Aspects of Geometric Mechanics and Control of Mechanical Systems

Thesis by
Andrew David Lewis

Technical Report 95-017
for the
Control and Dynamical Systems Option

California Institute of Technology
Pasadena, California

1995

(Thesis defended 21 April, 1995)

Abstract

Many interesting control systems are mechanical control systems. In spite of this, there has not been much effort to develop methods which use the special structure of mechanical systems to obtain analysis tools which are suitable for these systems. In this dissertation we take the first steps towards a methodical treatment of mechanical control systems.

First we develop a framework for analysis of certain classes of mechanical control systems. In the Lagrangian formulation we study "simple mechanical control systems" whose Lagrangian is "kinetic energy minus potential energy." We propose a new and useful definition of controllability for these systems and obtain a computable set of conditions for this new version of controllability. We also obtain decompositions of simple mechanical systems in the case when they are not controllable. In the Hamiltonian formulation we study systems whose control vector fields are Hamiltonian. We obtain decompositions which describe the controllable and uncontrollable dynamics. In each case, the dynamics are shown to be Hamiltonian in a suitably general sense.

Next we develop intrinsic descriptions of Lagrangian and Hamiltonian mechanics in the presence of external inputs. This development is a first step towards a control theory for general Lagrangian and Hamiltonian control systems. Systems with constraints are also studied. We first give a thorough overview of variational methods including a comparison of the "nonholonomic" and "vakonomic" methods. We also give a generalised definition for a constraint and, with this more general definition, we are able to give some preliminary controllability results for constrained systems.

Contents

List of Figures

List of Symbols

The symbols are listed in alphabetical order. Symbols with no alphabetical significance are listed first.

Symbol	: Description and page when applicable
$\sharp$	: the isomorphism from $\Omega^1(M)$ to $\mathfrak{X}(M)$ determined by a pseudo-Riemannian metric, 20
$\flat$	: the isomorphism from $\mathfrak{X}(M)$ to $\Omega^1(M)$ determined by a pseudo-Riemannian metric, 20
$\mathcal{E}$	: the set of sections of a vector bundle $\pi\colon E \to M$, 12
$\{f, g\}$	: the Poisson bracket of functions f and g, 23
0_x	: the point in the zero section of a vector bundle E corresponding to a point x in the base space, 12
$\otimes$	: the tensor product, 7
$\langle S\rangle_R$	: the submodule over R generated by the subset S of a module V
$[u, v]$	: the Lie bracket of elements u, v of a Lie algebra, 8
$\overset{\circ}{\bigcup\limits_{a\in A}} S_a$	: the disjoint union of the sets $\{S_a\}_{a\in A}$
$\alpha \wedge \beta$	: the wedge product of $\alpha, \beta \in \bigwedge(V)$, 7
U^0	: the subspace of V^* which annihilates $U \subset V$
$\bigwedge(V)$	: the set of skew-symmetric contravariant tensors on the vector space V, 7
$\bigwedge(V^*)$	: the set of skew-symmetric covariant tensors on the vector space V, 7
$\bigwedge^k(V)$	: the set of skew-symmetric k-contravariant tensors on the vector space V, 7
$\bigwedge^k(V^*)$	: the set of skew-symmetric k-covariant tensors on the vector space V, 7
$F^\perp$	: the set of vectors in V annihilated by $F \subset V^*$
$X \lrcorner\, \alpha$	: the interior product of the vector field X and the differential form α, 11
$A(X)$	: the free algebra generated by the set X, 8
$\mathrm{Aff}(TQ)$	: the bundle of affine subspaces of TQ, 123

$\mathrm{Aff}(V)$: the set of affine subspaces of the vector space V, 122
$\mathrm{Aff}^k(V)$: the set of k-dimensional affine subspaces of the vector space V, 122
$\beta(B)$: the symmetrisation of $B \in \mathrm{Br}(\boldsymbol{X})$, 47
$\mathrm{Br}(X)$: the subset of $L(X)$ consisting of brackets of elements in X, 9
C : the accessibility distribution, 32
C_0 : the strong accessibility distribution, 33
$C^2(q_1, q_2, [a,b], D, \gamma)$: the set of curves which connect q_1 and q_2 in time $[a,b]$ and satisfy the affine constraint (D,γ), 106
$C^\infty(M)$: the set of smooth $\mathbb{R}$-valued functions on a manifold M, 11
C_Σ : the characteristic distribution on a Poisson manifold with structure tensor Σ, 25
$\mathfrak{C}^0$: the set of all $\mathfrak{C}$-constraint forces, 124
$\mathcal{C}$: the accessibility algebra, 32
$\mathcal{C}_0$: the strong accessibility algebra, 33
$D_\mathcal{V}$: the distribution determined by a family of vector fields $\mathcal{V}$, 13
$\delta_a(B)$: the number of times the element X_a occurs in $B \in \mathrm{Br}(\boldsymbol{X})$, 9
$\boldsymbol{d}\alpha$: the exterior derivative of the differential form α, 11
$\boldsymbol{d}t$: the differential of the function t on both $J^m(\mathbb{R}, M)$ and $J^1(M, \mathbb{R})$, 27
$\mathfrak{E}(L)$: the set of equilibrium points for the Lagrangian L, 69
$\mathrm{Ev}(\phi)$: the homomorphism from $T(\mathbb{R}^X)$ to $\mathfrak{X}(M)$ determined by a bijection $\phi\colon X \to \mathcal{V}$, 13
$\mathrm{Ev}(\psi)$: the evaluation map for the symmetric product, 22
$\mathrm{Ev}_x(\phi)$: the linear map from $T(\mathbb{R}^X)$ to T_xM determined by a bijection $\phi\colon X \to \mathcal{V}$, 13
$\mathrm{Ev}_x(\psi)$: the evaluation map at $x \in M$ for the symmetric product, 22
F_c : an m-force field along the curve c, 96
F_c^* : a coforce field along the curve c, 97
F_L : the Lagrange force field, 100
$\mathcal{F}^*(\Lambda)$: the set of Λ-compatible coforces on Q, 96
$\mathcal{F}^m(\Lambda)$: the set of Λ-compatible m-forces on Q, 95
$\mathcal{F}_D$: the foliation corresponding to the integrable distribution D, 15
$G(T^*Q)$: the Grassmann bundle of T^*Q, 123
$G(V)$: the set of subspaces of the vector space V, 123
$G^k(V)$: the set of k-dimensional subspaces of the vector space V, 123
Γ^i_{jk} : Christoffel symbols for an affine connection in a given set of coordinates, 21
$\gamma_a(P)$: the number of times the element X_a occurs in $P \in \mathrm{Pr}(\boldsymbol{Y})$, 11

$\operatorname{grad} f$	:	the gradient of the function f, 20
$I(x)$	:	the subspace of $T^*_x M$ defined by the Pfaffian module $\mathcal{I}$, 16
$\mathcal{I}^{(i)}$	:	ith derived system, 17
I_Σ	:	the subbundle of $T^*(\mathbb{R}\times M\times\mathbb{R}^m)$ corresponding to the control system Σ on M, 38
$\mathcal{I}$	:	the differential ideal corresponding to the Pfaffian module $\mathcal{I}$, 16
$\mathcal{I}(x)$	:	the image of $\mathcal{I}$ in $\bigwedge(T^*_x M)$, 16
id_A	:	the identity map on the set A
$\mathcal{I}(L,F)$	:	the Pfaffian module on $J^1(\mathbb{R},Q)$ specified by the Lagrangian L and the 1-force field F, 103
$\mathcal{I}^{(\infty)}$	:	bottom derived system, 17
$\mathcal{I}_\Sigma$	:	the Pfaffian module on $\mathbb{R}\times M\times\mathbb{R}^m$ corresponding to the control system Σ, 38
$\mathcal{I}_{\Sigma,D}$	:	the Pfaffian module on $\mathbb{R}\times M\times\mathbb{R}^m$ generated by the control system Σ and the distribution D, 45
$\{\mathcal{I}\}$	:	the algebraic ideal corresponding to the Pfaffian module $\mathcal{I}$, 16
J	:	the classical functional in the calculus of variations, 108
$J^1(M,\mathbb{R})$	:	the set of 1 jets from M to $\mathbb{R}$, 28
$J^1(M,\mathbb{R})_{x,t}$	:	the set of 1 jets from M to $\mathbb{R}$ at $(x,t)\in M\times\mathbb{R}$, 28
$j^1 f$	:	defined by $x\mapsto[f]\in J^1(M,\mathbb{R})_{x,f(x)}$, 28
$J^m(\mathbb{R},M)$	:	the set of m-jets from $\mathbb{R}$ to M, 26
$J^m(\mathbb{R},M)_{t,x}$	:	the set of m-jets from $\mathbb{R}$ to M at $(t,x)\in\mathbb{R}\times M$, 26
$j^m c$	:	defined by $t\mapsto[c]_m\in J^m(\mathbb{R},M)_{t,c(t)}$, 27
$\ker(\phi)$	:	the kernel of the linear map ϕ
$L(X)$	:	the free Lie algebra generated by the set X, 9
Λ	:	a complete subset of T^*Q (also used to denote a general leaf of a foliation), 95
Λ_q	:	$\Lambda\cap T^*_q Q$ where Λ is a complete subset of T^*Q, 95
$\overline{\mathrm{Lie}}(\mathcal{V})$	:	the involutive closure of the family of vector fields $\mathcal{V}$, 13
X^{lift}	:	the vertical lift of the vector field X on Q, 50
$M(X)$	:	the free magma generated by the set X, 8
$M/\mathcal{F}$	:	the leaf space of a foliation $\mathcal{F}$ on M, 16
$\omega^1\equiv\omega^2 \bmod \mathcal{I}$	:	implies that $\omega^1-\omega^2\in\{\mathcal{I}\}$, 17
$\nabla_X Y$	:	the covariant derivative of Y with respect to X, 20
$\Omega(H,F^*)$	:	the two-form on $J^1(Q,\mathbb{R})$ which is defined by the Hamiltonian H and the coforce field F^*, 97
$\Omega(M)$	:	the exterior algebra of differential forms on the manifold M, 11
$\Omega^\flat$	:	the isomorphism from $\mathfrak{X}(P)$ to $\Omega^1(P)$ induced by the symplectic form Ω on P, 23

$\Omega^k(M)$	: the set of smooth k-forms on a manifold M, 11
$\Omega^\perp D$	: the skew-orthogonal complement of a distribution D, 22
$\Omega^\sharp$	: the isomorphism from $\Omega^1(P)$ to $\mathfrak{X}(P)$ induced by the symplectic form Ω on P, 23
p_Q	: the projection from $J^1(M,\mathbb{R})$ to T^*M, 29
$\pi_{1,0}$	: the projection from $J^1(M,\mathbb{R})$ to $M\times\mathbb{R}$, 28
$\pi_M\colon T^*M\to M$	: the cotangent bundle projection, 11
$\mathrm{Pr}(X)$	: the subset of $S(X)$ consisting of symmetric products of elements of X, 11
$\mathbb{R}$	: the set of real numbers
$\mathbb{R}^n$	: Euclidean n-space
$\mathbb{R}^X$	: the free vector space generated by the set X, 6
$\mathrm{range}(\phi)$	: the image of the linear map ϕ
$\mathcal{R}^U_Q(q_0,T)$	: the set of configurations reachable from q_0 in time exactly T, 68
$\mathcal{R}^U_Q(q_0,\leq T)$	: the set of configurations reachable from q_0 in time less than or equal to T, 68
$\mathcal{R}^V(x_0,\leq T)$	: the set of points reachable from x_0 in time less than or equal to T, 32
$\mathcal{R}^V(x_0,T)$	: the set of points reachable from x_0 in time exactly T, 32
$\rho(P)$	: the symmetrisation of $P\in\mathrm{Pr}(\boldsymbol{Y})$, 70
ρ_1^*	: the projection from $J^1(M,\mathbb{R})$ to M, 28
ρ_m	: the projection from $J^m(\mathbb{R},M)$ to M, 27
$s(U)$	: the subspace of V corresponding to the affine subspace U, 122
$S(X)$	: the free symmetric algebra generated by the set X, 11
$\mathbb{S}^n$	: the n-dimensional sphere
S_k	: the permutation group on k symbols, 7
$\Sigma^\sharp$	: the homomorphism from $\Omega^1(P)$ to $\mathfrak{X}(P)$ associated with the Poisson tensor Σ, 25
$\overline{\mathrm{Sym}}(\mathcal{V})$	: the symmetric closure of the family of vector fields $\mathcal{V}$, 22
$T(V)$	: the tensor algebra on the vector space V, 7
T^*M	: the cotangent bundle of a manifold M, 11
$\mathbb{T}^n$	: the n-dimensional torus
$\mathfrak{T}^r_s(M)$	: the set of smooth r-contravariant, s-covariant tensor fields on a manifold M, 11
$T^r_s(V)$	: the set of r-contravariant, s-covariant tensors on the vector space V, 7
$\tau_M\colon TM\to M$	: the tangent bundle projection, 11
$\tau_{m,l}$	: the projection from $J^m(\mathbb{R},M)$ to $J^l(\mathbb{R},M)$ for $m>l$, 27
TM	: the tangent bundle of a manifold M, 11

$T\phi\colon TM \to TN$	: the derivative of the map $\phi\colon M \to N$ between manifolds M and N, 11
$\mathrm{tr}(A)$	: the trace of the matrix A
$\mathcal{U}$	: the set of piecewise constant control inputs, 31
VM	: the vertical subbundle of a fibre bundle $\pi\colon M \to B$, 12
$\mathcal{V}_\Sigma$	: the family of vector fields corresponding to the control system Σ, 31
$\overline{\mathfrak{W}}$	: the set of all $\mathfrak{W}$-admissible forces, 124
$\mathfrak{X}(M)$	: the set of smooth vector fields on a manifold M, 11
$X_c(q_1, q_2, [a,b], D)$	: the subset of $T_cC^2(q_1, q_2, [a,b])$ consisting of virtual displacements, 107
X_L	: the Lagrangian vector field with Lagrangian L, 50
$\mathcal{Y}$	: the set of input vector fields $\{Y_1, \ldots, Y_m\}$, 63
$\mathbb{Z}$	: the set of integers
$Z(E)$	: the zero section of a vector bundle E, 12

Chapter 1

Introduction

Mechanical control systems form a large and interesting subset of all control systems. In spite of the proliferation of mechanical systems in the class of all control systems, very little fundamental work has been done to use the special structure of mechanical systems to build up a control theoretic tool bag which is suited to these systems. The structure in mechanical systems typically arises in two ways. In the Lagrangian framework, the structure is that of second-order dynamics on the tangent bundle of the configuration manifold. In the Hamiltonian setting, the structure is that of a symplectic manifold. Of particular interest in the class of Hamiltonian systems are those systems whose symplectic manifold is the cotangent bundle of the configuration manifold endowed with its canonical symplectic structure.

While one may view the work in this dissertation as an adaptation of the methods of nonlinear control theory to mechanical systems, one may also view it as an extension of the methods of geometric mechanics to systems with external inputs. The modern development of geometric mechanics has, for the most part, left out this important feature of mechanical systems. An example of work which *has* included external forces is that of (Yang, 1992). Another missing piece in geometric mechanics has been the inclusion of constraints in the formulation. This has received some recent attention in (Bloch *et al.*, 1994). A particularly interesting example of a system with constraints and inputs is the "Snakeboard" which was introduced in (Lewis *et al.*, 1994). In this example, one can ask interesting control theoretic questions which the existing tools are ill-suited to answering. Some initial results in the area may be found in (Ostrowski and Burdick, 1995).

It would be improper to give the impression that *no* work has been done in the area of mechanical control systems. In (Bloch and Crouch, 1992) a discussion is presented for mechanical control systems whose configuration space is a Riemannian manifold and whose Lagrangian is kinetic energy with respect to the Riemannian metric. With additional structure in the form of group symmetries and some assumptions on the inputs, a controllability result is given for this class of systems. The result relies on the work in (San Martin and Crouch, 1984) on controllability of systems on principal fibre bundles with compact structure group. These results, while interesting, lack generality since they require a priori knowledge of system symmetries. This knowledge is present in many systems, but in many more it is not.

There is a large body of work which is applicable to control problems whose control vector fields are the horizontal lifts of vector fields on the base space of a principal fibre bundle. A nice review of these results in the case where the bundle is trivial may be found in (Kelly and Murray, 1994). The discussion in that paper is geared towards controllability as it applies to locomotion. This motivated the authors to give two versions of controllability which they called *total controllability* and *fibre controllability*. The first corresponds to the usual notion of controllability, and the second is a weaker notion which does not take into account the final position in the base space. Thus locomotion problems are examples of systems which benefit from notions of controllability which are weaker than the standard versions from nonlinear control theory. We shall see this concept arise in mechanical systems as well. However, the work in (Kelly and Murray, 1994) does not consider important dynamical effects. Indeed, the class of problems studied is restricted exactly in such a manner that dynamics do not play a rôle.

In (Bloch *et al.*, 1992a) the stabilisability and controllability of mechanical systems with constraints is discussed. As kinematic systems (i.e., ones where the inputs are velocities rather than forces), systems with constraints may be viewed as members of a class of control systems known as "driftless" control systems. These systems are known to violate Brockett's necessary condition for stabilisability (see (Brockett, 1983)) and so cannot be stabilised under state feedback. In (Bloch *et al.*, 1992a) this is shown to also be the case when dynamics are taken into account. It is also shown in this paper that, with the assumption that forces are available from a set of forces which are complementary to the constraint forces, these systems are small-time locally controllable. We shall see in Section 6.2.2 that this is a very natural thing to expect.

