the

system for the controls at this step is $(H_{\langle Y_r, Y_s \rangle^V})_{rs}$, a symmetric matrix. Thus the reasoning about skew–symmetric matrices used in [Chitour et al. 2006; 2008] cannot be applied here.

6.8.1 Fully actuated

If the number of linearly independent input vector fields is the same as the dimension of the configuration manifold, the system (6.1.2) is said to be *fully actuated*. For time–optimal control problems associated with the systems, there are no abnormal extremals as is proved in [Ailon and Langholz 1985, Chen 1989, Chyba et al. 2003, Sontag 1989, Sontag and Sussmann 1986]. But in fact, this result can be proved for any optimal control problem.

Theorem 6.8.1. *If $\Sigma_{\mathcal{F}}$ is an optimal control problem such that Σ is a fully actuated affine connection control system, then it does not have abnormal extremals.*

(*Proof*) By Proposition 6.2.7, $\{[Z, \mathscr{Y}^V], \mathscr{Y}^V\}$ is a family of linear independent vector fields. As the system is fully actuated, these vector fields span the entire tangent space of TQ.

On the other hand, the abnormal constraint submanifold $N_1^{[0]}$ in Equation (6.6.16) determines the momenta as the annihilator of $\{[Z, \mathscr{Y}^V], \mathscr{Y}^V\}$. Thus, the unique possible abnormal momenta is zero, contradicting (*2b*) in Theorem 6.5.1. ∎

Recalling the kinds of extremals in Definition 4.1.15, the result in Theorem 6.8.1 is also true, replacing "abnormal" by "singular" for any optimal control problem associated with fully actuated affine connection control systems, because of Proposition 6.5.2 and Remark 6.5.3. This is proved for time–optimal control problems with control–affine mechanical systems in [Ailon and Langholz 1985] for small dimension and in [Chen 1989, Chyba et al. 2003, Sontag 1989, Sontag and Sussmann 1986] for more general cases.

From Definition 4.1.15 and Theorem 6.8.1 we immediately obtain the following result.

Corollary 6.8.2. *If $\Sigma_{\mathcal{F}}$ is an optimal control problem such that Σ is a fully actuated affine connection control system, then it does not have strict abnormal extremals.*

6.8.2 One control vector field in two dimension

In this case, the control–affine system on TQ is given by

$$\dot{\Upsilon}(t) = Z(\Upsilon(t)) + u(t) Y^V(\Upsilon(t)),$$

where $\Upsilon \colon I \to TQ$, Y^V denotes the vertical lift of the vector field $Y \in \mathfrak{X}(Q)$ and $\dim Q = 2$.

For abnormality, $p_0 = 0$, the first three steps of the constraint algorithm give

$$
\begin{aligned}
N_0^{[0]} &= \left\{ (\Lambda, u) \in T^*(TQ) \times U \mid H_{Y^V} = 0 \right\}, \\
N_1^{[0]} &= \left\{ (\Lambda, u) \in N_0^{[0]} \mid H_{[Z, Y^V]} = 0 \right\}, \\
N_2^{[0]} &= \left\{ (\Lambda, u) \in N_1^{[0]} \mid H_{[Z,[Z,Y^V]]} + u H_{\langle Y : Y \rangle^V} = 0 \right\}.
\end{aligned}
$$

Different cases have to be studied:

1. If $\langle Y : Y \rangle$ and Y are linearly independent on $N_1^{[0]}$, the control is completely determined as long as $H_{\langle Y:Y \rangle^V}(\Lambda) \neq 0$ for $\Lambda \in T^*(TQ)$.

2. If $\langle Y : Y \rangle$ and Y are linearly independent on $N_1^{[0]}$ and $H_{\langle Y:Y \rangle^V}(\Lambda) = 0$, then the momenta corresponding to the velocities vanishes identically because Y and $\langle Y: Y \rangle$ span the whole vertical tangent bundle. Moreover, from Hamilton's equations in (6.5.11), the momentum is zero. Thus, there are no abnormal extremals. That connects to Remark 6.4.5, the momenta corresponding to the velocities determine the momenta in the dual horizontal subbundle.

3. If $\langle Y : Y \rangle$ and Y are linearly dependent on $N_1^{[0]}$, then $N_2^{[0]}$ is defined in $N_1^{[0]}$ by $0 = H_{[Z,[Z,Y^V]]}$ because $(\Lambda, u) \in \operatorname{ann} \mathscr{Y}^V \times U$, see (2.1.1) for the definition of the annihilator. The algorithm must continue with the stabilization of the constraint $H_{[Z,[Z,Y^V]]} = 0$, which does not depend on the controls at all.

4. If $\langle Y : Y \rangle$ and Y are linearly dependent only on a submanifold S of $N_1^{[0]}$, then the control is not determined on S. The algorithm must go on with the stabilization of $0 = H_{[Z,[Z,Y^V]]}$ and also the stabilization of the constraints that define implicitly the submanifold S.

   In the open submanifold $N_1^{[0]} - S$, the control is completely determined if $H_{\langle Y:Y \rangle^V}(\Lambda) \neq 0$.

5. Same assumptions as in case 4, but $H_{\langle Y:Y \rangle^V}(\Lambda) = 0$ on $N_1^{[0]} - S$. Thus there are no abnormal extremals in $N_1^{[0]} - S$ because the momentum is zero, using the same reasoning as in case 2.

These different cases can be distinguished by the vector–valued quadratic form (6.6.18) defined at every $x \in Q$ by

$$
\begin{aligned}
B_x \colon \mathscr{Y}_x \times \mathscr{Y}_x &\longrightarrow T_x Q / \mathscr{Y}_x \\
(w_1, w_2) &\longmapsto \pi_{\mathscr{Y}_x}(\langle W_1 : W_2 \rangle),
\end{aligned}
$$

where W_1 and W_2 are vector fields in $\mathscr{Y}$ extending $w_1, w_2 \in \mathscr{Y}_x$, and $\pi_{\mathscr{Y}_x} \colon T_x Q \to T_x Q / \mathscr{Y}_x$ is the natural projection.

As $\dim Q = 2$ and there is only one input vector field, $\dim (T_x Q / \mathscr{Y}_x) = 1$. Thus the matrix B_x associated with the vector–valued quadratic form is 1×1. Then for any $\lambda \in (T_x Q / \mathscr{Y}_x)^* \simeq \operatorname{ann} \mathscr{Y}_x$, we define the real quadratic form analogous to (6.6.19):

$$
\begin{aligned}
(\lambda B)_x \colon \mathscr{Y}_x \times \mathscr{Y}_x &\longrightarrow \mathbb{R} \\
(w_1, w_2) &\longmapsto \langle \lambda, \langle W_1 : W_2 \rangle(x) \rangle = \lambda B a_1 a_2,
\end{aligned}
$$

where $W_i = a_i Y$ and $B = \langle \lambda, \langle Y : Y \rangle \rangle$.

According to definitions about the definiteness of vector–valued quadratic forms—see [Bullo and Lewis 2005a, Section 8.1] and Appendix C—every vector–valued quadratic form is either strongly semidefinite or essentially indefinite.

A vector–valued quadratic form at $x \in Q$ is strongly semidefinite if there exists $\lambda \in$ ann $\mathscr{Y}_x \backslash \{0\}_x$ such that $(\lambda B)_x$ is nonzero and positive–semidefinite. Observe that, according to this definition, a definite vector–valued quadratic form is also strongly semidefinite. If a vector–valued quadratic form is positive–definite, then it is also negative–definite because of Definition C.1.1. A vector–valued quadratic form is essentially indefinite if, for each $\lambda \in$ ann $\mathscr{Y}_x$, $(\lambda B)_x$ is either zero or neither positive–semidefinite nor negative–semidefinite.

Remark 6.8.3. As the image of B_x has dimension at most 1, the vector–valued quadratic form is never indefinite or semidefinite. It will be zero or positive–definite.

Let $\pi_1 \colon T^*TQ \times U \to T^*TQ$, the different previous cases correspond respectively with:

1. If $\langle Y : Y \rangle$ and Y are linearly independent on $N_1^{[0]}$ and $H_{\langle Y : Y \rangle^V}(\Lambda) \neq 0$, then B_x is definite at each $x \in (\tau_Q \circ \pi_{TQ} \circ \pi_1) \left(N_1^{[0]} \right) \subset Q$ because there exists a $\lambda \in$ ann $\mathscr{Y}_x$ such that $(\lambda B)_x$ is positive–definite since $\dim \text{Im } B_x = 1$.

2. If $\langle Y : Y \rangle$ and Y are linearly independent on $N_1^{[0]}$ and $H_{\langle Y : Y \rangle^V}(\Lambda) = 0$, B_x is strongly semidefinite. The hypotheses make Λ be the zero momentum. But another $\lambda_x \in$ ann $\mathscr{Y}_x$ can be chosen such that $(\lambda B)_x$ is positive–definite because $\langle Y : Y \rangle$ and Y are linearly independent.

3. If $\langle Y : Y \rangle$ and Y are linearly dependent on $N_1^{[0]}$, for every $x \in (\tau_Q \circ \pi_{TQ} \circ \pi_1) \left(N_1^{[0]} \right)$ in Q, B_x is essentially indefinite because B_x is zero; that is, for each $\lambda \in$ ann $\mathscr{Y}_x$, $(\lambda B)_x$ is zero.

4. If $\langle Y : Y \rangle$ and Y are linearly dependent on a submanifold S of $N_1^{[0]}$, B_x is essentially indefinite on S and strongly semidefinite on $N_1^{[0]} - S$.

5. If $\langle Y : Y \rangle$ and Y are linearly dependent on a submanifold S of $N_1^{[0]}$ and $H_{\langle Y : Y \rangle^V} = 0$, B_x is essentially indefinite on S and strongly semidefinite on $N_1^{[0]} - S$, as in the case 3.

Remark 6.8.4. Note that the vector–valued quadratic form cannot distinguish between cases 1 and 2. It is important to make clear, specially for assertion 2, that we only have one vector–valued quadratic form associated to the problem, but for every momentum in ann $\mathscr{Y}$ there exists an associated real quadratic form. The above cases are classified in terms of the vector–valued quadratic form.

Observe that if the vector–valued quadratic form B_x is essentially indefinite, then B_x is the zero mapping.

6.8.3 Underactuated by one input

Now consider the control–affine system with $n-1$ input vector fields,

$$\dot{\Upsilon}(t) = Z(\Upsilon(t)) + u_1(t)Y_1^V(\Upsilon(t)) + \ldots + u_{n-1}(t)Y_{n-1}^V(\Upsilon(t)),$$

where the dimension of Q is n.

The constraint submanifolds for abnormality are given by

$$
\begin{aligned}
N_0^{[0]} &= \left\{ (\Lambda, u) \in T^*(TQ) \times U \,|\, H_{Y_s^V} = 0 \,, \quad s = 1, \dots, n-1 \right\}, \\
N_1^{[0]} &= \left\{ (\Lambda, u) \in N_0^{[0]} \,|\, H_{[Z, Y_s^V]} = 0 \,, \quad s = 1, \dots, n-1 \right\}, \\
N_2^{[0]} &= \left\{ (\Lambda, u) \in N_1^{[0]} \,|\, H_{[Z, [Z, Y_s^V]]} + u_r H_{\langle Y_r : Y_s \rangle^V} = 0 \,, \ s = 1, \dots, n-1 \right\}.
\end{aligned}
$$

As the codimension of the distribution of the input vector fields is 1, in a neighbourhood of every $(\Lambda, u) \in N_1^{[0]}$ there exists a vector field Y on Q complementary to the subspace spanned by the control vector fields. Then, for $(\Lambda, u) \in N_1^{[0]} \subseteq \operatorname{ann} \mathscr{Y}^V \times U$,

$$
H_{\langle Y_r : Y_s \rangle^V}(\Lambda) = c_{rs} H_{Y^V}(\Lambda),
$$

where $\langle Y_r : Y_s \rangle = \sum_{i=1}^{n-1} c_{rs}^i Y_i + c_{rs} Y$ and $c_{rs}^i, c_{rs} \in \mathcal{C}^\infty(Q)$.

At present, keeping in mind that the matrix c_{rs} is a function on Q, the cases to be considered are:

1. The matrix c_{rs} has maximum rank on $N_1^{[0]}$ and $H_{Y^V}(\Lambda) \neq 0$, so that the controls are completely determined on $N_2^{[0]}$ and the algorithm stops.

2. The matrix c_{rs} has maximum rank on $N_1^{[0]}$ and $H_{Y^V}(\Lambda) = 0$, so that the controls cannot be determined on $N_2^{[0]}$, but the momenta corresponding with the velocities are zero because of the rank of $\mathscr{Y}$ and the choice of Y. Due to Hamilton's equations (6.5.11), the momentum is zero. Thus there are no abnormal extremals.

3. The matrix c_{rs} has nonzero rank, the rank is not maximum on $N_1^{[0]}$ and $H_{Y^V}(\Lambda) \neq 0$, so that some controls are determined but some conditions must still be stabilized.

4. The matrix c_{rs} has nonzero rank, the rank is not maximum on $N_1^{[0]}$ and $H_{Y^V}(\Lambda) = 0$, so that the controls cannot be determined, but the momentum corresponding with the velocities is zero. Due to Hamilton's equations (6.5.11) the momentum is zero. Thus there are no abnormal extremals.

5. The matrix c_{rs} is identically zero on $N_1^{[0]}$, so that $\langle \mathscr{Y} : \mathscr{Y} \rangle \subseteq \mathscr{Y}$. The controls remain undetermined on $N_2^{[0]}$ and the algorithm must go on with the stabilization of $H_{[Z, [Z, Y_s^V]]} = 0$ with $s = 1, \dots, n-1$.

6. The matrix c_{rs} is identically zero on a submanifold S of $N_1^{[0]}$, so that $\langle \mathscr{Y} : \mathscr{Y} \rangle \subseteq \mathscr{Y}$. The controls remain undetermined on $N_2^{[0]}$ and the algorithm must go on with the stabilization of $H_{[Z, [Z, Y_s^V]]} = 0$ with $s = 1, \dots, n-1$ and also with the stabilization of the constraints that determine implicitly the submanifold S.

   In $N_1^{[0]} - S$, the matrix c_{rs} may have maximum rank or not. See cases 2-5.

The associated vector–valued quadratic form for every $x \in Q$ is

$$
\begin{aligned}
B_x \colon \mathscr{Y}_x \times \mathscr{Y}_x &\longrightarrow T_x Q / \mathscr{Y}_x \\
(w_1, w_2) &\longmapsto \pi_{\mathscr{Y}_x}(\langle W_1 \colon W_2 \rangle)
\end{aligned}
$$

as defined in (6.6.18). Its associated matrix is a $(n-1) \times (n-1)$–matrix given by (c_{rs}), because $\dim T_x Q / \mathscr{Y}_x = 1$ and there are $n - 1$ input vector fields. For any $\lambda \in (T_x Q / \mathscr{Y}_x)^* \simeq \operatorname{ann} \mathscr{Y}_x$, we have the real quadratic form

$$
\begin{aligned}
(\lambda B)_x \colon \mathscr{Y}_x \times \mathscr{Y}_x &\longrightarrow \mathbb{R} \\
(w_1, w_2) &\longmapsto \langle \lambda, \langle W_1 \colon W_2 \rangle(x) \rangle = B_{rs} \, a_1^r a_2^s,
\end{aligned}
$$

where $W_i = a_i^r Y_r$ and $B_{rs} = \langle \lambda, \langle Y_r \colon Y_s \rangle \rangle = c_{rs} \langle \lambda, Y \rangle$.

The above different cases in terms of the vector–valued quadratic forms are:

1. If the matrix c_{rs} has maximum rank on $N_1^{[0]}$ and $H_{YV}(\Lambda) \neq 0$, then B_x is definite or indefinite on $N_1^{[0]}$.

2. If the matrix c_{rs} has maximum rank on $N_1^{[0]}$ and $H_{YV}(\Lambda) = 0$, then B_x can be strongly semidefinite or essentially indefinite, having in mind Remark 6.8.4. The zero momentum makes B_x essentially indefinite and a nonzero momentum can make B_x strongly semidefinite, although this momentum does not satisfy the hypotheses in this case.

3. If the matrix c_{rs} has nonzero rank, the rank is not maximum on $N_1^{[0]}$ and $H_{YV}(\Lambda) \neq 0$, then B_x is either strongly semidefinite or essentially indefinite on $N_1^{[0]}$, depending on the rank of the matrix c_{rs} and the sign of the eigenvalues.

4. If the matrix c_{rs} has nonzero rank, the rank is not maximum on $N_1^{[0]}$ and $H_{YV}(\Lambda) = 0$, then B_x is strongly semidefinite or essentially indefinite.

5. If $\langle \mathscr{Y} \colon \mathscr{Y} \rangle \subseteq \mathscr{Y}$, then B_x is essentially indefinite on $N_1^{[0]}$. To be more precise, it is zero.

6. If the matrix c_{rs} is zero on a submanifold S of $N_1^{[0]}$, B_x is essentially indefinite on S, and on $N_1^{[0]} - S$ the previous cases arise again.

Whenever B_x is essentially indefinite, every $(\lambda B)_x$ is indefinite. The property (i) in Lemma C.1.2 makes us believe that those abnormal extremals will not be optimal. The idea is to establish a connection between the tangent perturbation vectors and the image of the vector–valued quadratic form. Then, due to (i) in Lemma C.1.2 the necessary separation condition for optimality will not be satisfied if B_x is essentially indefinite. This result has not been proved yet.

Conjecture 6.8.5. *Let $\Sigma_{\mathcal{F}} = (Q, \nabla, \mathscr{Y}, U, \mathcal{F}, I)$ be an optimal control problem. If $(\Upsilon, u) \colon I \to TQ \times U$ is an abnormal optimal solution, then the vector–valued quadratic form $B_{(\tau_Q \circ \Upsilon)(t)}$ is either zero or semidefinite at every $t \in I$.*

This conjecture gives a necessary condition for having abnormal extremals. It does not discard the existence of normal lifts associated to these extremals.

This conjecture makes sense to Proposition 4.5.2 and results in [Hirschorn and Lewis 2002]. There, among other hypotheses if the vector–valued quadratic form is indefinite, then the control–affine system is STLC. In Proposition 4.5.2, a necessary condition for abnormality is not to be STLC.

Here we have studied carefully the constraint algorithm, twice applying the tangency conditions. The final submanifolds cannot be always obtained at this point. That is why a new research has already been started to connect the constraint algorithm with the high–order Maximum Principle [Krener 1977], in order to give a geometric version of the results stated by Krener [1977] through the constructions of mappings related to the vector–valued quadratic form at each step of the algorithm.

Chapter 7

Strict abnormal extremals in nonholonomic and kinematic control systems

We continue the study of abnormality in optimal control problems for mechanical control systems. The approach considered in this chapter consists of taking advantage of particular nonholonomic control mechanical systems, which are equivalent to kinematic control system. For more details in that equivalence see, for instance [Bloch 2003, Bullo and Lewis 2005a;c, Muñoz-Lecanda and Yániz-Fernández 2008].

The control system in subRiemannian geometry [Montgomery 1995; 2002] is control–linear, as with kinematic systems. As mentioned by Liu and Sussmann [1994b; 1995] and Montgomery [1994] there exist local strict abnormal minimizers for the problem of the shortest paths. As the kinematic control systems can be equivalent to nonholonomic control systems, we are going to use the strict abnormal minimizers in subRiemannian geometry to characterize the abnormal extremals for certain mechanical control systems. Dealing with a kinematic system is by far easier than dealing with a mechanical control system, which is either control–affine or nonlinear, because there is no drift and because more information is known about the control–linear systems [Liu and Sussmann 1995, Montgomery 1994; 1995; 2002].

Firstly, we investigate whether it is feasible to obtain any connection between the optimal control problems associated to the two control systems. Then, Pontryagin's Maximum Principle will be used to connect the abnormal extremals of both optimal control problems [Barbero-Liñán and Muñoz Lecanda 2008c].

This chapter is organized as follows: In §7.1 the different definitions and results associated with the optimal control problems for nonholonomic and kinematic systems are described, in particular, the possible equivalence between both problems. The Hamiltonian problems for both control problems are stated in §7.2 so as to apply the Maximum Principle. Definition 7.2.1 about the different kinds of extremals for the mechanical case is especially important as it gives a justification of the study made in [Bullo and Lewis 2005b]. In §7.2.4 it is shown how to use the strict abnormal minimizers in subRiemannian geometry to characterize the extremals for the corresponding optimal control problem with nonholonomic mechanical system by means of an example where there exists a local strict abnormal minimizer for the time–optimal control problem for the mechanical system.

7.1 Optimal control problem with nonholonomic mechanical systems versus kinematic systems

First, we study the nonholonomic and kinematic control systems from the viewpoint of control theory. Then, we study them from the approach of optimal control theory in §7.1.2.

7.1.1 Nonholonomic mechanical systems with control

Let (Q, g) be a Riemannian manifold of dimension n and ∇ be the Levi–Civita connection associated to the Riemannian metric g, see [do Carmo 1992] and §2 for more details in these notions. Let TQ be the tangent bundle with the natural projection $\tau_Q \colon TQ \to Q$. Consider $\mathcal{D} \subset TQ$, a nonintegrable distribution in Q with rank k and spanned by the *input or control vector fields* $\{Y_1, \ldots, Y_k\}$. The nonintegrability of the distribution is not necessary. But, the integrable case is not that interesting because, under the assumption of integrability, once the velocity of the trajectory starts in $\mathcal{D}$, it stays there.

Let $\mathcal{D}^\perp$ be the orthogonal distribution to $\mathcal{D}$ according to the metric g. Assume that $\mathcal{D}^\perp$ is spanned by $\{Z_1, \ldots, Z_{n-k}\}$, a family of vector fields on Q.

It is also possible to consider a vector field $F \in \mathfrak{X}(Q)$ describing an external force. Then, a *nonholonomic mechanical system with control* is given by $\Sigma_\mathcal{D} = (Q, g, F, \mathcal{D})$. A differentiable curve $\gamma \colon I \to Q$ is a *solution of* $\Sigma_\mathcal{D}$ for certain values of the *control functions* $u^i \in \mathcal{C}^\infty(TQ)$ if it satisfies the conditions

$$\begin{aligned} \nabla_{\dot\gamma}\dot\gamma &= F \circ \gamma + \sum_{r=1}^{n-k} \mu^r Z_r \circ \gamma + \sum_{s=1}^{k} u^s Y_s \circ \gamma \,, & (7.1.1) \\ \dot\gamma &\in \mathcal{D} \,. \end{aligned}$$

where $u \colon TQ \to U \subset \mathbb{R}^k$ being U an open set. The Lagrange multipliers μ^r are determined by the condition $\dot\gamma \in \mathcal{D}$ using D'Alembert principle [Bloch 2003].

The dynamical equations of mechanical systems are second–order differential equations in the configuration manifold Q, so they may be rewritten as first–order differential equations in TQ using the following vector field along the projection $\pi \colon TQ \times U \to TQ$

$$Y = Z_g + F^V + \sum_{r=1}^{n-k} \mu^r Z_r^V + \sum_{s=1}^{k} u^s Y_s^V, \qquad (7.1.2)$$

where Z_g is the geodesic spray associated with g and Y_s^V is the vertical lift of Y_s, analogously for F^V and Z_r^V. The vector field Y satisfies the second–order condition.

On the other hand, a differentiable curve $\gamma \colon I \to Q$ is a *solution of the kinematic system associated to (7.1.1)*, if there exist $w \colon I \to V \subset \mathbb{R}^k$ such that

$$\dot\gamma(t) = \sum_{s=1}^{k} w^s(t) Y_s(\gamma(t)) \,; \qquad (7.1.3)$$

that is, γ is an integral curve of the vector field $X = \sum_{s=1}^{k} w^s Y_s$.

The systems (7.1.1) and (7.1.3) are *equivalent* if and only if every solution of (7.1.1) is also a solution of (7.1.3) and vice versa. Notice that, in spite of the equivalence of the systems, a solution to both systems could have different control functions, but the curve on Q is exactly the same.

Remark 7.1.1. In this chapter, we consider the nonholonomic control system called *fully actuated* because the constraint distribution is exactly the distribution given by the input control vector fields. If the distribution of the input vector fields has rank strictly less than the rank of the constraint distribution, then we have underactuated systems. In this case (7.1.1) and (7.1.3) are not equivalent any more, but weak equivalent. See [Bloch 2003, Bullo and Lewis 2005a;c, Muñoz-Lecanda and Yániz-Fernández 2008] for more details.

Theorem 7.1.2. *[Bullo and Lewis 2005c, Muñoz-Lecanda and Yániz-Fernández 2008] Every fully actuated nonholonomic control system $\Sigma_{\mathcal{D}}$ is equivalent to the associated kinematic system.*

7.1.2 Associated optimal control problems

Given a control system we define an optimal control problem adding a cost function whose integral must be minimized over solutions of the control system. First, we consider an optimal control problem with a fully actuated nonholonomic mechanical control system. The equivalence of this system with a kinematic system, that is, a control–linear system, is known by Theorem 7.1.2. It should be useful to find a cost function for the kinematic system such that some connection between the optimal solutions to both problems may be established.

Let us point out the importance of relating those two optimal control problems. In [Liu and Sussmann 1994b; 1995] the strict abnormal minimizers have been described for the problem of shortest–paths in subRiemannian geometry. The control system in subRiemannian geometry is control–linear, so it can be understood as a kinematic system that comes from a nonholonomic mechanical control system as the example in §7.2.4 shows. Thus it might be expected to characterize abnormal extremals for mechanical control systems using the well–known abnormal minimizers in subRiemannian geometry.

Let us consider a cost function $\mathcal{F}\colon TQ \times U \to \mathbb{R}$ for the mechanical control system. The optimal control problem for (7.1.2) is stated as follows.

Statement 7.1.3. *(Nonholonomic optimal control problem) Given $x_a, x_b \in Q$, find $(\gamma, u)\colon I \to Q \times U$ such that*

    1. γ satisfies the endpoint conditions on Q, i.e. $\gamma(a) = x_a$, $\gamma(b) = x_b$;

    2. $\dot{\gamma}$ is an integral curve of Y in (7.1.2), i.e. $\ddot{\gamma}(t) = Y(\dot{\gamma}(t), u(t))$;

    3. $(\dot{\gamma}, u)$ gives the minimum of $\int_I \mathcal{F}(\dot{\gamma}(t), u(t))\mathrm{d}t$ among all the curves satisfying 1 and 2.

This optimal control problem is denoted by $\Sigma_m = (\Sigma_{\mathcal{D}}, \mathcal{F})$. In optimal control theory, it is common to consider the functional to be minimized as a new coordinate of the system;

see §4.1.2 for more details. In this way, all the elements in the optimal control problem are included in a control system, usually called the *extended system* [Athans and Falb 1966, Lee and Markus 1967, Pontryagin et al. 1962]. Nevertheless, the minimization of the functional must be included in the extended system, what turns out to be the minimization of the new coordinate; see Statement 4.1.2.

In the case of mechanical control systems, two new coordinates are added in order to maintain the second–order condition of the vector field (7.1.2). Let $\widehat{Q} = \mathbb{R} \times Q$, then the cost function is considered as a vector field along the projection $\widehat{\pi}_m \colon T\widehat{Q} \times U \to T\widehat{Q}$ with local expression $\mathcal{F}\partial/\partial x^0$. Then (7.1.1) becomes

$$\widehat{\nabla}_{\dot{\widehat{\gamma}}}\dot{\widehat{\gamma}} = F \circ \widehat{\gamma} + \sum_{r=1}^{n-k} \mu^r Z_r \circ \widehat{\gamma} + \sum_{s=1}^{k} u^s Y_s \circ \widehat{\gamma} + \mathcal{F} \circ (\dot{\widehat{\gamma}}, u) \left.\frac{\partial}{\partial x^0}\right|_{\dot{\widehat{\gamma}}} \qquad (7.1.4)$$

where $\widehat{\gamma} \colon I \to \widehat{Q}$ is a differentiable curve and $\widehat{\nabla}$ denotes the Levi–Civita connection extended to $\widehat{Q}$ considering the natural product connection on $\widehat{Q}$, taking ∇ and the trivial connection on $\mathbb{R}$. Moreover, we have $\pi_2 \circ \dot{\widehat{\gamma}} = \dot{\gamma} \in \mathcal{D}$ with the projection $\pi_2 \colon T\widehat{Q} = T\mathbb{R} \times TQ \to TQ$.

The second–order differential equation (7.1.4) admits a first–order differential equation given by the vector field

$$\widehat{Y} = v^0 \overbrace{\frac{\partial}{\partial x^0} + Z_g}^{\widehat{Z}_g} + F^V + \mathcal{F}\frac{\partial}{\partial v^0} + \sum_{r=1}^{n-k} \mu^r Z_r^V + \sum_{s=1}^{k} u^s Y_s^V \qquad (7.1.5)$$

along the projection $\widehat{\pi}_m \colon T\widehat{Q} \times U \to T\widehat{Q}$, where $\widehat{Z}_g$ is the geodesic spray of the above–mentioned extended connection $\widehat{\nabla}$. The differential equations added to (7.1.2) are

$$\begin{aligned} \dot{x}^0 &= v^0, \\ \dot{v}^0 &= \mathcal{F}, \end{aligned} \qquad (7.1.6)$$

taking into account the extension of the Levi–Civita connection to $\widehat{Q}$. The value that must be minimized in the optimal control problem is $v^0 = \int_I \mathcal{F}dt$.

Now consider the kinematic system (7.1.3) with a cost function $\mathcal{G} \colon Q \times V \to \mathbb{R}$ such that the problem to be solved is the following one.

Statement 7.1.4. *(Kinematic optimal control problem)* *Given $x_a, x_b \in Q$, find $(\gamma, w) \colon I \to Q \times V$ such that*

    1. γ satisfies the endpoint conditions on Q, i.e., $\gamma(a) = x_a$, $\gamma(b) = x_b$;

    2. γ is an integral curve of $X = \sum_{s=1}^{k} w^s Y_s$, i.e., $\dot{\gamma}(t) = X(\gamma(t), w(t))$;

    3. (γ, w) minimizes $\int_I \mathcal{G}(\gamma(t), w(t))\mathrm{d}t$ among all the curves satisfying 1 and 2.

This optimal control problem is denoted by $\Sigma_k = (\Sigma_m, \mathcal{G})$ because the nonholonomic

control system and the kinematic control system considered here are equivalent, see Theorem 7.1.2.

Remark 7.1.5. The problems in Statements 7.1.3 and 7.1.4 are called *fixed optimal control problems* because the domain of definition of the curves is given. However, the *free optimal control problems* may also be defined, analogously to Statement 4.3.1.

As before, let us extend the control system to the manifold $\widehat{Q} = \mathbb{R} \times Q$ such that we look for integral curves of the vector field

$$\widehat{X} = \mathcal{G} \frac{\partial}{\partial x^0} + \sum_{s=1}^{k} w^s Y_s \tag{7.1.7}$$

defined along $\widehat{\pi}_k \colon \widehat{Q} \times V \to \widehat{Q}$. The differential equation added to (7.1.3) is

$$\dot{x}^0 = \mathcal{G} \tag{7.1.8}$$

and the value to be minimized is $x^0 = \int_I \mathcal{G} \mathrm{d}t$.

By Theorem 7.1.2 we know that (7.1.1) and (7.1.3) are equivalent. We are interested in establishing a connection between (7.1.6) and (7.1.8) in such a way that it is obtained a relationship between the problems Σ_m and Σ_k in Statements 7.1.3 and 7.1.4, respectively.

In some sense, $\mathcal{G} = v^0 = \int \mathcal{F}$, but this equality must be properly understood. Observe that $\mathcal{G}$ is a function on $Q \times V$, meanwhile $\mathcal{F}$ is a function on $TQ \times U$. Hence, some simplifications must be considered. Before proceeding with the exact interpretation of $\mathcal{G} = \int \mathcal{F}$, note we also have to check what happens with the minimization conditions when $\mathcal{G} = \int \mathcal{F}$; that is, if the curves minimizing $\int \mathcal{G}$ determine the curves minimizing $\int \mathcal{F}$ and/or in the other way round.

Proposition 7.1.6. *Let $\mathcal{G} \colon I \times Q \to \mathbb{R}$. If $(\dot{\gamma}, u) \colon I \to TQ \times U$ is an optimal curve of a nonholonomic mechanical control system with cost function $\mathcal{F} = \partial \mathcal{G} / \partial t + v^i \partial \mathcal{G} / \partial x^i = \widehat{\mathrm{d}\mathcal{G}} \colon I \times TQ \to \mathbb{R}$, then there exists $w \colon I \to V$ such that (γ, w) is an optimal curve of the kinematic system with cost function $\mathcal{G}$.*

(*Proof*) If $(\dot{\gamma}, u) \colon I \to TQ \times U$ is an integral curve of (7.1.2), then by Theorem 7.1.2 there exist $w \colon I \to V$ such that (γ, w) is an integral curve of (7.1.3). Thus, it only remains to prove that the optimality condition for $\mathcal{F}$ implies the optimality condition for $\mathcal{G}$.

As $(\dot{\gamma}, u)$ minimizes $\int \mathcal{F}$, then, for any other integral curve $(\dot{\widetilde{\gamma}}, \widetilde{u})$ of the vector field (7.1.2) satisfying the endpoint conditions, we have

$$\mathcal{G}(t, \gamma(t)) - \mathcal{G}(a, \gamma(a)) = \int_a^t \widehat{\mathrm{d}\mathcal{G}}(s, \dot{\gamma}(s)) = \int_a^t \mathcal{F}(s, \dot{\gamma}(s)) \mathrm{d}s <$$

$$< \int_a^t \mathcal{F}(s, \dot{\widetilde{\gamma}}(s)) \mathrm{d}s = \int_a^t \widehat{\mathrm{d}\mathcal{G}}(s, \dot{\widetilde{\gamma}}(s)) = \mathcal{G}(t, \widetilde{\gamma}(t)) - \mathcal{G}(a, \widetilde{\gamma}(a)).$$

As γ and $\widetilde{\gamma}$ satisfy the endpoint conditions and none of the cost functions depends on the controls, we have

$$\mathcal{G}(t, \gamma(t)) < \mathcal{G}(t, \widetilde{\gamma}(t))$$

for any $t \in I$, then $\int_I \mathcal{G}(t, \gamma(t))\mathrm{d}t < \int_I \mathcal{G}(t, \widetilde{\gamma}(t))\mathrm{d}t$ by the monotony property of the integral.∎

The result just proved holds provided that the cost function for the nonholonomic optimal control problem is the total derivative of the cost function for the associated kinematic optimal control problem. Observe that both cost functions are independent of the controls.

Remark 7.1.7. Necessary conditions for a curve to be an optimal solution for a nonholonomic optimal control problem is to be an optimal solution to the optimal control problem for the associated kinematic system.

Remark 7.1.8. The inverse implication of Proposition 7.1.6 is not necessarily true. If (γ, w) is an optimal curve for the kinematic system, then for any other integral curve $(\widetilde{\gamma}, \widetilde{w})$ of the kinematic system

$$\int_I \mathrm{d}t \int_a^t \mathcal{F}(s, \dot{\gamma}(s))\mathrm{d}s = \int_I \mathcal{G}(t, \gamma(t))\mathrm{d}t < \int_I \mathcal{G}(t, \widetilde{\gamma}(t))\mathrm{d}t = \int_I \mathrm{d}t \int_a^t \mathcal{F}(s, \dot{\widetilde{\gamma}}(s))\mathrm{d}s \ .$$

The monotone property of the integral is satisfied only in one direction. We should think of conditions such that

$$\text{``} \int_I f < \int_I g \ \Rightarrow \ f < g \ , \quad \text{almost everywhere (a.e.)''}.$$

In general, we cannot expect better results than a.e., hence we will have optimal curves in a weak sense. For instance, if f and g are both positive or both negative, the implication is satisfied. Moreover, if f and g are continuous functions, then the inequality is satisfied everywhere.

Definition 7.1.9. *A nonholonomic optimal control problem $\Sigma_{\mathbf{m}}$ is **equivalent to** an optimal control problem $\Sigma_{\mathbf{k}}$ for the associated kinematic control system if there exists a curve γ on Q and controls $u: I \to U$ and $w: I \to V$ such that $(\dot{\gamma}, u)$ is a solution to Σ_m and (γ, w) is a solution to Σ_k.*

Once the equivalence between the optimal control problems Σ_m and Σ_k has been defined, we can give the following result.

Proposition 7.1.10. *The nonholonomic time–optimal control problem $\Sigma_m = (\Sigma_{\mathcal{D}}, 1)$ is equivalent to the associated kinematic optimal control problem $\Sigma_k = (\Sigma_m, \mathcal{G} = t)$.*

(*Proof*) A solution of the Σ_m gives a solution of Σ_k because of Proposition 7.1.6. Let us prove now that the optimal curves for kinematic systems with $\mathcal{G} = t$ are optimal curves for the nonholonomic time–optimal problem.

If $(\gamma, w): I \to Q \times V$ is a minimizer of $\int_I t\mathrm{d}t = t^2/2$ satisfying the kinematic system, then by Theorem 7.1.2 there exist $u: I \to U$ such that $(\dot{\gamma}, u)$ is an integral curve of the nonholonomic mechanical control system. For any other integral curve of the kinematic system with the same endpoint conditions as γ,

$$t^2/2 < \widetilde{t}^2/2.$$

As t, $\widetilde{t}$ are positive numbers, $t < \widetilde{t}$. That is $(\dot{\gamma}, u)$ is a minimizer of the nonholonomic time–optimal control problem because of the equivalence of integral curves of (7.1.2) and (7.1.3) given by Theorem 7.1.2 and because of the nature of the cost function. The cost function $\mathcal{G} = t$ is positive, so we are in one of the cases where the reverse implication of the monotony property of the integral in Remark 7.1.8 is satisfied. ∎

The optimal control problems considered in Proposition 7.1.10 are free in the sense given in Statement 4.3.1.

Remark 7.1.11. Indeed, it is feasible to consider the time–optimal problem for both control systems, and then $\Sigma_m = (\Sigma_\mathcal{D}, 1)$ and $\Sigma_k = (\Sigma_m, 1)$ are equivalent because the time is positive. Thus, to minimize the time or to minimize the square of the time is exactly the same. Moreover, the curve on the configuration manifolds are related to the same curve on Q since the equations defined by (7.1.2) and (7.1.3) also appear in the extended systems (7.1.5) and (7.1.7), respectively.

The following corollary links with the fact that some optimal control problems can be understood as time–optimal control problems, as for instance happens in the problem of shortest paths in subRiemannian geometry [Liu and Sussmann 1995]. There, to minimize the functional given by the square of the length of a path for a fixed optimal control problem generates the same solutions as to solve the time–optimal control problem with the same control system but with some restriction on the control set and assuming the control vector fields to be orthonormal by the metric.

Definition 7.1.12. *Two **optimal control problems for a control system on** $\mathbf{Q}$ ***are equivalent*** if the sets in Q given by the image of the solutions to both problems are the same.*

Corollary 7.1.13. *If $\Sigma_m = (\Sigma_\mathcal{D}, \mathcal{F})$ is an optimal control problem equivalent to the time–optimal control problem $\Sigma'_m = (\Sigma_\mathcal{D}, 1)$, then Σ_m is equivalent to the associated kinematic optimal control problem $\Sigma_k = (\Sigma_m, 1)$.*

(*Proof*) Observe that the first "equivalent" in the statement is the notion in Definition 7.1.12, but the second "equivalent" is related to the notion in Definition 7.1.9. Keeping this in mind, the proof of this corollary is obtained from Proposition 7.1.10 and Remark 7.1.11. ∎

7.2   Hamiltonian problems for nonholonomic mechanical systems versus kinematic systems

In order to make use of the optimal control problems defined in §7.1.2, let us associate them with a Hamiltonian problem in the sense of Pontryagin's Maximum Principle given in §4.1.4.

As explained in §4.2, the proof of this Principle consists of choosing the initial condition for the fibers of the cotangent bundle in a suitable way. In fact, at the final time b we ask that

$$
\begin{aligned}
\langle \widehat{\lambda}(b), \widehat{v}(b) \rangle &\leq 0, \\
\langle \widehat{\lambda}(b), (-1, \mathbf{0}) \rangle &\geq 0,
\end{aligned}
\tag{7.2.9}
$$

where $\widehat{\lambda}(b) \in T^*_{\widehat{\gamma}(b)}\widehat{M}$ and $\widehat{v}(b)$ are the perturbation vectors given by

$$\widehat{v}(b) = \widehat{X}(\widehat{\gamma}(b), u_b) - \widehat{X}(\widehat{\gamma}(b), u(b)) \qquad (7.2.10)$$

obtained from a determined variation of the control with value $u_b \in U$, see Definition 4.1.5 and [Barbero-Liñán and Muñoz Lecanda 2008b, Lee and Markus 1967, Lewis 2006, Pontryagin et al. 1962]. Moreover, $(-1, \mathbf{0})$ is the direction of decreasing in the functional, remember that as in Chapter 4 the $\mathbf{0}$ in bold points out that $\mathbf{0} \in T_{\gamma(b)}M$. Both vectors are in $T_{\widehat{\gamma}(t_b)}\widehat{M}$. Note that the initial condition for the momenta is, indeed, final since it is taken at final time.

7.2.1 Nonholonomic Pontryagin's Hamiltonian and extremals

For the extended mechanical system $\widehat{Y}$ given in Equation (7.1.5) we have Pontryagin's Hamiltonian function $H_m \colon T^*T\widehat{Q} \times U \to \mathbb{R}$ defined by

$$(\widehat{\Lambda}, u) \longmapsto \left\langle \widehat{\Lambda}, v^0 \frac{\partial}{\partial x^0} + \mathcal{F}\frac{\partial}{\partial v^0} + Z_g + F^V + \sum_{r=1}^{n-k} \mu^r Z_r^V + \sum_{s=1}^{k} u^s Y_s^V \right\rangle .$$

The Lagrange multipliers μ^r are fixed because they are chosen such that $\widehat{Y}$ in Equation (7.1.5) is tangent to $\mathcal{D}$. Another way to consider the Lagrange multipliers is modifying the connection, see [Lewis 1998].

For simplicity to write down the local expressions, we consider the system with zero connection and without external forces. Then the Lagrange multipliers are zero and the local expression of the Hamiltonian function is

$$H_m = p_0 v^0 + q_0 \mathcal{F} + p_i v^i + \sum_{s=1}^{k} q_i u^s Y_s^i ,$$

with Hamilton's equations

$$\begin{aligned}
\dot{x}^0 &= v^0, & \dot{p}_0 &= 0, \\
\dot{x}^i &= v^i, & \dot{p}_i &= -q_0\frac{\partial \mathcal{F}}{\partial x^i} - q_j u^s \frac{\partial Y_s^j}{\partial x^i}, \\
\dot{v}^0 &= \mathcal{F}, & \dot{q}_0 &= -p_0, \\
\dot{v}^i &= u^s Y_s^i, & \dot{q}_i &= -p_i - q_0\frac{\partial \mathcal{F}}{\partial v^i},
\end{aligned} \qquad (7.2.11)$$

where p_i, q_i are the momenta of the states and the velocities, respectively.

Observe that the Hamiltonian is autonomous. Pontryagin's Maximum Principle for this problem tells us that the elementary perturbation vector at time t for $u_1 \in U$ along an optimal curve is given by $\widehat{Y}(\dot{\widehat{\gamma}}(t), u_1) - \widehat{Y}(\dot{\widehat{\gamma}}(t), u(t))$, see (7.1.5) and (7.2.10),

$$\widehat{v}_m(t) = \sum_{s=1}^{k}(u_1^s - u^s(t))Y_s^V + (\mathcal{F}(\dot{\widehat{\gamma}}(t), u_1) - \mathcal{F}(\dot{\widehat{\gamma}}(t), u(t))) \left.\frac{\partial}{\partial v^0}\right|_{\dot{\widehat{\gamma}}(t)} . \qquad (7.2.12)$$

The covector $\widehat{\Lambda}$ associated to the optimal curve by Pontryagin's Maximum Principle satisfies a separation condition analogous to (7.2.9):

$$\langle \widehat{\Lambda}(t), \widehat{v}_m(t) \rangle = \langle \widehat{q}(t), \widehat{v}_m(t) \rangle \leq 0,$$
$$\langle \widehat{\Lambda}(t), (0, \mathbf{0}, -1, \mathbf{0}) \rangle = -q_0(t) \geq 0,$$

where $\widehat{\Lambda}(t) \in T^*_{\widehat{\widehat{\gamma}}(t)} T\widehat{Q}$. The vectors $\widehat{v}_m(t)$ and $(0, \mathbf{0}, -1, \mathbf{0})$ are in $T_{\widehat{\widehat{\gamma}}(t)} T\widehat{Q}$. The two zeroes in bold point out that $(\mathbf{0}, \mathbf{0}) \in T_{\dot{\gamma}(t)} TQ$. Here we do not use the vector $(-1, \mathbf{0})$, but $(0, \mathbf{0}, -1, \mathbf{0})$ as the direction of decreasing in value of the functional $\int_I \mathcal{F}$. Remember that the value to be minimized is v^0.

An analogous separation condition must be satisfied for the vector $(-1, \mathbf{0}, -1, \mathbf{0})$ in order not to contradict the hypothesis of optimality in Theorem 4.3.13; see the first part of the proof of Pontryagin's Maximum Principle in §4.2 and [Barbero-Liñán and Muñoz Lecanda 2008b, Lee and Markus 1967, Lewis 2006, Pontryagin et al. 1962] for the details of that contradiction. But if $(-1, \mathbf{0}, 0, \mathbf{0})$, the direction of decreasing in x^0, is in the same half–space as the perturbation vectors, we do not necessarily arrive at a contradiction because, in general, a decrease in x^0 does not imply a decrease in v^0.

Thus, in the mechanical case, the momenta must separate all the perturbation vectors from the vectors $(0, \mathbf{0}, -1, \mathbf{0})$ and $(-1, \mathbf{0}, -1, \mathbf{0})$, which implies the nonpositiveness of q_0. Taking into account Hamilton's equations (7.2.11), p_0 is constant and by normalization can be consider to be 0, -1 or 1. Then the different possibilities for the mechanical momenta are the following:

1. $p_0 = 0$ and $q_0 = 0$: Here the cost function does not take part in the computations. Note that in this case $(-1, \mathbf{0}, 0, \mathbf{0})$, $(-1, \mathbf{0}, -1, \mathbf{0})$ and $(0, \mathbf{0}, -1, \mathbf{0})$ are in the separating hyperplane defined by the kernel of the momenta.

2. $p_0 = 0$ and $q_0 = -1$: Here the cost function appears in the computations. We have $(-1, \mathbf{0}, 0, \mathbf{0})$ in the separating hyperplane, but $(-1, \mathbf{0}, -1, \mathbf{0})$ and $(0, \mathbf{0}, -1, \mathbf{0})$ are not.

3. $p_0 = -1$ and $q_0 = t + A$: The separation conditions will be satisfied or not, depending on the value of the final time and the constant A. It is necessary that $A < 0$ and $t_f \leq -A$. In this case, $(-1, \mathbf{0}, 0, \mathbf{0})$ is also separated from the perturbation vectors, but it is not in the separating hyperplane.

4. $p_0 = 1$ and $q_0 = -t + A$: The separation conditions will be satisfied or not, depending on the value of the final time and the constant A. It is necessary that $A > 0$ and $t_0 \geq A + 1$. In this case, $(-1, \mathbf{0}, 0, \mathbf{0})$ is contained in the half–space where the perturbation vectors are. Thus it could be associated with a perturbation vector, depending on the directions that are covered by the perturbations of the controls that give rise to the time perturbation cone in Definition 4.3.7.

To sum up, the last two previous cases cause more difficulty in chosing the initial condition for the momenta and the final time in the case of a free optimal control problem. Pontryagin's Maximum Principle guarantees the existence of a momenta, but without determining it. Hence, we

can chose the momenta that appear in the cases 1 and 2, as long as all the necessary conditions in Pontryagin's Maximum Principle are satisfied. Under this restriction, q_0 is a constant that plays a similar role as the constant momentum in Definition 4.1.15. Moreover, our mechanical Hamiltonian turns out to be the Hamiltonian considered in [Bullo and Lewis 2005b] to apply Pontryagin's Maximum Principle for affine connection control systems. Thus, the framework described here guarantees that the second–order condition is satisfied in the approach given in [Bullo and Lewis 2005b] and reviewed in §6.4 because it corresponds with our case of $p_0 = 0$.

In the extended problem for the mechanical system we have added two new coordinates, thus two new covectors have appeared. If we look at Definition 4.1.15, it is not clear how to define the extremals in this case. What we have to remember is that the abnormal extremals are characterized only using the geometry of the control system before extending it; that is, the cost function does not play any role in the computation of abnormal extremals. For the mechanical Hamiltonian H_m, this will happen if and only if p_0 and q_0 vanish simultaneously. Otherwise, the extremals are normal.

Definition 7.2.1. *A curve* $(\dot{\widehat{\gamma}}, u) \colon I \to T\widehat{Q} \times U$ *is*

1. *a **normal extremal for the nonholonomic optimal control problem in Statement 7.1.3** if it is an extremal with*

   (a) *either p_0 being a nonzero constant,*

   (b) *or $q_0 = -1$, then $p_0 = 0$;*

2. *an **abnormal extremal for the nonholonomic optimal control problem in Statement 7.1.3** if it is an extremal with $p_0 = q_0 = 0$;*

7.2.2   Kinematic Pontryagin's Hamiltonian and extremals

For the kinematic system, Pontryagin's Hamiltonian function is

$$
\begin{aligned}
H_k \colon \quad T^*\widehat{Q} \times V \quad &\longrightarrow \quad \mathbb{R} \\
(\widehat{a}, w) \quad &\longmapsto \quad \left\langle \widehat{a}, \mathcal{G}\frac{\partial}{\partial x^0} + \sum_{s=1}^{k} w^s Y_s \right\rangle,
\end{aligned}
$$

with local expression $H_k = a_0 \mathcal{G} + \sum_{l=1}^{k} a_i w^s Y_s^i$, and Hamilton's equations are given by

$$
\begin{aligned}
\dot{x}^0 &= \mathcal{G}, & \dot{a}_0 &= 0, \\
\dot{x}^i &= w^s Y_s^i, & \dot{a}_i &= -a_0 \frac{\partial \mathcal{G}}{\partial x^i} - a_j w^s \frac{\partial Y_s^j}{\partial x^i}.
\end{aligned}
\tag{7.2.13}
$$

The elementary perturbation vector along the optimal curve at t for $w_1 \in V$ is

$$
\widehat{v}_k(t) = \sum_{s=1}^{k} (w_1^s - w^s(t)) Y_s + (\mathcal{G}(\widehat{\gamma}(t), w_1) - \mathcal{G}(\widehat{\gamma}(t), w(t))) \left.\frac{\partial}{\partial x^0}\right|_{\widehat{\gamma}(t)}
\tag{7.2.14}
$$

according to (7.2.10). The covector $\widehat{a}$ defined along the optimal curve that comes from Pontryagin's Maximum Principle satisfies a separation condition analogous to (7.2.9):

$$\begin{aligned}
\langle \widehat{a}(t), \widehat{v}_k(t) \rangle &\leq 0, \\
\langle \widehat{a}(t), (-1, \mathbf{0}) \rangle = -a_0 &\geq 0,
\end{aligned}$$

where $\widehat{v}_k(t)$ and $(-1, \mathbf{0})$ are in $T_{\widehat{\gamma}(t)}\widehat{Q}$, and $\widehat{a}(t) \in T^*_{\widehat{\gamma}(t)}\widehat{Q}$. Here the definition of extremals is exactly the same as in Definition 4.1.15 because there is only one more momentum, just as in Theorem 4.1.14.

7.2.3 Connection between nonholonomic and kinematic momenta

Thus we have two different Hamiltonian problems, one defined on $T^*T\widehat{Q} \times U$ and the other one defined on $T^*\widehat{Q} \times V$. We investigate if there is any way to relate the momenta of both problems such that not only satisfies Hamilton's equations, but also the necessary conditions of Pontryagin's Maximum Principle. Using the Tulczyjew diffeomorphism $\phi_{\widehat{Q}}$ defined in [Tulczyjew 1974], there is a natural way to go from $T^*T\widehat{Q}$ to $T^*\widehat{Q}$ with local expression,

$$\begin{array}{ccccc}
T^*(T\widehat{Q}) & \xrightarrow{\phi_{\widehat{Q}}} & T(T^*\widehat{Q}) & \xrightarrow{\tau_{T^*\widehat{Q}}} & T^*\widehat{Q} \\
(x, v, p, q) & \longmapsto & (x, q, v, p) & \longmapsto & (x, q),
\end{array}$$
(7.2.15)

and vice versa only for a curve $\widehat{a} \colon I \to T^*\widehat{Q}$

$$\begin{array}{ccc}
T^*\widehat{Q} & T(T^*\widehat{Q}) & \xrightarrow{\phi_{\widehat{Q}}^{-1}} & T^*(T\widehat{Q}) \\
(x, q) & (x, q, \dot{x}, \dot{q}) & & (x, \dot{x}, \dot{q}, q) \\
\Big\uparrow \widehat{a} & \nearrow & & \\
I & & &
\end{array}$$
(7.2.16)

where all the coordinates are functions of t and $(x, q, \dot{x}, \dot{q})$ is the canonical lift of a curve $(x(t), q(t))$ to the tangent bundle.

Due to (7.2.15), we could think that, knowing the momenta for the mechanical system, the covector for the kinematic system is given by the momenta of the velocities. But this is not true in general because the momenta for the kinematic system we are looking for must also satisfy the other necessary conditions of Pontryagin's Maximum Principle. Moreover, both Hamilton's equations are not exactly the same as shown by (7.2.11) and (7.2.13).

In the sequel we consider optimal control problems for nonholonomic control systems in Statement 7.1.3 only with the cost function equal to 1 or the cost function given in Proposition 7.1.6; that is, $\mathcal{F} = \widehat{d\mathcal{G}}$.