Another area of research in control of mechanical systems that has received a great deal of attention is stabilisation of satellites and related problems. These problems have a configuration space which is a Lie group. Because of invariance of the mechanical properties with respect to the group action, it is often possible to reduce the system to the Lie algebra in the Lagrangian case, and to the dual of the Lie algebra in the Hamiltonian case. Some examples of work in this area are (Meyer, 1971; Jurdjevic and Sussmann, 1972; Krishnaprasad, 1985; Aeyels and Szafranski, 1988; Wang and Krishnaprasad, 1992; Bloch *et al.*, 1992b).

Below we outline the dissertation chapter-by-chapter and state what is new in each chapter.

Chapter 2

In Chapter 2 we give the necessary mathematical preliminaries. The purpose of this chapter is twofold. First, it serves to review the relevant areas of mathematics, and second, it is used to present various new or uncommon technical results which will be needed later. The most significant new object we introduce is a "symmetric algebra" which we shall use in Section 4.1 to discuss control theory for Lagrangian systems.

Chapter 3

Here we review some basic concepts from nonlinear control theory since a good understanding of these ideas is essential for a clear presentation of our results for mechanical systems. We present both the distribution and exterior differential systems viewpoints for representing control systems. In particular, we give precise statements of the conditions for local accessibility and strong local accessibility in terms of Pfaffian modules. These results are known, but, to our knowledge, have not appeared in the literature. We also give a thorough presentation of invariant distributions in this chapter. Some new results are presented for characterising integrable invariant distributions using exterior differential systems.

Chapter 4

In this chapter we present the main results of this dissertation. The aim is to generalise the basic ideas from nonlinear control theory presented in Chapter 3 to specific classes of mechanical control systems in both the Lagrangian and Hamiltonian framework.

In the Lagrangian framework we consider what we call "simple mechanical control systems." These systems are characterised primarily by having the Lagrangian be of the form "kinetic energy minus potential energy." We introduce a new notion of controllability in terms of the configuration space, as this is often what is most interesting. We are then able to determine computable conditions for our new version of controllability. Our computations rely in an interesting way on the structure of simple mechanical control systems. In particular, the covariant derivative with respect to the Riemannian metric which defines the kinetic energy plays an important rôle in our computations.

We also discuss Hamiltonian control systems which fully utilise the structure of the underlying symplectic manifold. The results we derive in this area are partially present in (Nijmeijer and van der Schaft, 1990). We give the results more structure by exploiting the various reductions which may be performed on these systems. We are able to generate some clean results for Hamiltonian control systems in this way.

To conclude the chapter, we go through the computations for a few examples and we see how the machinery relates to our intuition for the given problems.

Chapter 5

In this chapter, we discuss means of representing general mechanical systems in the presence of external forces, but in the absence of constraints. We present both the Lagrangian and Hamiltonian points of view as they are similar. Under regularity conditions, we are able to show that the two formulations are equivalent, generalising the classical results. We also introduce a new object which we call the "Lagrange force field." This object establishes Lagrange's equations as the components of a geometric entity. This chapter may be regarded as one where we establish a solid framework for future work which may be done in controlling fairly general

Lagrangian and Hamiltonian control systems.

Chapter 6

In this chapter results are presented for systems with constraints. We begin our discussion by introducing the variational principles associated with mechanical systems with constraints. When the constraints are absent, the variational formulation (Hamilton's Principle) is well-known and accepted as standard. However, when constraints are introduced, the variational formulation is less obvious as there are at least two viable ways to formulate a variational principle in this case. We investigate the "nonholonomic" and "vakonomic" methods and show that they are equivalent when the constraints are holonomic. A list of the pros and cons of the nonholonomic and vakonomic methods is given and, using this, we give strong arguments on behalf of the nonholonomic method being the proper way to represent constrained systems.

In Section 6.2 we present a general definition of constraints which takes up where we left off with external forces in Chapter 5. With these general notions of external forces and constraints, we are able to give some preliminary control theoretic results for systems with constraints. The Lagrange force field introduced in Section 5.6 is useful in establishing these results.

Chapter 7

In this chapter we summarise the new results in this dissertation and suggest some avenues for future work based upon these results.

Chapter 2

Mathematical Preliminaries

In this chapter we present the mathematical tools which will be useful in our discussion of mechanics and mechanical control systems. We begin with some algebraic concepts in Sections 2.1 and 2.2. Of particular interest here is the concept of a symmetric algebra introduced in Section 2.2.5. Some basic terminology from differential geometry is presented in Section 2.3. The intent here is to establish our notation. Then, in Section 2.4, we discuss some ideas from the theory of (geometric) distributions. This leads naturally to a discussion of exterior differential systems in Section 2.5. Next we discuss the mathematical structures which are important for describing mechanical control systems. In Section 2.7 we present the basic notions from Riemannian geometry as we will need them to analyse Lagrangian control systems. It is in this section that we introduce the symmetric product on the set of vector fields on a Riemannian manifold. This product becomes very important in determining conditions for controllability of Lagrangian systems in Section 4.1. We see the symmetric product as one of the most intriguing developments in this dissertation. Sections 2.8 and 2.9 are devoted to symplectic and Poisson manifolds, respectively. Both of these structures are useful in describing Hamiltonian mechanics. Finally, in Section 2.10, some concepts from the theory of jet bundles are presented. We shall use these ideas in formulating basic descriptions of mechanical systems in Chapter 5.

The following mathematical notation shall be used.

$\square$	: end of remark, example, or definition
$\blacksquare$	: end of proof
$\blacktriangledown$	: proof of subresult is done, but the proof of the main result continues
$a \in A$	: a is an element of the set A
$A \subset B$	: A is a subset of B (the same as $A \subseteq B$)
$A \setminus B$	: the points in A that are not in B
$A \cup B$	: the union of sets A and B
$A \cap B$	: the intersection of sets A and B

2.1 Algebra

We assume the reader to have a basic understanding of linear algebra. However, we shall need to establish some notation and present some ideas which will be useful later.

2.1.1 Algebras

We begin with introductory definitions concerning algebras. We will consider only objects over the field of real numbers although general definitions may be made over a commutative ring with unit. The algebra definitions come from, for example, (Lang, 1984).

An *algebra* is a vector space, A, with a product. The product must have the property that

$$a(uv) = (au)v = u(av)$$

for every $a \in \mathbb{R}$ and $u, v \in A$. A map, $\phi \colon A \to A'$, between algebras is called an *algebra homomorphism* if $\phi(u \cdot v) = \phi(u) \cdot \phi(v)$. A vector subspace, I, of an algebra A is called a *left ideal* (resp. *right ideal*) if it is closed under algebra multiplication and if $u \in A$ and $i \in I$ implies that $ui \in I$ (resp. $iu \in I$). A subspace, I, is said to be a *two-sided ideal* if it is both a left and right ideal. An ideal may not be an algebra itself, but the quotient of an algebra by a two-sided ideal inherits an algebra structure from A.

2.1.2 Free Vector Spaces

We will need the notion of a free vector space. Let X be a nonempty set. We define $\mathbb{R}^X$ to be the *free vector space* generated by X. It is the set of finite length, associative, and commutative sums of elements in X. Thus a typical element of $\mathbb{R}^X$ is formally written as

$$a^1 u_1 + \cdots + a^n u_n$$

for $a^1, \ldots, a^n \in \mathbb{R}$ and $u_1, \ldots, u_n \in X$. By definition of $\mathbb{R}^X$, for any vector space V and any map $\phi \colon X \to V$, there exists a unique linear map $\bar{\phi} \colon \mathbb{R}^X \to V$ which extends ϕ. Thus X forms a basis for $\mathbb{R}^X$.

2.1.3 The Tensor Algebra of a Vector Space

What we discuss in this section may be found in (Abraham *et al.*, 1988).

Let V be a $\mathbb{R}$-vector space with dual V^*. A *tensor* of contravariant order r and covariant order s is a multilinear map $t \colon V^* \times \cdots \times V^* \times V \times \cdots \times V \to \mathbb{R}$ (with r copies of V^* and s copies of V). We may define a product on the set of tensors as follows. Let t_1 be a tensor of contravariant order r_1 and covariant order s_1 and let t_2 be a tensor of contravariant order r_2 and covariant order s_2. We define a tensor

$t_1 \otimes t_2$ of contravariant order $r_1 + r_2$ and covariant order $s_1 + s_2$ by

$$t_1 \otimes t_2(\alpha_1, \dots, \alpha_{r_1+r_2}, u_1, \dots, \dots, u_{s_1+s_2}) = \\ t_1(\alpha_1, \dots, \alpha_{r_1}, u_1, \dots, u_{s_1}) \cdot t_2(\alpha_{r_1+1}, \dots, \alpha_{r_1+r_2}, u_{s_1+1}, \dots, u_{s_1+s_2}).$$

The set of r-contravariant, s-covariant tensors on V shall be denoted $T^r_s(V)$. The set of all tensors on V will be denoted by $T(V)$ and they form a $\mathbb{R}$-algebra with the product given by $\otimes$.

Certain subsets of $T(V)$ will be of particular interest to us. In particular, we mention that the set of contravariant tensors on an n-dimensional vector space V is isomorphic to the non-commutative polynomials over $\mathbb{R}$ in n variables. The isomorphism of these algebras is fixed by determining a basis for V. In particular, if $V = \mathbb{R}^X$ for an ordered set X of n elements, the isomorphism is natural. We shall need these notions in Section 2.2.3 when we discuss free Lie algebras.

The set of skew-symmetric k-covariant (resp. k-contravariant) tensors on V will be denoted by $\bigwedge^k(V^*)$ (resp. $\bigwedge^k(V)$). The set of all skew-symmetric covariant (resp. contravariant) tensors is denoted by $\bigwedge(V^*)$ (resp. $\bigwedge(V)$). We shall call elements of $\bigwedge^k(V^*)$ *k-forms* on V and elements of $\bigwedge^k(V)$ *k-multivectors* on V. Often $\bigwedge(V)$ is called the *tensor algebra* of V. We shall be performing similar operations on $\bigwedge^k(V^*)$ as on $\bigwedge^k(V)$ so we shall present the specifics for $\bigwedge^k(V^*)$ only.

On $\bigwedge^k(V^*)$ we may define a special product which preserves the skew-symmetry of these tensors. This product is called the *wedge product*. To define it we first define the *alternation mapping* on $T^0_k(V)$ as follows:

$$\boldsymbol{A}t(u_1, \dots, u_k) = \frac{1}{k!} \sum_{\sigma \in S_k} (\operatorname{sign} \sigma) t(u_{\sigma(1)}, \dots, u_{\sigma(k)}).$$

Here S_k is the permutation group on k symbols. Restricted to $\bigwedge^k(V^*)$, the mapping $\boldsymbol{A}$ is the identity. We now define the wedge product between $\alpha \in \bigwedge^k(V^*)$ and $\beta \in \bigwedge^l(V^*)$ by

$$\alpha \wedge \beta = \frac{(k+l)!}{k!l!} \boldsymbol{A}(\alpha \otimes \beta).$$

It may be verified that $\alpha \wedge \beta \in \bigwedge^{k+l}(V^*)$.

2.2 Lie Algebras and Symmetric Algebras

When studying control systems in Chapter 3, we will need some basic notions of Lie algebras. In particular, we will need the notion of a *free Lie algebra* and generators for this Lie algebra. In Section 4.1 we will need the notion of what we shall call a *symmetric algebra*.

2.2.1 Lie Algebra Definitions

We begin with introductory definitions concerning Lie algebras. We will consider only objects over the field of real numbers although general definitions may be made over a commutative ring with unit. The basic Lie algebra concepts are from (Serre, 1992).

A *Lie algebra* is an algebra, A, where the multiplication (usually denoted by $(u,v) \mapsto [u,v]$) has the following properties:

LA1. $[u,u] = 0$ for every $u \in A$, and

LA2. $[u,[v,w]] + [w,[u,v]] + [v,[w,u]] = 0$ for all $u, v, w \in A$.

The condition LA2 is typically referred to as *Jacobi's identity.* A subspace $E \subset A$ of a Lie algebra is called a *Lie subalgebra* if $[u,v] \in E$ for every $u, v \in E$. A map $\phi\colon A \to A'$ between Lie algebras is called a *Lie algebra homomorphism* if $\phi([u,v]) = [\phi(u), \phi(v)]$ for each $u, v \in A$.

2.2.2 Free Algebras

A *magma* is a set M with a map from $M \times M$ to M. We shall use "$\cdot$" to denote this map. Thus the image of (m_1, m_2) under the magma map is $m_1 \cdot m_2$. If M and N are magmas, a map $\phi\colon M \to N$ is called a *magma morphism* if $\phi(m_1 \cdot m_2) = \phi(m_1) \cdot \phi(m_2)$. If X is a set, we may generate the *free magma on* X as follows. Define $X_1 = X$ and inductively define $X_n = \coprod_{p+q=n} X_p \times X_q$ for $n \geq 2$. The free magma on X is the set

$$M(X) = \coprod_{n=1}^{\infty} X_n.$$

The map from $M(X) \times M(X)$ to $M(X)$ which makes this a magma is specified by $(m_1, m_2) \mapsto m_1 \cdot m_2$ where $m_1 \in X_p$, $m_2 \in X_q$, and $m_1 \cdot m_2 \in X_{p+q}$ by the inclusion of $X_p \times X_q$ in X_{p+q} specified in the construction. The name *free* comes from the fact that the image of the magma $M(X)$ in another magma is uniquely determined by the image of the set $X \subset M(X)$. Note that any $u \in M(X) \setminus X$ may be uniquely written as $u = v \cdot w$ for some $v, w \in M(X)$. Also note that each $u \in M(X)$ is in X_n for some uniquely defined positive integer n. We shall call n the *length* of u.

Now we define the *free algebra* associated with a set X. This $\mathbb{R}$-algebra is denoted $A(X)$ and consists of all finite linear combinations

$$\sum_{m \in M(X)} a_m m$$

where $a_m \in \mathbb{R}$. The product of two elements in $A(X)$ is given by

$$\left(\sum_{m \in M(X)} a_m m \right) \cdot \left(\sum_{m \in M(X)} b_m m \right) = \sum_{m_1 \cdot m_2 \in M(X)} a_{m_1} b_{m_2} m_1 \cdot m_2.$$

2.2.3 Free Lie Algebras

To construct the free Lie algebra generated by X, let I be the two-sided ideal of $A(X)$ generated by elements of the form $a \cdot a$ and $a \cdot (b \cdot c) + c \cdot (a \cdot b) + b \cdot (c \cdot a)$ for $a, b, c \in A(X)$. The *free Lie algebra* generated by X is the quotient algebra, $L(X) = A(X)/I$. The inherited product on $L(X)$ is typically denoted by $[\cdot, \cdot]$. We denote by $\mathrm{Br}(X)$ the subset of $L(X)$ containing products of elements in X. This subset generates $L(X)$ as a $\mathbb{R}$-vector space. However, it is not a linearly independent subset since, for example, $[u, v] = -[u, v]$ for each $u, v \in L(X)$. In Section 2.2.4 we construct sets of generators which are contained in $\mathrm{Br}(X)$.

The set X is canonically included in the free vector space $\mathbb{R}^X$. In turn, $\mathbb{R}^X$ in canonically included in the tensor algebra $T(\mathbb{R}^X)$. Therefore we have a canonical inclusion of X in $T(\mathbb{R}^X)$. This inclusion induces a magma morphism from $L(X)$ to $T(\mathbb{R}^X)$ which is, in fact, an algebra homomorphism. In (Serre, 1992) it is shown that the image of $L(X)$ under this homomorphism is the algebra of multivectors described in Section 2.1.3. It may be shown that $T(\mathbb{R}^X)$ is isomorphic to the *universal enveloping algebra* of $L(X)$.

We will need the notion of what we shall call the components of an element $u \in L(X)$. Every such element u has a unique decomposition as $u = [u_1, u_2]$. In turn, each of u_1 and u_2 may be uniquely expressed as $u_1 = [u_{11}, u_{12}]$ and $u_2 = [u_{21}, u_{22}]$. This process may be continued until we end up with elements whose lengths are one. All such elements $u_{i_1 \cdots i_m}$, $i_a \in \{1, 2\}$, shall be called *components* of u.

Of special interest to us is the case where the set X is finite. We shall denote $\boldsymbol{X} = \{X_0, \ldots, X_l\}$ as a finite set with $l + 1$ elements. In this case we develop some extra notation. Let $B \in \mathrm{Br}(\boldsymbol{X})$. We define $\delta_a(B)$ to be the number of times the element X_a occurs in B for $a = 0, \ldots, l$. The *degree* of B is the sum the δ_a's.

2.2.4 Generators for Free Lie Algebras

We will find it helpful to write down a generating set for $L(X)$. It is possible to determine linearly independent generating sets, called *Philip Hall bases* in the literature (Serre, 1992). However, we shall not need such sophisticated techniques and it is good enough to just determine a generating set without the condition that it be linearly independent.

We shall present two methods for determining generators for the free Lie algebra $L(X)$.

Proposition 2.1 *Every element of $L(X)$ is a linear combination of repeated brackets of the form*

$$[X_k, [X_{k-1}, [\cdots, [X_2, X_1] \cdots]]] \tag{2.1}$$

where $X_i \in X$, $i = 1, \ldots, k$.

Proof: Denote by $\bar{L}(X)$ the subspace of $L(X)$ generated by brackets of the form (2.1). It is clear that $\bar{L}(X) \subset L(X)$ by definition. Also, $X \subset \bar{L}(X)$. Thus, to show that $\bar{L}(X) = L(X)$, we need only show that $\bar{L}(X)$ is a subalgebra of $L(X)$

since $L(X)$ is the smallest subalgebra containing X. Note that k in (2.1) is the *degree* of the expression. Now consider two such expressions of degree j and l,

$$U = [U_j, [U_{j-1}, [\cdots, [U_2, U_1] \cdots]]] \tag{2.2a}$$
$$V = [V_l, [V_{l-1}, [\cdots, [V_2, V_1] \cdots]]]. \tag{2.2b}$$

We shall prove by induction that $[U, V] \in \bar{L}(X)$ for any j and l. Note that $[U, V] \in \bar{L}(X)$ for all V and l, and for $j = 1$. Now suppose this is true for $j = 1, \ldots, k$. Then, taking $j = k + 1$ in (2.2a), we have

$$[U, V] = [[U_{k+1}, U^1], V]$$

where $U^1 = [U_{j-1}, [\cdots, [U_2, U_1] \cdots]]$. By the Jacobi identity we have

$$[[U_{k+1}, U^1], V] + [[V, U_{k+1}], U^1] + [[U^1, V], U_{k+1}] = 0.$$

This gives

$$[U, V] = [U^1, [V, U_{k+1}]] + [U_{k+1}, [U^1, V]].$$

By the induction hypothesis, $[U^1, [U_{k+1}, V]] \in \bar{L}(X)$ since the degree of U^1 is k. Also $[U^1, V] \in \bar{L}(X)$ so the second term is in $\bar{L}(X)$. Thus $\bar{L}(X)$ is a subalgebra and hence $\bar{L}(X) = L(X)$. ■

Another method of constructing a generating set for $L(X)$ is given by the following proposition.

Proposition 2.2 *For $k \in \mathbb{Z}^+$ define $L_k(X)$ to be the subset of* $\mathrm{Br}(X)$ *given by all brackets of the form (2.1). Then every element of $L(X)$ is a linear combination of repeated brackets of the form*

$$[Z_k, [Z_{k-1}, [\cdots, [Z_2, Z_1] \cdots]]] \tag{2.3}$$

where $Z_i \in L_j(X)$ with $j \leq i$, $i = 1, \ldots, k$.

Proof: It is clear that $\bar{L}(X)$ is a subset of the set of brackets given by (2.3). The proposition then follows from Proposition 2.1. ■

2.2.5 Symmetric Algebra Definitions

As far as we know, the idea of a symmetric algebra does not appear in the literature. However, the concept is a very natural one and shall be useful to us.

A *symmetric algebra* is an algebra, A, where the multiplication (which we shall denote by $(u, v) \mapsto \langle u : v \rangle$) is symmetric. Thus $\langle u : v \rangle = \langle v : u \rangle$ for $u, v \in A$. A map, $\phi\colon A \to A'$, between symmetric algebras is called a *symmetric algebra homomorphism* if $\phi(\langle u : v \rangle) = \langle \phi(u) : \phi(v) \rangle$ for each $u, v \in A$.

2.2.6 Free Symmetric Algebras

In this section we construct a symmetric algebra which is generated by a given set X. To construct this algebra, let X be a set and recall that $A(X)$ is the free algebra on X. The *free symmetric algebra* on X, denoted $S(X)$, is the quotient algebra obtained by taking the quotient of $A(X)$ by the two-sided ideal generated by all elements of the form $a \cdot b - b \cdot a$ where $a, b \in A(X)$. We shall denote the product in $S(X)$ by $\langle u : v\rangle$. Note that, by construction, $\langle u : v\rangle = \langle v : u\rangle$ for every $u, v \in S(X)$. We denote by $\mathrm{Pr}(X)$ the subset of $S(X)$ consisting of the symmetric products whose elements are in X.

As with free Lie algebras, the finitely generated case is the most interesting to us. Let $\boldsymbol{Y} = \{X_1, \ldots, X_{l+1}\}$ (the reason for the slightly unusual enumeration will become clear in Section 4.1.7). For $P \in \mathrm{Pr}(\boldsymbol{Y})$ define $\gamma_a(P)$ to be the number of times the element X_a occurs in $P \in \mathrm{Pr}(\boldsymbol{Y})$ for $a = 1, \ldots, l+1$. We shall call the sum of the γ_a's the *degree* of P.

2.3 Differential Geometry

A basic understanding of differential geometry is assumed. In this section we first quickly review the notation which will be used. Then we discuss some concepts from fibre bundle theory which will be useful to us.

The manifolds we deal with in this dissertation will be assumed to belong to the C^∞ category. We shall further suppose that all manifolds are finite-dimensional, paracompact, and Hausdorff unless otherwise stated. For the most part, the notation we use is from (Abraham *et al.*, 1988). The tangent bundle of a manifold M is denoted TM and the cotangent bundle by T^*M. The tangent bundle and cotangent bundle projections are denoted by $\tau_M : TM \to M$ and $\pi_M : T^*M \to M$, respectively. If $\phi \colon M \to N$ is a smooth mapping from a manifold M to a manifold N, we will denote its derivative by $T\phi \colon TM \to TN$. The set of all smooth mappings from M to N will be denoted $C^\infty(M, N)$. We reserve special notation for the case when $N = \mathbb{R}$. In this case, $C^\infty(M)$ denotes the set of real-valued smooth functions on M.

The set of r-contravariant, s-covariant tensor fields on M is denoted by $\mathfrak{T}^r_s(M)$. Elements of $\mathfrak{T}^1_0(M)$ are called vector fields on M and we shall denote the set of vector fields on M by $\mathfrak{X}(M)$. The set of vector fields forms a Lie algebra when equipped with the Lie bracket which we shall denote by $[\cdot,\cdot] \colon \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$. The skew-symmetric k-covariant tensors are also interesting and are given the name k-forms. We shall denote the set of k-forms on M by $\Omega^k(M)$. By convention we take $\Omega^0(M) = C^\infty(M)$. We denote by $\Omega(M) \triangleq \oplus_{k=0}^{\infty} \Omega^k(M)$ the entire exterior algebra on M. This is made an exterior algebra by the wedge product which may be taken between any two elements of $\Omega(M)$. We denote the wedge product of $\alpha, \beta \in \Omega(M)$ by $\alpha \wedge \beta$. The exterior algebra of differential forms also comes equipped with the exterior derivative which we denote by $\boldsymbol{d}$. Recall that the exterior derivative of a k-form is a $(k+1)$-form. We say that a k-form α is *closed* if $\boldsymbol{d}\alpha = 0$ and *exact* if $\alpha = \boldsymbol{d}\beta$ for some $(k-1)$-form β. Given a vector field X and a k-form α, the interior product of X and α is a $(k-1)$-form which we denote by $X \lrcorner\, \alpha$.

Now we discuss some basic notions for fibre bundles. A *fibre bundle* is given by a surjective submersion $\pi\colon M \to B$ which has the property of being locally trivial. Thus, there exists a manifold F such that, for each point $b \in B$, there exists a neighborhood U of b and a diffeomorphism $\phi\colon \pi^{-1}(U) \to U \times F$. The diffeomorphism ϕ must have the further property that $\pi \circ \phi^{-1} \mid U = \mathrm{id}_U$. The *vertical subbundle* for a fibre bundle $\pi\colon M \to B$ is the subbundle of TM defined by $VM = \ker(T\pi)$. We shall call a vector field on M *vertical* if it takes its values in VM. In a similar manner we define the *horizontal subbundle* of T^*M to be the subbundle H^*M which annihilates VM. A one-form on M will be called *horizontal* if it takes its values in H^*M.

A special class of fibre bundles are *vector bundles* whose fibres have a vector space structure. A *section* of a vector bundle $\pi\colon E \to M$ is a map $\gamma\colon M \to E$ so that $\pi \circ \gamma = \mathrm{id}_M$. The set of sections of a vector bundle E will typically denoted by $\mathcal{E}$. If $\pi\colon E \to M$ is a vector bundle, then M can be naturally realised as a submanifold of E by identifying $m \in M$ with the zero vector in $\pi^{-1}(m)$. We will denote this submanifold by $Z(E)$ and call it the *zero section* of E. For each point $x \in M$, we denote by 0_x the corresponding point in the zero section of E.

2.4 Distributions and Foliations

We will at times need to be fairly precise about some concepts from the theory of (geometric) distributions. In this section we present the relevant concepts in some detail.

2.4.1 Distributions

Here we present the basic definitions for distributions.

Definition 2.3 Let M be a differentiable manifold. A *distribution* on M is a subbundle of TM. We shall call the dimension of $D(x)$ over $\mathbb{R}$ the *rank* of D at x. A distribution D is said to be *involutive* if $[X, Y] \in \mathcal{D}$ for each $X, Y \in \mathcal{D}$. A function $f \in C^\infty(M)$ is called an *integral* of D if $\boldsymbol{d}f(x) \in D^0(x)$ for each $x \in M$. An *integral manifold* of D is a submanifold, N, of M so that $T_xN \subset D(x)$ for each $x \in N$. A distribution is said to be *integrable* if, for each $x \in M$, there exists an integral manifold N of D through x whose dimension is the same as the rank of D at x. We shall call such a submanifold the *maximal integral manifold* through x. □

Frobenius' Theorem asserts that involutivity and integrability of a distribution are equivalent notions, at least locally.

We shall often ask that a distribution have *constant rank* by which we mean that its rank be a function independent of $x \in M$.

2.4.2 Distributions Generated by a Family of Vector Fields

A common way to arrive at a distribution is via a family of vector fields. A *family of vector fields* on a differentiable manifold M is simply a subset $\mathcal{V} \subset \mathfrak{X}(M)$. Given a family of vector fields $\mathcal{V}$, we may define a distribution on M by

$$D_{\mathcal{V}}(x) = \langle X(x) \mid \quad X \in \mathcal{V} \rangle_{\mathbb{R}}.$$

Since $\mathfrak{X}(M)$ is a Lie algebra, we may ask for the smallest Lie subalgebra of $\mathfrak{X}(M)$ which contains a family of vector fields $\mathcal{V}$. This will be the set of vector fields on M generated by repeated Lie brackets of elements in $\mathcal{V}$. It is most convenient to describe this subalgebra using the ideas from free Lie algebras presented in Section 2.2.3.

Let X be a set which is bijective to $\mathcal{V}$. Thus each element of X is in 1–1 correspondence with a vector field in $\mathcal{V}$. Recall that $T(\mathbb{R}^X)$ is the tensor algebra of the free vector space on X. Thus each element of $T(\mathbb{R}^X)$ is an associative, but not necessarily commutative, product of finite linear combinations of elements from X. Given a bijection $\phi \colon X \to \mathcal{V}$, we may define a $\mathbb{R}$-algebra homomorphism from $T(\mathbb{R}^X)$ to $\mathfrak{X}(M)$ by "plugging in" the vector field $\phi(u)$ for the element $u \in X$ in expressions in $T(\mathbb{R}^X)$. The map is explicitly given by

$$\begin{aligned} \mathrm{Ev}(\phi) \colon \; T(\mathbb{R}^X) &\to \mathfrak{X}(M) \\ u_1 \otimes \cdots \otimes u_k &\mapsto \phi(u_1) \circ \cdots \circ \phi(u_k). \end{aligned}$$

Here we are using the algebra structure on $\mathfrak{X}(M)$ given by its being the set of derivations on $C^\infty(M)$. Since elements of $L(X)$ may be regarded naturally as elements of $T(\mathbb{R}^X)$, the map $\mathrm{Ev}(\phi)$ restricts to $L(X)$ and so defines a Lie algebra homomorphism from $L(X)$ to $\mathfrak{X}(M)$.

The smallest Lie subalgebra of $\mathfrak{X}(M)$ which contains $\mathcal{V}$ may now be stated in a simple manner. It is simply the image of $L(X)$ under the homomorphism $\mathrm{Ev}(\phi)$. We shall denote this subalgebra by $\overline{\mathrm{Lie}}(\mathcal{V})$ and call it the *involutive closure* of $\mathcal{V}$.

For $x \in M$ we define the map $\mathrm{Ev}_x(\phi) \colon T(\mathbb{R}^X) \to T_xM$ by

$$\mathrm{Ev}_x(\phi)(u) = (\mathrm{Ev}(\phi)(u))(x).$$

We shall say that $\mathcal{V}$ satisfies the *Lie algebra rank condition* (LARC) at x if $\mathrm{Ev}_x(\phi)(L(X)) = T_xM$.

It is often helpful to be able to compute generators for $\overline{\mathrm{Lie}}(\mathcal{V})$, so we shall present two common ways of doing this. The first construction goes as follows. Let $\mathcal{V}^{(0)} = \mathcal{V}$ and then iteratively define a sequence of families of vector fields by

$$\mathcal{V}^{(i+1)} = \mathcal{V}^{(i)} \cup \{[X, Y] \mid \quad X \in \mathcal{V} \text{ and } Y \in \mathcal{V}^{(i)}\}.$$

First we show that this does indeed generate $\overline{\mathrm{Lie}}(\mathcal{V})$. The following result is proved in (Nijmeijer and van der Schaft, 1990). However, we have essentially proved this result in Section 2.2.4.

Proposition 2.4 *Every element of* $\overline{\mathrm{Lie}}(\mathcal{V})$ *is a linear combination of vector fields of the form*

$$[Z_k, [Z_{k-1}, [\cdots, [Z_2, Z_1] \cdots]]]$$

where $Z_i \in \mathcal{V}$, $i = 1, \ldots, k$.

Proof: Follows the same method as the proof of Proposition 2.1. ■

This shows that we may use our first iterative procedure to make a set of generators for $\overline{\mathrm{Lie}}(\mathcal{V})$.

Now we present another method of producing a set of generators. This method will enable us to make connections between Pfaffian modules and families of vector fields in Section 2.5.2. In this construction we define $\overline{\mathcal{V}}^{(0)} = \mathcal{V}$ and iteratively define

$$\overline{\mathcal{V}}^{(i+1)} = \overline{\mathcal{V}}^{(i)} \cup \{[X,Y] \mid \quad X, Y \in \overline{\mathcal{V}}^{(i)}\}.$$

We may show that this procedure also generates $\overline{\mathrm{Lie}}(\mathcal{V})$.

Proposition 2.5 *Every element of* $\overline{\mathrm{Lie}}(\mathcal{V})$ *is a linear combination of repeated Lie brackets of the form*

$$[Z_k, [Z_{k-1}, [\cdots, [Z_2, Z_1] \cdots]]]$$

where $Z_i \in \overline{\mathcal{V}}^{(i)}$, $i = 1, \ldots, k$.

Proof: Follows along the lines of the proof of Proposition 2.2. ■

This verifies that our two methods of constructing generators for $\overline{\mathrm{Lie}}(\mathcal{V})$ are equivalent.

We may define a nested sequence of distributions

$$D_{\mathcal{V}} = D_{\mathcal{V}}^{(0)} \supset \cdots \supset D_{\mathcal{V}}^{(i)} \supset \cdots$$

on M. If it is the case that each of these distributions is of constant rank, it is then easy to see that this sequence of distributions must terminate at some finite integer. We will denote the largest distribution generated in this way by $D_{\mathcal{V}}^{(\infty)}$. We may think of this distribution as being the smallest integrable distribution on M which contains $D_{\mathcal{V}}$.

In a similar manner we may construct a sequence of distributions

$$D_{\overline{\mathcal{V}}} = D_{\overline{\mathcal{V}}}^{(0)} \supset \cdots \supset D_{\overline{\mathcal{V}}}^{(i)} \cdots$$

which, when each has constant rank, will terminate in a distribution which is denoted by $D_{\overline{\mathcal{V}}}^{(\infty)}$. It is clear that $D_{\mathcal{V}}^{(\infty)} = D_{\overline{\mathcal{V}}}^{(\infty)}$ when both are defined.

We shall call a distribution D *controllable* if the sequence $D_{\mathcal{D}}^{(0)} \supset \cdots \supset D_{\mathcal{D}}^{(i)} \supset \cdots$ terminates in a finite number of steps at TM. The following proposition justifies this terminology.

Proposition 2.6 *Let* D *be a controllable distribution on a connected manifold. Then, for each* $x_1, x_2 \in M$, *there exists a piecewise differentiable curve,* $c\colon [0,T] \to M$, *so that* $c(0) = x_1$, $c(T) = x_2$, *and* $c'(t) \in D(c(t))$ *for each* $t \in [0,T]$.