Proposition 7.2.2. *Let $\widehat{\Lambda} \colon I \to T^*(T\widehat{Q})$ be a covector curve along a solution $\dot{\widehat{\gamma}}$ to the nonholonomic optimal control problem $\Sigma_m = (\Sigma_{\mathcal{D}}, \widehat{d\mathcal{G}})$ ($\Sigma_m = (\Sigma_{\mathcal{D}}, 1)$). If there exists a $t_1 \in I$ such that the momenta q for the velocities satisfy $\langle \widehat{q}(t_1), \widehat{v}_k(t_1) \rangle \leq 0$ for every elementary perturbation vector of the associated kinematic optimal control problem $\Sigma_k = (\Sigma_m, \mathcal{G})$*

($\Sigma_k = (\Sigma_m, 1)$), *then $\widehat{q}(t_1)$ is the initial condition for the kinematic covector along $\widehat{\gamma}$ to solve the Hamilton's equation (7.2.13) for $\Sigma_k = (\Sigma_m, \mathcal{G})$ ($\Sigma_k = (\Sigma_m, 1)$).*

(*Proof*) As $\dot{\widehat{\gamma}}$ is the optimal solution to the problem Σ_m, there exist controls such that $\widehat{\gamma}$ is solution to Σ_k by Proposition 7.1.6 (Remark 7.1.11). Thus, we can apply Pontryagin's Maximum Principle that assures the existence of kinematic momenta. But if for some $t_1 \in I$, we have $\langle \widehat{q}(t_1), \widehat{v}_k(t_1) \rangle \leq 0$, $\widehat{q}(t_1)$ is taking to be the initial condition for the momenta to integrate Hamilton's equations. For this choice of the initial condition, all the necessary conditions of kinematic Pontryagin's Maximum Principle are satisfied. The sign of the above inequality is unchanged because of the property of the integral curves of the complete lift and the cotangent lift of a vector field on $\widehat{Q}$ given in Proposition 2.2.6. Thus at every time $t \in I$ all the perturbation vectors in (7.2.14) and their transport lie in one half–space determined by the hyperplane given by the kernel of the momenta. ∎

Corollary 7.2.3. *The abnormal optimal curves for the problem $\Sigma_m = (\Sigma_{\mathcal{D}}, \widehat{\mathrm{d}\mathcal{G}})$ ($\Sigma_m = (\Sigma_{\mathcal{D}}, 1)$) with covectors satisfying the hypothesis in Proposition 7.2.2 determine abnormal optimal curves for the problem $\Sigma_k = (\Sigma_m, \mathcal{G})$ ($\Sigma_k = (\Sigma_m, 1)$).*

(*Proof*) The momenta of abnormal solutions to the nonholonomic optimal control problems are given by $p_0 = q_0 = 0$ because of Definition 7.2.1. If the hypothesis in Proposition 7.2.2 are satisfied, then the initial condition for the kinematic momenta is $\widehat{q}(t_0)$; that is, $a_0(t_0) = 0$. As a_0 is constant because of Hamilton's equations (7.2.13), the abnormal solutions to the nonholonomic optimal control problem determine abnormal solutions to the associated kinematic optimal control problem. ∎

Remark 7.2.4. There is an analogous result for the normal solutions as long as the momentum for p_0 is taken to be equal to 0. In other words, if we consider the case of normal solutions to nonholonomic optimal control problems with momenta given by $p_0 = 0$ and q_0 to be a nonzero negative constant, then these momenta determine a normal solution to the associated kinematic optimal control problem.

Remark 7.2.5. Observe that the extremals for the kinematic system are extremals for the nonholonomic optimal control problem, considering as necessary conditions for optimality the ones given in Proposition 7.2.2. But from the kinematic momenta is not necessarily possible to specify the mechanical momenta.

7.2.4 Example

Let us prove that the strict abnormal minimizer given by Liu and Sussmann [1995] can be understood as a solution to a nonholonomic control mechanical system, that turns out to be a local strict abnormal minimizer.

Let $Q = \mathbb{R}^3$ with coordinates (x, y, z). We consider the distribution given by

$$\mathcal{D} = \ker \omega = \ker(x^2 \mathrm{d}y - (1 - x)\mathrm{d}z) = \mathrm{span}\{X, Y\}, \qquad (7.2.17)$$

where

$$X = \frac{\partial}{\partial x}, \text{ and } Y = (1-x)\frac{\partial}{\partial}y + x^2\frac{\partial}{\partial z}.$$

Consider the Riemannian metric on Q, $g = dx \otimes dx + \psi(x)(dy \otimes dy + dz \otimes dz)$, where $\psi(x) = ((1-x)^2 + x^4)^{-1}$. Observe that X and Y are a g–orthonormal basis of sections of Q.

Pontryagin's Hamiltonian function for the time–optimal control problem for the kinematic control system associated to $\mathcal{D}$ is

$$H_k(\widehat{a}, w_1, w_2) = a_0 + a_1 w_1 + a_2 w_2 (1-x) + a_3 x^2 w_2.$$

The curve $(\widehat{\gamma}, w)\colon [0,1] \to \widehat{Q}\times[0,1]\times[0,1], t \mapsto (t,0,t,0,0,1)$ satisfying the initial conditions $\gamma(0) = (0,0,0)$ and $\gamma(1) = (0,1,0)$ is a local strict abnormal minimizer for the time–optimal problem. It is impossible to find momenta with $a_0 = -1$ verifying all the necessary conditions of Pontryagin's Maximum Principle. Let us check this. The corresponding Hamilton equations for abnormality and normality are the same for this particular problem:

$$
\begin{aligned}
\dot{x}_0 &= 1, & \dot{a}_0 &= 0, \\
\dot{x} &= w_1, & \dot{a}_1 &= a_2 w_2 - 2xw_2 a_3, \\
\dot{y} &= w_2(1-x), & \dot{a}_2 &= 0, \\
\dot{z} &= x^2 w_2, & \dot{a}_3 &= 0.
\end{aligned}
\tag{7.2.18}
$$

The kinematic Hamiltonian along an abnormal biextremal $\widehat{a}(t) \in T^*_{\widehat{\gamma}(t)}\widehat{Q}$ is

$$H_k(\widehat{a}, w_1, w_2) = a_1 w_1 + a_2 w_2.$$

As the maximum of this Hamiltonian over the controls must be zero and must be given by $w_1 = 0$, $w_2 = 1$, it is necessary to have $a_1 = a_2 = 0$. Then, the abnormal momentum is $\widehat{a}\colon [0,1] \to T^*\widehat{Q}, t \mapsto (0,0,0,a_3)$ along $\widehat{\gamma}(t)$, with a_3 being a nonzero constant. Observe that $H_k(\widehat{a}(t), w_1, w_2) = 0$ for all $t \in [0,1]$.

For the normal case, $a_0 = -1$, in order to guarantee the maximization of the Hamiltonian over the controls along $\widehat{a}(t) \in T^*_{\widehat{\gamma}(t)}\widehat{Q}$ it is necessary $a_1 = 0$, $a_2 = 0$. But then for all $t \in [0,1]$ $H_k(\widehat{a}(t), w_1, w_2) = -1 \neq 0$, contradicting a necessary condition of Pontryagin's Maximum Principle; see Theorem 4.3.13. Thus, γ is a strict abnormal extremal, as claimed. The local optimality is proved in [Liu and Sussmann 1995].

According to the metric, the Christoffel symbols that do not vanish are

$$
\begin{aligned}
\Gamma^1_{22} = \Gamma^1_{33} &= \frac{-1 + x + 2x^3}{((1-x)^2 + x^4)^2}, \\
\Gamma^2_{12} = \Gamma^2_{21} &= \frac{1 - x - 2x^3}{(1-x)^2 + x^4}, \\
\Gamma^3_{13} = \Gamma^3_{31} &= \frac{1 - x - 2x^3}{(1-x)^2 + x^4},
\end{aligned}
$$

where 1 stands for coordinate x and so on. We use numerical subscripts to make easier the

notation in the subsequent paragraphs. Observe that the connection associated to the metric has zero torsion.

Having this in mind, Pontryagin's Hamiltonian function for the time–optimal control problem for the nonholonomic control system is

$$
\begin{aligned}
H_m(\widehat{\Lambda}, u_1, u_2) \;=\;& p_0 v_0 + q_0 + p_1 v_1 + p_2 v_2 + p_3 v_3 \\
& + \; q_1(-\Gamma^1_{22} v_2^2 - \Gamma^1_{33} v_3^2 + u_1) \\
& + \; q_2(-2\Gamma^2_{21} v_2 v_1 + u_2(1-x)) + q_3(-2\Gamma^3_{13} v_1 v_3 + x^2 u_2).
\end{aligned}
$$

Hamilton's equations are

$$
\begin{array}{ll}
\dot{x}_0 = v_0, & \dot{p}_0 = 0, \\[4pt]
\dot{x} = v_1, & \dot{p}_1 = \dfrac{\partial \Gamma^i_{jk}}{\partial x} q_i v^j v^k + q_2 u_2 - 2x u_2 q_3, \\[6pt]
\dot{y} = v_2, & \dot{p}_2 = 0, \\[2pt]
\dot{z} = v_3, & \dot{p}_3 = 0, \\[2pt]
\dot{v}_0 = 1, & \dot{q}_0 = -p_0, \\[2pt]
\dot{v}_1 = -\Gamma^1_{22} v_2^2 - \Gamma^1_{33} v_3^2 + u_1, & \dot{q}_1 = -p_1 + 2q_2\Gamma^2_{21} v_2 + 2q_3\Gamma^3_{13} v_3, \\[2pt]
\dot{v}_2 = -2\Gamma^2_{21} v_2 v_1 + u_2(1-x), & \dot{q}_2 = -p_2 + 2q_1\Gamma^1_{22} v_2 + 2q_2\Gamma^2_{21} v_1, \\[2pt]
\dot{v}_3 = -2\Gamma^3_{13} v_1 v_3 + x^2 u_2, & \dot{q}_3 = -p_3 + 2q_1\Gamma^1_{33} v_3 + 2q_3\Gamma^3_{13} v_1.
\end{array}
\tag{7.2.19}
$$

The strict abnormal minimizer for the kinematic system gives rise to the extremal $\dot{\widehat{\gamma}}(t) = (t^2/2, 0, t, 0, t, 0, 1, 0)$ for the mechanical system. Substituting into the Hamilton's equations for the states along $\widehat{\gamma}$ we have $u_1 = -1$ and $u_2 = 0$.

Remark 7.2.6. The controls are different for the equivalent control systems, which possibility was mentioned in §7.1.1.

For a biextremal $\widehat{\Lambda}(t) \in T^*_{\dot{\widehat{\gamma}}(t)} T\widehat{Q}$,

$$
H_m(\widehat{\Lambda}(t), u_1, u_2) = p_0 t + q_0 + p_2 + q_1(1 + u_1) + q_2 u_2.
$$

The maximum of this Hamiltonian over the controls must be given by $u_1 = -1$ and $u_2 = 0$ for every $t \in [0,1]$, thus $q_1 = q_2 = 0$. From the Hamilton's equations for the momenta we have

$$
\dot{p}_1 = 0, \; p_1 = 0, \; p_2 = 0, \; \dot{q}_3 = -p_3,
$$

where p_3 is constant. These restrictions on the momenta are valid for abnormal and normal extremals because of the cost function considered.

Then, the abnormal momenta, $p_0 = q_0 = 0$, is

$$
\widehat{\Lambda}(t) = (0, 0, 0, p_3, 0, 0, 0, -p_3 t + A)
\tag{7.2.20}
$$

with p_3 and A being constants, that cannot vanish simultaneously. If now we evaluate the Hamiltonian, $H_m(\widehat{\Lambda}(t), u_1, u_2) = 0$. Thus, the abnormal minimizer for the kinematic system gives an abnormal extremal in the nonholonomic case.

Let us try to find the normal momenta; that is, either $q_0 = -1$ or $p_0 = \pm 1$. The different cases are:

1. $p_0 = \pm 1$: then by Hamilton's equations $q_0(t) = \mp t + B$ with a constant B;

2. $p_0 = 0$: then $q_0 = -1$.

Thus, the normal momenta along $\dot{\widehat{\gamma}}$ is either

$$\widehat{\Lambda}_1(t) = (\pm 1, 0, 0, p_3, \mp t + B, 0, 0, -p_3 t + A)$$

or

$$\widehat{\Lambda}_2(t) = (0, 0, 0, p_3, -1, 0, 0, -p_3 t + A).$$

If we evaluate the Hamiltonian H_m at these covectors,

$$
\begin{aligned}
H_m(\widehat{\Lambda}_1(t), u_1, u_2) &= \pm t \mp t + B = B, \\
H_m(\widehat{\Lambda}_2(t), u_1, u_2) &= -1.
\end{aligned}
$$

The Hamiltonian evaluated at $\widehat{\Lambda}_1$ does not vanish because the discussion about the values of p_0 and q_0 in §7.2 and the fact that the initial time is zero imply that $B < 0$. Thus the strict abnormal minimizer for the kinematic system does not give a normal extremal for the mechanical case. Therefore, we have a strict abnormal extremal for the nonholonomic optimal control problem because of Corollary 7.2.3, and Remarks 7.2.4 and 7.2.5.

As for the elementary perturbation vectors (7.2.12) and (7.2.14) along the extremals considered, we have

$$
\begin{aligned}
\widehat{v}_k(t) &= \tilde{w}_1 \frac{\partial}{\partial x} + (\tilde{w}_2 - 1) \frac{\partial}{\partial y}, \\
\widehat{v}_m(t) &= (\tilde{u}_1 - 1) \frac{\partial}{\partial v_1} + \tilde{u}_2 \frac{\partial}{\partial v_2}.
\end{aligned}
$$

For the momenta $\widehat{a}$ and $\widehat{\Lambda}$ calculated for the kinematic and the mechanical systems respectively, the separation conditions (7.2.9) are

$$
\begin{aligned}
\langle (0, 0, 0, a_3), \widehat{v}_k(t) \rangle &= 0, \\
\langle (0, 0, 0, p_3, 0, 0, 0, -p_3 t + A), \widehat{v}_m(t) \rangle &= 0.
\end{aligned}
$$

Thus all the elementary perturbation vectors for the kinematic and the mechanical case are in the separating hyperplane for the momenta found.

Observe that the kinematic momenta and the mechanical momenta are related through (7.2.15) and (7.2.16). At time $t_1 \in I$ the kinematic momenta $\widehat{a}(t_1) = (0, 0, 0, a_3)$ gives the nonholonomic momenta

$$(0, 0, 0, 0, 0, 0, 0, a_3)$$

because of the relation (7.2.16). This is the initial condition at $t_1 \in I$ for the nonholonomic Hamilton's equations (7.2.19). After integrating, it turns out that the nonholonomic momenta is

constant along $\dot{\widehat{\gamma}}$. That is, in (7.2.20) we take $p_3 = 0$ and $A = a_3$ in order to have the nonholonomic momenta determined by the kinematic one. But observe that not all the nonholonomic momenta must be constant, see Equation (7.2.20). Only the constant nonholonomic momenta can be related with a kinematic momenta, as we have just seen.

Vice versa, at time $t_1 \in I$ the nonholonomic momenta in (7.2.20) is

$$(0, 0, 0, p_3, 0, 0, 0, -p_3 t_1 + A).$$

Using (7.2.15) the corresponding kinematic momenta at $t_1 \in I$ is $(0, 0, 0, -p_3 t_1 + A)$. This is the initial condition to integrate the kinematic Hamilton's equations (7.2.18). Then, the kinematic momenta along $\widehat{\gamma}$ is $\widehat{a}(t) = (0, 0, 0, -p_3 t_1 + A)$. For every time $t_1 \in I$, we have a different initial condition for the kinematic momenta and we obtain different kinematic momenta after integrating equations (7.2.18).

Observe that using (7.2.15), we obtain all the possible kinematic momenta. But using (7.2.16) we do not obtain all the possible nonholonomic momenta. Thus, it must be highlighted the fact that the mappings (7.2.15), (7.2.16) defined using Tulczyjew's diffeomorphism do not establish a one–to–one relation between the momenta of both Hamilton's equations for every time.

Remark 7.2.7. Due to Proposition 7.1.10 and Remark 7.1.11, the strict abnormal extremal found for the nonholonomic time–optimal control problem defined by $\mathcal{D}$ in Equation (7.2.17) is a local strict abnormal minimizer.

<p style="text-align: right;">Chapter 8</p>

Skinner–Rusk unified formalism for optimal control systems and applications

A geometric approach to time–dependent optimal control problems is proposed. This formulation is based on the Skinner and Rusk formalism for Lagrangian and Hamiltonian systems, which seems to be a natural geometric setting for Pontryagin's Maximum Principle [Barbero-Liñán et al. 2007]. Following [Cortés et al. 2002b, Echeverría-Enríquez et al. 2004, León et al. 2003], we adapt the Skinner–Rusk formalism [Skinner and Rusk 1983] to study time–dependent optimal control problems. In this way we obtain a geometric version of the weak Maximum Principle given in Chapter 5 that can be applied to a wide range of control systems, provided that the differentiability with respect to controls is assumed and the space of controls is open. For instance, these techniques allow us to tackle geometrically implicit optimal control systems; that is, those where the control equations are implicit, $F(t, x, \dot{x}, u) = 0$ where the x's denote the state variables and the u's the control variables. In fact, systems of differential–algebraic equations appear frequently in control theory, as for instance the descriptor systems in §8.3.2 and [Müller 1998; 1999]. Usually, in the literature, it is assumed that it is possible to rewrite the problem as an explicit system of differential equations, $\dot{x} = G(t, x, u)$, perhaps using the algebraic conditions to eliminate some variables, as in the case of holonomic constraints.

Taking into account the formulation in §2.5 and [Barbero-Liñán et al. 2008], we develop a unified formalism for explicit time–dependent optimal control problems, giving a geometric Pontryagin's Maximum Principle in a weak form in §8.1, and analogously for implicit optimal control systems in §8.2. These two sections contain the main contributions of this chapter. Finally, §8.3 is devoted to examples and applications. First we study the optimal control of Lagrangian systems with controls; that is, systems defined by a Lagrangian and external forces depending on controls [Agrawal and Fabien 1999, Blankenstein et al. 2002, Bloch 2003, Bloch et al. 2000]. These are considered as implicit systems defined by the Euler–Lagrange equations. Second, we analyze a quadratic optimal control problem for a descriptor system [Müller 1998]. We point out the importance of these kinds of systems in engineering problems, see [Müller 1999] and references therein.

8.1   Non–autonomous optimal control problems

In contrast with the approach in Chapter 4, we consider a non–autonomous control system such that the dependence on the time comes not only from the controls, but explicitly. For a m–dimensional manifold M, a control set U in $\mathbb{R}^k$ and a time interval $I \subset \mathbb{R}$, this class of systems are determined by a vector field

$$Y \colon I \times M \times U \to TM \tag{8.1.1}$$

defined along the projection $\pi \colon I \times M \times U \to M$, see Definition 2.2.1. Locally, the set of differential equations given the *state equations* is

$$\dot{x}^i = f^i(t, x^j(t), u^l(t)) \,, \ 1 \le i \le m \,,$$

where t is time, x^j denote the state variables and u^l, $1 \le l \le k$, are the control inputs of the system that must be determined. Prescribing initial conditions of the state variables and fixing control inputs, we know completely the trajectory of the state variables $x^j(t)$. In the sequel, all the functions are assumed to be at least $\mathcal{C}^2$. The objective is the following.

Statement 8.1.1. *(Non–autonomous optimal control problem) Find a $\mathcal{C}^2$–piecewise smooth curve $\gamma \colon I \to I \times M \times U$ and $T \in \mathbb{R}^+$ satisfying the endpoint conditions for the state variables, being an integral curve of the control system (8.1.1), and minimizing the functional*

$$\mathcal{S}[\gamma, u] = \int_0^T \mathcal{F}(\gamma(t)) \, \mathrm{d}t \,.$$

The solutions to this problem are called *optimal trajectories or curves*. The necessary conditions to obtain the solutions to such a problem are provided by *Pontryagin's Maximum Principle* for non–autonomous systems [Pontryagin et al. 1962]. In this case, considering the time as another state variable we have the same statement of the Principle as in Theorem 4.1.14. Now, the control system is given by the vector field $\widehat{Y} \colon (\mathbb{R} \times \widehat{M}) \times U \to T(\mathbb{R} \times \widehat{M})$ along the projection $\pi_t \colon (\mathbb{R} \times \widehat{M}) \times U \to \mathbb{R} \times \widehat{M}$ with local expression

$$\widehat{Y}(t, x, u) = \frac{\partial}{\partial t} + \mathcal{F}(t, x, u)\frac{\partial}{\partial x^0} + f^i(t, x, u)\frac{\partial}{\partial x^i}.$$

Pontryagin's Hamiltonian is a function $H \colon T^*(\mathbb{R} \times M) \times U \to \mathbb{R}$ defined by

$$H(t, x, p, u) = \langle p, \widehat{Y} \rangle,$$

where p includes the momentum for the time—i.e., the new variable—and the momentum for the value of the functional $\mathcal{S}$. In local coordinates (p, p_0, p_i) for the momenta,

$$H(t, x^0, x^i, p, p_0, p_i, u^l) = p + p_0 \mathcal{F}(t, x^i, u^l) + p_j f^j(t, x^i, u^l). \tag{8.1.2}$$

Observe that p denotes the fiber in $T^*(\mathbb{R} \times \widehat{M})$ and also the momentum for the time. It is assumed that the meaning in each situation will be clear from the context.

As explained in the weak Pontryagin's Maximum Principle, Theorem 5.1.1, the maximization of the Hamiltonian over the controls is replaced by

$$\varphi_l \equiv \frac{\partial H}{\partial u^l} = 0, \quad 1 \le l \le k \,, \tag{8.1.3}$$

when the optimal controls are in the interior of the control set and all the functions are differentiable with respect to the controls.

Remark 8.1.2. Despite the interest in abnormality during the whole dissertation, here we focus on the normal case, $p_0 = -1$, for the sake of simplicity. Analogous formalism can be constructed for abnormality.

An optimal control problem is said to be *regular* if the following matrix has maximal rank

$$\left(\frac{\partial \varphi_l}{\partial u^s} \right) = \left(\frac{\partial^2 H}{\partial u^l \partial u^s} \right) \,. \tag{8.1.4}$$

In the following sections we develop a geometric formulation of the weak Maximum Principle, similar to the Skinner–Rusk approach to non–autonomous mechanics as was explained in §2.5 and references therein.

8.1.1   Unified geometric framework for optimal control theory

In a global description, we have a fiber bundle structure $\pi^C \colon C \longrightarrow E$ and $\pi \colon E \longrightarrow \mathbb{R}$, where E is equipped with natural coordinates (t, x^i) and C is the bundle of controls, with coordinates (t, x^i, u^l).

The state equations can be geometrically described as a smooth map $X : C \longrightarrow J^1\pi$ such that the following diagram is commutative:
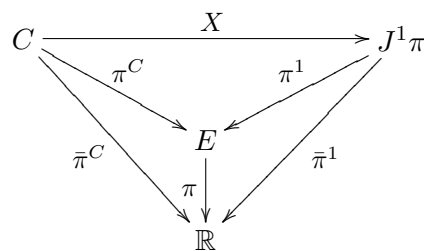
which means that X is a jet field along π^C and also along the projection $\bar{\pi}^C$. Locally we have $X(t, x^i, u^l) = (t, x^i, f^i(t, x^j, u^l))$.

Geometrically, we will assume that an *optimal control system* is determined by the pair $(\mathbf{L}, X)$, where $\mathbf{L} \in \Omega^1(C)$ is a $\bar{\pi}^C$–semibasic 1–form; that is, $\mathbf{L} = \mathcal{F}\mathrm{d}t$, with $\mathcal{F} \in \mathcal{C}^\infty(C)$ representing the cost function; and X is the previous jet field along π^C.

In this framework, Theorem 4.3.14 can be restated as follows.

Theorem 8.1.3. *If $\gamma : I \to C$ is a solution to the non–autonomous optimal problem in State-*

ment 8.1.1, then there exists a curve $\lambda : I \to C \times_E T^*E$ and $p_0 \in \{-1, 0\}$ such that

1. *λ satisfies Hamilton's equations for the Hamiltonian (8.1.2);*

2. *the maximization of this Hamiltonian over the controls is equal to the Hamiltonian along the optimal curve;*

3. *the Hamiltonian vanishes along the optimal curve a.e.*

Now we develop the geometric model of optimal control theory according to the Skinner–Rusk formulation.

The graph of the mapping X, Graph X, is a subset of $C \times_E J^1\pi$ and allows us to define the *extended* and the *restricted control–jet–momentum bundles*, respectively:

$$\mathcal{W}^X = \text{Graph } X \times_E T^*E \,, \quad \mathcal{W}_r^X = \text{Graph } X \times_E J^1\pi^* \,,$$

which are submanifolds of $C \times_E \mathcal{W} = C \times_E J^1\pi \times_E T^*E$ and $C \times_E \mathcal{W}_r = C \times_E J^1\pi \times_E J^1\pi^*$, respectively.

In $\mathcal{W}^X$ and $\mathcal{W}_r^X$ we have natural coordinates (t, x^i, u^l, p, p_i) and (t, x^i, u^l, p_i), respectively. We have the immersions

$$
\begin{aligned}
i^X &: \quad \mathcal{W}^X \hookrightarrow C \times_E \mathcal{W} \,, \quad i^X(t, x^i, u^l, p, p_i) = (t, x^i, u^l, f^i(t, x^j, u^s), p, p_i) \,, \\
i_r^X &: \quad \mathcal{W}_r^X \hookrightarrow C \times_E \mathcal{W}_r \,, \quad i_r^X(t, x^i, u^l, p_i) = (t, x^i, u^l, f^i(t, x^j, u^s), p_i) \,;
\end{aligned}
$$

see Diagram (8.1.5).

$$(8.1.5)$$

Taking the natural projection

$$\sigma_{\mathcal{W}} \colon C \times_E \mathcal{W} \longrightarrow \mathcal{W},$$

and keeping in mind Definition 2.5.1, we can construct the pullback of the coupling 1–form $\hat{\mathcal{C}}$ and of the forms $\Theta_{\mathcal{W}}$ and $\Omega_{\mathcal{W}}$ to $\mathcal{W}^X$:

$$\mathcal{C}_{\mathcal{W}^X} = (\sigma_{\mathcal{W}} \circ i^X)^*\hat{\mathcal{C}} \,, \quad \Theta_{\mathcal{W}^X} = (\sigma_{\mathcal{W}} \circ i^X)^*\Theta_{\mathcal{W}} \,, \quad \Omega_{\mathcal{W}^X} = (\sigma_{\mathcal{W}} \circ i^X)^*\Omega_{\mathcal{W}} = (\rho_2^X)^*\Omega,$$

whose local expressions are:

$$\mathcal{C}_{\mathcal{W}^X} = (p + p_i f^i(t, x^j, u^l))\mathrm{d}t \,, \quad \Theta_{\mathcal{W}^X} = p_i \mathrm{d}x^i + p\mathrm{d}t \,, \quad \Omega_{\mathcal{W}^X} = -\mathrm{d}p_i \wedge \mathrm{d}x^i - \mathrm{d}p \wedge \mathrm{d}t \,.$$

Hence, we can draw the commutative Diagram (8.1.5), where ρ_2^X, ρ_2, $\mu_{\mathcal{W}^X}$ and $\sigma_{\mathcal{W}_r}$ are natural projections.