Proof: Let $x \in M$ and let U be a sufficiently small neighborhood of x. We shall construct a sequence of submanifolds of U, $N_1, \ldots, N_n$ where $\dim(N_j) = j$. Since $\operatorname{rank}(D(x)) \neq 0$ we may choose $X_1 \in \mathcal{D}$ so that $X_1(x) \neq 0$. For $\epsilon_1 > 0$ sufficiently small,

$$N_1 \triangleq \{X_1^{t_1}(x) \mid \ 0 < t_1 < \epsilon_1\}$$

is a submanifold of M of dimension 1 which is contained in U. Here $X_1^{t_1}$ is the flow of X_1. We now construct N_j for $j > 1$ by induction. Suppose that $N_{j-1} \subset U$ is given by

$$N_{j-1} = \{X_{j-1}^{t_{j-1}} \circ \cdots \circ X_1^{t_1}(x) \mid \ 0 \leq \sigma_i < t_i < \epsilon_i, \ \ i = 1, \ldots, j-1\}.$$

Here X_i, $i = 1, \ldots, j-1$ are vector fields in $\mathcal{D}$ and $\sum_{i=1}^{j-1} \sigma_i$ is sufficiently small. If $j-1 < n$ we may find $X_j \in \mathcal{D}$ and $x' \in N_{j-1}$ so that $X_j(x') \notin T_{x'}N_{j-1}$. If this were not possible then this would violate the assumption that $\operatorname{rank}(D_{\mathcal{D}}^{(\infty)}) = n$ in U. For the same reason we may choose x' as close to x as we like. Thus the map

$$(t_j, \ldots, t_1) \mapsto X_j^{t_j} \circ X_{j-1}^{t_{j-1}} \circ \cdots \circ X_1^{t_1}(x)$$

has rank j for $0 \leq \sigma_i < t_i < \epsilon_i$ for $i = 1, \ldots, j$. Therefore, the image of this map is a j-dimensional submanifold of W for ϵ_i sufficiently small. We may continue this process until $n = j$ at which time it will terminate. Observe that N_n is a non-empty open subset of M and all points in N_n are reachable from x.

Now note that if $X \in \mathcal{D}$ then $-X \in \mathcal{D}$. For $(s_1, \ldots, s_n)$ which satisfy the relation $\sigma_i < s_i < \epsilon_i$, $i = 1, \ldots, n$, consider the map

$$(t_n, \ldots, t_n) \mapsto (-X_1)^{s_1} \circ \ldots \circ (-X_n)^{s_n} \circ X_j^{t_j} \circ X_{j-1}^{t_{j-1}} \circ \cdots \circ X_1^{t_1}(x).$$

Since $(-X_i)^{s_i} = X_i^{-s_i}$, the image of this map must contain a neighborhood of x since x is clearly in the interior of the image. Thus we have shown that we may reach a neighborhood of x from x.

Let $\mathcal{R}(x)$ be the set of points reachable from x. This set is open by our above calculations. Now suppose that $\mathcal{R}(x) \subsetneq M$ and let y be a point on the boundary of $\mathcal{R}(x)$. Clearly $\mathcal{R}(y)$ contains a neighborhood of y. Hence $\mathcal{R}(y) \cap \mathcal{R}(x) \neq \emptyset$ which contradicts y being a boundary point for $\mathcal{R}(x)$. This completes the proof. ■

2.4.3 Foliations

Related to integrable distributions are foliations. Without getting too involved with the technical definition, a *foliation*, $\mathcal{F}$, of a differentiable manifold M is a collection of disjoint immersed submanifolds of M whose disjoint union equals M. We call each connected submanifold of $\mathcal{F}$ a *leaf* of the foliation. Given an integrable distribution D, the collection of maximal integral manifolds for D defines a foliation of M. We shall denote this foliation by $\mathcal{F}_D$.

A foliation, $\mathcal{F}$, of M defines an equivalence relation on M whereby two points in M are equivalent if they lie in the same leaf of $\mathcal{F}$. The set of equivalence classes is denoted $M/\mathcal{F}$ and will be called the *leaf space* of $\mathcal{F}$. A foliation $\mathcal{F}$ is said to be *simple* if $M/\mathcal{F}$ inherits a manifold structure so that the projection from M to $M/\mathcal{F}$ is a surjective submersion.

2.5 Exterior Differential Systems

When we discuss our formulations for mechanical systems, we shall call on the tools of exterior differential systems. A discussion of these techniques may be found in (Bryant *et al.*, 1991).

2.5.1 Pfaffian Modules

We shall be interested primarily in particular types of exterior differential systems, namely those which are generated by Pfaffian modules. In this situation one is interested in a submodule of $\Omega^1(M)$. We do not deal with Pfaffian *systems* since these do not naturally arise in the applications we encounter.

Definition 2.7 Let M be a differentiable manifold. A *Pfaffian module* on M is a submodule, $\mathcal{I}$, of $\Omega^1(M)$. We denote by $\{\mathcal{I}\}$ the subset of $\Omega(M)$ given by

$$\{\mathcal{I}\} = \left\{ \omega \in \Omega(M) \mid \quad \omega = \sum_{i=1}^{k} \alpha^i \wedge \theta^i \text{ for } \alpha^1, \ldots, \alpha^k \in \mathcal{I} \text{ and } \theta^1, \ldots, \theta^k \in \Omega(M) \right\}.$$

We shall call $\{\mathcal{I}\}$ the *algebraic ideal* corresponding to $\mathcal{I}$. We denote by $\mathscr{I}$ the smallest submodule of $\Omega(M)$ containing $\mathcal{I}$ which is closed under exterior differentiation. We will call $\mathscr{I}$ the *differential ideal* corresponding to $\mathcal{I}$. □

Given a Pfaffian module $\mathcal{I}$, we may define a subbundle of T^*M, or a *codistribution* on M, by

$$I(x) = \langle \alpha(x) \mid \quad \alpha \in \mathcal{I} \rangle_{\mathbb{R}}.$$

It may be shown that $\mathscr{I}$ is the *algebraic* ideal generated by the set $\{\alpha, \boldsymbol{d}\alpha \mid \quad \alpha \in \mathcal{I}\}$. We will denote by $\mathscr{I}(x)$ the algebraic ideal of $\bigwedge(T_x^*M)$ generated by $\{\alpha(x), \boldsymbol{d}\alpha(x) \mid \alpha \in \mathcal{I}\}$. It is the differential ideal $\mathscr{I}$ that is the actual exterior differential system. Since we are dealing with the particular case of Pfaffian modules, things simplify somewhat.

Now we turn to defining integral manifolds of a Pfaffian module.

Definition 2.8 Let $\mathcal{I}$ be a Pfaffian module on M. A submanifold N of M is called an *integral manifold* of $\mathcal{I}$ if $T_xN \subset I(x)^{\perp}$ for every $x \in N$. A curve $c\colon I \to M$ is called an *integral curve* of $\mathcal{I}$ if $c'(t) \in I(c(t))^{\perp}$ for every $t \in I$. □

Corresponding to a Pfaffian module $\mathcal{I}$, we have a distribution, D, on M defined by $D(x) = I(x)^{\perp}$. Thus we may speak of integrability of Pfaffian modules.

Definition 2.9 Let $\mathcal{I}$ be a Pfaffian module on M and let D be the corresponding distribution on M. We say that $\mathcal{I}$ is *integrable* if D is integrable. $\square$

For $\omega^1, \omega^2 \in \Omega(M)$, we say that $\omega^1 \equiv \omega^2 \bmod \mathcal{I}$ if $\omega^1 - \omega^2 \in \{\mathcal{I}\}$.

2.5.2 The Derived Flag for a Pfaffian Module

Now we turn to the derived flag which will be an important tool when we discuss control theory. Denote $\mathcal{I}^{(0)} \triangleq \mathcal{I}$, and define

$$\mathcal{I}^{(1)} = \{\omega \in \mathcal{I} \mid \ \boldsymbol{d}\omega \equiv 0 \bmod \mathcal{I}\}.$$

In this way we can inductively define a sequence of Pfaffian modules, called the *derived flag*, denoted

$$\mathcal{I} = \mathcal{I}^{(0)} \supset \cdots \mathcal{I}^{(i)} \supset \cdots .$$

If $\mathcal{I}^{(i)}$ is the set of sections of a constant rank subbundle, $I^{(i)} \subset T^*M$, for each i, then the sequence can be shown to terminate for some integer N called the *derived length*. We may think of $I^{(N)}$ as the smallest integrable codistribution contained in I. We shall call $\mathcal{I}^{(i)}$ the ith *derived system*, and we denote by $\mathcal{I}^{(\infty)}$ the Pfaffian module which generates the smallest integrable codistribution when this is defined. We call $\mathcal{I}^{(\infty)}$ the *bottom derived system*.

The following result makes connections with the sequences of families of vector fields considered in Section 2.4.2.

Lemma 2.10 *Let $\mathcal{I}$ be a Pfaffian module on M and let $D = I^{\perp}$ be the corresponding distribution on M. Then D is controllable if and only if the bottom derived system of $\mathcal{I}$ is zero.*

Proof: By Proposition 2.6 the lemma will be proved if we can show that

$$(\mathcal{D} + [\mathcal{D}, \mathcal{D}])^0 = \{\omega \in \mathcal{I} \mid \ \boldsymbol{d}\omega = 0 \bmod \mathcal{I}\}.$$

Suppose that $\beta \in (\mathcal{D} + [\mathcal{D}, \mathcal{D}])^0$. Then

$$\begin{aligned} & \beta([X,Y] + Z) = 0 && \forall\ X, Y, Z \in \mathcal{D} \\ \Longrightarrow\ & \beta([X,Y]) + \beta(Z) = 0 && \forall\ X, Y, Z \in \mathcal{D} \\ \Longrightarrow\ & \beta([X,Y]) = 0 && \forall\ X, Y \in \mathcal{D}. \end{aligned}$$

Now we use the formula

$$\boldsymbol{d}\beta(X,Y) = X \cdot \beta(Y) - Y \cdot \beta(X) + \beta([X,Y])$$

for $X, Y \in \mathfrak{X}(M)$. If we choose $X, Y \in \mathcal{D}$ we obtain

$$\beta \in (\mathcal{D} + [\mathcal{D}, \mathcal{D}])^0 \quad \forall\ X, Y \in \mathcal{D} \quad \Longrightarrow \quad \boldsymbol{d}\beta(X,Y) = 0 \quad \forall\ X, Y \in \mathcal{D}.$$

Now choose a basis, $\{\omega^1, \ldots, \omega^k, \omega^{k+1}, \ldots, \omega^n\}$, for $\Omega^1(M)$ over $C^\infty(M)$ (perhaps only locally) so that the first k elements form a basis for $\mathcal{I}$. Then we have

$$\boldsymbol{d}\beta = B_{ij}\omega^i \wedge \omega^j \tag{2.4}$$

for some skew-symmetric matrix $\boldsymbol{B}$. Now choose $\{\xi_1, \ldots, \xi_n\}$ to be the basis for $\mathfrak{X}(M)$ over $C^\infty(M)$ dual to the given basis for $\Omega^1(M)$. Then we must have

$$\boldsymbol{d}\beta(\xi_l, \xi_l) = B_{ij}(\omega^i(\xi_l)\omega^j(\xi_m) - \omega^i(\xi_m)\omega^j(\xi_l)) = B_{lm} - B_{ml} = 0$$

for $l, m = k+1, \ldots, n$. Since $\boldsymbol{B}$ is skew-symmetric this means that $B_{lm} = 0$ for $l, m = k+1, \ldots, n$. Therefore, each term in the sum in (2.4) contains an element of $\{\omega^1, \ldots, \omega^k\}$. In other words $\boldsymbol{d}\beta = 0 \bmod \mathcal{I}$.

Now suppose that $\beta \in \mathcal{I}$ is such that $\boldsymbol{d}\beta = 0 \bmod \mathcal{I}$. Let $\{\omega^1, \ldots, \omega^k\}$ be a basis for $\mathcal{I}$ over $C^\infty(M)$. Then

$$\boldsymbol{d}\beta = \sum_{i=1}^{k} \theta^i \wedge \omega^i$$

for some $\theta^1, \ldots, \theta^k \in \Omega^1(M)$. Now let $X, Y \in \mathcal{D}$. Then

$$\boldsymbol{d}\beta(X, Y) = X \cdot \beta(Y) - Y \cdot \beta(X) + \beta([X, Y]) = \beta([X, Y]).$$

Since $\beta \in \mathcal{I}$ we obtain

$$\beta([X, Y] + Z) = 0 \quad \forall \ X, Y, Z \in \mathcal{D}.$$

This completes the proof of the lemma. ■

2.5.3 Pfaffian Modules with Independence Conditions

We shall need the notion of an independence condition to formulate mechanical systems in the language of exterior differential systems.

Definition 2.11 A *Pfaffian module with independence condition* on M is a pair, $(\mathcal{I}, [\omega])$, where $\mathcal{I}$ is a Pfaffian module and $[\omega]$ is an equivalence class of l-forms on M such that

i) ω and ω' are equivalent if $\omega \equiv \omega' \bmod \mathcal{I}$,

ii) locally we may write any representative in $[\omega]$ as

$$\omega = \omega^1 \wedge \cdots \wedge \omega^l$$

for one-forms $\omega^1, \ldots, \omega^l$, and

iii) $\omega(x) \notin \mathcal{I}(x)$ for all $x \in M$. □

We will be interested in integral manifolds of Pfaffian modules with independence conditions.

Definition 2.12 Let $(\mathcal{I}, [\omega])$ be a Pfaffian module with independence condition on M. We say that an l-dimensional submanifold, N, of M is an *integral manifold* of $(\mathcal{I}, [\omega])$ if N is an integral manifold of $\mathcal{I}$ and if ω restricted to N is nowhere zero. □

2.6 Some Constructions with Differential Two-Forms

In this section we give two constructions which may be applied to a given two-form on a manifold M. One construction determines a distribution on M and the other determines a Pfaffian module on M.

If $\Omega \in \Omega^2(M)$ we define the *characteristic distribution* corresponding to Ω by

$$D_\Omega(x) \triangleq \{v \in T_xM \mid \quad \Omega(u, v) = 0 \text{ for all } u \in T_xM\}.$$

We now may state a result.

Lemma 2.13 *Let $\Omega \in \Omega^2(M)$ be given by*

$$\Omega = \sum_{i=1}^{r} \alpha^i \wedge \beta^i$$

where $\{\alpha^1, \ldots, \alpha^r, \beta^1, \ldots, \beta^r\}$ is linearly independent. Then $v \in D_\Omega(x)$ if and only if v is annihilated by the Pfaffian module generated by

$$\{\alpha^1, \ldots, \alpha^r, \beta^1, \ldots, \beta^r\}.$$

Proof: Let $\mathcal{I}$ be the Pfaffian module generated by $\{\alpha^1, \ldots, \alpha^r, \beta^1, \ldots, \beta^r\}$ and suppose that $v \in D_\Omega(x)$. Then

$$\begin{aligned}
0 =& \Omega(u, v) && \text{for all } u \in T_xM \\
=& \sum_{i=1}^{r} \alpha^i \wedge \beta^i(u, v) && \text{for all } u \in T_xM \\
=& \sum_{i=1}^{r} (\alpha^i(u)\beta^i(v) - \alpha^i(v)\beta^i(u)) && \text{for all } u \in T_xM.
\end{aligned}$$

Since $\{\alpha^1, \ldots, \alpha^r, \beta^1, \ldots, \beta^r\}$ is linearly independent, we may select u so that $\alpha^i(u) = 0$ unless $i = j$ when $\alpha^j(u) = 1$, and so that $\beta^i(u) = 0$ for $i = 1, \ldots, r$. In this case we have $\beta^j(v) = 0$. Similarly we may show that $\alpha^j(v) = 0$ for $j = 1, \ldots, r$. This shows that v is annihilated by $I(x)$.

Now suppose that $v \in T_xM$ is annihilated by I. It is then clear by reversing the above argument that $v \in D_\Omega(x)$. This completes the proof of the lemma. ■

Note that if Ω is of constant rank r, then it is always possible to locally write it as in the hypothesis of Lemma 2.13.

Now we define the *Cartan system* of a two-form Ω. This is the Pfaffian module on M given by

$$C_\Omega = \{X \lrcorner \Omega \mid \quad X \in \mathfrak{X}(M)\}.$$

The following characterisation of the Cartan system follows in much the same way as Lemma 2.13.

Lemma 2.14 *Let $\Omega \in \Omega^2(M)$ be given by*

$$\Omega = \sum_{i=1}^{r} \alpha^i \wedge \beta^i$$

where $\{\alpha^1, \ldots, \alpha^r, \beta^1, \ldots, \beta^r\}$ is linearly independent. Then $\gamma \in C_\Omega$ if and only if γ is a linear combination over $C^\infty(M)$ of elements of $\{\alpha^1, \ldots, \alpha^r, \beta^1, \ldots, \beta^r\}$.

2.7 Riemannian Geometry

The subject of Riemannian geometry is a vast one and here we shall present only that part of it which bears upon the subjects in mechanics which are of interest to us. A detailed discussion of Riemannian geometry may be found in (Klingenberg, 1982). In Section 2.7.2 we introduce the notion of a *symmetric family of vector fields* on a Riemannian manifold. This concept will be important for Lagrangian control theory in Section 4.1.

2.7.1 Riemannian Geometry Definitions

A *pseudo-Riemannian metric* on a manifold M is a symmetric nondegenerate section of $\mathfrak{T}^0_2(M)$. A pseudo-Riemannian metric is *Riemannian* if it is also positive-definite on each fibre. A *Riemannian manifold* is a pair, (M, g), where M is a differentiable manifold and g is a Riemannian metric on M. Given a pseudo-Riemannian metric, we may define two isomorphisms of $C^\infty(M)$ modules; $\sharp\colon \Omega^1(M) \to \mathfrak{X}(M)$ and $\flat\colon \mathfrak{X}(M) \to \Omega^1(M)$. The map $\flat$ is defined by

$$X^\flat = \{Y \mapsto g(X, Y)\}$$

and $\sharp$ is its inverse. These isomorphisms are sometimes called the "musical isomorphisms." In particular, if f is a function on Q, we define its *gradient* by $\operatorname{grad} f = (\boldsymbol{d}f)^\sharp$.

A Riemannian manifold is endowed with an *affine connection*. In general an affine connection is a map from $\mathfrak{X}(M) \times \mathfrak{X}(M)$ to $\mathfrak{X}(M)$ denoted by $\nabla_X Y$ which has the following properties:

1. It is $\mathbb{R}$-linear in both X and Y, and
2. $\nabla_{fX} Y = f\nabla_X Y$ and $\nabla_X fY = f\nabla_X Y + (\boldsymbol{\mathcal{L}}_f X)Y$ for each $f \in C^\infty(M)$.

We shall call $\nabla_X Y$ the *covariant derivative* of Y with respect to X. Given an affine connection and a set of coordinates $(x^1, \ldots, x^n)$ for M, we define the *Christoffel symbols* for the affine connection in these coordinates by

$$\nabla_{\partial/\partial x^j}\left(\frac{\partial}{\partial x^k}\right) = \Gamma^i_{jk}\frac{\partial}{\partial x^i}.$$

Given the properties of an affine connection, it may be easily verified that

$$\nabla_X Y = \left(\frac{\partial Y^i}{\partial x^j}X^j + \Gamma^i_{jk}X^j Y^k\right)\frac{\partial}{\partial x^i}.$$

Given a curve $c\colon [0,T] \to M$ on M and $X_0 \in T_{c(0)}M$, there is a unique vector field $X(t)$ along c with the property that $\nabla_X c'(t) = 0$. This then defines a map from $T_{c(s)}M$ to $T_{c(t)}M$ for $s, t \in [0,T]$ which sends $X(s)$ to $X(t)$. This map is called *parallel translation.*

If (M, g) is a Riemannian manifold, there exists a unique affine connection on M with the properties that $\nabla_X Y - \nabla_Y X = [X, Y]$ and that parallel translation with respect to this affine connection is an isometry. This affine connection is often called the *Levi-Civita* connection. It may be verified that the Christoffel symbols of the Levi-Civita connection are given by

$$\Gamma^i_{jk} = \frac{1}{2}g^{il}\left(\frac{\partial g_{lj}}{\partial q^k} + \frac{\partial g_{lk}}{\partial q^j} - \frac{\partial g_{jk}}{\partial q^l}\right).$$

A curve $c\colon [0,T] \to M$ on a Riemannian manifold is said to be a *geodesic* if $\nabla_{c'(t)}c'(t) = 0$. In local coordinates, a geodesic is given by the solution of the following second-order differential equation:

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

This differential equation is, of course, the local representative of a vector field on TM. This vector field is called the *geodesic spray* or simply the *spray.* We shall denote it by Z_g. In local coordinates

$$Z_g = v^i\frac{\partial}{\partial q^i} - \Gamma^i_{jk}v^j v^k\frac{\partial}{\partial v^i}.$$

There are other topics in Riemannian geometry which are interesting in mechanics. In particular, the curvature of the Levi-Civita connection has important dynamical consequences. See (Ong, 1975) for some interesting results in this area. We will not, however, find the curvature tensor necessary.

2.7.2 The Symmetric Algebra Generated by a Family of Vector Fields

We shall need the concept of a "symmetric subalgebra" of $\mathfrak{X}(M)$ which is generated by a family of vector fields $\mathcal{V} \subset \mathfrak{X}(M)$. This construction relies on the covariant derivative discussed in Section 2.7.1. We may make $\mathfrak{X}(M)$ into a symmetric algebra by defining the symmetric product

$$\langle X : Y \rangle = \nabla_X Y + \nabla_Y X.$$

Let $\mathcal{V}$ be a family of vector fields on M and let X be a set which is bijective to $\mathcal{V}$ with bijection $\psi\colon X \to \mathcal{V}$. As in Section 2.2.6, let $S(X)$ be the free symmetric algebra on X and let $\mathrm{Pr}(X)$ be the symmetric products with elements in X. We may define a define a symmetric algebra homomorphism from $S(X)$ to $\mathfrak{X}(M)$ by extending ψ in the natural way (i.e., $\psi(\langle P_1 : P_1 \rangle) \mapsto \langle \psi(P_1) : \psi(P_2) \rangle$) to yield a map from $\mathrm{Pr}(X)$ to $\mathfrak{X}(M)$. This map may then be extended by $\mathbb{R}$-linearity to take values from $S(X)$. We denote the resulting map from $S(X)$ to $\mathfrak{X}(M)$ by $\mathrm{Ev}(\psi)$. We also define $\mathrm{Ev}_x(\psi)(P) = (\mathrm{Ev}(\psi)(P))(x)$ for $x \in M$. We denote by $\overline{\mathrm{Sym}}(\mathcal{V})$ the image of $S(X)$ under this homomorphism.

2.8 Symplectic Manifolds

When studying Hamiltonian mechanics, the basic mathematical tool is the symplectic manifold. In this section we give the definition of a symplectic manifold as well as a description of some symplectic concepts which shall be useful to us.

Definition 2.15 An *almost symplectic manifold* is a pair, (P, Ω), where P is a differentiable manifold and Ω is a nondegenerate two-form on P. We shall say that an almost symplectic manifold is *symplectic* if $\boldsymbol{d}\Omega = 0$. □

Now we turn to defining important distributions on symplectic manifolds. We shall make the necessary rank assumptions so that all objects defined are subbundles. Given a subbundle D, we define its *skew-orthogonal complement* by

$$\Omega^{\perp} D(p) = \{ v \in T_pP \mid \; \Omega(p)(v,u) = 0 \quad \forall \; u \in D(p) \}.$$

We say that D is

i) *isotropic* if $\Omega^{\perp} D \subset D$,
ii) *coisotropic* if $D \subset \Omega^{\perp} D$,
iii) *Lagrangian* if $D = \Omega^{\perp} D$, and
iv) *symplectic* if $D \cap \Omega^{\perp} D = \{0\}$.

Remarks 2.16

1. If the dimension of the manifold P is $2n$, then all isotropic subbundles have rank less than or equal to n, and all coisotropic subbundles have rank greater than or equal to n.

2. By the above remark, a Lagrangian subbundle will have rank half the dimension of P.

3. The above definitions may be applied to submanifolds of P by placing the requirements on the tangent spaces of the submanifold. □

Since the symplectic form Ω is nondegenerate, the map

$$X \mapsto X \lrcorner \Omega$$

from $\mathfrak{X}(P)$ to $\Omega^1(P)$ is an isomorphism. We denote this map by $\Omega^\flat$ and denote its inverse by $\Omega^\sharp$. Given a function f on P, we define the corresponding *Hamiltonian vector field* by

$$X_f = \Omega^\sharp \boldsymbol{d} f.$$

It is well-known that Hamiltonian vector fields leave the symplectic form invariant. That is to say, $\boldsymbol{\mathcal{L}}_{X_f}\Omega = 0$ for every $f \in C^\infty(P)$. Any vector field which has the property of leaving the symplectic form invariant is called a *locally Hamiltonian vector field.*

We define the Poisson bracket between two functions on P as follows:

$$\{f, g\} = \Omega(X_f, X_g).$$

Some authors use a different sign for the Poisson bracket than the convention we have chosen.

Now we gather some results which we shall need.

Lemma 2.17 *Let (P, Ω) be a symplectic manifold and let $\{,\}$ be the corresponding Poisson bracket.*

i) $\Omega^\sharp \boldsymbol{d}\{f, g\} = -[X_f, X_g]$.

ii) If D is an integrable distribution and f is an integral of D, $\Omega^\sharp \boldsymbol{d} f$ is a section of $\Omega^\perp D$.

Proof: i) This is just a restatement of the identity $X_{\{f,g\}} = -[X_f, X_g]$.

ii) We must show that $\Omega(\Omega^\sharp \boldsymbol{d} f, X) = 0$ for all sections X of D. We have

$$\Omega(\Omega^\sharp \boldsymbol{d} f, X) = \boldsymbol{d} f \cdot X.$$

Since f is an integral of D and X is a section of D, we get the result. ■

2.9 Poisson Manifolds

The concept of a Poisson manifold, $(P, \{,\})$, generalises a symplectic manifold by retaining only the structure of a Poisson bracket between functions. Thus the map $\{,\}: C^\infty(P) \times C^\infty(P) \to C^\infty(P): (f, g) \mapsto \{f, g\}$ is skew-symmetric, $\mathbb{R}$-linear, satisfies the Jacobi identity, and the map $g \mapsto \{f, g\}$ is a derivation on the $\mathbb{R}$-algebra $C^\infty(P)$. The Poisson bracket makes $C^\infty(P)$ into a $\mathbb{R}$-Lie algebra.

Definition 2.18 Let $(P, \{,\}_P)$ and $(N, \{,\}_N)$ be Poisson manifolds. A map $\phi: P \to N$ is *Poisson* if $\{f, g\}_N \circ \phi = \{f \circ \phi, g \circ \phi\}_P$ for all $f, g \in C^\infty(N)$. A vector field X on P is said to be a *Poisson vector field* if its flow defines a one-parameter family of Poisson mappings. □

There is a useful infinitesimal condition for checking that a vector field is Poisson.

Lemma 2.19 *Let $(P, \{,\})$ be a Poisson manifold and let $X \in \mathfrak{X}(P)$. Then X is Poisson if and only if*

$$\mathfrak{L}_X\{f, g\} = \{\mathfrak{L}_X f, g\} + \{f, \mathfrak{L}_X g\}. \tag{2.5}$$

Proof: Let X_t denote the flow of X. For $t_0 \in \mathbb{R}$ and $f, g \in C^\infty(P)$ we have

$$\frac{d}{dt} X^*_{-t}\{X^*_t f, X^*_t g\}\Big|_{t=t_0} = -X^*_{-t_0}(\mathfrak{L}_X\{X^*_{t_0} f, X^*_{t_0} g\}) + \\ X^*_{-t_0}(\{\mathfrak{L}_X X^*_{t_0} f, X^*_{t_0} g) + X^*_{-t_0}(\{X^*_{t_0} f, \mathfrak{L}_X X^*_{t_0} g).$$

If X is Poisson then the left hand side of the equation is zero, so (2.5) is true. Conversely, if (2.5) holds, then the right hand side of the equation is zero and so X is Poisson. ■

On Poisson manifolds it is possible to define analogues of Hamiltonian vector fields. Given a differentiable function f on P, we define the *Hamiltonian vector field* X_f by defining it as the derivation on $C^\infty(P)$ given by

$$X_f(g) = \{g, f\}.$$

The map $f \mapsto X_f$ is an anti-homomorphism from the Lie algebra $C^\infty(P)$ to the Lie algebra $\mathfrak{X}(P)$. Sometimes one sees Hamiltonian vector fields defined to be of opposite sign to the definition we have given. A Hamiltonian vector field is Poisson, but the converse is not necessarily true.

On a Poisson manifold we have an associated section of the bundle $\mathfrak{T}^2_0(P)$ of bivector fields on P. The following result is proved in (Libermann and Marle, 1987).

Proposition 2.20 *Let $(P, \{,\})$ be a Poisson manifold. Then there exists a differentiable, skew-symmetric section Σ of $\mathfrak{T}^2_0(P)$ with the property that*

$$\{f, g\} = \Sigma(\boldsymbol{d}f, \boldsymbol{d}g).$$

We shall call Σ the ***Poisson tensor*** *for the given Poisson structure.*

Proof: We first show that $\{f, g\}$ at a point $p \in P$ depends only on the values of $\boldsymbol{d}f$ and $\boldsymbol{d}g$ at that point. First fix f. Then we have

$$\{f, g\}(p) = -\boldsymbol{d}g(p) \cdot X_f(p).$$

Therefore, $\{f,g\}(p)$ depends only on $\boldsymbol{d}g(p)$. Similarly we may show that $\{f,g\}(p)$ depends only upon $\boldsymbol{d}f(p)$. Now note that the map $f \mapsto \boldsymbol{d}f(p)$ is a surjective map from $C^\infty(P)$ to T_p^*P. Also observe that, by definition, the map $(f,g) \mapsto \{f,g\}(p)$ is skew-symmetric and bilinear. This all combines to exhibit the existence of $\Sigma(p) \in \bigwedge^2(T_pP)$ so that

$$\{f,g\}(p) = \Sigma(p)(\boldsymbol{d}f(p), \boldsymbol{d}g(p)).$$

Differentiability of Σ follows from observing that, in a coordinate chart, the components of Σ are the Poisson brackets of the coordinates. ■

In the sequel we shall refer to a Poisson manifold by its structure tensor and so will write it as (P,Σ). The tensor field Σ allows us to define a bundle map, $\Sigma^\sharp$, from T^*P to TP by

$$\alpha \mapsto \{\beta \mapsto \Sigma(\alpha,\beta)\}.$$

The image of T^*P under $\Sigma^\sharp$ defines a subset of TP which we shall call the *characteristic distribution* of Σ. We shall denote this distribution by C_Σ. We assume that the dimension of the characteristic distribution is independent of $p \in P$. This occurs exactly when the rank of Σ is independent of p.

Proposition 2.21 *If the rank of the characteristic distribution is constant, then it is an integrable distribution.*

Idea of Proof: Note that $C_\Sigma(p)$ is generated by the set of all Hamiltonian vector fields passing trough p. Since the bracket of two Hamiltonian vector fields is again a Hamiltonian vector field, the distribution is integrable. ■

Remarks 2.22

1. Proposition 2.21 is true in a more general sense even when the rank of Σ is *not* constant. See (Libermann and Marle, 1987).
2. Let C_Σ^* be the maximal subbundle of T^*P with the property that $\Sigma^\sharp \mid C_\Sigma^*$ is a bijection onto C_Σ. In this case we have an almost symplectic structure on the leaves of C_Σ (i.e., a two-form of maximal rank). In fact, this almost symplectic structure can be shown to be symplectic. This defines a foliation of P into symplectic manifolds. Furthermore, since the tangent spaces to the symplectic leaves are generated by Hamiltonian vector fields, the integral curves of Hamiltonian vector fields leave the leaves of the symplectic foliation invariant. □

2.10 Jet Bundles

We begin with some introductory definitions which we will be using to formulate Lagrangian and Hamiltonian mechanics. The notation for jet bundles is from (Golubitsky and Guillemin, 1973).

2.10.1 The Bundle of Jets from $\mathbb{R}$ to M

We first need to say what we mean when two curves have the same derivative up to some order at a point. Let $c_1\colon [a,b] \to M$ and $c_2\colon [a,b] \to M$ be two curves on M so that $c_1(t) = c_2(t) = x$. Let $(x^1,\ldots,x^n)$ be a coordinate chart around x. We shall say that c_1 and c_2 *agree at order k at x* if the kth time derivatives of the components $(x^1(s),\ldots,x^n(s))$ agree at $s=t$. It may be seen that this definition of equivalence is independent of coordinate chart. If c_1 and c_2 agree at order k at x we shall write

$$c_1^{(k)}(t) = c_2^{(k)}(t).$$

Definition 2.23 Let M be a differentiable manifold, let $t \in \mathbb{R}$, and let $c_1, c_2\colon \mathbb{R} \to M$ be curves on M such that $c_1(t) = c_2(t) = x$. We say that c_1 and c_2 are *equivalent to order m at t* if

$$c_1^{(k)}(t) = c_2^{(k)}(t)$$

for $k = 1,\ldots,m$. We will write $c_1 \sim_m c_2$ at t and denote the equivalence class by $[c_1]_m$. We denote the set of all such equivalence classes by $J^m(\mathbb{R},M)_{t,x}$. The set

$$J^m(\mathbb{R},M) \triangleq \mathop{\overset{\circ}{\bigcup}}_{(t,x)\in\mathbb{R}\times M} J^m(\mathbb{R},M)_{t,x}$$

is called the set of *m-jets from $\mathbb{R}$ to M*. By definition we take $J^0(\mathbb{R},M) = \mathbb{R}\times M$. $\square$

We will be interested in the sets of 1-jets and 2-jets for the most part. If $(x^1,\ldots,x^n)$ is a coordinate chart for M, we have natural coordinates for $J^1(\mathbb{R},M)$ given by

$$(t, x^1,\ldots,x^n, v^1,\ldots,v^n).$$

Explicitly, if $c\colon \mathbb{R} \to M$ maps $t \in \mathbb{R}$ to $(x^1,\ldots,x^n) \in M$ in coordinates, then $[c]_1$ in natural coordinates for $J^1(\mathbb{R},M)$ is given by

$$[c]_1 = \left(t, x^1,\ldots,x^n, v^1 = \frac{dx^1}{ds}(t),\ldots, v^n = \frac{dx^n}{ds}(t)\right).$$

In a similar manner we have coordinates

$$(t, x^1,\ldots,x^n, v^1,\ldots,v^n, a^1,\ldots,a^n)$$

for $J^2(\mathbb{R},M)$. Explicitly, if $c\colon \mathbb{R} \to M$ maps $t \in \mathbb{R}$ to $(x^1,\ldots,x^n) \in M$ in coordinates, then $[c]_2$ in natural coordinates for $J^2(\mathbb{R},M)$ is given by

$$[c]_2 = \Big(t, x^1,\ldots,x^n, v^1 = \frac{dx^1}{ds}(t),\ldots, v^n = \frac{dx^n}{ds}(t), \\ a^1 = \frac{d^2x^1}{ds^2}(t),\ldots, a^n = \frac{d^2x^n}{ds^2}(t)\Big).$$

Elements of $J^1(\mathbb{R}, M)$ and $J^2(\mathbb{R}, M)$ transform in natural coordinates in specific ways according to the change of coordinates on M. To be specific, if $(X^1, \ldots, X^n)$ are coordinates for M different than $(x^1, \ldots, x^n)$, we have, with the obvious notation,

$$\begin{aligned} V^i &= \frac{\partial X^i}{\partial x^j} v^j, \\ A^i &= \frac{\partial X^i}{\partial x^j} a^j + \frac{\partial^2 X^i}{\partial x^j \partial x^k} v^j v^k. \end{aligned}$$

The fact that the accelerations do not transform linearly is a reflection of the fact that the 2-jets form an *affine bundle* over the 1-jets. This is discussed in (Goldschmidt, 1967).