Furthermore we can define the unique function $H_{\mathcal{W}^X} : \mathcal{W}^X \longrightarrow \mathbb{R}$ by the condition

$$\mathcal{C}_{\mathcal{W}^X} - (\rho_1^X)^* \mathbf{L} = H_{\mathcal{W}^X} \mathrm{d}t \,,$$

where $\rho_1^X : \mathcal{W}^X \longrightarrow C$ is the natural projection. This function $H_{\mathcal{W}^X}$ is locally described as

$$H_{\mathcal{W}^X}(t, x^i, u^l, p, p_i) = p + p_i f^i(t, x^j, u^l) - \mathcal{F}(t, x^j, u^l). \tag{8.1.6}$$

This is exactly the natural Pontryagin's Hamiltonian function (8.1.2), cf. Equation (2.5.20).

Let $\mathcal{W}_0^X$ be the submanifold of $\mathcal{W}^X$ defined by the vanishing of $H_{\mathcal{W}^X}$:

$$\mathcal{W}_0^X = \left\{ w \in \mathcal{W}^X \mid H_{\mathcal{W}^X}(w) = 0 \right\} \,.$$

In local coordinates, $\mathcal{W}_0^X$ is given by the constraint

$$p + p_i f^i(t, x^j, u^l) - \mathcal{F}(t, x^j, u^l) = 0.$$

Observe that, in this way, we recover the fact that Pontryagin's Hamiltonian is zero along the optimal curves. An obvious set of coordinates in $\mathcal{W}_0^X$ is (t, x^i, u^l, p_i). We denote by

$$j_0^X : \mathcal{W}_0^X \longrightarrow \mathcal{W}^X$$

the natural embedding; in local coordinates,

$$j_0^X(t, x^i, u^l, p_i) = (t, x^i, u^l, \mathcal{F}(t, x^j, u^s) - p_i f^i(t, x^j, u^s), p_j).$$

In a way similar to that of Proposition 2.5.2, we may prove the following result.

Proposition 8.1.4. $\mathcal{W}_0^X$ *is a codimension 1, $\mu_{\mathcal{W}^X}$–transverse submanifold of $\mathcal{W}^X$, diffeomorphic to $\mathcal{W}_r^X$.*

As a consequence, the submanifold $\mathcal{W}_0^X$ induces a section of the projection $\mu_{\mathcal{W}^X}$,

$$\hat{h}^X : \mathcal{W}_r^X \longrightarrow \mathcal{W}^X \,. \tag{8.1.7}$$

Locally, $\hat{h}^X$ is specified by giving the local *Hamiltonian function* $\hat{H}^X = p_j f^j - \mathcal{F}$; that is, $\hat{h}^X(t, x^i, u^l, p_i) = (t, x^i, u^l, p = -\hat{H}^X, p_i)$. The map $\hat{h}^X$ is called a *Hamiltonian section* of $\mu_{\mathcal{W}^X}$.

Thus, we can draw the commutative Diagram (8.1.8) where all the projections are natural.

Finally we define the forms

$$\Theta_{\mathcal{W}_0^X} = (\jmath_0^X)^* \Theta_{\mathcal{W}^X} , \quad \Omega_{\mathcal{W}_0^X} = (\jmath_0^X)^* \Omega_{\mathcal{W}^X},$$

with local expressions

$$\Theta_{\mathcal{W}_0^X} = p_i \mathrm{d}x^i + (\mathcal{F} - p_i f^i)\mathrm{d}t , \quad \Omega_{\mathcal{W}_0^X} = -\mathrm{d}p_i \wedge \mathrm{d}x^i - \mathrm{d}(\mathcal{F} - p_i f^i) \wedge \mathrm{d}t .$$

$$(8.1.8)$$
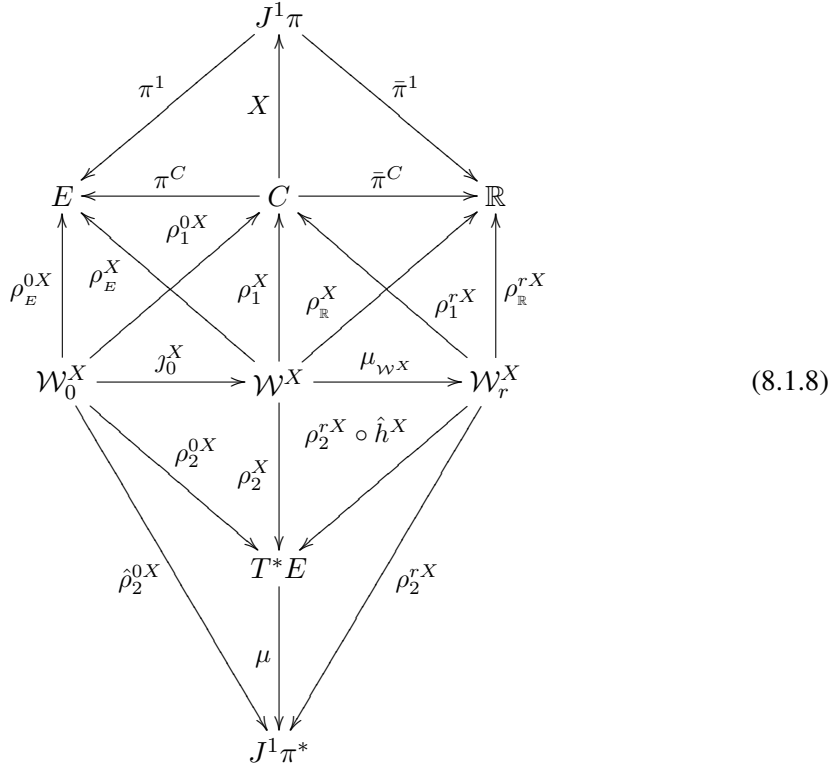
8.1.2   Optimal control equations

Now we are going to establish the dynamical problem for the system $(\mathcal{W}_0^X, \Omega_{\mathcal{W}_0^X})$, and as a consequence we obtain a geometrical version of the weak form of the Maximum Principle.

Proposition 8.1.5. *If* $(\mathbf{L}, X)$ *is a regular optimal control problem, then there exist a submanifold* $\mathcal{W}_1^X$ *of* $\mathcal{W}_0^X$ *and a unique vector field* $Z \in \mathfrak{X}(\mathcal{W}_0^X)$ *tangent to* $\mathcal{W}_1^X$ *such that*

$$[i_Z \Omega_{\mathcal{W}_0^X}]|_{\mathcal{W}_1^X} = 0, \tag{8.1.9}$$
$$[i_Z \mathrm{d}t]|_{\mathcal{W}_1^X} = 1. \tag{8.1.10}$$

The integral curves λ *of* Z *satisfy the necessary conditions of Theorem 8.1.3.*

(*Proof*) In a natural coordinate system we have

$$Z = f\frac{\partial}{\partial t} + A^i\frac{\partial}{\partial x^i} + B^l\frac{\partial}{\partial u^l} + C_i\frac{\partial}{\partial p_i} \, ,$$

where f, A^i, B^l, C_i are functions on $\mathcal{W}_0^X$ to be determined. Then Equation (8.1.10) leads to $f = 1$, and from (8.1.9) we have

coefficients in $\mathrm{d}p_i$: $f^i - A^i = 0,$ (8.1.11)

coefficients in $\mathrm{d}u^l$: $\dfrac{\partial \mathcal{F}}{\partial u^l} - p_j\dfrac{\partial f^j}{\partial u^l} = 0,$ (8.1.12)

coefficients in $\mathrm{d}x^i$: $\dfrac{\partial \mathcal{F}}{\partial x^i} - p_j\dfrac{\partial f^j}{\partial x^i} - C_i = 0,$ (8.1.13)

coefficients in $\mathrm{d}t$: $-A^i\dfrac{\partial \mathcal{F}}{\partial x^i} + A^i p_j\dfrac{\partial f^j}{\partial x^i} - B^l\dfrac{\partial \mathcal{F}}{\partial u^l} + B^l p_j\dfrac{\partial f^j}{\partial u^l} + C_i f^i = 0 \, .$

Now, if $\lambda(t) = (t, x^i(t), u^l(t), p_i(t))$ is an integral curve of Z, we have

$$A^i = \frac{\mathrm{d}x^i}{\mathrm{d}t}, \; B^l = \frac{\mathrm{d}u^l}{\mathrm{d}t}, \; C_i = \frac{\mathrm{d}p_i}{\mathrm{d}t}.$$

Pontryagin's Hamiltonian function is $H = p + p_i f^i - \mathcal{F}$. As we are in $\mathcal{W}_0^X$, the condition $H = 0$ is satisfied. Furthermore, we make the following observations

- From (8.1.11) we deduce that $A^i = f^i$; that is, $\dfrac{\mathrm{d}x^i}{\mathrm{d}t} = \dfrac{\partial H}{\partial p_i}$, from Hamilton's equations.

- Equations (8.1.12) determine a new set of conditions

$$\varphi_l = \frac{\partial \mathcal{F}}{\partial u^l} - p_j\frac{\partial f^j}{\partial u^l} = \frac{\partial H}{\partial u^l} = 0 \qquad (8.1.14)$$

which are equations (8.1.3). We assume that they define the new submanifold $\mathcal{W}_1^X$ of $\mathcal{W}_0^X$. We denote by $\jmath_1^X : \mathcal{W}_1^X \hookrightarrow \mathcal{W}_0^X$ the natural embedding.

- From (8.1.13) we completely determine the functions $C_i = \dfrac{\mathrm{d}p_i}{\mathrm{d}t} = -\dfrac{\partial H}{\partial x^i}$ from Hamilton's equations.

- Finally, using (8.1.11), (8.1.12) and (8.1.13) it is easy to prove that equations for the coefficient of $\mathrm{d}t$ holds identically.

Furthermore Z must be tangent to $\mathcal{W}_1^X$; that is,

$$Z(\varphi_l) = Z\left(\frac{\partial H}{\partial u^l}\right) = 0 \qquad (\text{on } \mathcal{W}_1^X).$$

In other words,

$$0 = \frac{\partial^2 H}{\partial t \partial u^l} + f^i \frac{\partial^2 H}{\partial x^i \partial u^l} + B^s \frac{\partial^2 H}{\partial u^s \partial u^l} - \frac{\partial H}{\partial x^i} \frac{\partial^2 H}{\partial p_i \partial u^l} \qquad (\text{on } \mathcal{W}_1^X). \qquad (8.1.15)$$

However, as the optimal control problem is regular, the matrix $\dfrac{\partial^2 H}{\partial u^s \partial u^l}$ has maximum rank. Then Equations (8.1.15) determine all the coefficients B^s. ∎

As a direct consequence of this proposition, we state the intrinsic version of Theorem 8.1.3.

Theorem 8.1.6. *(Geometric weak Pontryagin's Maximum Principle) If $\gamma\colon I \to C$ is a solution to the regular optimal control problem given by $(\mathbf{L}, X)$, then there exists an integral curve of a vector field $Z \in \mathfrak{X}(\mathcal{W}_0^X)$, whose projection to C is γ, and such that Z is a solution to the equations*

$$i_Z \Omega_{\mathcal{W}_0^X} = 0, \quad i_Z \mathrm{d}t = 1,$$

in a submanifold $\mathcal{W}_1^X$ of $\mathcal{W}_0^X$, which is locally given by the condition (8.1.14).

Remark 8.1.7. In fact, Equation (8.1.10) could be relaxed to the condition

$$i_Z \mathrm{d}t \neq 0,$$

which determines vector fields transversal to π whose integral curves are equivalent to those obtained above, up to reparametrization.

Note that, using the implicit function theorem on the equations $\varphi_l = 0$, we get the functions $u^l = u^l(t, x, p)$. Therefore, for regular control problems, we can choose local coordinates (t, x^i, p_i) on $\mathcal{W}_1^X$, and $H|_{\mathcal{W}_1^X}$ is locally a function of these coordinates.

If the control problem is not regular, then one has to implement a constraint algorithm to obtain a final constraint submanifold $\mathcal{W}_f^X$ (if it exists) where the vector field Z is tangent; see, for instance, §2.3.2, Chapter 5 and [Barbero-Liñán and Muñoz Lecanda 2008a, Delgado-Téllez and Ibort 2003].

If $\jmath_1\colon \mathcal{W}_1^X \to \mathcal{W}_0^X$ is the natural embedding, the form $\Omega_{\mathcal{W}_1^X} = (\jmath_1^X)^* \Omega_{\mathcal{W}_0^X}$ is locally written as

$$\Omega_{\mathcal{W}_1^X} = -\mathrm{d}p_i \wedge \mathrm{d}x^i - \mathrm{d}H|_{\mathcal{W}_1^X} \wedge \mathrm{d}t.$$

Hence, for optimal control problems, taking into account the regularity of the matrix (8.1.4), we have the following result.

Proposition 8.1.8. *If the optimal control problem is regular, then $(\mathcal{W}_1^X, \Omega_{\mathcal{W}_1^X}, \mathrm{d}t)$ is a cosymplectic manifold; that is, $(\Omega_{\mathcal{W}_1^X})^m \wedge \mathrm{d}t$ is a volume form; see [León and Rodrigues 1989].*

8.2 Implicit optimal control problems

The formalism presented in §8.1.1 is valid for a more general class of optimal control problems not previously considered from a geometric perspective: optimal control problems whose state

equations are *implicit*; that is,

$$\Psi^\alpha(t, q, \dot{q}, u) = 0 \, , \ 1 \le \alpha \le s \, , \ \text{with } \mathrm{d}\Psi^1 \wedge \ldots \wedge \mathrm{d}\Psi^s \ne 0 \, . \tag{8.2.16}$$

There are several examples of these kinds of optimal control problems, some of them coming from engineering applications. In §8.3 we study two specific examples: the controlled Lagrangian systems which play a relevant role in robotics and the descriptor systems which appear in electrical engineering.

8.2.1 Unified geometric framework for implicit optimal control problems

From a more geometric point of view, we may interpret Equations (8.2.16) as constraint functions determining a submanifold M_C of $C \times_E J^1\pi$, with natural embedding $\jmath^{M_C} \colon M_C \hookrightarrow C \times_E J^1\pi$. We will also assume that $(\pi^C \times \pi^1) \circ \jmath^{M_C} \colon M_C \longrightarrow E$ is a surjective submersion.

In this situation, the techniques presented in §8.1 are still valid. Now the implicit optimal control system is determined by the data $(\mathbf{L}, M_C)$, where $\mathbf{L} \in \Omega^1(M_C)$ is a semibasic form with respect to the projection $\tau^{M_C} \colon M_C \longrightarrow \mathbb{R}$, and hence it can be written as $\mathbf{L} = \mathcal{F}\mathrm{d}t$, for some $\mathcal{F} \in \mathcal{C}^\infty(M_C)$. First define the *extended control–jet–momentum manifold* and the *restricted control–jet–momentum manifold*

$$\mathcal{W}^{M_C} = M_C \times_E T^*E, \quad \mathcal{W}_r^{M_C} = M_C \times_E J^1\pi^*,$$

which are submanifolds of $C \times_E \mathcal{W} = C \times_E J^1\pi \times_E T^*E$ and $C \times_E \mathcal{W}_r = C \times_E J^1\pi \times_E J^1\pi^*$, respectively.

We have the canonical immersions (embeddings)

$$i^{M_C} \colon \mathcal{W}^{M_C} \hookrightarrow C \times_E \mathcal{W}, \quad i_r^{M_C} \colon \mathcal{W}_r^{M_C} \hookrightarrow C \times_E \mathcal{W}_r.$$

So we can draw a diagram analogous to (8.1.5) replacing the core of the diagram by

$$\begin{array}{ccc}
 & M_C & \\
{\scriptstyle \rho_1^{M_C}} \nearrow & & \nwarrow {\scriptstyle \rho_1^{rM_C}} \\
\mathcal{W}^{M_C} & \xrightarrow{\quad \mu_{\mathcal{W}^{M_C}} \quad} & \mathcal{W}_r^{M_C}
\end{array}$$

where all the projections are natural.

Now, keeping in mind Definition 2.5.1, consider the pullback of the coupling 1–form $\hat{\mathcal{C}}$ and the forms $\sigma_{\mathcal{W}}^*\Theta_{\mathcal{W}}$ and $\sigma_{\mathcal{W}}^*\Omega_{\mathcal{W}}$ to $\mathcal{W}^{M_C}$ by the map $i^{M_C} \colon \mathcal{W}^{M_C} \longrightarrow C \times_E \mathcal{W}$; that is

$$\mathcal{C}_{\mathcal{W}^{M_C}} = (\sigma_{\mathcal{W}} \circ i^{M_C})^*\hat{\mathcal{C}}, \ \Theta_{\mathcal{W}^{M_C}} = (\sigma_{\mathcal{W}} \circ i^{M_C})^*\Theta_{\mathcal{W}}, \ \Omega_{\mathcal{W}^{M_C}} = (\sigma_{\mathcal{W}} \circ i^{M_C})^*\Omega_{\mathcal{W}},$$

and denote by $\hat{C} \in \mathcal{C}^\infty(\mathcal{W}^{M_C})$ the unique function such that $\mathcal{C}_{\mathcal{W}^{M_C}} = \hat{C}\mathrm{d}t$. Finally, let $H_{\mathcal{W}^{M_C}} \colon \mathcal{W}^{M_C} \longrightarrow \mathbb{R}$ be the unique function such that $\mathcal{C}_{\mathcal{W}^{M_C}} - (\rho_1^{M_C})^*\mathbf{L} = H_{\mathcal{W}^{M_C}}\mathrm{d}t$. Observe that $H_{\mathcal{W}^{M_C}} = \hat{C} - \hat{\mathcal{F}}$, where $\hat{\mathcal{F}} = (\rho_1^{M_C})^*\mathcal{F}$, and remember that $H_{\mathcal{W}^{M_C}}$ is the

Pontryagin's Hamiltonian function; see (8.1.6).

Let $\mathcal{W}_0^{M_C}$ be the submanifold of $\mathcal{W}^{M_C}$ defined by the vanishing of $H_{\mathcal{W}^{M_C}}$, i.e.,

$$\mathcal{W}_0^{M_C} = \{w \in \mathcal{W}^{M_C} \mid H_{\mathcal{W}^{M_C}}(w) = (\hat{C} - \hat{\mathcal{F}})(w) = 0\}, \tag{8.2.17}$$

and denote by $\jmath_0^{M_C} : \mathcal{W}_0^{M_C} \hookrightarrow \mathcal{W}^{M_C}$ the natural embedding. As in Proposition 2.5.2 we may prove the following.

Proposition 8.2.1. $\mathcal{W}_0^{M_C}$ *is a codimension 1, $\mu_{\mathcal{W}^{M_C}}$–transverse submanifold of $\mathcal{W}^{M_C}$, diffeomorphic to $\mathcal{W}_r^{M_C}$.*

As a consequence, the submanifold $\mathcal{W}_0^X$ induces a section of the projection $\mu_{\mathcal{W}^{M_C}}$,

$$\hat{h}^{M_C} : \mathcal{W}_r^{M_C} \longrightarrow \mathcal{W}^{M_C}.$$

Then we can draw the following diagram,
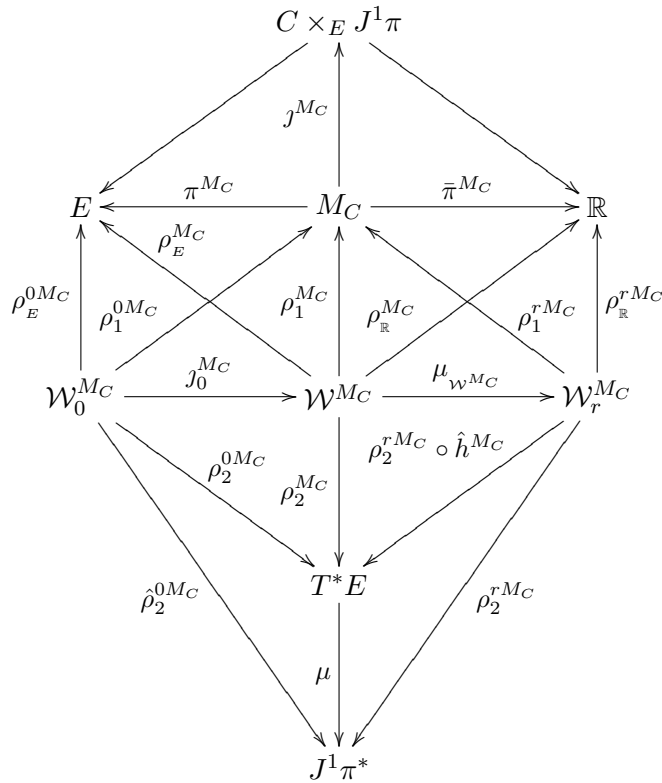
which is analogous to (8.1.8), where all the projections are natural.

Finally, we define the forms

$$\Theta_{\mathcal{W}_0^{M_C}} = (\jmath_0^{M_C})^* \Theta_{\mathcal{W}^{M_C}}, \quad \Omega_{\mathcal{W}_0^{M_C}} = (\jmath_0^{M_C})^* \Omega_{\mathcal{W}^{M_C}}.$$

8.2.2   Optimal control equations

Now, we will see how the dynamics of the optimal control problem $(\mathbf{L}, M_C)$ is determined by the solutions (when they exist) of the equations

$$i_Z \Omega_{\mathcal{W}_0^{M_C}} = 0, \tag{8.2.18}$$
$$i_Z \mathrm{d}t = 1, \tag{8.2.19}$$

for $Z \in \mathfrak{X}(\mathcal{W}_0^{M_C})$.

As in §8.1.2, the Equation (8.2.19) can be relaxed to the condition

$$i_Z \mathrm{d}t \neq 0 \ .$$

In order to work in local coordinates we need the following proposition, whose proof is obvious:

Proposition 8.2.2. *For a given $w \in \mathcal{W}_0^{M_C}$, the following conditions are equivalent:*

1. *there exists a vector $Z_w \in T_w \mathcal{W}_0^{M_C}$ verifying that*

$$\Omega_{\mathcal{W}_0^{M_C}}(Z_w, Y_w) = 0 \ , \text{ for every } Y_w \in T_w \mathcal{W}_0^{M_C} \ ;$$

2. *there exists a vector $Z_w \in T_w(C \times_E \mathcal{W})$ satisfying*

   (i) $Z_w \in T_w \mathcal{W}_0^{M_C}$ and
   (ii) $i_{Z_w}(\sigma_{\mathcal{W}}^* \Omega_{\mathcal{W}})_w \in (T_w \mathcal{W}_0^{M_C})^0$, where $(T_w \mathcal{W}_0^{M_C})^0$ is the annihilator of $T_w \mathcal{W}_0^{M_C}$.

As a consequence of this last proposition, we can obtain the implicit optimal control equations using condition 2 as follows. There exists $Z \in \mathfrak{X}(C \times_E \mathcal{W})$ such that

(i) Z is tangent to $\mathcal{W}_0^{M_C}$ and

(ii) the 1–form $i_Z \sigma_{\mathcal{W}}^* \Omega_{\mathcal{W}}$ is null on the vector fields tangent to $\mathcal{W}_0^{M_C}$.

As $\mathcal{W}_0^{M_C}$ is defined in (8.2.17), and the constraints are $\Psi^\alpha = 0$ and $\hat{C} - \hat{\mathcal{F}} = 0$, then there exist $\mu_\alpha, \mu \in \mathcal{C}^\infty(C \times_E \mathcal{W})$ such that

$$(i_Z \sigma_{\mathcal{W}}^* \Omega_{\mathcal{W}})|_{\mathcal{W}_0^{M_C}} = (\mu_\alpha \mathrm{d}\Psi^\alpha + \mu \mathrm{d}(\hat{C} - \hat{\mathcal{F}}))|_{\mathcal{W}_0^{M_C}} \ .$$

As usual, the undetermined functions μ_α's and μ are called Lagrange multipliers.

Now using coordinates $(t, x^i, u^l, v^i, p, p^i)$ in $C \times_E \mathcal{W}$, we look for a vector field

$$Z = \frac{\partial}{\partial t} + A^i \frac{\partial}{\partial x^i} + B^l \frac{\partial}{\partial u^l} + C^i \frac{\partial}{\partial v^i} + D_i \frac{\partial}{\partial p_i} + E \frac{\partial}{\partial p} \ ,$$

where A^i, B_l, C^i, D_i, E are unknown functions in $\mathcal{W}_0^{M_C}$ satisfying

$$
\begin{aligned}
0 &= i_Z\left(\mathrm{d}x^i \wedge \mathrm{d}p_i + \mathrm{d}t \wedge \mathrm{d}p\right) - \mu_\alpha \mathrm{d}\Psi^\alpha - \mu \mathrm{d}(p + p_i v^i - \mathcal{F}(t,x,u)) \\
&= \left(-E - \mu_\alpha \frac{\partial \Psi^\alpha}{\partial t} + \mu \frac{\partial \mathcal{F}}{\partial t}\right)\mathrm{d}t + \left(\mu \frac{\partial \mathcal{F}}{\partial x^i} - \mu_\alpha \frac{\partial \Psi^\alpha}{\partial x^i} - D_i\right)\mathrm{d}x^i \\
&\quad + \left(\mu \frac{\partial \mathcal{F}}{\partial u^l} - \mu_\alpha \frac{\partial \Psi^\alpha}{\partial u^l}\right)\mathrm{d}u^l + \left(-\mu p_i - \mu_\alpha \frac{\partial \Psi^\alpha}{\partial v^i}\right)\mathrm{d}v^i \\
&\quad + (A^i - \mu v^i)\mathrm{d}p_i + (1 - \mu)\mathrm{d}p \,.
\end{aligned}
$$

Thus, we obtain $\mu = 1$, and

$$
A^i = v^i \,,\; D_i = \frac{\partial \mathcal{F}}{\partial x^i} - \lambda_\alpha \frac{\partial \Psi^\alpha}{\partial x^i} \,,\; E = \frac{\partial \mathcal{F}}{\partial t} - \mu_\alpha \frac{\partial \Psi^\alpha}{\partial t} \,,\; p_i = -\mu_\alpha \frac{\partial \Psi^\alpha}{\partial v^i} \,,\; 0 = \frac{\partial \mathcal{F}}{\partial u^l} - \mu_\alpha \frac{\partial \Psi^\alpha}{\partial u^l} \,,
$$

together with the tangency conditions

$$
\begin{aligned}
0 &= Z(\Psi^\alpha)|_{\mathcal{W}_0^{M_C}} = \left.\left(\frac{\partial \Psi^\alpha}{\partial t} + A^i \frac{\partial \Psi^\alpha}{\partial x^i} + B^l \frac{\partial \Psi^\alpha}{\partial u^l} + C^i \frac{\partial \Psi^\alpha}{\partial v^i}\right)\right|_{\mathcal{W}_0^{M_C}} \,, \\
0 &= Z(p + p_i v^i - \mathcal{F}(t,x,u))|_{\mathcal{W}_0^{M_C}} \,.
\end{aligned}
$$

Therefore, the equations of motion are:

$$
\frac{\mathrm{d}}{\mathrm{d}t}\left(\mu_\alpha(t)\frac{\partial \Psi^\alpha}{\partial v^i}(t,x(t),\dot{x}(t),u(t))\right) + \frac{\partial \mathcal{F}}{\partial x^i}(t,x(t),u(t)) - \mu_\alpha(t)\frac{\partial \Psi^\alpha}{\partial x^i}(t,x(t),\dot{x}(t),u(t)) = 0,
$$

$$
\frac{\partial \mathcal{F}}{\partial u^l}(t,x(t),u(t)) - \mu_\alpha(t)\frac{\partial \Psi^\alpha}{\partial u^l}(t,x(t),\dot{x}(t),u(t)) = 0,
$$

$$
\Psi^\alpha(t,x(t),\dot{x}(t),u(t)) = 0.
$$
$$(8.2.20)$$

Let $\mathcal{F}_0 = \mathcal{F} - \mu_\alpha \Psi^\alpha$ be the classical extended Lagrangian for constrained systems. Then these last equations are the usual dynamical equations in optimal control obtained by applying the Lagrange multipliers method to the constrained variational problem. The first equation in (8.2.20) are the Euler–Lagrange equations for $\mathcal{F}_0$, the second one is the extremum necessary condition at interior points, and the third one are the constraints.