We now define a family of projections from "higher" jet bundles to "lower" jet bundles. For $l < m$ there is a canonical projection, $\tau_{m,l}\colon J^m(\mathbb{R}, M) \to J^l(\mathbb{R}, M)$, which "forgets" the higher order of equivalence. We also define projections $\rho_m\colon J^m(\mathbb{R}, M) \to M$ by $\rho_m \triangleq pr_2 \circ \tau_{m,0}$ where $pr_2\colon \mathbb{R} \times M \to M$ is the projection onto the second factor. Note that in natural coordinates for $J^1(\mathbb{R}, M)$ we have

$$\rho_1(t, x^1, \ldots, x^n, v^1, \ldots, v^n) = (x^1, \ldots, x^n)$$

and in natural coordinates for $J^2(\mathbb{R}, M)$ we have

$$\rho_2(t, x^1, \ldots, x^n, v^1, \ldots, v^n, a^1, \ldots, a^n) = (x^1, \ldots, x^n).$$

If $c\colon \mathbb{R} \to M$ is a map, $j^m c\colon \mathbb{R} \to J^m(\mathbb{R}, M)$ will denote the map which assigns to t the equivalence class $[c]_m \in J^m(\mathbb{R}, M)_{t,c(t)}$. If the map c is given by

$$s \mapsto (x^1(s), \ldots, x^n(s)),$$

then the map $j^1 c$ is given by

$$j^1 c(t) = \left(t, x^1(t), \ldots, x^n(t), \frac{dx^1}{ds}(t), \ldots, \frac{dx^n}{ds}(t)\right)$$

and the map $j^2 c$ is given by

$$j^2 c(t) = \left(t, x^1(t), \ldots, x^n(t), \frac{dx^1}{ds}(t), \ldots, \frac{dx^n}{ds}(t), \frac{d^2x^1}{ds^2}(t), \ldots, \frac{d^2x^n}{ds^2}(t)\right).$$

For each $t \in \mathbb{R}$ and $x \in M$ we have a canonical identification of $T_x M$ with $J^1(\mathbb{R}, M)_{t,x}$. We will implicitly utilise this identification at times.

Note that there is an intrinsically defined function, τ, on $J^m(\mathbb{R}, M)$ defined by $\tau([c]_m) = t$ if $[c] \in J^m(\mathbb{R}, M)_{t,x}$. We shall use the notation $\boldsymbol{d}t \triangleq \boldsymbol{d}\tau$.

Note on Notation It is common to see natural coordinates for $J^1(\mathbb{R}, M)$ written as

$$(t, x^1, \ldots, x^n, \dot{x}^1, \ldots, \dot{x}^n).$$

We will stick to using v's instead of $\dot{x}$'s unless there is a specific curve on M which we are considering and so we wish to think of v^i as dx^i/dt. In this case we will use $\dot{x}^i$. Similar remarks hold for using a^i as opposed to using $\ddot{x}^i$. □

2.10.2 The Bundle of Jets from M to $\mathbb{R}$

In this case we will only be interested in first order equivalence.

Definition 2.24 Let M be a differentiable manifold, let $x \in M$, and let $f_1, f_2 \colon M \to \mathbb{R}$ be two functions on M such that $f_1(x) = f_2(x) = t$. We say that f_1 and f_2 are *equivalent at* $x \in M$ if $\boldsymbol{d}f_1(x) = \boldsymbol{d}f_2(x)$. We will write $f_1 \sim_1 f_2$ at x and denote the equivalence class by $[f_1]$. We denote the set of all such equivalence classes by $J^1(M, \mathbb{R})_{x,t}$. The set

$$J^1(M, \mathbb{R}) \triangleq \mathop{\overset{\circ}{\bigcup}}_{(x,t)\in M\times\mathbb{R}} J^1(M, \mathbb{R})_{x,t}$$

is called the set of *one-jets from* M *to* $\mathbb{R}$. □

If $(x^1, \ldots, x^n)$ is a coordinate chart for M, we have an associated set of natural coordinates for $J^1(M, \mathbb{R})$ given by

$$(x^1, \ldots, x^n, p_1, \ldots, p_n, t).$$

Explicitly, if $f \colon M \to \mathbb{R}$ maps $(x^1, \ldots, x^n) \in M$ to $t \in \mathbb{R}$ in coordinates, then $[f]$ in natural coordinates for $J^1(M, \mathbb{R})$ is given by

$$[f] = \left(x^1, \ldots, x^n, p_1 = \frac{\partial f}{\partial x^1}(x), \ldots, p_n = \frac{\partial f}{\partial x^n}(x), t\right).$$

The map $\pi_{1,0} \colon J^1(M, \mathbb{R}) \to M \times \mathbb{R}$ will denote the projection defined as follows: Let $[f] \in J^1(M, \mathbb{R})$ and let $x \in M$ and $t \in \mathbb{R}$ be such that $[f] \in J^1(M, \mathbb{R})_{x,t}$. We let $\pi_{1,0}([f]) = (x, t)$. In natural coordinates for $J^1(M, \mathbb{R})$ we have

$$\pi_{1,0}(x^1, \ldots, x^n, p_1, \ldots, p_n, t) = (x^1, \ldots, x^n, t).$$

We may also define the projection $\rho_1^* \colon J^1(M, \mathbb{R}) \to M$ by $\rho_1^* = pr_1 \circ \pi_{1,0}$ where $pr_1 \colon M \times \mathbb{R} \to M$ is projection onto the first factor. If f is a function on M then $j^1 f \colon M \to J^1(M, \mathbb{R})$ will denote the map which assigns to x the equivalence class $[f] \in J^1(M, \mathbb{R})_{x,f(x)}$. In coordinates, the map $j^1 f$ is given by

$$j^1 f(x^1, \ldots, x^n) = \left(x^1, \ldots, x^n, \frac{\partial f}{\partial x^1}(x), \ldots, \frac{\partial f}{\partial x^n}(x), f(x)\right).$$

For each $x \in M$ and each $t \in \mathbb{R}$ we have a canonical identification of $T_x^* M$ with $J^1(M, \mathbb{R})_{x,t}$. This identification will be used implicitly below.

There is a canonical projection of $J^1(M, \mathbb{R})$ onto T^*M which “forgets” the value of the function f in the equivalence class $[f] \in J^1(M, \mathbb{R})$. We shall call this projection

p_Q. In natural coordinates we have

$$p_Q(x^1, \ldots, x^n, p_1, \ldots, p_n, t) = (x^1, \ldots, x^n, p_1, \ldots, p_n).$$

As with $J^m(\mathbb{R}, M)$ we may define the one-form $\boldsymbol{d}t$ on $J^1(M, \mathbb{R})$.

Chapter 3

Nonlinear Control Theory

In this chapter we review some well-known results for general (i.e., not necessarily mechanical) control systems. Since our results for mechanical systems require much familiarity with these concepts, they are presented in some detail so that the reader may refer to them as needed. We also wish to develop the well-known results in the language of exterior differential systems. In Chapter 5 we shall formulate mechanics in the presence of external forces in terms of exterior differential systems. It is our opinion that these methods will be useful for future developments in mechanical control systems. Therefore, the results we present here for using exterior differential systems in nonlinear control theory may prove to have some significance in any further work we do in the arena of control of mechanical systems.

The control system we consider has state space M, a smooth n-dimensional manifold, and is affine in the controls. Thus it has the form

$$\dot{x} = X(x) + u^a Y_a(x) \tag{3.1}$$

where $X, Y_1, \ldots, Y_m$ are vector fields on M. The vector field X is called the *drift* vector field and the vector fields $Y_1, \ldots, Y_m$ are called the *control* vector fields. For the purpose of notation, we will denote by Σ the control system defined by the manifold M and the vector fields $X, Y_1, \ldots, Y_m$. To fully specify the control system properly, one should also specify the type of control actions to be considered. In this dissertation we consider our controls to be taken from the set

$$\mathcal{U} = \{u \colon \mathbb{R} \to \mathbb{R}^m \mid \ u \text{ is piecewise constant}\}.$$

It is mentioned in (Sontag, 1990) that this class of controls is sufficient to deal with all analytic control systems. More generally, one may wish to consider measurable functions which take their values in a subset of $\mathbb{R}^m$.

We point out that, given a control system of the form (3.1), it is possible to define a family of vector fields on M by

$$\mathcal{V}_\Sigma = \{X + u^a Y_a \mid \ u \in \mathbb{R}^m\}.$$

We shall be using ideas about families of vector fields from Section 2.4.2. We shall also be using the related ideas from Pfaffian modules from Section 2.5.

3.1 Nonlinear Controllability via Distributions

In this section we begin our review of nonlinear controllability by defining the appropriate notions of accessibility and giving tests for these in terms of distributions. When the system is *not* accessible, there is a local splitting of the state space. Most of what we say in this section is extracted from (Nijmeijer and van der Schaft, 1990).

3.1.1 Definitions

Let us review the definitions for local accessibility and strong local accessibility. A *solution* of (3.1) is a pair, (c, u), where $c\colon [0,T] \to M$ is a piecewise smooth curve on M and $u \in \mathcal{U}$ such that

$$c'(t) = X(c(t)) + u^a(t) Y_a(c(t))$$

for each $t \in [0,T]$. For $x_0 \in M$, a neighborhood V of x_0, and $T > 0$ denote

$$\begin{aligned}\mathcal{R}^V(x_0, T) = \{x \in M \mid \ & \text{there exists a solution } (c,u) \text{ of } (3.1)\\ & \text{such that } c(0) = x_0,\ c(t) \in V \text{ for } t \in [0,T], \text{ and } c(T) = x\}\end{aligned}$$

and denote

$$\mathcal{R}^V(x_0, \leq T) = \bigcup_{0 \leq t \leq T} \mathcal{R}^V(x_0, t).$$

Now we can define the versions of controllability.

Definition 3.1 The system (3.1) is *locally accessible* from x_0 if there exists $T > 0$ so that $\mathcal{R}^V(x_0, \leq t)$ contains a non-empty open set of M for all neighborhoods V of x_0 and all $0 < t \leq T$. If this holds for any $x_0 \in M$ then the system is called *locally accessible*.

The system (3.1) is *strongly locally accessible* from x_0 if there exists $T > 0$ so that $\mathcal{R}^V(x_0, t)$ contains a non-empty open set of M for all neighborhoods V of x_0 and all $0 < t \leq T$. If this holds for any $x_0 \in M$ then the system is called *strongly locally accessible*. $\square$

3.1.2 The Associated Distributions

Now we define the necessary distributions. We begin with the accessibility distribution.

Definition 3.2 The *accessibility algebra*, $\mathcal{C}$, corresponding to (3.1) is the smallest subalgebra of $\mathfrak{X}(M)$ which contains $X, Y_1, \ldots, Y_m$. The *accessibility distribution*, C, is the distribution on M defined by

$$C(x) = \langle \{ Z(x) \mid \ Z \in \mathcal{C} \} \rangle_{\mathbb{R}}\,. \qquad \square$$

This distribution may be computed, for example, by using the methods described in Section 2.4.2.

Now we turn to the strong accessibility distribution.

Definition 3.3 The *strong accessibility algebra*, $\mathcal{C}_0$, corresponding to (3.1) is the smallest subalgebra of $\mathfrak{X}(M)$ which contains $Y_1, \dots, Y_m$ and for which $[X, Z] \in \mathcal{C}_0$ for all $Z \in \mathcal{C}_0$. The *strong accessibility distribution*, C_0, is the distribution on M defined by

$$C_0(x) = \langle \{Z(x) \mid \quad Z \in \mathcal{C}_0\} \rangle_{\mathbb{R}}. \qquad \square$$

We have the following result which describes the form of the strong accessibility algebra.

Proposition 3.4 *Every element of* $\mathcal{C}_0$ *is a linear combination of vector fields of the form*

$$[Z_k, [Z_{k-1}, [\cdots, [Z_1, Y_a] \cdots]]]$$

for $a = 1, \dots, m$ *and where* $Z_i \in \{X, Y_1, \dots, Y_m\}$, $i = 1, \dots, k$.

Proof: The proof of this proposition mirrors that of Proposition 2.1. ■

The accessibility distribution and the strong accessibility distribution are related.

Lemma 3.5 *Let* $x \in M$. *Then* $C(x) = C_0(x) + \langle X(x) \rangle_{\mathbb{R}}$.

Proof: Let $Z \in \mathcal{C}_0$. By Propositions 2.4 and 3.4 we clearly have $Z + X \in \mathcal{C}$. Thus $C_0(x) + \langle X(x) \rangle_{\mathbb{R}} \subset C(x)$. Now suppose that $Z \in \mathcal{C}$. Then Z may have the form

$$Z = [Z_k, [Z_{k-1}, [\cdots, [Z_1, Y_a] \cdots]]]$$

for $a = 1, \dots, m$ and where $Z_i \in \{X, Y_1, \dots, Y_m\}$, $i = 1, \dots, k$. In this case $Z \in \mathcal{C}_0$. It is also possible that $Z = X$ in which case $Z(x) \in \langle X(x) \rangle_{\mathbb{R}}$. Thus $C(x) \subset C_0(x) + \langle X(x) \rangle_{\mathbb{R}}$ which completes the proof of the lemma. ■

3.1.3 Controllability Tests Using Distributions

Since the distributions C and C_0 have names associated to them which indicate that they have something to do with accessibility, we need to make this association clear. The results in this section are from (Nijmeijer and van der Schaft, 1990).

Proposition 3.6 *For the system (3.1) suppose that* $\operatorname{rank}(C(x_0)) = n$. *Then, for any neighborhood* V *of* x_0 *and* $T > 0$, *the set* $\mathcal{R}^V(x_0, \leq T)$ *contains a non-empty open subset of* M.

Proof: By continuity, there is a neighborhood U of x_0 so that $\operatorname{rank}(C) = n$ in U. We may construct a sequence of submanifolds of U, $N_1, \dots, N_n$ where $\dim(N_j) = j$, exactly as we did in the proof of Proposition 2.6. Just as was the case in that proof,

N_n is a non-empty open subset of M and all points in N_n are reachable from x_0. This proves the proposition. ■

Remark 3.7 By Proposition 2.6 we can see that if $X = 0$ (i.e., the system is driftless) then local accessibility implies controllability. Thus driftless systems have much more structure than do their counterparts with drift. However, it is not true that controllability is not in general possible for systems with drift. Indeed, linear systems are controllable if they are locally accessible. We will speak more about controllability for nonlinear systems in Section 3.4. □

Now we prove what amounts to the converse of Proposition 3.6.

Proposition 3.8 *If the system (3.1) is locally accessible then* $\operatorname{rank}(C(x)) = n$ *for* x *in an open dense subset of* M.

Proof: First note that if $\operatorname{rank}(C(x_0)) = n$ then $\operatorname{rank}(C(x)) = n$ for x in a neighborhood of x_0. Thus the set of points where $\operatorname{rank}(C(x)) = n$ is open in M. Now suppose that $\operatorname{rank}(C(x)) < n$ for x in some open subset, U, of M. Then there exists an open subset, $\bar{U}$, of U so that $\operatorname{rank}(C(x)) = k < n$ for all $x \in \bar{U}$. However, this contradicts local accessibility by Proposition 3.12. Therefore, there can be no open subset of M on which $\operatorname{rank}(C) < n$ and so the set of points x where $\operatorname{rank}(C(x)) = n$ must be dense. ■

Now we prove the analogous results for the strong accessibility distribution.

Proposition 3.9 *Consider the system (3.1). Suppose that* $\operatorname{rank}(C_0(x_0)) = n$. *Then, for any neighborhood* V *of* x_0 *and any* $T > 0$, *the set* $\mathcal{R}^V(x_0, T)$ *contains a non-empty open subset of* M.