Here we recover again that idea that the optimal control problems come from the calculus of variations, as was roughly pointed out in §4.6. For more details about it see [Lewis 2006, Sussmann and Willems 1997].

Remark 8.2.3. In the particular case that $\Psi^j = v^j - f^j = 0$, the vector field Z obtained above is just the image of the vector field obtained in §8.1.2 by the Hamiltonian section (8.1.7), as a simple calculation in coordinates shows.

Remark 8.2.4. If k is any integer, another obvious but significant remark is that we can take $\bar{\pi}^k \colon J^k\pi \longrightarrow \mathbb{R}$ (the bundle of k–jets of π) instead of $\pi \colon E \longrightarrow \mathbb{R}$, and hence $J^k\bar{\pi}^k$ and $T^*J^k\pi$ instead of $J^1\bar{\pi}^1$ and T^*E, respectively. These changes allows us to address those optimal control problems where we have $\Phi^{kC} \colon C \longrightarrow J^k\pi$; that is, we deal with higher–order equations, and their solutions must satisfy that $(\gamma(t), j^{k+1}(\pi^k \circ \Phi^{kC} \circ \gamma)(t)) \in M$, where M

is a submanifold of $C \times_{J^k \pi} J^{k+1} \pi$.

8.3   Applications and examples

To conclude this chapter we study two different optimal control problems as application of the theory described in §8.2 for implicit optimal control problems.

8.3.1   Optimal control of Lagrangian systems with controls

See §2.6 and [Barbero-Liñán et al. 2007] for previous geometric concepts which are needed in this section related with Tulczyjew's operators. For a complete study of these systems see [Blankenstein et al. 2002, Bloch et al. 2000] and references therein.

First, recall that, associated with every jet bundle $J^1 \pi$, we have the *contact system*, which is a subbundle $\mathcal{C}_\pi$ of $T^* J^1 \pi$ whose fibres at every $j^1 \phi(t) \in J^1 \pi$ are defined as

$$\mathcal{C}_\pi(j^1 \phi(t)) = \{\alpha \in T^*_{j^1 \phi(t)}(J^1 \pi) \mid \alpha = (T_{j^1 \phi(t)} \pi^1 - T_{j^1 \phi(t)}(\phi \circ \bar{\pi}^1))^* \beta, \ \beta \in V^*_{\phi(t)} \pi\}) \ .$$

One may readily see that a local basis for the sections of this bundle is given by $\{dq^i - v^i dt\}$.

Now we provide a definition of a *controlled–force*, which allows dependence on time, configuration, velocities and control inputs. In a global description, one assumes a fiber bundle structure $\Phi^{1C} : C \longrightarrow J^1 \pi$, where C is the bundle of controls, with coordinates (t, x, v, u). Then a controlled–force is a smooth map $\mathbf{F} : C \longrightarrow \mathcal{C}_\pi$, so that $\pi_{J^1 \pi} \circ \mathbf{F} = \Phi^{1C}$; see diagram (2.6.24). In a natural chart, a controlled–force is represented by

$$\mathbf{F}(t, x, v, u) = \mathbf{F}_i(t, x, v, u)(dx^i - v^i dt) \ .$$

According to notation in §2.6.3, a *controlled Lagrangian system* is a pair $(\mathcal{L}, \mathbf{F})$ which determines an implicit control system described by the subset D_C of $C \times_{J^1 \pi} J^2 \pi$:

$$
\begin{aligned}
D_C &= \{(c, \hat{p}) \in C \times_{J^1 \pi} J^2 \pi \mid (\iota_1^* d_T \Theta_{\mathcal{L}} - (\pi_1^2)^* dL)(\hat{p}) = ((\pi_1^2)^* \mathbf{F})(c)\} \\
&= \{(c, \hat{p}) \in C \times_{J^1 \pi} J^2 \pi \mid \mathcal{E}_{\mathcal{L}}(\hat{p}) = ((\pi_1^2)^* \mathbf{F})(c)\} \\
&= \{(c, \hat{p}) \in C \times_{J^1 \pi} J^2 \pi \mid (\mathcal{E}_{\mathcal{L}} \circ pr_2 - (\pi_1^2)^* \mathbf{F} \circ pr_1)(c, \hat{p}) = 0\},
\end{aligned}
$$

where pr_1 and pr_2 are the natural projections from $C \times_{J^1 \pi} J^2 \pi$ onto the factors. In fact, D_C is not necessarily a submanifold of $C \times_{J^1 \pi} J^2 \pi$. There are a lot of cases where this does happen. In local coordinates

$$
\begin{aligned}
D_C = \Big\{ &(t, q, v, w, u) \in C \times_{J^1 \pi} J^2 \pi \ \Big| \ \frac{\partial^2 L}{\partial v^i \partial v^j}(t, q, v) w^j + \frac{\partial^2 L}{\partial v^i \partial q^j}(t, q, v) v^j \\
&+ \frac{\partial^2 L}{\partial v^i \partial t}(t, q, v) - \frac{\partial L}{\partial q^i}(t, q, v) - \mathbf{F}_i(t, q, v, u) = 0 \Big\} \ .
\end{aligned}
$$

A solution to the controlled Lagrangian system $(\mathcal{L}, \mathbf{F})$ is a map $\gamma \colon \mathbb{R} \longrightarrow C$ satisfying:

(i) $\Phi^{1C} \circ \gamma = j^1 (\pi^1 \circ \Phi^{1C} \circ \gamma)$ and

(ii) $(\gamma(t), j^2(\pi^1 \circ \Phi^{1C} \circ \gamma)(t)) \in D_C$, for every $t \in \mathbb{R}$.

The condition (i) means that $\Phi^{1C} \circ \gamma$ is holonomic—i.e., the canonical lift of the curve $\pi^1 \circ \Phi^{1C} \circ \gamma$—and (ii) is the condition (2.6.25); that is, (ii) are the Euler–Lagrange equations for the controlled Lagrangian system $(\mathcal{L}, \mathbf{F})$.

Now, consider the map $(\mathrm{Id}, \Upsilon) \colon C \times_{J^1\pi} J^2\pi \longrightarrow C \times_{J^1\pi} J^1\bar{\pi}^1$, where $\Upsilon \colon J^2\pi \longrightarrow J^1\bar{\pi}^1$ is defined in (2.6.23) and let $M_C = (\mathrm{Id}, \Upsilon)(D_C)$. As (Id, Υ) is an injective map, we can identify $D_C \subset C \times_{J^1\pi} J^2\pi$ with this subset M_C of $C \times_{J^1\pi} J^1\bar{\pi}^1$. Observe that there is a natural projection from M_C to $J^1\pi$.

If $\mathcal{F} \colon M_C \longrightarrow \mathbb{R}$ is a cost function, we may consider the implicit optimal control system determined by the pair $(\mathbf{L}, M_C)$, where $\mathbf{L} = \mathcal{F}\mathrm{d}t$, and apply the method developed in §8.2.

Let $\overline{\mathcal{W}}^{M_C} = M_C \times_{J^1\pi} T^*J^1\pi$ and $\overline{\mathcal{W}}^C = C \times_{J^1\pi} J^1\bar{\pi}^1 \times_{J^1\pi} T^*J^1\pi$. The natural projection from $\overline{\mathcal{W}}^C$ to $T^*J^1\pi$ allows us to pullback the canonical 2–form $\Omega_{J^1\pi}$ to a presymplectic form $\Omega_{\overline{\mathcal{W}}^C} \in \Omega^2\left(\overline{\mathcal{W}}^C\right)$. Furthermore, in $J^1\bar{\pi}^1 \times_{J^1\pi} T^*J^1\pi$ there is the natural coupling form $\bar{\bar{\mathcal{C}}}$, see Definition 2.5.1. We denote by $\bar{\mathcal{C}}$ its pullback to $\overline{\mathcal{W}}^C$. We denote by $\mathbf{L}$ and $\mathcal{F}$ the pullback of $\mathbf{L}$ and $\mathcal{F}$ from M_C to $\overline{\mathcal{W}}^C$, for the sake of simplicity.

Then, let $\bar{H}_{\mathcal{W}^C} \colon \overline{\mathcal{W}}^C \longrightarrow \mathbb{R}$ be the unique function such that $\bar{\mathcal{C}} - \mathbf{L} = \bar{H}_{\mathcal{W}^C}\mathrm{d}t$, whose local expression is $\bar{H}_{\mathcal{W}^C} = p + p_i\bar{v}^i + \bar{p}_iw^i - \mathcal{F}$, and consider the submanifold

$$\overline{\mathcal{W}}_0 = \{\tilde{q} \in \overline{\mathcal{W}}^C \mid \bar{H}_{\mathcal{W}^C}(\tilde{q}) = 0\}.$$

The pullback of $\bar{H}_{\mathcal{W}^C}$ to $\overline{\mathcal{W}}^{M_C}$ is Pontryagin's Hamiltonian, denoted by $\bar{H}_{\mathcal{W}^{M_C}}$.

Finally, the dynamics evolve in the submanifold $\overline{\mathcal{W}}_0^{M_C} = \overline{\mathcal{W}}^{M_C} \cap \overline{\mathcal{W}}_0$ of $\overline{\mathcal{W}}^C$, where $\jmath_1^{M_C}$ is the natural embedding. $\overline{\mathcal{W}}_0^{M_C}$ is endowed with the presymplectic form $\Omega_{\overline{\mathcal{W}}_0^{M_C}} = (\jmath_1^{M_C})^*\Omega_{\overline{\mathcal{W}}^C}$. Therefore, the motion is determined by a vector field $Z \in \mathfrak{X}(\overline{\mathcal{W}}_0^{M_C})$ satisfying the equations

$$i_Z\Omega_{\overline{\mathcal{W}}_0^{M_C}} = 0, \quad i_Z\mathrm{d}t = 1.$$

A local chart in $\overline{\mathcal{W}}^C$ is $(t, q^i, v^i, \bar{v}^i, w^i, u_l, p, p_i, \bar{p}_i)$, where $(\bar{v}^i, w^i)$ and $(p, p_i, \bar{p}_i)$ are the natural fiber coordinates in $J^1\bar{\pi}^1$ and $T^*J^1\pi$, respectively. The manifold $\overline{\mathcal{W}}^{M_C}$ is given locally by the $2m$ constraints:

$$
\begin{aligned}
\varphi_i(t, x^i, v^i, \bar{v}^i, w^i, u^l, p, p_i, \bar{p}_i) &= w^j\frac{\partial^2 L}{\partial v^i\partial v^j}(t, x, v) + \bar{v}^j\frac{\partial^2 L}{\partial v^i\partial x^j}(t, x, v) + \frac{\partial^2 L}{\partial v^i\partial t}(t, x, v)\\
&\quad - \frac{\partial L}{\partial x^i}(t, x, v) - \mathbf{F}_i(t, x, v, u) = 0,\\
\bar{\varphi}^i(t, x^i, v^i, \bar{v}^i, w^i, u^l, p, p_i, \bar{p}_i) &= v^i - \bar{v}^i = 0,
\end{aligned}
$$

$\overline{\mathcal{W}}_0$ is given by

$$\phi(t, x^i, v^i, \bar{v}^i, w^i, u^l, p, p_i, \bar{p}_i) = \bar{H}_{\mathcal{W}^C}(t, x^i, v^i, \bar{v}^i, w^i, u^l, p, p_i, \bar{p}_i)$$

$$= p + p_i \bar{v}^i + \bar{p}_i w^i - \mathcal{F}(t, x, v, u) = 0,$$

and

$$\Omega_{\overline{\mathcal{W}}_0^{M_C}} = \mathrm{d}x^i \wedge \mathrm{d}p_i + \mathrm{d}v^i \wedge \mathrm{d}\bar{p}_i + \mathrm{d}t \wedge \mathrm{d}(\mathcal{F} - p_i \bar{v}^i - \bar{p}_i w^i).$$

Following Proposition 8.2.2, we look for a vector field $Z \in \mathfrak{X}\left(\overline{\mathcal{W}}^C\right)$ such that, for every $\mathbf{w} \in \overline{\mathcal{W}}_0^{M_C}$,

(i) $Z_{\mathbf{w}} \in \mathrm{T}_{\mathbf{w}}\overline{\mathcal{W}}_0^{M_C}$,

(ii) $i_{Z_{\mathbf{w}}}\Omega_{\overline{\mathcal{W}}^C} \in \left(\mathrm{T}_{\mathbf{w}}\overline{\mathcal{W}}_0^{M_C}\right)^0$,

or, equivalently

(i) $(\jmath_1^{M_C})^*(Z(\varphi_i)) = 0, \quad (\jmath_1^{M_C})^*(Z(\bar{\varphi}^i)) = 0, \quad (\jmath_1^{M_C})^*(Z(\phi)) = 0.$

(ii) $(\jmath_1^{M_C})^* \left(i_Z \Omega_{\overline{\mathcal{W}}^C}\right) = 0.$

Remember that the constraints are $\varphi_i = 0, \quad \bar{\varphi}^i = 0, \quad \phi = 0$.

If Z is given locally by

$$Z = \frac{\partial}{\partial t} + A^i \frac{\partial}{\partial x^i} + \mathcal{A}^i \frac{\partial}{\partial v^i} + \bar{A}^i \frac{\partial}{\partial \bar{v}^i} + \bar{\mathcal{A}}^i \frac{\partial}{\partial w^i} + B^l \frac{\partial}{\partial u^l} + D \frac{\partial}{\partial p} + C_i \frac{\partial}{\partial p_i} + \bar{C}_i \frac{\partial}{\partial \bar{p}_i},$$

then $A^i, \mathcal{A}^i, \bar{A}^i, \bar{\mathcal{A}}^i, B^l, D, C_i, \bar{C}_i$ are unknown functions in $\overline{\mathcal{W}}^C$ such that

$$i_Z \Omega_{\overline{\mathcal{W}}^C} = \mu^i \mathrm{d}\varphi_i + \bar{\mu}_i \mathrm{d}\bar{\varphi}^i + \mu \mathrm{d}(p + p_i \bar{v}^i + \bar{p}_i w^i - \mathcal{F}(t, x, v, u)),$$

$Z(\varphi_i) = 0, Z(\bar{\varphi}^i) = 0$, and $Z(p + p_i \bar{v}^i + \bar{p}_i w^i - \mathcal{F}(t, x, v, u)) = 0$. From these equations we obtain

$$
\begin{array}{llll}
\mu &= 1, & A^i &= \bar{v}^i, & \mathcal{A}^i &= w^i, \\
C_i &= \frac{\partial \mathcal{F}}{\partial x^i} - \mu^j \frac{\partial \varphi_j}{\partial x^i}, & \bar{C}_i &= \frac{\partial \mathcal{F}}{\partial v^i} - \mu^j \frac{\partial \varphi_j}{\partial v^i} - \bar{\mu}_i, & D &= \frac{\partial \mathcal{F}}{\partial t} - \mu^j \frac{\partial \varphi_j}{\partial t}, \\
0 &= \frac{\partial \mathcal{F}}{\partial u^l} + \mu^i \frac{\partial \mathbf{F}_i}{\partial u^l}, & p_i &= \bar{\mu}_i - \mu^j \frac{\partial^2 L}{\partial v^j \partial q^i}, & \bar{p}_i &= -\mu^j \frac{\partial^2 L}{\partial v^i \partial v^j},
\end{array}
$$

and the tangency conditions

$$
\begin{array}{rl}
Z(\varphi_i) &= \frac{\partial \varphi_i}{\partial t} + \bar{v}^j \frac{\partial \varphi_i}{\partial x^j} + w^j \frac{\partial \varphi_i}{\partial v^j} + \bar{A}^j \frac{\partial^2 L}{\partial v^i \partial x^j} - B^l \frac{\partial \mathbf{F}_i}{\partial u^l} + \bar{\mathcal{A}}^j \frac{\partial^2 L}{\partial v^i \partial v^j} = 0, \\
Z(\bar{\varphi}^i) &= w^i - \bar{A}^i = 0, \\
Z(\phi) &= Z(p + p_i \bar{v}^i + \bar{p}_i w^i - \mathcal{F}(t, x, v, u)) = 0,
\end{array}
$$

where the third condition is satisfied identically using the previous equations.

Assuming that the Lagrangian L is regular, that is, $\det(W_{ij}) = \det\left(\frac{\partial^2 L}{\partial v^i \partial v^j}\right) \neq 0$, then from the above equations for p_i and $\bar{p}_i$ we obtain explicit values of the Lagrange multipliers μ^i and $\bar{\mu}_i$. Therefore, the remaining equations above are now rewritten as the new set of constraints

$$\psi_l(t, x, v, u, \bar{p}) = \frac{\partial \mathcal{F}}{\partial u^l} - W^{ij} \bar{p}_i \frac{\partial \mathbf{F}_j}{\partial u^l} = 0 \,, \tag{8.3.21}$$

which corresponds to $\dfrac{\partial \bar{H}_{\mathcal{W}^{MC}}}{\partial u^l} = 0$.

The new compatibility condition is

$$Z(\psi_l) = \frac{\partial \psi_l}{\partial t} + \bar{v}^j \frac{\partial \psi_l}{\partial x^j} + w^j \frac{\partial \psi_l}{\partial v^j} + B^s \frac{\partial \psi_l}{\partial u^s} + \bar{C}_i \frac{\partial \psi_l}{\partial \bar{p}_i} = 0 \,. \tag{8.3.22}$$

Furthermore, we assume that

$$\det\left(\frac{\partial \psi_l}{\partial u^s}\right) \neq 0.$$

Then, from the tangency condition $Z(\varphi_i) = 0$ and Equation (8.3.22), we obtain the remaining components $\bar{A}^i$ and B^l, and we determine completely the vector field Z.

The equations of motion for a curve are determined by the system of implicit–differential equations:

$$\dot{p}_i(t) = \frac{\partial \mathcal{F}}{\partial x^i}(t, x(t), \dot{x}(t), u(t)) - \mu^j(t, x(t), \dot{x}(t), \bar{p}(t)) \frac{\partial \varphi_j}{\partial x^i}(t, x(t), \dot{x}(t), \ddot{x}(t), u(t)),$$

$$\dot{\bar{p}}_i(t) = \frac{\partial \mathcal{F}}{\partial v^i}(t, x(t), \dot{x}(t), u(t)) - p_i(t)$$
$$- \mu^j(t, x(t), \dot{x}(t), \bar{p}(t)) \left(\frac{\partial \varphi_j}{\partial v^i}(t, x(t), \dot{x}(t), \ddot{x}(t), u(t)) + \frac{\partial^2 L}{\partial v^j \partial x^i}(t, x(t), \dot{x}(t)) \right),$$

$$0 = \frac{d}{dt}\left(\frac{\partial L}{\partial v^i}(t, x(t), \dot{x}(t)) \right) - \frac{\partial L}{\partial x^i}(t, x(t), \dot{x}(t)) - \mathbf{F}_i(t, x(t), \dot{x}(t), u(t)),$$

$$0 = \frac{\partial \mathcal{F}}{\partial u^a}(t, x(t), \dot{x}(t), u(t)) - W^{ij}(t, x(t), \dot{x}(t)) \bar{p}_i(t) \frac{\partial \mathbf{F}_j}{\partial u^a}(t, x(t), \dot{x}(t), u(t)).$$

The last equation is the explicit expression of (8.3.21).

Agrawal and Fabien [1999] study optimal control of Lagrangian systems with controls in a more restrictive situation using higher–order dynamics, obtaining that the states are determined by a set of fourth–order differential equations. First it is necessary to assume that the system is *fully actuated*, that is $k = m$, and $\operatorname{rank}(\Xi_{ij}) = \operatorname{rank}\left(\frac{\partial \mathbf{F}_i}{\partial u^j}\right) = m$. Moreover, in the sequel we assume that the system is affine in controls. That is, we assume that

$$\mathbf{F}_i(t, x, \dot{x}, u) = A_i(t, x, \dot{x}) + A_{ij}(t, x, \dot{x}) u^j \,.$$

Therefore, $\Xi_{ij} = A_{ij}$. Then, from the constraint equations in the previous system of im-

plicit–differential equations, applying the implicit function theorem, we deduce that

$$u^i(t) = u^i(t, x(t), \dot{x}(t), \ddot{x}(t))$$

$$= A^{ij}\left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial v^j}(t, x(t), \dot{x}(t))\right) - \frac{\partial L}{\partial x^j}(t, x(t), \dot{x}(t)) - A_j(t, x(t), \dot{x}(t))\right],$$

$$\bar{p}_i(t) = \mathcal{H}_i^j(t, x(t), \dot{x}(t))\frac{\partial \mathcal{F}}{\partial u^j}(t, x(t), \dot{x}(t), u(t, x(t), \dot{x}(t), \ddot{x}(t))),$$

where $(\mathcal{H}_i^j)$ are the components of the inverse matrix of the regular matrix $(W^{is}A_{sj})$.

Taking the derivative with respect to time of the equation for $\dot{\bar{p}}_i(t)$, and substituting the value of $\dot{p}_i(t)$ we obtain a fourth–order differential equation depending on the states. After some computations we deduce that

$$\mathcal{H}_i^j(t, x(t), \dot{x}(t))\frac{\partial^2 \mathcal{F}}{\partial u^j \partial u^s}(t, x(t), \dot{x}(t), \ddot{x}(t))\frac{\mathrm{d}^4 x^s}{\mathrm{d}t^4}(t) = G_i(t, x(t), \dot{x}(t), \ddot{x}(t), \dddot{x}(t)).$$

Finally, under the assumption that the matrix $\left(\frac{\partial^2 \mathcal{F}}{\partial u^j \partial u^s}\right)$ is invertible, we obtain a explicit fourth–order system of differential equations:

$$\frac{\mathrm{d}^4 x^i}{\mathrm{d}t^4}(t) = \bar{G}^i(t, x(t), \dot{x}(t), \ddot{x}(t), \dddot{x}(t)),$$

as appears in [Agrawal and Fabien 1999].

8.3.2   Optimal control problems for descriptor systems

Control systems given in a descriptor form have many applications in engineering, as for instance in electrical, mechanical or chemical problems. In [Müller 1998] there are illustrative examples of these controls systems and, in particular, the academic example we study here. Consider the problem of minimizing the functional

$$S = \frac{1}{2}\int_0^{+\infty}\left[a_i(x^i)^2 + ru^2\right]\,\mathrm{d}t,$$

$1 \le i \le 3$, with control equations

$$\dot{x}^2 = x^1 + b_1 u \quad , \quad \dot{x}^3 = x^2 + b_2 u \quad , \quad 0 = x^3 + b_3 u$$

with parameters $a_i, b_i \ge 0$ and $r > 0$.

As in the previous section, the geometric framework developed in §8.1.1 is also valid for this class of systems. Let $E = \mathbb{R} \times \mathbb{R}^3$ with coordinates (t, x^i), and $C = \mathbb{R} \times \mathbb{R}^3 \times \mathbb{R}$ with coordinates (t, x^i, u). The submanifold $M_C \subset C \times_E J^1\pi$ is given by

$$M_C = \{(t, x^1, x^2, x^3, v^1, v^2, v^3, u) \mid v^2 = x^1 + b_1 u \,, v^3 = x^2 + b_2 u \,, 0 = x^3 + b_3 u\}\,.$$

The cost function is

$$\mathcal{F}: \quad\quad\quad C \;\longrightarrow\; \mathbb{R}$$
$$(t, x^1, x^2, x^3, u) \;\longmapsto\; \frac{1}{2}\left[a_1(x^1)^2 + a_2(x^2)^2 + a_3(x^3)^2 + ru^2 \right]$$

We analyze the dynamics of the implicit optimal control system determined by the pair $(\mathbf{L}, M_C)$.