Proof: Let us extend the system (3.1) to $\mathbb{R} \times M$ to obtain the system

$$\begin{aligned} \dot{s} &= 1 \\ \dot{x} &= X(x) + u^a Y_a(x). \end{aligned} \tag{3.2}$$

This is a control system on $\mathbb{R} \times M$ with drift vector field $\tilde{X} = X + \frac{\partial}{\partial s}$ and with control vector fields $\tilde{Y}_a = Y_a$ for $a = 1, \ldots, m$. Since $T(\mathbb{R} \times M) \simeq T\mathbb{R} \times TM$ we shall identify a vector field on M with a vector field on $\mathbb{R} \times M$.

Denote by $\tilde{C}$ and $\tilde{C}_0$ the accessibility and strong accessibility distributions corresponding to the extended system (3.2). We have the following lemma.

Lemma 3.10 $\tilde{C}(x_0, t_0) = C_0(x_0) + \langle X(x_0) + \frac{\partial}{\partial s} \rangle_{\mathbb{R}}$.

Proof: Note that by Lemma 3.5 we have

$$\tilde{C}_0(x_0, t_0) = \tilde{C}(x_0, t_0) + \langle X(x_0) + \tfrac{\partial}{\partial s} \rangle_{\mathbb{R}}.$$

A typical element of $\tilde{C}_0(x_0, t_0)$ has the form

$$[\tilde{Z}_k, [\tilde{Z}_{k-1}, [\cdots, [\tilde{Z}_1, \tilde{Y}_a] \cdots]]], \quad a = 1, \ldots, m$$

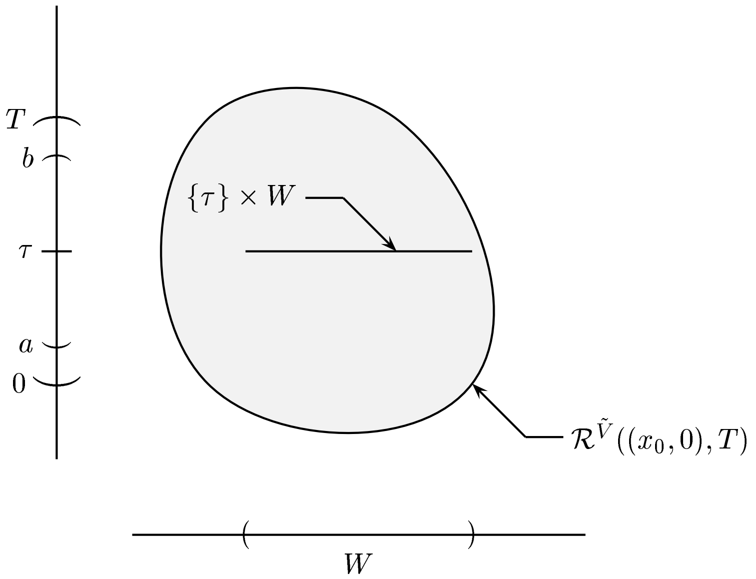

Figure 3.1 Schematic for strong local accessibility proof

for $\tilde{Z}_i \in \{\tilde{X}, \tilde{Y}_1, \ldots, \tilde{Y}_m\}$, $i = 1, \ldots, k$. However, since $[\frac{\partial}{\partial s}, X] = 0$ and $[\frac{\partial}{\partial s}, Y_a] = 0$ for $a = 1, \ldots, m$, we have

$$[\tilde{Z}_k, [\tilde{Z}_{k-1}, [\cdots, [\tilde{Z}_1, \tilde{Y}_a] \cdots]]] = [Z_k, [Z_{k-1}, [\cdots, [Z_1, Y_a] \cdots]]], \quad a = 1, \ldots, m$$

for $Z_i \in \{X, Y_1, \ldots, Y_m\}$, $i = 1, \ldots, k$. This proves that $\tilde{C}_0(x_0, t_0) = C_0(x_0)$. ▼

From the lemma and the assumption that $\operatorname{rank}(C_0(x_0)) = n$, we have

$$\operatorname{rank}(\tilde{C}(x_0, 0)) = \operatorname{rank}(C_0(x_0)) + 1 = n + 1.$$

This implies that (3.2) is locally accessible at $(x_0, 0)$. Thus, for any neighborhood V of x_0 and any $T > 0$, $\mathcal{R}^{\tilde{V}}((x_0, 0), \leq T)$ contains a non-empty open subset of $\mathbb{R} \times M$ where $\tilde{V} = (-\epsilon, T + \epsilon) \times V$ for some $\epsilon > 0$. Thus there is a non-empty open subset W of M and an interval $(a, b) \subset (0, T]$ so that $(a, b) \times W \subset \mathcal{R}^{\tilde{V}}((x_0, 0), \leq T)$. Therefore, for any $\tau \in (a, b)$ we have

$$\{\tau\} \times W \subset \mathcal{R}^{\tilde{V}}((x_0, 0), \leq T).$$

From this we may conclude that $W \subset \mathcal{R}^V(x_0, \tau)$. See Figure 3.1. Now let u be an admissible input with $Z = X + u^a Y_a$ the corresponding vector field on M. Then the map $x \mapsto Z^{T-\tau}(x)$ maps W onto an open subset $\tilde{W} \subset \mathcal{R}^U(x_0, T)$ for some neighborhood U of x_0. If we choose T small enough, $\tilde{W} \cap \mathcal{R}^V(x_0, T)$ will contain a non-empty open subset of M and so (3.1) is strongly locally accessible. ■

The converse of this is the following result.

Proposition 3.11 *If the system (3.1) is strongly locally accessible then* $\operatorname{rank}(C_0(x)) = n$ *for* x *in an open dense subset of* M.

Proof: The proof goes like the proof of Proposition 3.8. ■

3.1.4 Local Decompositions

Corresponding to both the accessibility distribution and the strong accessibility distribution there are useful decompositions of the state space. Indeed, the following two results are proved in (Nijmeijer and van der Schaft, 1990). First we present the result for local accessibility.

Proposition 3.12 *Suppose that* C *has constant rank* k *in a neighborhood of* $x_0 \in M$. *Then there exists a coordinate chart,* (U, ϕ), *about* x_0 *such that the submanifold*

$$S_{x_0} = \{x \in U \mid \; x^i(x) = x^i(x_0), \;\; i = k+1, \ldots, n\}$$

is an integral manifold of C. *Then, for any neighborhood* $V \subset U$ *of* x_0 *and for all* $T > 0$, $\mathcal{R}^V(x_0, \leq T)$ *is contained in* S_{x_0}. *Furthermore,* $\mathcal{R}^V(x_0, \leq T)$ *contains a non-empty open set of the integral manifold* S_{x_0}. *Hence the system restricted to* S_{x_0} *is locally accessible.*

Proof: Since $\mathcal{C}$ contains $X, Y_1, \ldots, Y_m$, we may restrict the system to S_{x_0}. The restricted system is locally accessible by Proposition 3.6 since $\dim(S_{x_0}) = \operatorname{rank}(C \mid S_{x_0})$. ■

The analogous result for strong local accessibility is given in the following proposition.

Proposition 3.13 *Suppose that* C_0 *has constant rank* k *in a neighborhood of* $x_0 \in M$. *Then there exists a coordinate chart,* (U, ϕ), *about* x_0 *such that the submanifolds*

$$S = \{x \in U \mid \; x^i(x) = a^i, \;\; i = k+1, \ldots, n\}$$

for $|a_i| < \epsilon$ *are integral manifolds of* C_0 *and such that the integral manifold through* x_0 *is*

$$S_{x_0} = \{x \in U \mid \; x^i(x) = 0, \;\; i = k+1, \ldots, n\}.$$

There are now two possibilities:

i) If $X(x_0) \in C_0(x_0)$, *then* $X(x) \in C_0(x)$ *for all* $x \in S_{x_0}$ *and* $\mathcal{R}^U(x_0, T) \subset S_{x_0}$ *for all* $T > 0$. *In this case, the system restricted to* S_{x_0} *is locally strongly accessible.*

ii) If $X(x_0) \notin C_0(x_0)$, then, by continuity, $X(x) \notin C_0(x)$ for all $x \in \tilde{U}$ for some neighborhood $\tilde{U} \subset U$ of x_0, and $\operatorname{rank}(C(x)) = \operatorname{rank}(C_0(x)) + 1$ for all $x \in \tilde{U}$. In this case we may choose coordinates $\bar{x}_{k+1}, \ldots, \bar{x}_n$ on $\tilde{U}$ so that

$$S_{x_0} = \{x \in \tilde{U} \mid \ \bar{x}^{k+1}(x) = \cdots = \bar{x}^n(x) = 0\},$$

and, if we let

$$S_{x_0}^T = \{x \in \tilde{U} \mid \ \bar{x}^{k+1}(x) = T, \bar{x}^{k+2}(x) = \cdots = \bar{x}^n(x) = 0\},$$

then $\mathcal{R}^{\tilde{U}}(x_0, T)$ is contained in $S_{x_0}^T$ for any $T > 0$ and, moreover, $\mathcal{R}^{\tilde{U}}(x_0, T)$ contains a non-empty open subset of $S_{x_0}^T$ for any $T > 0$ sufficiently small.

Proof: Recall that $[X, Z] \in \mathcal{C}_0$ for $Z \in \mathcal{C}_0$. By Frobenius' Theorem we may choose coordinates $(x^1, \ldots, x^k, x^{k+1}, \ldots, x^n)$ so that the leaves of $\mathcal{C}_0$ are given by $x^{k+1} = \cdots = x^n = constant$. Note that

$$C_0(x_0) = \left\langle \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^k} \right\rangle_{\mathbb{R}}.$$

For $i = 1, \ldots, n$ we have

$$\left[X, \frac{\partial}{\partial x^i}\right] = -\frac{\partial X^j}{\partial x^i}\frac{\partial}{\partial x^j}.$$

If we evaluate this for $i = 1, \ldots, k$ we obtain

$$\frac{\partial X^i}{\partial x^j} = 0, \quad i = k+1, \ldots, n, \ j = 1, \ldots, k. \tag{3.3}$$

i: Now suppose that $X(x_0) \in C_0(x_0)$. From (3.3) we see that $X(x) \in C_0(x)$ for all $x \in S_{x_0}$. Thus (3.1) leaves S_{x_0} invariant and, if we apply Proposition 3.9 to the restricted system, we see that it is strongly locally accessible.

ii: Now suppose that $X(x_0) \notin C_0(x_0)$. By continuity $X(x) \notin C_0(x)$ for $x \in \tilde{U} \subset U$. Using the local coordinates $(x^1, \ldots, x^k, x^{k+1}, \ldots, x^n)$, we may define a vector field on $\mathbb{R}^{n-k}$ by

$$\bar{X}(x^{k+1}, \ldots, x^n) = \sum_{i=k+1}^{n} X^i(x^{k+1}, \ldots, x^n)\frac{\partial}{\partial x^i}.$$

Since this vector field is not zero on $\tilde{U}$, we may make a change of coordinates from $(x^{k+1}, \ldots, x^n)$ to $(\bar{x}^{k+1}, \ldots, \bar{x}^n)$ so that

$$\bar{X}(\bar{x}^{k+1}, \ldots, \bar{x}^n) = \frac{\partial}{\partial \bar{x}^{k+1}}.$$

It is now clear that $\mathcal{R}^{\tilde{U}}(x_0, T) \subset S_{x_0}^T$. To prove that $\mathcal{R}^{\tilde{U}}(x_0, T)$ is non-empty and open in $S_{x_0}^T$, we may proceed as in Proposition 3.9 since the system in the coordinates

$(x^1, \ldots, x^k, \bar{x}^{k+1}, \ldots, \bar{x}^n)$ is essentially the same (the same plus trivial dynamics) as the system (3.2). ■

Remarks 3.14

1. Strong local accessibility always implies local accessibility.
2. In Proposition 3.13 i we see that in the case when $\mathrm{rank}(C) = \mathrm{rank}(C_0)$, local accessibility implies strong local accessibility.
3. In Proposition 3.13 ii we see that we may locally regard the leaves of the accessibility distribution as the leaves of the strong accessibility distribution "cross" time. □

3.2 Nonlinear Controllability via Exterior Differential Systems

In this section we cast the results of Section 3.1 in the language of exterior differential systems. This gives us a taste of how we may use the derived flag to say things about nonlinear controllability. For background on exterior differential systems see Section 2.5 and (Bryant *et al.*, 1991).

3.2.1 The Pfaffian Module Corresponding to a Control System

For a few moments, to ease the notation, let us consider control systems of the form

$$\dot{x} = Z(x, u) \tag{3.4}$$

where $x \in M$, as usual, and $u \in \mathbb{R}^m$. For the sake of name calling, let us denote this control system by Σ in the usual way. Corresponding to this system we may define a Pfaffian module, $\mathcal{I}_\Sigma$, on $\mathbb{R} \times M \times \mathbb{R}^m$ by defining

$$I_\Sigma(t, x, u) = \left(\frac{\partial}{\partial s} + Z(x, u) \right)^0 .$$

If we choose coordinates $(x^1, \ldots, x^n)$ for M we have

$$I_\Sigma(t, x, u) = \left\langle dx^1 - Z^1(x, u)dt, \ldots, dx^n - Z^n(x, u)dt \right\rangle_{\mathbb{R}} .$$

Notice that integral curves of $\mathcal{I}_\Sigma$ are solutions of the control system. More precisely, say that a pair, (c, u), is a *solution* of (3.4) if $c\colon [0, T] \to M$ and $u \in \mathcal{U}$ are such that (3.4) is satisfied.

Lemma 3.15 *Suppose that (c, u) is a solution of (3.4). Then the curve $\sigma\colon s \mapsto (s, c(s), u(s))$ is an integral curve of $(\mathcal{I}_\Sigma, \boldsymbol{d}t)$.*

Conversely, let N be an integral manifold of $(\mathcal{I}_\Sigma, \boldsymbol{d}t)$. Then there exists curves $c\colon [0, T] \to M$ and $u \in \mathcal{U}$ so that N is the locally image of the curve $\sigma\colon s \mapsto (s, c(s), u(s))$. Furthermore, so defined, (c, u) is a solution of (3.4).

Proof: Suppose that $c\colon [0,T] \to M$ and $u \in \mathcal{U}$ is a solution of (3.4). Then

$$(dx^i - Z^i(s, c(s), u(s))dt) \cdot \sigma'(s) = \dot{x}^i(s) - Z^i(s, c(s), u(s)) = 0.$$

Also note that $\boldsymbol{dt}(s) \cdot \sigma'(s) = 1$. Thus σ is an integral curve of $(\mathcal{I}_\Sigma, \boldsymbol{dt})$.

Now suppose that N is an integral manifold of $(\mathcal{I}_\Sigma, \boldsymbol{dt})$. Since $\boldsymbol{dt} \neq 0$ on N, we may regard N as a graph over t. Thus N has the form

$$(s, c(s), u(s)).$$

It remains to show that the curve $\sigma\colon s \mapsto (s, c(s), u(s))$ is annihilated by $\mathcal{I}_\Sigma$. But this is clear since N is an integral manifold of $\mathcal{I}_\Sigma$. ■

3.2.2 Local Decompositions

Because of Lemma 3.15, we expect that the derived system will give us some information about the controllability of the system (3.1). We will use the local decomposition result of Proposition 3.13. We shall now revert back to the control affine system specified by (3.1).

Proposition 3.16 *Consider the control system (3.1). Suppose that C_0 has constant rank k in a neighborhood of $x_0 \in M$. There are two cases:*

i) Suppose that $X(x_0) \in C_0(x_0)$. Then, in the coordinate chart guaranteed by Proposition 3.13 i,

$$I_\Sigma^{(\infty)}(x_0) = \left\langle dx^{k+1}, \dots, dx^n \right\rangle_{\mathbb{R}}.$$

ii) Suppose that $X(x_0) \notin C_0(x_0)$. Then, in the coordinate chart guaranteed by Proposition 3.13 ii,

$$I_\Sigma^{(\infty)}(x_0) = \left\langle dx^{k+1} - X^{k+1}dt, \dots, dx^n - X^n dt \right\rangle_{\mathbb{R}}.$$

Proof: i: In the coordinates guaranteed by Proposition 3.13, (3.1) takes the form

$$\dot{x}^i = X^i(x^1, \dots, x^n) + u^a Y_a^i(x^1, \dots, x^n), \quad i = 1, \dots, k \tag{3.5a}$$

$$\dot{x}^b = X^b(x^{k+1}, \dots, x^n), \quad b = k+1, \dots, n. \tag{3.5b}$$

(See Proposition 3.26.) However, since $X(x_0) \in C_0(x_0)$, we have $X^{k+1} = \cdots = X^n = 0$. Therefore, we immediately have

$$\left\langle dx^{k+1}, \dots, dx^n \right\rangle_{\mathbb{R}} \subset I_\Sigma^{(\infty)}(x_0).$$

We now make a little computation. For simplicity let

$$Z^i(x, u) = X^i(x) + u^a Y_a(x), \quad i = 1, \dots, k.$$

For $i = 1, \ldots, k$ we have

$$\begin{aligned}\boldsymbol{d}(dx^i - Z^i dt) &= \sum_{j=1}^{k} \frac{\partial Z^i}{\partial x^j} dx^j \wedge dt + \sum_{b=k+1}^{n} \frac{\partial Z^i}{\partial x^b} dx^b \wedge dt + \sum_{a=1}^{m} \frac{\partial Z^i}{\partial u^a} du^a \wedge dt \\ &= \left(\sum_{j=1}^{k} \frac{\partial Z^i}{\partial x^j} dx^j \wedge dt + \sum_{a=1}^{m} \frac{\partial Z^i}{\partial u^a} du^a \wedge dt \right) \quad \mathrm{mod}\ \mathcal{I}_\Sigma^{(\infty)}.\end{aligned}$$

Therefore, to prove this part of the proposition, we need only show that, for $x^{k+1}, \ldots, x^n$ regarded as constants, the bottom derived system for the Pfaffian module

$$\mathcal{I}' \triangleq \left\langle dx^1 - Z^1(x,u)dt, \ldots, dx^k - Z^k(x,u)dt \right\rangle_{C^\infty(\mathbb{R} \times M \times \mathbb{R}^m)}$$

is zero. Suppose that it is not. Then all integral curves of $\mathcal{I}'$, and hence all integral curves of $(\mathcal{I}', \boldsymbol{dt})$, must lie on a submanifold of $\mathbb{R} \times \mathbb{R}^k$ of codimension at least 1. But this contradicts Proposition 3.26 which says that (3.5a) is strongly locally accessible for each fixed $x^{k+1}, \ldots, x^n$.

ii: The proof here is the same as for i except that $X^{k+1}, \ldots, X^n$ are not all zero. ■

With this proven we may easily prove the following result.