Let $\mathcal{W}^{M_C} = M_C \times_E T^*E$ and $\mathcal{W}^C = C \times_E J^1\pi \times_E T^*E$ with coupling form $\mathcal{C}$ inherited from the natural coupling form in $J^1\pi \times T^*E$. Let $H_{\mathcal{W}^C} \colon \mathcal{W}^C \longrightarrow \mathbb{R}$ be the unique function such that $\mathcal{C} - \mathbf{L} = H_{\mathcal{W}^C}\,dt$, and consider the submanifold $\mathcal{W}_0 = \{\tilde{q} \in \mathcal{W}^C \,|\, H_{\mathcal{W}^C}(\tilde{q}) = 0\}$. Finally, the dynamics evolves in the submanifold $\mathcal{W}_0^{M_C} = \mathcal{W}^{M_C} \cap \mathcal{W}_0$ of $\mathcal{W}^C$. Locally,

$$\mathcal{W}_0^{M_C} \;=\; \{(t, x^1, x^2, x^3, v^1, v^2, v^3, u, p, p_1, p_2, p_3) \mid v^2 = x^1 + b_1 u \,,\, v^3 = x^2 + b_2 u \,,$$
$$x^3 + b_3 u = 0 \,,\, p + p_1 v^1 + p_2 v^2 + p_3 v^3 - \mathcal{F} = 0\} \,.$$

Therefore, the motion is determined by a vector field $Z \in \mathfrak{X}(\mathcal{W}_0^{M_C})$ satisfying the Equations (8.2.18) and (8.2.19), which according to Proposition 8.2.2 is equivalent to finding a vector field $Z \in \mathfrak{X}(\mathcal{W}^C)$ (if it exists):

$$Z \;=\; \frac{\partial}{\partial t} + A^1 \frac{\partial}{\partial x^1} + A^2 \frac{\partial}{\partial x^2} + A^3 \frac{\partial}{\partial x^3} + C^1 \frac{\partial}{\partial v^1} + C^2 \frac{\partial}{\partial v^2} + C^3 \frac{\partial}{\partial v^3} + B \frac{\partial}{\partial u}$$
$$+\; D_1 \frac{\partial}{\partial p_1} + D_2 \frac{\partial}{\partial p_2} + D_3 \frac{\partial}{\partial p_3} + E \frac{\partial}{\partial p}$$

such that

$$i_Z \Omega_{\mathcal{W}^C} = \mu_1 d(x^1 + b_1 u - v^2) + \mu_2 d(x^2 + b_2 u - v^3) + \mu_3 d(x^3 + b_3 u) + \mu\, dH_{\mathcal{W}^C} \,,$$
$$Z(x^1 + b_1 u - v^2) = 0 \,, \quad Z(x^2 + b_2 u - v^3) = 0 \,, \quad Z(x^3 + b_3 u) = 0 \,, \quad Z(H_{\mathcal{W}^C}) = 0,$$

where $\Omega_{\mathcal{W}^C} \in \Omega^2(\mathcal{W}^C)$ is the 2–form with local expression

$$\Omega_{\mathcal{W}^C} = dx^1 \wedge dp_1 + dx^2 \wedge dp_2 + dx^3 \wedge dp_3 + dt \wedge dp \,.$$

After some straightforward computations, we obtain that

$$
\begin{array}{llllll}
A^1 &=& v^1, & A^2 &=& x^1 + b_1 u, \qquad A^3 \;=\; x^2 + b_2 u, \\
\mu &=& 1, & E &=& 0, \qquad\qquad\quad\; 0 \;=\; ru - b_1 p_2 - b_2 p_3 - b_3 \mu_3, \\
C^2 &=& v^1 + b_1 B, & C^3 &=& A^2 + b_2 B, \qquad\; 0 \;=\; A^3 + b_3 B, \\
p_1 &=& 0, & p_2 &=& \mu_1, \qquad\qquad\; p_3 \;=\; \mu_2, \\
D_1 &=& a_1 x^1 - p_2, & D_2 &=& a_2 x^2 - p_3, \qquad D_3 \;=\; a_3 x^3 - \mu_3.
\end{array}
$$

We deduce that

$$\mu_3 = \frac{1}{b_3}(ru - b_1 p_2 - b_2 p_3) \,, \quad B = -\frac{1}{b_3}(x^2 + b_2 u) \,.$$

Therefore, the new constraint submanifold $\mathcal{W}_1^{M_C} \hookrightarrow \mathcal{W}_0^{M_C}$ is

$$\mathcal{W}_1^{M_C} = \{(t, x^1, x^2, v^1, u, p_1, p_2, p_3) \mid p_1 = 0\}\,.$$

Consistency of the dynamics implies that

$$0 = Z(p_1) = D_1 = a_1 x^1 - p_2\,.$$

Thus,

$$\mathcal{W}_2^{M_C} = \{(t, x^1, x^2, v^1, u, p_2, p_3) \mid a_1 x^1 - p_2 = 0\}$$

and once again we impose the tangency to the new constraints:

$$0 = Z(a_1 x^1 - p_2) = a_1 v^1 - a_2 x^2 + p_3,$$

which implies that

$$\mathcal{W}_3^{M_C} = \{(t, x^1, x^2, v^1, u, p_3) \mid a_1 v^1 - a_2 x^2 + p_3 = 0\}\,.$$

From the compatibility condition

$$0 = Z(a_1 v^1 - a_2 x^2 + p_3)$$

and the constraints we determine the remaining component C^1 of Z:

$$C^1 = \frac{1}{a_1 b_3} \left((a_2 b_3 - a_1 b_1) x^1 - b_2 a_2 x^2 + (a_2 b_1 b_3 + a_3 b_3^2 + r) u + b_2 a_1 v^1 \right)\,.$$

Therefore the equations of motion of the optimal control problem are:

$$\ddot{x}^1(t) = \frac{1}{a_1 b_3} \left((a_2 b_3 - a_1 b_1) x^1(t) - a_2 b_2 x^2(t) + (a_2 b_1 b_3 + a_3 b_3^2 + r) u(t) + a_1 b_2 \dot{x}^1(t) \right),$$
$$\dot{x}^2(t) = x^1(t) + b_1 u(t),$$
$$0 = x^2(t) + b_2 u(t) - b_3 \dot{u}(t).$$

From the first equation we deduce that

$$u(t) = \frac{1}{a_2 b_1 b_3 + a_3 b_3^2 + r} \left((a_1 b_1 - a_2 b_3) x^1(t) + a_2 b_2 x^2(t) - a_1 b_2 \dot{x}^1(t) + a_1 b_3 \ddot{x}^1(t) \right)\,.$$

This is the result obtained by Müller [1998], where the optimal feedback control depends on the state variables and also on their derivatives (non–casuality).

Choosing local coordinates (t, x^1, x^2, v^1, u) on $\mathcal{W}_3^{M_C}$, if $\jmath_3 : \mathcal{W}_3^{M_C} \mapsto \mathcal{W}^C$ is the canonical embedding, then $\Omega_{\mathcal{W}_3^{M_C}} = \jmath_3^* \Omega_{\mathcal{W}^C}$ is locally written as

$$\Omega_{\mathcal{W}_3^{M_C}} = -a_1 \mathrm{d} x^1 \wedge \mathrm{d} x^2 + a_2 b_3 \mathrm{d} x^2 \wedge \mathrm{d} u - a_1 b_3 \mathrm{d} v^1 \wedge \mathrm{d} u + \mathrm{d} t \wedge \mathrm{d} \jmath_3^* p,$$

where $\jmath_3^* p : \mathcal{W}_3^{M_C} \longrightarrow \mathbb{R}$ is the function

$$\jmath_3^* p = -\frac{1}{2}a_1(x^1)^2 - \frac{1}{2}a_2(x^2)^2 + \frac{1}{2}(r + a_3 b_3^2)u^2 - a_1 b_1 x^1 u - a_2 b_2 x^2 u + a_1 b_2 v^1 u + a_1 x^2 v^1 \, .$$

Obviously, $(\Omega_{\mathcal{W}_3^{M_C}}, \mathrm{d}t)$ is a cosymplectic structure on $\mathcal{W}_3^{M_C}$—see Proposition 8.1.8—and there exists a unique vector field $\bar{Z} \in \mathfrak{X}(\mathcal{W}_3^{M_C})$ satisfying

$$i_{\bar{Z}}\Omega_{\mathcal{W}_3^{M_C}} = 0, \quad i_{\bar{Z}}\mathrm{d}t = 1 \, .$$

Chapter 9
Conclusion and future work

The closing chapter of this dissertation is a summary of the contributions of the work. A guideline for the future line research is also described.

9.1   Summary of the contributions

After clarifying some aspects in differential geometry related with the time–dependent variational equations in §2.2.2 in an intrinsic way, we move to optimal control theory.

The importance of Pontryagin's Maximum Principle in optimal control theory justifies that the entirety of Chapter 4 is devoted to giving a geometric proof and to the understanding of the Maximum Principle. A key point in that proof is the selection of the initial condition for the momenta to integrate Hamilton's equations such that the separation conditions pictured in §4.6.2 and in Figure 4.4 are fulfilled. Another key point is the linear approximation of the trajectories so that the reachable set and the tangent perturbation cone are identified locally, as explained in §4.5. All this new approach to the proof of Pontryagin's Maximum Principle including all the necessary details from the different areas of mathematics appears in [Barbero-Liñán and Muñoz Lecanda 2008b], paper accepted in *Acta Applicandae Mathematicae* in September 2008.

After a first symplectic approach to optimal control problems, we consider the presymplectic framework. Then, under the assumption of having an open control set and differentiability with respect to the controls, a new adaptation of the presymplectic constraint algorithm is described to study the different kinds of extremals for an optimal control problem in §5.2 and §5.3. The presymplectic formalism generates two different Hamiltonians, thus we have two different presymplectic equations that must be solved. How to take advantage of the final constraint submanifolds to characterize the extremals is summarized in Proposition 5.2.9. In §5.5, we are able to find a strict abnormal extremal for an academic optimal mechanical control problem by means of the geometric method given in Chapter 5. The contents in Chapter 5 correspond with the preprint [Barbero-Liñán and Muñoz Lecanda 2008a], which has been submitted for publication.

Then, in Chapter 6, we focus on mechanical systems modeled by affine connection control systems and optimal control problems for them. The geometric method described in Chapter 5 is applied for these particular mechanical control systems. We mainly study specific optimal control problems, such as problems with a control–quadratic cost function and the time–optimal control problem, obtaining new results about the extremals in Propositions 6.7.1, 6.7.2 and 6.7.4. Finally, in §6.8, we establish connections between the constraints in the algorithm and vector–valued quadratic forms for particular examples. From the careful study of

these examples we guess a feasible way to characterize abnormal extremals in terms of vector–valued quadratic forms, as stated in Conjecture 6.8.5. The description of the adaptation of the presymplectic constraint algorithm for the study of the extremals in the mechanical case and the application to time–optimal control problem were explained in a talk in the XVI International Workshop on Geometry and Physics held in Lisbon (Portugal) from 5 to 8 September 2007. This talk is published in the Proceedings of the conference [Barbero-Liñán and Muñoz Lecanda 2008d].

In our search for strict abnormal minimizers motivated by R. Montgomery, W. Liu and H. J. Sussmann, we decide to attack the problem for the mechanical case using the existent equivalence between nonholonomic mechanical control systems and kinematic control systems. In Chapter 7, we establish connections between the optimal control problems associated with both control systems, and in particular, between the extremals for both problems; Propositions 7.1.6, 7.1.10 and §7.2.3. As a result, a local strict abnormal minimizer is found for a nonholonomic control system in §7.2.4. This chapter was an invited talk in the Special Session Nonholonomic constraints in Mechanics and Optimal Control Theory organized by M. de León, J.C. Marrero, D. Martín de Diego in the 7th AIMS International Conference on Dynamical Systems, Differential Equations and Applications in the University of Texas (Arlington , USA) from 18 to 21 May 2008. These results have been submitted to a Meeting–Issue number in a journal, independent of the Proceedings of the conference.

Finally, in Chapter 8, we present a different formalism for the non–autonomous optimal control problems based on the Skinner–Rusk formalism that gives rise to a geometric formulation of the weak Pontryagin's Maximum Principle in Proposition 8.1.5 and in Theorem 8.1.6. That approach is also valid for the implicit control systems as showed in §8.2. The theory developed in Chapter 8 is applied to the implicit optimal control problems given by the controlled Lagrangian systems in §8.3.1 and by the descriptor systems in §8.3.2, that is a common system in engineering. This formalism, its results and applications are published in *Journal of Physics A: Mathematical and Theoretical* [Barbero-Liñán et al. 2007].

9.2   Future work

Once we have laid the foundations of the geometric optimal control theory to understand the abnormality of optimal trajectories, there remains much to be done.

9.2.1   Search for strict abnormal minimizers

Our study of optimal control theory has provided some new insights into abnormal extremals. Some particular examples have been obtained for strict abnormal extremals and local strict abnormal minimizers in the mechanical case in §5.5 and §7.2.4, respectively.

The study of strict abnormal minimizers forces us to consider different cost functions, because only the property of being an abnormal extremal depends exclusively on the geometry of the control system. That makes the problem much harder and the possible forthcoming results

will be valid only for specific optimal control problems.

Moreover, it will be interesting to find strict abnormal minimizers in real mechanical systems. So far we know the existence of abnormal extremals for the planar rigid body, but we have not found any cost function that guarantees this abnormal extremal to be strict. A method of finding strict abnormal minimizers consists of first finding an abnormal extremal for a given control system. Then, to find a cost function that makes this extremal to be strict abnormal.

Liu and Sussmann [1995] proved the existence of strict abnormal minimizers, but also their density in some sense. It will be useful for understanding the importance of strict abnormal minimizers to find similar results in particular optimal mechanical control problems. Although there are some results about the non–existence of strict abnormal extremals in a generic sense for control–affine systems satisfying particular conditions, [Chitour et al. 2006; 2008], these conditions are not fulfilled by the affine connection control systems. Thus, there is still hope.

9.2.2   Lie algebroids and abnormality

The presymplectic constraint algorithm described in Chapter 5 has been applied to characterize the extremals for any optimal control problem. However, the control systems considered can be included in a bigger class of control systems defined on Lie algebroids [Cortés et al. 2006, León et al. 2005, Iglesias et al. 2007]. Lie algebroids also make it possible to study the mechanical case considered in Chapter 6.

In some sense the Lie algebroids provide a splitting associated with a vector bundle different from the one considered in §6.4.1.1. It is expected that the constraints obtained from the presymplectic constraint algorithm will give different information about the abnormal extremals and might make easier the study of some examples, as for the instance the ones in [Sachkov 2006].

Pontryagin's Maximum Principle has already been considered in the framework of Lie algebroids in [Cortés and Martínez 2004, Martínez 2004]. A first approach about how to apply the presymplectic constraint algorithm on Lie algebroids, in particular for abnormal extremals, is given in the poster titled "Lie algebroids and optimal control: abnormality" presented in XVII International Workshop on Geometry and Physics held in Castro Urdiales (Cantabria, Spain) from 3 to 6 September 2008. This work has been submitted by publication to the proceedings of the conference. This line research was started as a result of a short stay in Instituto de Ciencias Matemáticas (CSIC–UAM–UC3M–UCM) in Madrid (Spain) under the supervision of Professor David Martín de Diego from 30 March 2008 to 12 April 2008.

9.2.3   Geometric high–order Maximum Principle

Pontryagin's Maximum Principle provides first–order necessary conditions for optimality. As shown in this work, these conditions are not enough to determine the controls for abnormal and singular extremals. That is why in Chapters 5 and 6, when we focus on abnormality, the stabilization steps must continue. Thus, an analogy must exist between the constraint algorithm

and the high–order Maximum Principle [Bianchini 1998, Kawski 2003, Knobloch 1981, Krener 1977].

Pontryagin's Maximum Principle works with linear approximation of the trajectories, whereas in the high–order Maximum Principle high–order perturbations must be considered [Bianchini 1998, Bianchini and Stefani 1993, Gabasov and Kirillova 1972, Kawski 2003, Knobloch 1981, Krener 1977]. The way to construct the proof is the same as in Pontryagin's Maximum Principle, but now the tangent perturbation cones are bigger since not only linear approximation of the trajectories are considered.

In the same way, we have provided a geometric meaning to most of the elements in Pontryagin's Maximum Principle, we expect to give a geometric version of high–order Maximum Principle suggested by Krener [1977], focusing on abnormality. First studies have been done in that and presented in a talk titled "Abnormality for affine connection control systems" in the Special Session Control Theory and Mechanics organized by Professor Andrew D. Lewis in the 18th Mathematical Theory of Networks and Systems (MTNS) held in Virginia Tech (Blacksburg, Virginia, USA) from 28 July to 1 August 2008.

9.2.4   The algebra of accessibility without zero velocity

In the literature on control theory, there exist necessary and sufficient conditions for accessibility, and only sufficient conditions for controllability for the affine connection control systems when it is assumed to start with zero velocity [Cortés and Martínez 2003, Lewis and Murray 1997, Nijmeijer and van der Schaft 1990, Ostrowski and Burdick 1997, Sussmann 1987, Žefran et al. 1999]. Under this assumption, the previous conditions can be written in such a way that only constructions on the manifold Q are involved. But if the initial velocity is nonzero, the conditions will be written in terms of constructions on the tangent bundle TQ.

During the stay at Queen's University (Kingston, Ontario, Canada) from 21 October to 21 December 2007, the algebra of accessibility for mechanical systems at nonzero initial velocity was studied. It was expected to obtain useful results to describe the constraints that appear as a result of applying the adapted presymplectic contraint algorithm to optimal control problems for ACCS in §6.5 and §6.6. Although this effort has not been fruitful in that sense, we have been able to give a recursive way to obtain the algebra of accessibility for ACCS at nonzero initial velocity. Thus we have tackled a more general problem than the one in [Lewis and Murray 1997] where the initial velocity is assumed to be zero.

The splitting explained in §6.4.1.1 is useful to describe the accessibility distribution on TQ for ACCS, as already used in [Lewis and Murray 1997]. Unless the velocity is assumed to be zero, the computations for the accessibility distribution become slightly involved because it is necessary the use of geometric elements defined along the tangent projection bundle τ_Q, such as connections and their curvature [Martínez et al. 1992; 1993, Szilasi 2003]. All these constructions make it possible to obtain the horizontal and vertical submodules of the algebra of accessibility in the splitting given by a linear connection on τ_Q associated with an affine connection on Q.

Proposition 9.2.1. *Let m be a natural number and Σ be an ACCS. The horizontal submodule $H(D_m)$ of degree m of the smallest involutive distribution containing Z and $\mathscr{Y}^V = \{Y_1, \ldots, Y_k\}$ is spanned over $\mathbb{R}$ by*

$$\{\nabla^H(v, X) - W, \mathrm{d}^V_{Y_j} X \mid X \in H(D_{m-1}), W \in V(D_{m-1}), Y_j \in \mathscr{Y}\},$$

and the vertical submodule $V(D_m)$ of degree m is spanned over $\mathbb{R}$ by

$$\{R(X, v)v + \nabla^H(v, W), \mathrm{d}^V_{Y_j} W - \nabla^H(X, Y_j) \quad | \quad X \in H(D_{m-1}),$$
$$W \in V(D_{m-1}), Y_j \in \mathscr{Y}\},$$

where ∇^H denotes an affine connection along τ_Q, R is its curvature and d^V is a vertical derivative according to τ_Q.

From this result, research must continue in order to try to give the horizontal and vertical subbundles as distributions invariant by some geometric element. We are interested in finding geometric results analogous to Theorem 6.2.6 and the ones in [Lewis and Murray 1997], so that conditions about accessibility and controllability of mechanical control systems from nonzero initial velocity can be established.

Appendix A
Lebesgue times

$\mathbf{I}$n this Appendix, we focus on some necessary technicalities, mainly for Chapter 4. These are related with results from analysis and the notion of a Lebesgue point for a real function that is applied for this dissertation for vector fields as pointed out in Remark A.2.7. For more general details see [Royden 1963, Varberg 1965, Zaanen 1989].

A.1   Results on real functions

Definition A.1.1. *Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f\colon X \to Y$ is **Lipschitz** if there exists $K \in \mathbb{R}$ such that $d_Y(f(x_1), f(x_2)) \leq K\, d_X(x_1, x_2)$ for all $x_1, x_2 \in X$.*

*A function $f\colon X \to Y$ is **locally Lipschitz** if, for every $x \in X$ there exists an open neighbourhood V of x and $K \in \mathbb{R}^+$ such that $d_Y(f(x_1), f(x_2)) \leq K\, d_X(x_1, x_2)$ for all x_1 and x_2 in V.*

If M is a differentiable manifold, g is a Riemannian metric on M and $d_g\colon M \times M \to \mathbb{R}$ is the induced distance; then (M, d_g) is a metric space where the notion of Lipschitz on M can be defined. A real–valued function $F\colon M \to \mathbb{R}$ is locally Lipschitz if, for every $p \in M$ we take the local chart (V, ϕ) such that $\phi(p) = 0$, $\phi(V) = B(0, r)$ is the open ball centered at the origin with radius $r > 0$ in the standard Euclidean space, and $F \circ \phi^{-1}\colon B(0, r) \to \mathbb{R}$ is Lipschitz. That is, there exists $K \in \mathbb{R}^+$ with

$$|F(p_1) - F(p_2)| = |(F \circ \phi^{-1})(\phi(p_1)) - (F \circ \phi^{-1})(\phi(p_2))| \leq Kd(\phi(p_1), \phi(p_2)), \ \forall p_1, p_2 \in V.$$

Hence, given the local chart (V, ϕ), we define a distance $d_\phi\colon V \times V \to \mathbb{R}$ on V, $d_\phi(p_1, p_2) = d(\phi(p_1), \phi(p_2))$. Consequently, (V, ϕ) is a metric space with the topology induced by the open set V in M. This distance is equivalent to the distance induced by the Riemannian metric on V. Observe that the notion of locally Lipschitz for functions on manifolds depends on the local chart, but $\mathcal{C}^1$ functions are always locally Lipschitz. See [Ferreira 2006] for more details about the notion of Lipschitz functions on a manifold.

Now, we give two extensions to the notion of continuity.

Definition A.1.2. *A function $f\colon [a, b] \to \mathbb{R}$ is **uniformly continuous on** $[a, b]$ if, for every $\epsilon > 0$, there exists $\delta > 0$ such that for any $x, y \in [a, b]$ with $|x - y| < \delta$, we have $|f(x) - f(y)| < \epsilon$.*

Definition A.1.3. *A function $f\colon [a, b] \to \mathbb{R}$ is **absolutely continuous on** $[a, b]$ if, for every $\epsilon > 0$, there exists $\delta > 0$ such that for every finite number of nonoverlapping subintervals (a_i, b_i) of $[a, b]$ with $\sum_{i=1}^{n} |b_i - a_i| < \delta$, we have $\sum_{i=1}^{n} |f(b_i) - f(a_i)| < \epsilon$.*

We consider an interval $I = [a, b]$ in $\mathbb{R}$ with the usual Lebesgue measure. A statement is said to be satisfied *almost everywhere* if it is fulfilled in I except on a zero measure set. A measurable subset $A \subset I$ is said of full measure if $I - A$ has measure zero. Recall that if $A, B \subset I$ and $I - A, I - B$ have measure zero, then $A \cap B$ is not empty.

Results in [Royden 1963, pp. 96, 100, 105] allow one to prove the following result.

Proposition A.1.4. *If f is absolutely continuous, then f has a derivative almost everywhere.*

Theorem A.1.5. *[[Royden 1963, pp.105], [Varberg 1965, pp.836]] If f is absolutely continuous and $f'(x) = 0$ almost everywhere on $[a, b]$, then f is a constant function.*

Definition A.1.6. *A real–valued function f on a metric space (X, d) is called **lower semicontinuous at** $x_0 \in X$ if, for every $\epsilon > 0$, there exists $\delta(\epsilon, x_0) > 0$ such that $f(x) \geq f(x_0) - \epsilon$ whenever $d(x, x_0) \leq \delta(\epsilon, x_0)$.*

*If f is lower semicontinuous at every point of (X, d), it is said to be **lower semicontinuous on** (X, d).*

The following result is stated in [Pontryagin et al. 1962, pp. 102], but it is neither proved nor stated as a proposition. We believe it is appropriate to write it with more detail because it is used in §4.2.

Proposition A.1.7. *Let f and g be real functions, $f, g \colon [a, b] \to \mathbb{R}$. If f is continuous, g is lower semicontinuous, $f \leq g$ and $f = g$ almost everywhere, then $f = g$ everywhere.*

(*Proof*) Let $t_0 \in [a, b]$. As g is lower semicontinuous on $[a, b]$, for every $\epsilon > 0$ there exists $\delta(\epsilon, t_0) = \delta > 0$ such that

$$g(t) \geq g(t_0) - \epsilon$$

whenever $|t - t_0| < \delta(\epsilon, t_0)$.

Since f and g coincide almost everywhere on $[a, b]$, there exists $t_1 \in (t_0 - \delta, t_0 + \delta)$ such that $f(t_1) = g(t_1)$. Moreover, $f \leq g$, so

$$f(t_0) \leq g(t_0) \leq g(t_1) + \epsilon = f(t_1) + \epsilon. \tag{A.1.1}$$

The continuity of f guarantees that for every $\epsilon' > 0$, there exists $\delta' > 0$ such that if $|t_1 - t_0| < \delta'$, then $f(t_1) - \epsilon' < f(t_0) < f(t_1) + \epsilon'$. Hence Equation (A.1.1) is rewritten as follows:

$$f(t_0) \leq g(t_0) \leq f(t_0) + \epsilon' + \epsilon.$$

As this inequality is valid for every $\epsilon, \epsilon' > 0$, $g(t_0) = f(t_0)$ for every $t_0 \in [a, b]$. Thus $f = g$ everywhere. ∎

A.2 Lebesgue points for a real function

After introducing the concept of measurable function and some properties of such functions, we will state Lebesgue's differentiation theorem, which enables us to distinguish certain points

for a measurable function. In the entire work we consider the Lebesgue measure in $\mathbb{R}$. See [Zaanen 1989] for more details.

Definition A.2.1. *A function $f\colon [a,b] \to \mathbb{R}$ is **measurable** if the set $\{x \in [a,b] : f(x) > \alpha\}$ is measurable for every $\alpha \in \mathbb{R}$.*

Definition A.2.2. *A function $f\colon [a,b] \to \mathbb{R}$ is **Lebesgue integrable** over each Lebesgue measurable set of finite measure if $\nu(x) = \int_a^x f d\mu$ is well defined for every $x \in [a,b]$.*

Theorem A.2.3. *(**Lebesgue's Differentiation Theorem [Zaanen 1989]**) Let μ be the Lebesgue measure. If $f\colon [a,b] \to \mathbb{R}$ is a Lebesgue integrable function over every Lebesgue measurable set of finite measure, then for $\nu(x) = \int_a^x f d\mu$*

$$D\nu(x_+) = D\nu(x_-) = f(x)$$

holds for μ–almost every $x \in [a,b]$, where $D\nu(x_+)$, $D\nu(x_-)$ are the right and left derivatives of ν, respectively.

The equality $D\nu(x_-) = f(x)$ may be rewritten in the following four different equivalent ways, almost everywhere for $h > 0$

$$\lim_{h \to 0} \frac{\nu(x-h) - \nu(x)}{-h} = f(x) \quad \text{a.e.;} \quad \lim_{h \to 0} \frac{\int_a^{x-h} f(t)dt - \int_a^x f(t)dt}{-h} = f(x) \quad \text{a.e.;}$$

$$\lim_{h \to 0} \frac{\int_{x-h}^x f(t)dt}{h} = f(x) \quad \text{a.e.;} \quad \int_{x-h}^x f(t)dt = hf(x) + o(h) \quad \text{a.e..} \tag{A.2.2}$$

Definition A.2.4. *If $f\colon [a,b] \to \mathbb{R}$ is a integrable function, $x \in (a,b)$ is a **Lebesgue point for f** if,*

$$\lim_{h \to 0} \int_{x-h}^x \frac{f(t) - f(x)}{h} \, dt = 0.$$

Remark A.2.5. As Theorem A.2.3 is true almost everywhere, the set of Lebesgue points for a measurable function has full measure.

Remark A.2.6. Observe that if $u\colon I \to U$ is measurable and bounded, then it is integrable and the set of Lebesgue points for u has full measure. If $f\colon U \to \mathbb{R}$ is continuous, then $f \circ u\colon I \to \mathbb{R}$ is integrable, and the intersection of Lebesgue points for u and $f \circ u$ has full measure.

Remark A.2.7. Assume we have a manifold M, a set $U \subset \mathbb{R}^k$ and a continuous vector field X along the projection $\pi\colon M \times U \to M$. If $(\gamma, u)\colon I \subset \mathbb{R} \to M \times U$, where γ is absolutely continuous and u is integrable and bounded, then $X \circ (\gamma, u)\colon I \to TM$ is an integrable vector field along (γ, u), in the sense that in any coordinate system its coordinate functions are integrable. A point $t \in I$ is a *Lebesgue point for u* if

$$\int_{t-h}^t X(\gamma(s), u(s))ds = hX(\gamma(t), u(t)) + o(h). \tag{A.2.3}$$

The Lebesgue points for a vector field will be useful in §4.1.3 and §4.3.2 to guarantee the differentiability of some curves; that is, the existence of its tangent vector. See [Cañizo–Rincón

2004, Coddington and Levinson 1955, Filippov 1988] for more details about differential equations and measurability.

A.3   One corollary of Brouwer Fixed–Point Theorem

From the statement of Brouwer Fixed–Point Theorem, it is possible to prove a result in [Lee and Markus 1967, Scholium, pp. 251] useful for the proof of Proposition 4.1.12.

Theorem A.3.1. *(**Brouwer Fixed–Point Theorem**) Let B_1^m be the closed unit ball in $\mathbb{R}^m$. Any continuous function $G\colon B_1^m \to B_1^m$ has a fixed point.*

Corollary A.3.2. *Let $g\colon B_1^m \to \mathbb{R}^m$ be a continuous map. Let P be an interior point of B_1^m. If $\|g(x) - x\| < \|x - P\|$ for every x in the boundary ∂B_1^m, then the image $g(B_1^m)$ covers P.*

(*Proof*)   Without loss of generality, we assume that P is the origin of $\mathbb{R}^m$. Consider the mapping g as a continuous vector field on the unit ball B_1^m.

As $\|g(x) - x\| < \|x\|$, we are going to show that $g(x)$ makes an acute angle with the outward ray from the origin through x for every $x \in \partial B^m$. Let us consider the equality

$$\|y - z\|^2 + \|z - x\|^2 = \|y - x\|^2 + 2\langle y - z, x - z\rangle,$$

and take $y = g(x)$ and $z = 0$. Then

$$\begin{aligned}
2\langle g(x), x\rangle &= \|g(x)\|^2 + \|x\|^2 - \|g(x) - x\|^2 > \|g(x)\|^2 + \|x\|^2 - \|x\|^2 \\
&= \|g(x)\|^2 \geq 0.
\end{aligned}$$

Thus $g(x)$ makes an acute angle with x. So $g(x)$ has an outward radial component at every point $x \in \partial B_1^m$. The vector $-g(x)$ has a negative radial component. For a sufficiently small positive number α the function $x \to x - \alpha\, g(x)$ goes from B_1^m to B_1^m. By Theorem A.3.1 there exists a fixed point x_0 such that $x_0 = x_0 - \alpha\, g(x_0)$, then $\alpha\, g(x_0) = 0$ and $g(x_0) = 0$ since $\alpha \in \mathbb{R}^+$. As g is continuous and $g(x_0) = 0$, the image of a neighbourhood of x_0 covers the origin.   ∎

Appendix B
Convex sets, cones and hyperplanes

We study some properties satisfied by convex sets and cones; see [Bertsekas et al. 2001, Rockafellar and Wets 1998] for details. Unless otherwise stated, we suppose that all the sets are in a n–dimensional vector space E.

B.1 Convex sets and cones

We need to define the different kinds of cones and linear combinations used mainly in Chapter 4.

Definition B.1.1. *A **cone** C with vertex at $0 \in E$ satisfies that if $v \in C$, then $\lambda v \in C$ for every $\lambda \geq 0$.*

Definition B.1.2. *Given a family of vectors $V \subset E$.*

1. *A **conic non–negative combination** of elements in V is a vector of the form $\lambda_1 v_1 + \ldots + \lambda_r v_r$, with $\lambda_i \geq 0$ and $v_i \in V$ for all $i \in \{1, \ldots, r\}$.*

2. *The **convex cone** generated by V is the set of all conic non–negative combinations of vectors in V.*

3. *An **affine combination** of elements in V is a vector of the form $\lambda_1 v_1 + \ldots + \lambda_r v_r$, with $v_i \in V$, $\lambda_i \in \mathbb{R}$ for all $i \in \{1, \ldots, r\}$ and $\sum_{i=1}^r \lambda_i = 1$.*

4. *A **convex combination** of elements in V is a vector of the form $\lambda_1 v_1 + \ldots + \lambda_r v_r$, with $v_i \in V$, $0 \leq \lambda_i \leq 1$ for all $i \in \{1, \ldots, r\}$ and $\sum_{i=1}^r \lambda_i = 1$.*

Remember that a set $A \subset E$ is *convex* if, given two different elements in A, then any convex combination of them is contained in A. Thus, all the convex combination of elements in A are in A.

Definition B.1.3. *The **convex hull** of a set $A \subset E$, $\mathrm{conv}(A)$, is the smallest convex subset containing A.*

Let us prove another characterization of the convex hull that will be useful.

Proposition B.1.4. *The convex hull of a set A is the set of the convex combinations of elements in A.*

(*Proof*) Let us denote by C the set of all convex combinations of elements in A. First, we prove that C is a convex set. If x, y are in C, then they are convex combinations of elements in A; that is, $x = \sum_{i=1}^{l} \lambda_i v_i$, $y = \sum_{i=1}^{r} \mu_i w_i$, with $\sum_{i=1}^{l} \lambda_i = 1$, $\sum_{i=1}^{r} \mu_i = 1$. For $s \in (0,1)$, consider

$$sx + (1-s)y = s\left(\sum_{i=1}^{l} \lambda_i v_i\right) + (1-s)\left(\sum_{i=1}^{r} \mu_i w_i\right),$$

that will be in C if the sum of the coefficients is equal to 1 and each of the coefficients lies in $[0,1]$. Observe that $s\sum_{i=1}^{l} \lambda_i + (1-s)\sum_{i=1}^{r} \mu_i = s + (1-s) = 1$ and the other condition is satisfied trivially. As C is convex and contains A, the convex hull of A is a subset of C.

Second, we prove that $C \subset \mathrm{conv}(A)$ by induction on the number of vectors in the convex combinations of elements in A. Trivially, when the convex combination is given by an element in A, it lies in the convex hull of A.

Now, suppose that a convex combination of $l-1$ elements of A is in $\mathrm{conv}(A)$. Let

$$x = \sum_{i=1}^{l} \mu_i v_i = \sum_{i=1}^{l-1} \mu_i v_i + \mu_l v_l.$$

If $\sum_{i=1}^{l-1} \mu_i = 0$, then $\mu_l = 1$. By the first step of the induction, x is in $\mathrm{conv}(A)$. If $\sum_{i=1}^{l-1} \mu_i \in (0,1]$, then $\mu_l \in [0,1)$ and we can rewrite x as

$$x = (1-\mu_l)\left(\sum_{i=1}^{l-1} \mu_i (1-\mu_l)^{-1} v_i\right) + \mu_l v_l.$$

Observe that $\sum_{i=1}^{l-1} \mu_i (1-\mu_l)^{-1} = (1-\mu_l)(1-\mu_l)^{-1} = 1$, and so $\sum_{i=1}^{l-1} \mu_i (1-\mu_l)^{-1} v_i$ is in $\mathrm{conv}(A)$. By the first step of the induction, v_l is in $\mathrm{conv}(A)$. As $(1-\mu_l) + \mu_l = 1$, x is in $\mathrm{conv}(A)$. Thus $C \subset \mathrm{conv}(A)$ and so $C = \mathrm{conv}(A)$. ∎

Proposition B.1.5. *Let C be a convex set. If $\overline{C}$ and $\mathrm{int}\, C$ are the topological closure and the interior of C, respectively, we have:*

(a) *for every $x \in \mathrm{int}\, C$, if $y \in \overline{C}$, then $(1-\lambda)x + \lambda y \in \mathrm{int}\, C$ for all $\lambda \in [0,1)$;*

(b) *$\overline{C} = \overline{\mathrm{int}\, C}$;*

(c) *the interior of C is empty if and only if the interior of $\overline{C}$ is empty;*

(d) *$\mathrm{int}\, C = \mathrm{int}\, \overline{C}$.*

(*Proof*) **(a)** If $x \in \mathrm{int}\, C$, then there exists $\epsilon_x > 0$ such that $B(x, \epsilon_x) \subset C$, where $B(x, \epsilon_x)$ denotes the open ball centered at x of radius ϵ_x.

Observe that if $y \in \overline{C}$, for any $\epsilon > 0$, $y \in C + \epsilon B(0,1) = \{x + \epsilon z \mid x \in C, \, z \in B(0,1)\}$.

For every $\lambda \in [0,1)$, we consider $x_\lambda = (1-\lambda)x + \lambda y$. Let us compute the value of ϵ_λ such

that $x_\lambda + \epsilon_\lambda B(0,1) \subset C$.

$$x_\lambda + \epsilon_\lambda B(0,1) = (1-\lambda)x + \lambda y + \epsilon_\lambda B(0,1)$$

$$\subset (1-\lambda)x + \lambda C + \lambda \epsilon B(0,1) + \epsilon_\lambda B(0,1) = (1-\lambda)x + (\lambda\epsilon + \epsilon_\lambda)B(0,1) + \lambda C.$$

If $\epsilon_\lambda = (1-\lambda)\epsilon_x - \lambda\epsilon$, then

$$(1-\lambda)x + (\lambda\epsilon + \epsilon_\lambda)B(0,1) \subset (1-\lambda)C,$$

and $x_\lambda + \epsilon_\lambda B(0,1) \subset C$. For $\epsilon > 0$ small enough, ϵ_λ is positive. Here we use the sum operation of convex sets, which is well–defined if the coefficients are positive (if C_1 and C_2 are convex sets, $\mu_1 C_1 + \mu_2 C_2$ is a convex set for all $\mu_1, \mu_2 \geq 0$).

(b) As $\operatorname{int} C \subset C$, $\overline{\operatorname{int} C} \subset \overline{C}$.

On the other hand, each point in the closure of C can be approached along a line segment by points in the interior of C by (a). Thus $\overline{C} \subset \overline{\operatorname{int} C}$.

(c) As $\operatorname{int} C \subset \operatorname{int} \overline{C}$, if $\operatorname{int} \overline{C}$ is empty, then $\operatorname{int} C$ is empty.

Conversely, if $\operatorname{int} C$ is empty, then by (b) $\overline{C}$ is empty. So C is empty and $\operatorname{int} C$ is also empty.

(d) Trivially $\operatorname{int} C \subset \operatorname{int} \overline{C}$.

As the equality of the sets is true when they are empty because of (c), let us suppose that $\operatorname{int} C$ is not empty. If $z \in \operatorname{int} \overline{C}$ and take $x \in \operatorname{int} C$. Then there exists a small enough positive number δ such that $y = z + \delta(z - x) \in \operatorname{int} \overline{C} \subset \overline{C}$.

Hence,

$$z = \frac{1}{1+\delta}y + \frac{\delta}{1+\delta}x.$$

Note that

$$0 < \frac{1}{1+\delta} < 1, \quad 0 < \frac{\delta}{1+\delta} < 1, \quad \frac{1}{1+\delta} + \frac{\delta}{1+\delta} = 1.$$

As $y \in \overline{C}$, $x \in \operatorname{int} C$ and $1/(1+\delta)$ lies in $(0,1)$. By (a), $z \in \operatorname{int} C$. ∎

Remark B.1.6. Consequently, if C is convex and dense, then C is the whole space.

B.2 Distinguished hyperplanes

The following paragraphs introduce elements playing an important role in the proof of Pontryagin's Maximum Principle, §4.2 and §4.4.

Definition B.2.1. *Let C be a cone with vertex at $0 \in E$. A **supporting hyperplane to C at** 0 is a hyperplane such that C is contained in one of the half–spaces defined by the hyperplane.*

Remark B.2.2. In a geometric framework, we will define a hyperplane in E as the kernel of a

nonzero 1–form α in the dual space E^* of E. Then the hyperplane P_α associated to α is $\ker \alpha$ and the supporting hyperplane to C at 0 is a hyperplane P_α such that $\alpha(v) \leq 0$ for all $v \in C$. A supporting hyperplane to C at 0 is not necessarily unique.

From now on, we consider that all the cones have vertex at 0.

Definition B.2.3. *Let C be a cone, the **polar of** C is*

$$C^* = \{\alpha \in E^* \mid \alpha(v) \leq 0, \ \forall\, v \in C\}.$$

Note that the polar of a cone is a closed and convex cone in E^*.

Definition B.2.4. *Let C be a cone, the set*

$$C^{**} = \{w \in E \mid \alpha(w) \leq 0 \ , \ \forall\, \alpha \in C^*\}$$

*is called the **polar of the polar of** C.*

Observe that $C \subset C^{**}$. The following lemma is used in the proof of the existence of a supporting hyperplane to a cone with vertex at 0.

Lemma B.2.5. *The cone C is closed and convex if and only if $C^{**} = C$.*

(*Proof*) Observe that

$$C^{**} = \{w \in E \mid \alpha(w) \leq 0 \ , \ \forall\, \alpha \in C^*\} = \bigcap_{\alpha \in C^*} \{w \in E \mid \alpha(w) \leq 0\}.$$

Rockafellar and Wets [1998, Theorem 6.20] prove that the closure of the convex hull of a set is the intersection of all the closed half–spaces containing the set. Then $C^{**} = \overline{\mathrm{conv}(C)}$. Now, the result is immediate. ∎

The following proposition guarantees the existence of a supporting hyperplane to a cone with vertex at 0. This result is used throughout the proof of Pontryagin's Maximum Principle in Chapter 4.

Proposition B.2.6. *If C is a convex and closed cone that is not the whole space, then there exists a supporting hyperplane to C at 0.*

(*Proof*) If there is no supporting hyperplane containing the cone in one of the two half–spaces, then for all $\alpha \in E^*$ there exist $v_1, v_2 \in C$ with $\alpha(v_1) \leq 0$ and $\alpha(v_2) \geq 0$. Thus $C^* = \{0\}$ and $C^{**} = E$. Then, by Lemma B.2.5, $C = C^{**} = E$ in contradiction with the hypothesis on C. ∎

Corollary B.2.7. *If C is a convex cone that is not the whole space, then there exists a supporting hyperplane to C at 0.*

(*Proof*) If $C \neq E$, then $\overline{C} \neq E$ by Proposition B.1.5 (d). Hence, by Proposition B.2.6, there exists a supporting hyperplane to $\overline{C}$ which is also a supporting hyperplane to C. ∎

Definition B.2.8. *Let C_1 and C_2 be **cones** with common vertex 0. They **are separated** if there exists a hyperplane P such that each cone lies in a different closed half–space defined by P. P is called a **separating hyperplane of C_1 and C_2**.*

A point x is a *relative interior point of a set C*, if $x \in C$ and there exists a neighbourhood V of x such that $V \cap \text{aff}(U) \subseteq U$. Then, a useful characterization of separated convex cones is the following:

Proposition B.2.9. *The convex cones C_1 and C_2, with common vertex 0, are separated if and only if one of the two following conditions are satisfied:*

(1) there exists a hyperplane containing both C_1 and C_2,

(2) there is no point that is a relative interior point of both C_1 and C_2.

(*Proof*) $\Rightarrow$ If C_1 and C_2 are separated then there exists a separating hyperplane P_α such that
$$\alpha(v_1) \leq 0 \quad \forall\, v_1 \in C_1, \ \alpha(v_2) \geq 0 \quad \forall\, v_2 \in C_2.$$
If $\alpha(v_i) = 0$ for all $v_i \in C_i$ and $i = 1, 2$, then we are in the first case.

If some $v_i \in C_i$ satisfies the strict inequality, then both sets do not lie in the hyperplane P_α. They lie in a different closed half–space. If the convex cones intersect, the intersection lies in the boundary of the cones and in the hyperplane. Hence, there is no point that is a relative interior point of both C_1 and C_2 since the $\text{aff}(C_i)$ is equal to the subspace that contains the whole cone, for $i = 1, 2$.

$\Leftarrow$ First, we are going to prove that if (1) is true, then C_1 and C_2 are separated. As there exists a hyperplane determined by α such that $\alpha(v_i) = 0$ for all $v_i \in C_i$, α determines a separating hyperplane of C_1 and C_2.

Now, we are going to prove that if (2) is true, then C_1 and C_2 are separated. As C_1 and C_2 are convex cones,
$$C_1 - C_2 = \{u \in E \mid u = v_1 - v_2, \ v_1 \in C_1, \ v_2 \in C_2\}$$
is a convex cone. Since there is no relative interior point of both C_1 and C_2, 0 does not lie in $C_1 - C_2$. By Corollary B.2.7 there exists a supporting hyperplane P_α to $C_1 - C_2$ such that $\alpha(v_1 - v_2) \leq 0$, that is, $\alpha(v_1) \leq \alpha(v_2)$, for all $v_1 \in C_1, v_2 \in C_2$.

Observe that a supporting hyperplane to $C_1 - C_2$ is a supporting hyperplane to C_1, because, taking $v_2 = 0$, $\alpha(v_1) \leq \alpha(v_2) = 0$ for all $v_1 \in C_1$.

As $\partial(C_1 - C_2) \cap C_1 \subset \partial C_1$, we consider a supporting hyperplane P_α to $C_1 - C_2$ such that $\alpha(v_1) = 0$ for some $v_1 \in \partial C_1$. Hence $\alpha(v_2) \geq \alpha(v_1) = 0$ for all $v_2 \in C_2$. As $\alpha(v_1) \leq 0$ for all $v_1 \in C_1$, α determines a separating hyperplane of C_1 and C_2. ∎

This proposition gives us necessary and sufficient conditions for the existence of a separating hyperplane of two convex cones with common vertex. Observe that a separating hyperplane of two cones with common vertex is also a supporting hyperplane to each cone at the vertex.

Corollary B.2.10. *If the convex cones C_1 and C_2 with common vertex 0 are not separated, then $E = C_1 - C_2$.*

(*Proof*) If the cones are not separated, by Proposition B.2.9 there exists no any hyperplane containing both and the intersection of their relative interior is not empty.

Let us suppose that the convex cone $C_1 - C_2 \neq E$. Then, by Corollary B.2.7, there exists a supporting hyperplane determined by λ at the vertex such that $\lambda(v) \geq 0$ for every v in $C_1 - C_2$.

Because of the definition of cones, if $v_1 \in C_1$, then $v_1 \in C_1 - C_2$ and $\lambda(v_1) \geq 0$. Analogously, if $v_2 \in C_2$, then $-v_2 \in C_1 - C_2$ and $\lambda(-v_2) \geq 0$, that is, $\lambda(v_2) \leq 0$. ∎

Appendix C
Vector–valued quadratic forms

$\mathbf{R}$ecently, there has been a new interest in vector–valued quadratic forms to try to give more conditions about the controllability of the systems. This is a field still under research as shown in [Basto-Gonçalves 1998, Bullo and Lewis 2005a, Hirschorn and Lewis 2002].

C.1   Definition and properties

Let V and W be finite–dimensional $\mathbb{R}$–vector spaces. The set of symmetric $\mathbb{R}$–bilinear maps from $V \times V \to W$ is denoted by $\Sigma_2(V; W)$ and also called the set of *vector–valued quadratic forms*. For $B \in \Sigma_2(V; W)$, we define the mapping $Q_B \colon V \to W$ by $Q_B(v) = B(v, v)$ and is called the *quadratic form associated with B*.

For $\lambda \in W^*$, we define $\lambda B \colon V \times V \to \mathbb{R}$, $\lambda B(v_1, v_2) = \langle \lambda, B(v_1, v_2) \rangle$.

Definition C.1.1. *Let $B \in \Sigma_2(V; W)$.*

*(i) B is **definite** if there exists $\lambda \in W^*$ such that λB is positive–definite;*

*(ii) B is **semidefinite** if there exists $\lambda \in W^* \setminus \{0\}$ such that λB is positive–semidefinite;*

*(iii) B is **strongly semidefinite** if there exists $\lambda \in W^* \setminus \{0\}$ such that λB is nonzero and positive–semidefinite;*

*(iv) B is **indefinite** if, for each $\lambda \in W^* \setminus \{0\}$, λB is neither positive nor negative–semidefinite;*

*(v) B is **essentially indefinite** if, for each $\lambda \in W^*$, λB is either*

- *zero or*
- *neither positive nor negative–semidefinite.*

Observe that the notion of (semi)definite can be also defined saying that the real quadratic form is negative–(semi)definitive. It is enough to consider $-\lambda \in W^*$.

A useful result from [Bullo and Lewis 2005a] is the following:

Lemma C.1.2. *Let V and W be finite–dimensional $\mathbb{R}$–vector spaces with $B \colon V \times V \to W$ being a nonzero vector–valued quadratic form. Define $Q_B \colon V \to W$ by $Q_B(v) = B(v; v)$. The following statements hold:*

(i) B is indefinite if and only if

$$0 \in \mathrm{int}_{\mathrm{aff}(\mathrm{Im}(Q_B))}(\mathrm{conv}(\mathrm{Im}(Q_B)));$$

(ii) B is definite if and only if there exists a hyperplane P through $0 \in W$ such that

    (a) $\mathrm{Im}(Q_B)$ lies on one side of P and

    (b) $\mathrm{Im}(Q_B) \cap P = \{0\}$;

(iii) if Q_B is surjective, then B is indefinite.

C.2   A particular vector–valued quadratic form

In this dissertation we are interested in the following vector–valued quadratic form. Let Q be a manifold of dimenstion n and a $\mathscr{Y}$ be distribution on Q. We define the **vector–valued quadratic form at** $x \in Q$ as follows:

$$
\begin{aligned}
B_{\mathscr{Y}_x}(\mathscr{Y}_x)\colon \mathscr{Y}_x \times \mathscr{Y}_x &\longrightarrow T_x Q / \mathscr{Y}_x \\
(w_1, w_2) &\longmapsto \pi_{\mathscr{Y}_x}(\langle W_1 : W_2 \rangle)
\end{aligned}
$$

where W_1 and W_2 are vector fields on Q spanned by $Y_1, \ldots, Y_k$ and extending $w_1, w_2 \in \mathscr{Y}_x$ and $\pi_{\mathscr{Y}_x}\colon T_x Q \to T_x Q / \mathscr{Y}_x$ is the natural projection onto the quotient space. This vector–valued quadratic form is well–defined; that is, it does not depend on the extensions considered. Let W_i and V_i be two different extensions for w_i, $i = 1, 2$, we need to prove that $\pi_{\mathscr{Y}_x}(\langle W_1 : W_2 \rangle) = \pi_{\mathscr{Y}_x}(\langle V_1 : V_2 \rangle)$. If $W_i = a_i^j Y_j$,

$$\nabla_{W_1} W_2 = a_1^j(Y_j(a_2^l))Y_l + a_1^j a_2^l \nabla_{Y_j} Y_l,$$

then $\pi_{\mathscr{Y}_x}(\langle W_1 : W_2 \rangle) = a_1^j a_2^l \pi_{\mathscr{Y}_x}(\nabla_{Y_j} Y_l + \nabla_{Y_l} Y_j) = \pi_{\mathscr{Y}_x}(\langle V_1 : V_2 \rangle)$.

Assuming to have a regular distribution on Q with rank k, the matrix of the vector–valued quadratic form is given by $n - k$ symmetric $(k \times k)$–matrices denoted by $B^1, \ldots, B^{n-k}$ such that $B_{jl}^i = \langle \eta_x^i, \pi_{\mathscr{Y}_x}(\langle Y_j : Y_l \rangle) \rangle$, where η_x^i is a basis of $(T_x Q / \mathscr{Y}_x)^*$.

For any $\lambda \in (T_x Q / \mathscr{Y}_x)^* \simeq \mathrm{ann}\, \mathscr{Y}_x$, we have the following real quadratic form at $x \in Q$

$$
\begin{aligned}
(\lambda\, B)_x = \lambda B_{\mathscr{Y}_x}(\mathscr{Y}_x)\colon \mathscr{Y}_x \times \mathscr{Y}_x &\longrightarrow \mathbb{R} \\
(w_1, w_2) &\longmapsto \langle \lambda, \langle W_1 : W_2 \rangle(x) \rangle = \lambda_i B_{jl}^i a_1^j a_2^l,
\end{aligned}
\tag{C.2.1}
$$

where $\lambda_i B_{jl}^i = \langle \lambda, \langle Y_j : Y_l \rangle \rangle$.

According to Bullo and Lewis [2005b], Hirschorn and Lewis [2002], any vector–valued quadratic form $B_{\mathscr{Y}_x}(\mathscr{Y}_x) = B$ is

- either **strongly semidefinite**; that is, there exists $\lambda \in \mathrm{ann}\, \mathscr{Y}_x \backslash \{0\}$ such that the real quadratic form $(\lambda\, B)_x$ is nonzero and positive semidefinite;

- or **essentially indefinite**; that is, for all $\lambda \in \text{ann}\,\mathscr{Y}_x$ the real quadratic form $(\lambda\,B)_x$ is zero or nondefinite.

In [Bullo and Lewis 2005a, Chapter 8] there are results of first–order controllability related with the vector–valued quadratic form.

Bibliography

R. Abraham and J. E. Marsden. *Foundations of Mechanics*. Benjamin/Cummings Publishing Co. Inc. Advanced Book Program, Reading, Mass., second edition, 1978.

R. Abraham, J. E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications*, volume 75 of *Applied Mathematical Sciences*. Springer–Verlag, New York, second edition, 1988.

A. Agrachev. Introduction to optimal control theory. In *Mathematical Control Theory, Part 1, 2 (Trieste, 2001)*, ICTP Lect. Notes, VIII, page 453–513 (electronic). Abdus Salam Int. Cent. Theoret. Phys., Trieste, 2002a.

A. Agrachev. Differential geometry and optimal control problems. Lectures given on the school "Geometric control theory" and written by A. Panasyuk, Warsaw, 2002b.

A. Agrachev. Is it possible to recognize local controllability in a finite number of differentiations? In *Open Problems in Mathematical Systems and Control Theory*, Comm. Control Engrg. Ser., page 15–18. Springer, London, 1999.

A. Agrachev and J.-P. Gauthier. On the subanalyticity of Carnot–Caratheodory distances. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 18(3):359–382, 2001.

A. Agrachev and Y. L. Sachkov. *Control Theory from the Geometric Viewpoint*, volume 87 of *Encyclopaedia of Mathematical Sciences*. Springer–Verlag, Berlin, 2004. Control Theory and Optimization, II.

A. Agrachev and A. V. Sarychev. Strong minimality of abnormal geodesics for 2–distributions. *J. Dynam. Control Systems*, 1(2):139–176, 1995a.

A. Agrachev and A. V. Sarychev. On abnormal extremals for Lagrange variational problems. *J. Math. Systems Estim. Control*, 5(1):31 pp. (electronic), 1995b.

A. Agrachev and A. V. Sarychev. Abnormal sub–Riemannian geodesics: Morse index and rigidity. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 13(6):635–690, 1996.

A. Agrachev and A. V. Sarychev. Sub–Riemannian metrics: Minimality of abnormal geodesics versus subanalyticity. *ESAIM Control Optim. Calc. Var.*, 4:377–403 (electronic), 1999.

A. Agrachev and I. Zelenko. On feedback classification of control–affine systems with one–and two–dimensional inputs. *SIAM J. Control Optim.*, 46(4):1431–1460 (electronic), 2007.

S. K. Agrawal and B. C. Fabien. *Optimization of Dynamic Systems*, volume 70 of *Solid Mechanics and Its Applications*. Kluwer Academic Publishers, 1999.

C. Aguilar and A. D. Lewis. A jet bundle setting for studying the reachable set. In *Proceedings of the 18th Mathematical Theory of Networks and Systems in Blacksburg (Virginia)*, 2008. to appear.

A. Ailon and G. Langholz. On the existence of time–optimal control of mechanical manipulators. *J. Optim. Theory Appl.*, 46(1):1–21, 1985.

D. A. Anisi. Optimal motion control of a ground vehicle. Technical report, Swedish Defence Research Agency, System Technology Division, Stockholm, Jul 2003.

M. Athans and P. L. Falb. *Optimal Control. An Introduction to the Theory and its Applications.* McGraw–Hill Book Co., New York, 1966.

M. Barbero-Liñán and M. C. Muñoz Lecanda. Constraint algorithm for extremals in optimal control problems. Preprint, arXiv:0708.0102 math.OC, Feb 2008a.

M. Barbero-Liñán and M. C. Muñoz Lecanda. Geometric approach to Pontryagin's maximum principle. To be published in *Acta Appl. Math.*, arxiv:0805.1169 math.OC, May 2008b.

M. Barbero-Liñán and M. C. Muñoz Lecanda. Strict abnormal extremals in nonholonomic and kinematic control systems. Preprint, arxiv:0806.2814 math.OC, Jun 2008c.

M. Barbero-Liñán and M. C. Muñoz Lecanda. Optimal control problems for affine connection control systems: Characterization of extremals. In *Geometry and Physics: XVI International Fall Workshop, R. L. Fernandes and R. Picken (eds.), AIP Conference Proceedings*, volume 1023, page 127–131, 2008d.

M. Barbero-Liñán, A. Echeverría-Enríquez, D. Martín de Diego, M. C. Muñoz-Lecanda, and N. Román-Roy. Skinner–Rusk unified formalism for optimal control systems and applications. *J. Phys. A*, 40(40):12071–12093, 2007.

M. Barbero-Liñán, A. Echeverría-Enríquez, D. Martín de Diego, M. C. Muñoz-Lecanda, and N. RománRoy. Unified formalism for nonautonomous mechanical systems. *J. Math. Phys.*, 49(6), 2008.

J. Basto-Gonçalves. Second–order conditions for local controllability. *Systems Control Lett.*, 35(5):287–290, 1998.

A. Bensoussan. On the stochastic maximum principle for infinite-dimensional equations and application to the control of Zakai equation. In *Filtering and Control of Random Processes (Paris, 1983)*, volume 61 of *Lecture Notes in Control and Inform. Sci.*, page 13–23. Springer, Berlin, 1984.

D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, Koksal. Convexity, Duality and Lagrange Multipliers, 2001. http://citeseer.ist.psu.edu/499868.html, Lectures Notes for the Massachussets Institute of Technology.

R. M. Bianchini. High order necessary optimality conditions. *Rend. Sem. Mat. Univ. Politec. Torino*, 56(4):41–51 (2001), 1998. Control theory and its applications (Grado, 1998).

R. M. Bianchini and G. Stefani. Controllability along a trajectory: A variational approach. *SIAM J. Control Optim.*, 31(4):900–927, 1993.

G. Blankenstein, R. Ortega, and A. J. van der Schaft. The matching conditions of controlled Lagrangians and IDA–passivity based control. *Internat. J. Control*, 75(9):645–665, 2002.

A. M. Bloch. *Nonholonomic Mechanics and Control*, volume 24 of *Interdisciplinary Applied Mathematics*. Springer–Verlag, New York, 2003. With the collaboration of J. Baillieul, P. Crouch and J. Marsden, With scientific input from P. S. Krishnaprasad, R. M. Murray and D. Zenkov, Systems and Control.

A. M. Bloch, N. E. Leonard, and J. E. Marsden. Controlled Lagrangians and the stabilization of mechanical systems. I. The first matching theorem. *IEEE Trans. Automat. Control*, 45(12): 2253–2270, 2000.

V. Boltyanski, H. Martini, and V. Soltan. *Geometric Methods and Optimization Problems*, volume 4 of *Combinatorial Optimization*. Kluwer Academic Publishers, Dordrecht, 1999.

B. Bonnard and J.-B. Caillau. Introduction to nonlinear optimal control. In *Advanced Topics in Control Systems Theory*, volume 328 of *Lecture Notes in Control and Inform. Sci.*, page 1–60. Springer, London, 2006.

B. Bonnard and M. Chyba. *Singular Trajectories and their Role in Control Theory*, volume 40 of *Mathématiques & Applications (Berlin)*. Springer–Verlag, Berlin, 2003.

B. Bonnard and E. Trélat. On the role of abnormal minimizers in sub–Riemannian geometry. *Ann. Fac. Sci. Toulouse Math. (6)*, 10(3):405–491, 2001.

U. Boscain and B. Piccoli. On automaton recognizability of abnormal extremals. *SIAM J. Control Optim.*, 40(5):1333–1357 (electronic), 2002.

U. Boscain and B. Piccoli. *Optimal Syntheses for Control Systems on 2–D Manifolds*, volume 43 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer–Verlag, Berlin, 2004.

A. Bressan and B. Piccoli. *Introduction to the Mathematical Theory of Control*, volume 2 of *AIMS Series on Applied Mathematics*. American Institute of Mathematical Sciences (AIMS), Springfield, MO, 2007.

F. Bullo and A. D. Lewis. On the homogeneity of the affine connection model for mechanical control systems. In *Proceedings of the 39th IEEE Conference on Decision and Control*, volume 2, page 1260–1265, 2000.

F. Bullo and A. D. Lewis. *Geometric Control of Mechanical Systems*, volume 49 of *Texts in Applied Mathematics*. Springer–Verlag, New York, 2005a. Modeling, analysis, and design for simple mechanical control systems.

F. Bullo and A. D. Lewis. Supplementary chapters of geometric Control of Mechanical Systems. http://penelope.mast.queensu.ca/smcs/, 2005b.

F. Bullo and A. D. Lewis. Low–order controllability and kinematic reductions for affine connection control systems. *SIAM J. Control Optim.*, 44(3):885–908 (electronic), 2005c.

F. Bullo and A. D. Lewis. Reduction, linearization, and stability of relative equilibria for mechanical systems on Riemannian manifolds. *Acta Appl. Math.*, 99(1):53–95, 2007.

J. A. Cañizo–Rincón. Ecuaciones diferenciales ordinarias en el sentido de Carathéodory. http://www.ceremade.dauphine.fr/~canizo/tex/ordinarias-debil.pdf, Jul 2004.

J. F. Cariñena. Theory of singular Lagrangians. *Fortschr. Phys.*, 38(9):641–679, 1990.

Y.-C. Chen. On the structure of the time–optimal controls for robotic manipulators. *IEEE Trans. Automat. Control*, 34(1):115–116, 1989.

Y. Chitour, F. Jean, and E. Trélat. Genericity results for singular curves. *J. Differential Geom.*, 73(1):45–73, 2006.

Y. Chitour, F. Jean, and E. Trélat. Singular trajectories of control–affine systems. *SIAM J. Control Optim.*, 47(2):1078–1095, 2008.

M. Chyba and T. Haberkorn. Designing efficient trajectories for underwater vehicles using geometric control theory. In *Proceedings of OMAE05 2005, 24rd International Conference on Offshore Mechanics and Artic Engineering Halkidiki, Greece, June 12–17, 2005*, 2005.

M. Chyba, N. E. Leonard, and E. D. Sontag. Singular trajectories in multi–input time–optimal problems: Application to controlled mechanical systems. *J. Dynam. Control Systems*, 9(1): 103–129, 2003.

M. Chyba, E. Hairer, and G. Vilmart. The role of symplectic integrators in optimal control. *Optim. Control Appl. Meth.*, 2008. http://dx.doi.org/10.1002/oca.855.

E. A. Coddington and N. Levinson. *Theory of Ordinary Differential Equations*. McGraw–Hill Book Company, Inc., New York–Toronto–London, 1955.

L. Conlon. *Differentiable Manifolds: a First Course*. Birkhäuser Advanced Texts: Basel Textbooks. Birkhäuser Boston Inc., Boston, MA, 1993.

A. T. Coombs. Time–optimal control of two simple mechanical systems with three degrees of freedom and two inputs.
http://qspace.library.queensu.ca/bitstream/1974/48/1/2000k_letter.pdf, 2000.

J. Cortés and E. Martínez. Mechanical control systems on Lie algebroids. *IMA J. Math. Control Inform.*, 21(4):457–492, 2004.

J. Cortés and S. Martínez. Configuration controllability of mechanical systems underactuated by one control. *SIAM J. Control Optim.*, 41(6):1901–1921 (electronic), 2003.

J. Cortés, M. de León, D. Martín de Diego, and S. Martínez. Geometric description of vakonomic and nonholonomic dynamics. Comparison of solutions. *SIAM J. Control Optim.*, 41 (5):1389–1412 (electronic), 2002a.

J. Cortés, S. Martínez, and F. Cantrijn. Skinner–Rusk approach to time–dependent mechanics. *Phys. Lett. A*, 300(2-3):250–258, 2002b.

J. Cortés, M. de León, J. C. Marrero, D. Martín de Diego, and E. Martínez. A survey of Lagrangian mechanics and control on Lie algebroids and groupoids. *Int. J. Geom. Methods Mod. Phys.*, 3(3):509–558, 2006.

J. Cortés Monforte. *Geometric, Control and Numerical Aspects of Nonholonomic Systems*, volume 1793 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002.

M. Crampin. On the differential geometry of the Euler–Lagrange equations, and the inverse problem of Lagrangian dynamics. *J. Phys. A*, 14(10):2567–2575, 1981.

M. Delgado-Téllez and A. Ibort. A panorama of geometrical optimal control theory. *Extracta Math.*, 18(2):129–151, 2003.

M. P. do Carmo. *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

A. Echeverría-Enríquez, M. C. Muñoz Lecanda, and N. Román-Roy. Connections and jet fields. DGDSA Report.

A. Echeverría-Enríquez, M. C. Muñoz-Lecanda, and N. Román-Roy. Geometrical setting of time –dependent regular systems. Alternative models. *Rev. Math. Phys.*, 3(3):301–330, 1991.

A. Echeverría-Enríquez, J. Marín-Solano, M. C. Muñoz-Lecanda, and N. Román-Roy. Geometric reduction in optimal control theory with symmetries. *Rep. Math. Phys.*, 52(1):89–113, 2003.

A. Echeverría-Enríquez, C. López, J. Marín-Solano, M. C. Muñoz-Lecanda, and N. Román-Roy. Lagrangian–Hamiltonian unified formalism for field theory. *J. Math. Phys.*, 45(1): 360–380, 2004.

O. P. Ferreira. Proximal subgradient and a characterization of Lipschitz function on Riemannian manifolds. *J. Math. Anal. Appl.*, 313(2):587–597, 2006.

A. F. Filippov. On certain questions in the theory of optimal control. *J. SIAM Control Ser. A*, 1: 76–84, 1962.

A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides*, volume 18 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1988. Translated from the Russian.

R. Gabasov and F. M. Kirillova. High order necessary conditions for optimality. *SIAM J. Control*, 10:127–168, 1972.

M. Giaquinta and S. Hildebrandt. *Calculus of Variations. I The Lagrangian Formalism*, volume 310 of *Fundamental Principles of Mathematical Sciences*. Springer-Verlag, Berlin, 1996a.

M. Giaquinta and S. Hildebrandt. *Calculus of Variations. II The Hamiltonian Formalism*, volume 311 of *Fundamental Principles of Mathematical Sciences*. Springer-Verlag, Berlin, 1996b.

M. J. Gotay and J. M. Nester. Presymplectic Lagrangian systems. I. The constraint algorithm and the equivalence theorem. *Ann. Inst. H. Poincaré Sect. A (N.S.)*, 30(2):129–142, 1979.

M. J. Gotay, J. M. Nester, and G. Hinds. Presymplectic manifolds and the Dirac–Bergmann theory of constraints. *J. Math. Phys.*, 19(11):2388–2399, 1978.

X. Gràcia and R. Martín. Geometric aspects of time–dependent singular differential equations. *Int. J. Geom. Methods Mod. Phys.*, 2(4):597–618, 2005.

X. Gràcia and J. M. Pons. A generalized geometric framework for constrained systems. *Differential Geom. Appl.*, 2(3):223–247, 1992.

X. Gràcia, J. Marín-Solano, and M. C. Muñoz-Lecanda. Some geometric aspects of variational calculus in constrained systems. *Rep. Math. Phys.*, 51(1):127–148, 2003.

V. Guibout and A. Bloch. A discrete maximum principle for solving optimal control problems. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, volume 2, page 1806–1811, 2004.

E. Hairer. Analyse II, Partie B. Lecture Notes in Université de Genève. http://www.unige.ch/~hairer/polycop.html, 1999.

U. Hamenstädt. Some regularity theorems for Carnot–Carathéodory metrics. *J. Differential Geom.*, 32(3):819–850, 1990.

U. G. Haussmann. *A Stochastic Maximum Principle for Optimal Control of Diffusions*, volume 151 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow, 1986.

R. M. Hirschorn and A. D. Lewis. Geometric local controllability: Second–order conditions. Preprint, http://penelope.mast.queensu.ca/~andrew/, 2002.

C. L. Hwang and L. T. Fan. A discrete version of Pontryagin's maximum principle. *Operations Res.*, 15:139–146, 1967.

D. Iglesias, J. C. Marrero, D. Martín de Diego, E. Martínez, and E. Padrón. Reduction of symplectic Lie algebroids by a Lie subalgebroid and a symmetry Lie group. *SIGMA Symmetry Integrability Geom. Methods Appl.*, 3:Paper 049, 28 pp. (electronic), 2007.

F. Jean. Sub–Riemannian geometry. Lectures at the Trimester on Dynamical and Control Systems, Trieste, September–December, 2003.

V. Jurdjevic. *Geometric Control Theory*, volume 52 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.

M. Kawski. High–order maximal principles. In *New Trends in Nonlinear Dynamics and Control, and their Applications*, volume 295 of *Lecture Notes in Control and Inform. Sci.*, page 313–326. Springer, Berlin, 2003.

D. Kirschner, S. Lenhart, and S. Serbin. Optimal control of the chemotherapy of HIV. *J. Math. Biol.*, 35:775–792, 1997.

H.-W. Knobloch. *Higher Order Necessary Conditions in Optimal Control Theory*, volume 34 of *Lecture Notes in Computer Science*. Springer–Verlag, Berlin, 1981.

S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry. Vol. I and II.* Wiley Classics Library. John Wiley & Sons Inc., New York, 1996. Reprint of the 1963 original, A Wiley-Interscience Publication.

I. Kolář, P. W. Michor, and J. Slovák. *Natural Operations in Differential Geometry*. Springer–Verlag, Berlin, 1993.

A. J. Krener. The high order maximal principle and its application to singular extremals. *SIAM J. Control Optim.*, 15(2):256–293, 1977.

Olga Krupková. *The Geometry of Ordinary Variational Equations*, volume 1678 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1997.

I. Kupka. Geometric theory of extremals in optimal control problems. I. The fold and Maxwell case. *Trans. Amer. Math. Soc.*, 299(1):225–243, 1987.

R. Kuwabara. Time–dependent mechanical symmetries and extended Hamiltonian systems. *Rep. Math. Phys.*, 19(1):27–38, 1984.

B. Langerock. Connections in sub–Riemannian geometry. In *Differential Geometry and its Applications (Opava, 2001)*, volume 3 of *Math. Publ.*, page 105–113. Silesian Univ. Opava, Opava, 2001.

B. Langerock. Geometric aspects of the maximum principle and lifts over a bundle map. *Acta Appl. Math.*, 77(1):71–104, 2003a.

B. Langerock. *Generalised Connections and Applications to Control Theory*. PhD thesis, Universiteit Gent, 2003b. https://archive.ugent.be/retrieve/957/801001362505.pdf.

B. Langerock. A connection theoretic approach to sub–Riemannian geometry. *J. Geom. Phys.*, 46(3-4):203–230, 2003c.

B. Langerock. Conic distributions and accessible sets. *J. Dyn. Control Syst.*, 14(1):43–70, 2008.

E. B. Lee and L. Markus. *Foundations of Optimal Control Theory*. John Wiley and Sons, 1967.

J. M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer–Verlag, New York, 2003.

M. de León and D. Martín de Diego. On the geometry of non–holonomic Lagrangian systems. *J. Math. Phys.*, 37(7):3389–3414, 1996.

M. de León and Paulo R. Rodrigues. *Methods of Differential Geometry in Analytical Mechanics*, volume 158 of *North–Holland Mathematics Studies*. North–Holland Publishing Co., Amsterdam, 1989.

M. de León, J. C. Marrero, and D. Martín de Diego. A new geometric setting for classical field theories. In *Classical and Quantum Integrability (Warsaw, 2001)*, volume 59 of *Banach Center Publ.*, page 189–209. Polish Acad. Sci., Warsaw, 2003.

M. de León, J. Cortés, D. Martín de Diego, and S. Martínez. General symmetries in optimal control. *Rep. Math. Phys.*, 53(1):55–78, 2004.

M. de León, J. C. Marrero, and E. Martínez. Lagrangian submanifolds and dynamics on Lie algebroids. *J. Phys. A*, 38(24):R241–R308, 2005.

A. D. Lewis. A brief on controllability of nonlinear systems. Preprint, http://penelope.mast.queensu.ca/∼andrew/, 2001.

A. D. Lewis. The maximum principle of Pontryagin in control and in optimal control, 2006. Course held in Department of IV Applied Mathematics in Technical University of Catalonia, http://penelope.mast.queensu.ca/MP-course.

A. D. Lewis. Affine connections and distributions with applications to nonholonomic mechanics. *Rep. Math. Phys.*, 42(1-2):135–164, 1998. Pacific Institute of Mathematical Sciences Workshop on Nonholonomic Constraints in Dynamics (Calgary, AB, 1997).

A. D. Lewis and R. M. Murray. Configuration controllability of simple mechanical control systems. *SIAM J. Control Optim.*, 35(3):766–790, 1997.

W. Liu and H. J. Sussmann. Abnormal sub–Riemannian minimizers. In *Differential Equations, Dynamical Systems, and Control Science*, volume 152 of *Lecture Notes in Pure and Appl. Math.*, page 705–716. Dekker, New York, 1994a.

W. Liu and H. J. Sussmann. Abnormal extremals and optimality in sub–Riemannian manifolds. In *Proceedings of the 33rd Conference on Decision and Control, Lake Buena Vista, FL*, Dec 1994b.

W. Liu and H. J. Sussmann. Shortest paths for sub–Riemannian metrics on rank–two distributions. *Mem. Amer. Math. Soc.*, 118(564), 1995.

P. D. Loewen. The rocket car. Lecture notes in UBC Math 403, Nov 2004.

C. López and E. Martínez. Sub–Finslerian metric associated to an optimal control system. *SIAM J. Control Optim.*, 39(3):798–811 (electronic), 2000.

L. Mangiarotti and G. Sardanashvily. *Gauge Mechanics*. World Scientific Publishing Co. Inc., River Edge, NJ, 1998.

E. Martínez. Reduction in optimal control theory. *Rep. Math. Phys.*, 53(1):79–90, 2004.

E. Martínez, J. F. Cariñena, and W. Sarlet. Derivations of differential forms along the tangent bundle projection. *Differential Geom. Appl.*, 2(1):17–43, 1992.

E. Martínez, J. F. Cariñena, and W. Sarlet. Derivations of differential forms along the tangent bundle projection. II. *Differential Geom. Appl.*, 3(1):1–29, 1993.

E. J. McShane. On multipliers for Lagrange problems. *Amer. J. Math.*, 61:809–819, 1939.

R. Montgomery. *A Tour of sub–Riemannian Geometries, their Geodesics and Applications*, volume 91 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2002.

R. Montgomery. A survey of singular curves in sub–Riemannian geometry. *J. Dynam. Control Systems*, 1(1):49–90, 1995.

R. Montgomery. Abnormal minimizers. *SIAM J. Control Optim.*, 32(6):1605–1620, 1994.

P. C. Müller. Linear–quadratic optimal control of descriptor systems. *J. Braz. Soc. Mech. Sci.*, 21(3):423–432, 1999.

P. C. Müller. Stability and optimal control of nonlinear descriptor systems: A survey. *Appl. Math. Comput. Sci.*, 8(2):269–286, 1998.

M. C. Muñoz-Lecanda and N. Román-Roy. Lagrangian theory for presymplectic systems. *Ann. Inst. H. Poincaré Phys. Théor.*, 57(1):27–45, 1992.

M. C. Muñoz-Lecanda and F. J. Yániz-Fernández. Control and kinematic systems. To appear in IEEE Trans. Automat. Control, 2008.

H. Nijmeijer and A. van der Schaft. *Nonlinear Dynamical Control Systems*. Springer–Verlag, New York, 1990.

J. P. Ostrowski and J. W. Burdick. Controllability for mechanical systems with symmetries and constraints. *Appl. Math. Comput. Sci.*, 7(2):305–331, 1997. Recent developments in robotics.

M. Pelletier. Time–optimality of nonsmooth singular trajectories of affine systems with a scalar input in $\mathbb{R}^3$. In *Geometric Control Theory (Russian) (Moscow, 1998)*, volume 64 of *Itogi Nauki Tekh. Ser. Sovrem. Mat. Prilozh. Temat. Obz.*, page 197–221. Vseross. Inst. Nauchn. i Tekhn. Inform. (VINITI), Moscow, 1999.

L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. *The Mathematical Theory of Optimal Processes*. Translated from the Russian by K. N. Trirogoff; edited by L. W. Neustadt. Interscience Publishers John Wiley & Sons, Inc. New York–London, 1962.

M. F. Rañada. Extended Legendre transformation approach to the time–dependent Hamiltonian formalism. *J. Phys. A*, 25(14):4025–4035, 1992.

R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Fundamental Principles of Mathematical Sciences*. Springer–Verlag, Berlin, 1998.

H. L. Royden. *Real Analysis*. The Macmillan Co., New York, 1963.

Y. L. Sachkov. Control theory on Lie groups, 2006. Lectures given in SISSA, Trieste, Italy, http://digitallibrary.sissa.it/retrieve/1131/15-2006M.pdf.

D. J. Saunders. *The Geometry of Jet Bundles*, volume 142 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1989.

R. Skinner and R. Rusk. Generalized Hamiltonian dynamics. I. Formulation on $T^*Q \oplus TQ$. *J. Math. Phys.*, 24(11):2589–2594, 1983.

E. D. Sontag. Remarks on the time–optimal control of a class of Hamiltonian systems. In *Proceedings of the 28th IEEE Conference on Decision and Control, Vol. 1–3 (Tampa, FL, 1989)*, page 217–221, New York, 1989. IEEE.

E. D. Sontag and H. J. Sussmann. Time–optimal control of manipulators. In *Proceedings of the IEEE International Conference on Robotics and Automation, San Francisco, CA*, volume 3, page 1692–1697. IEEE, 1986.

R. S. Strichartz. Sub–Riemannian geometry. *J. Differential Geom.*, 24(2):221–263, 1986.

J. Struckmeier. Hamiltonian dynamics on the symplectic extended phase space for autonomous and non–autonomous systems. *J. Phys. A*, 38(6):1257–1278, 2005.

H. J. Sussmann. A sufficient condition for local controllability. *SIAM J. Control Optim.*, 16(5): 790–802, 1978.

H. J. Sussmann. A cornucopia of four–dimensional abnormal sub–Riemannian minimizers. In *Sub–Riemannian Geometry*, volume 144 of *Progr. Math.*, page 341–364. Birkhäuser, Basel, 1996.

H. J. Sussmann. Geometry and optimal control. In *Mathematical Control Theory*, page 140–198. Springer, New York, 1999.

H. J. Sussmann. Course on optimal control of finite–dimensional deterministic systems. http://www.math.rutgers.edu/∼sussmann/weizmann-course-2000.html, 2000.

H. J. Sussmann. A very non–smooth maximum principle with state constraints. In *Proceedings of the 44th IEEE Conference on Decision and Control, Sevilla, Spain, December 12–15, 2005*, 2005.

H. J. Sussmann. Lie brackets and local controllability: A sufficient condition for scalar–input systems. *SIAM J. Control Optim.*, 21(5):686–713, 1983.

H. J. Sussmann. A general theorem on local controllability. *SIAM J. Control Optim.*, 25(1): 158–194, 1987.

H. J. Sussmann. An introduction to the coordinate–free maximum principle. In *Geometry of Feedback and Optimal Control*, volume 207 of *Monogr. Textbooks Pure Appl. Math.*, page 463–557. Dekker, New York, 1998.

H. J. Sussmann and V. Jurdjevic. Controllability of nonlinear systems. *J. Differential Equations*, 12:95–116, 1972.

H. J. Sussmann and G. Tang. Shortest paths for the Reeds–Shepp car: A worked out example of the use of geometric techniques in nonlinear optimal control. Technical report, Rutgers Center for Systems and Control, SYCON Report, Sep 1991.

H. J. Sussmann and J. C. Willems. 300 years of optimal control: From the brachystochrone to the maximum principle. *IEEE Control Syst.*, 1997. Historical Perspectives.

J. Szilasi. A setting for spray and Finsler geometry. In *Handbook of Finsler Geometry. Vol. 1, 2*, page 1183–1426. Kluwer Acad. Publ., Dordrecht, 2003.

E. Trélat. Some properties of the value function and its level sets for affine control systems with quadratic cost. *J. Dynam. Control Systems*, 6(4):511–541, 2000.

E. Trélat. Asymptotics of accessibility sets along an abnormal trajectory. *ESAIM Control Optim. Calc. Var.*, 6:387–414 (electronic), 2001.

J. L. Troutman. *Variational Calculus and Optimal Control. Optimization with Elementary Convexity*. Undergraduate Texts in Mathematics. Springer–Verlag, New York, second edition, 1996. With the assistance of William Hrusa.

W. M. Tulczyjew. Hamiltonian systems, Lagrangian systems and the Legendre transformation. In *Symposia Mathematica, Vol. XIV (Convegno di Geometria Simplettica e Fisica Matematica, INDAM, Rome, 1973)*, page 247–258. Academic Press, London, 1974.

D. Tyner. *Geometric Jacobian linearization*. PhD thesis, Queen's University (Kingston, Ontario, Canada), 2007. http://penelope.mast.queensu.ca/~andrew/.

D. E. Varberg. On absolutely continuous functions. *Amer. Math. Monthly*, 72:831–841, 1965.

M. Žefran, F. Bullo, and J. Radford. An investigation into non–smooth locomotion. In *Proceedings of the IEEE International Conference on Robotics and Automation, Detroit, Michigan*, 1999.

A. C. Zaanen. *Continuity, Integration and Fourier Theory*. Universitext. Springer–Verlag, Berlin, 1989.

E. Zeidler. *Nonlinear Functional Analysis and its Applications III. Variational Methods and Optimization*. Springer–Verlag, 1985.

I. Zelenko. Nonregular abnormal extremals of 2–distribution: Existence, second variation, and rigidity. *J. Dynam. Control Systems*, 5(3):347–383, 1999.