Proposition 3.17 *The system (3.1) is*

i) strongly locally accessible if and only if $\mathcal{I}_\Sigma^{(\infty)} = \{0\}$, *and*

ii) locally accessible if and only if $\mathrm{rank}(\mathcal{I}_\Sigma^{(\infty)}) \leq 1$ *and the integral manifolds of the bottom derived system are time-dependent.*

Proof: i: Suppose that (3.1) is strongly locally accessible. Then $\mathrm{rank}(C_0) = n$ and so, by Proposition 3.16, we must have $\mathcal{I}_\Sigma^{(\infty)} = \{0\}$. Now suppose that $\mathcal{I}_\Sigma^{(\infty)} = \{0\}$. Then, again by Proposition 3.16, $\mathrm{rank}(C_0) = n$ and so (3.1) is strongly locally accessible.

ii: Suppose that (3.1) is locally accessible. Then $\mathrm{rank}(C) = n$. Then, either $\mathrm{rank}(C_0) = n$ (case i of Proposition 3.16) or $\mathrm{rank}(C_0) = n - 1$ (case ii of Proposition 3.16). In the first case $\mathcal{I}_\Sigma^{(\infty)} = \{0\}$. In the second case $\mathcal{I}_\Sigma^{(\infty)} = \langle dx^n - X^n(x^n)dt \rangle_{C^\infty(\mathbb{R} \times M \times \mathbb{R}^m)}$. Thus, in the second case, $\mathrm{rank}(\mathcal{I}_\Sigma^{(\infty)}) = 1$ and the integral manifolds of the bottom derived system are time-dependent. Now suppose that $\mathrm{rank}(\mathcal{I}_\Sigma^{(\infty)}) \leq 1$. Then, by Proposition 3.16, $\mathrm{rank}(C_0) \geq n - 1$ and so $\mathrm{rank}(C) = n$. Thus (3.1) is locally accessible. ■

3.3 Invariant Distributions

In this section we introduce invariant distributions and show how they may be used to simplify control systems. The notion of an invariant distribution has several interpretations, all of which will be useful to us when we come to using them for simplifying control systems. We begin our discussion with the case of a distribution being invariant under a vector field.

3.3.1 Distributions Invariant Under a Vector Field

We have the following definition.

Definition 3.18 Let Y be a vector field on M and let D be a distribution on M. We will say that D is *invariant* under Y if $[Y, Z] \in \mathcal{D}$ for every $Z \in \mathcal{D}$. □

The case when the distribution is integrable and defines a foliation $\mathcal{F}_D$ is especially interesting. We shall suppose that the quotient space, $M/\mathcal{F}_D$, has a differentiable structure which makes the projection a submersion. If D has constant rank, this is always true locally. The computations we present below are extracted from various locations in (Marmo *et al.*, 1985).

First we prove a technical lemma.

Lemma 3.19 *Let $\pi\colon M \to B$ be a surjective submersion and let X be a projectable vector field on M. Thus $T\pi \circ X(x) \in T_bB$ is independent of $x \in \pi^{-1}(b)$ for each $b \in B$. Then $\mathcal{L}_X f$ is constant on fibres of π for every function f which is constant on fibres of π. Conversely, if $\mathcal{L}_X f$ is constant on fibres of π for every function f which is constant on fibres of π, then X is projectable.*

Proof: Let f be a function which is constant on the fibres of π. Then $f = \tilde{f} \circ \pi$ for some function $\tilde{f}$ on B. Therefore, $\boldsymbol{d}f = \boldsymbol{d}(\pi^* \tilde{f})$. By definition $\mathcal{L}_X f = \boldsymbol{d}f \cdot X$. If X is projectable then both X and $\boldsymbol{d}f$ are constant on fibres of π, and so the function $\mathcal{L}_X f$ is constant on fibres of π.

Now suppose that $\mathcal{L}_X \pi^* \tilde{f}$ is constant on fibres of π for every function $\tilde{f}$ on B. Thus $\mathcal{L}_X \pi^* \tilde{f} = \pi^* \tilde{g}$ for some function g on B. For convenience, if $g \in C^\infty(M)$ is constant on the fibres of π, let us denote by $\tilde{g} \in C^\infty(B)$ the function which satisfies $g = \pi^* \tilde{g}$. We shall also denote $g' = \mathcal{L}_X g$. We claim that the map

$$\begin{aligned} \tilde{X}\colon C^\infty(B) &\to C^\infty(B) \\ \tilde{g} &\mapsto \{\tilde{f} \mid \ \mathcal{L}_X g = f\} \end{aligned}$$

is a derivation. We denote $\tilde{g}' = \tilde{X}(\tilde{g})$. Now let $\tilde{g}_1, \tilde{g}_2 \in C^\infty(B)$ and let $\tilde{g} = \tilde{g}_1 \tilde{g}_2$. We then have

$$\begin{aligned} \pi^* \tilde{g}' \triangleq g' =& \mathcal{L}_X g = \mathcal{L}_X (g_1 g_2) = (\mathcal{L}_X g_1) g_2 + (\mathcal{L}_X g_2) g_1 \\ =& g_1' g_2 + g_2' g_1 = \pi^* (\tilde{g}_1' \tilde{g}_2 + \tilde{g}_2' \tilde{g}_1). \end{aligned}$$

This verifies that $\tilde{X}$ is a derivation on $C^\infty(B)$ and hence a vector field on B. We only need to show now that $\tilde{X}$ is the projection of X. Denote the projection of X by X'. We compute

$$\pi^* (\mathcal{L}_{\tilde{X}} \tilde{g}) = \pi^* \tilde{g}' = g' = \mathcal{L}_X g = \pi^* (X' \lrcorner \tilde{g}) = \pi^* (\mathcal{L}_{X'} \tilde{g})$$

for any $\tilde{g} \in C^\infty(B)$. This completes the proof. ■

Now we state what happens for invariant distributions.

Proposition 3.20 *Let D be an integrable distribution which gives rise to a simple foliation, $\mathcal{F}_D$, of M, and let X be a vector field on M. Then D is invariant under X if and only if there exists a vector field $\tilde{X}$ on $B = M/\mathcal{F}_D$ so that the following diagram commutes.*

$$\begin{tikzcd} M \arrow[r, "X"] \arrow[d, "\pi"'] & TM \arrow[d, "T\pi"] \\ B \arrow[r, "\tilde{X}"'] & TB \end{tikzcd}$$

Here $\pi\colon M \to B$ is the projection.

Proof: First note that there exists a vector field $\tilde{X}$ which makes the diagram commute if and only if $T\pi \circ X(x) \in T_bB$ is independent of $x \in \pi^{-1}(b)$ for each $b \in B$. By Lemma 3.19 this implies that $\mathcal{L}_X f$ is constant on fibres for every f which is constant on fibres. Since sections of D are tangent to the fibres, for every function f which is constant on fibres, and for every vector field Y which is a section of D we have $\mathcal{L}_Y f = 0$. Therefore, $\mathcal{L}_X\mathcal{L}_Y f = 0$ for every function f which is constant on the fibres of π and vector field Y which is a section of D. Thus we have shown that $\mathcal{L}_X\mathcal{L}_Y f = 0$ for every function f which is constant on fibres of π and for every section Y of D if and only if there exists a vector field $\tilde{X}$ such that the diagram above commutes.

Now suppose that D is *not* invariant under X. Then there exists a section Y' of D so that $[X, Y']$ is not a section of D. Locally we may suppose that we may find a function f with the property that $\mathcal{L}_Y f = 0$ if and only if Y is a section of D. Therefore,

$$\mathcal{L}_{[X,Y']} f = \mathcal{L}_X\mathcal{L}_{Y'} f - \mathcal{L}_{Y'}\mathcal{L}_X f = -\mathcal{L}_{Y'}\mathcal{L}_X f \neq 0.$$

Thus $\mathcal{L}_X f$ is not constant on leaves of π for every function f which is constant on leaves of π. Therefore, there is no vector field $\tilde{X}$ on B which makes the above diagram commute.

Now suppose that D is invariant under X. Thus

$$\mathcal{L}_{[X,Y]} f = \mathcal{L}_X\mathcal{L}_Y f - \mathcal{L}_Y\mathcal{L}_X f = -\mathcal{L}_Y\mathcal{L}_X f = 0$$

for every function f constant on the fibres of π and every section Y of D. This implies that $\mathcal{L}_X f$ is constant on fibres of π which means that a vector field $\tilde{X}$ exists which makes the above diagram commute. This completes the proof. ■

The following result gives an important local decomposition for invariant distributions.

Proposition 3.21 *Let Y be a vector field on M and let D be an integrable distribution on M which is invariant under Y. Then, in a neighborhood of $x \in M$ where* rank$(D) = k$, *there exists coordinates* $(x^1, \dots, x^k, x^{k+1}, \dots, x^n) \triangleq (x_1, x_2)$ *so that* Y *has representative* $(Y_1(x_1, x_2), Y_2(x_2))$.

Proof: By Frobenius' Theorem we choose coordinates $(x^1, \dots, x^k, x^{k+1}, \dots, x^n)$ so that

$$D(x) = \left\langle \frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^k} \right\rangle_{\mathbb{R}}.$$

Then, as in the proof of Proposition 3.13, we obtain

$$\frac{\partial Y^i}{\partial x^j} = 0$$

for $i = k+1, \dots, n$ and $j = 1, \dots, k$. The result follows directly. ■

Note that Proposition 3.21 is a local consequence of Proposition 3.20, but we did not use this in the proof.

3.3.2 Distributions Invariant Under a Control System

First we define the concept of an invariant distribution for a control system of the form (3.1).

Definition 3.22 Let D be a distribution on M. We say that D is *invariant* for (3.1) if $[X, \mathcal{D}] \in \mathcal{D}$ and $[Y_a, \mathcal{D}] \in \mathcal{D}$ for $a = 1, \dots, m$. □

Now we prove an easy lemma.

Lemma 3.23 *A distribution* D *is invariant under* Σ *if and only if* D *is invariant under the vector field*

$$Z_u \triangleq X + u^a Y_a$$

for every $u \in \mathbb{R}^m$.

Proof: Suppose that D is invariant under Σ. Then $[X, Z], [Y_1, Z], \dots, [Y_m, Z] \in \mathcal{D}$ for every $Z \in \mathcal{D}$. Thus

$$[X, Z] + u^1[Y_1, Z] + \cdots + u^m[Y_m, Z] \in \mathcal{D}$$

for every $Z \in \mathcal{D}$ and $u \in \mathbb{R}^m$. Since u is constant, it now follows that $[Z_u, Z] \in \mathcal{D}$ for every $Z \in \mathcal{D}$ and $u \in \mathbb{R}^m$. Thus D is invariant under Z_u.

Now suppose that D is invariant under Z_u for every $u \in \mathbb{R}^m$. Choosing $u = 0$ gives $[X, Z] \in \mathcal{D}$ for every $Z \in \mathcal{D}$. Choosing $u = e_a$, the ath standard basis element for $\mathbb{R}^m$, for $a = 1, \dots, m$ gives $[Y_a, Z] \in \mathcal{D}$ for every $Z \in \mathcal{D}$. Thus D is invariant under Σ. ■

Now we give some decompositions for control systems possessing invariant distributions. First we give an analogue of Proposition 3.21.

Proposition 3.24 *Suppose that D is a constant rank, integrable distribution which is invariant for (3.1). Then there are coordinates, $(x^1, \dots, x^k, x^{k+1}, \dots, x^n) \triangleq (x_1, x_2)$, for M so that*

$$\begin{aligned}\dot{x}_1 &= X_1(x_1, x_2) + u^a Y_{a,1}(x_1, x_2)\\ \dot{x}_2 &= X_2(x_2) + u^a Y_{a,2}(x_2).\end{aligned}$$

Proof: As in the proof of Proposition 3.21, we choose coordinates guaranteed by Frobenius' Theorem and determine that

$$\frac{\partial X^i}{\partial x^j} = 0, \qquad \frac{\partial Y_a^i}{\partial x^j} = 0$$

for $a = 1, \dots, m$, $i = k+1, \dots, n$ and $j = 1, \dots, k$. The result follows directly. ■

Thus the presence of an integrable invariant distribution allows us to decouple some variables, the x_2 variables, from the others. The accessibility distribution and the strong accessibility distribution are special examples of invariant distributions. The following results are from (Nijmeijer and van der Schaft, 1990). For the accessibility distribution we have the following result.

Proposition 3.25 *The accessibility distribution is the smallest integrable invariant distribution which contains $X, Y_1, \dots, Y_m$. In the coordinates guaranteed by Proposition 3.24 the control system (3.1) assumes the form*

$$\dot{x}_1 = X_1(x_1, x_2) + u^a Y_a(x_1, x_2) \qquad (3.6a)$$

$$\dot{x}_2 = 0. \qquad (3.6b)$$

Furthermore, for each fixed value of x_2, the control system (3.6a) is locally accessible.

Proof: By definition, the accessibility distribution is the smallest integrable distribution containing $X, Y_1, \dots, Y_m$. By virtue of its being invariant, it is also the smallest integrable invariant distribution containing $X, Y_1, \dots, Y_m$. The form of the equations in the coordinates given by Proposition 3.24 follows from these being the same coordinates given by Proposition 3.12. Recall that in these coordinates the submanifolds of M defined by $x^{k+1} = \cdots = x^n = \textit{constant}$ are invariant under (3.1). ■

For the strong accessibility distribution we have the following result.

Proposition 3.26 *The strong accessibility distribution is the smallest integrable distribution, invariant under X, which contains $Y_1, \dots, Y_m$. In the coordinates guaranteed by Proposition 3.24 the control system (3.1) assumes the form*

$$\dot{x}_1 = X_1(x_1, x_2) + u^a Y_a(x_1, x_2) \qquad (3.7a)$$

$$\dot{x}_2 = X_2(x_2). \qquad (3.7b)$$

Furthermore, for each fixed value of x_2, the control system (3.7a) is strongly locally accessible.

Proof: It is clear that C_0 is the smallest integrable distribution containing $Y_1, \ldots, Y_m$ which is invariant under X. The coordinates given by Proposition 3.24 are the same as those given by Proposition 3.13. This gives the form of the equations as presented. ∎

3.3.3 An Exterior Differential Systems Interpretation

Now we give an interpretation of invariant distributions in terms of the Pfaffian module $\mathcal{I}_\Sigma$ on $\mathbb{R} \times M$. Recall that $\mathcal{I}_\Sigma$ is the subset of $\Omega^1(\mathbb{R} \times M \times \mathbb{R}^m)$ given by

$$\mathcal{I}_\Sigma = \left\langle \beta - (\beta \cdot Z)\boldsymbol{d}t \mid \quad \beta \in \Omega^1(M) \right\rangle_{C^\infty(\mathbb{R}\times M\times\mathbb{R}^m)}$$

where

$$Z = X + u^a Y_a.$$

Given a distribution D on M, we define a subset of $\mathcal{I}_\Sigma$ by

$$\mathcal{I}_{\Sigma,D} \triangleq \left\langle \beta - (\beta \cdot Z)\boldsymbol{d}t \mid \quad \beta \in \mathcal{D}^0 \right\rangle_{C^\infty(\mathbb{R}\times M\times\mathbb{R}^m)}.$$

Observe that for fixed u we may regard $\mathcal{I}_\Sigma$ and $\mathcal{I}_{\Sigma,D}$ as Pfaffian modules on $\mathbb{R} \times M$. For $u \in \mathbb{R}^m$ we shall denote these modules by $\mathcal{I}_{\Sigma_u}$ and $\mathcal{I}_{\Sigma_u,D}$, respectively.

Now we may state an intermediate result.

Proposition 3.27 *Let D be an integrable distribution on M and let Y be a vector field on M. Denote by Σ' the trivial control system consisting of the vector field Y (i.e., no controls). Then D is invariant under Y if and only if $\mathcal{I}_{\Sigma',D}$ is integrable.*

Proof: Let $(x^1, \ldots, x^k, x^{k+1}, \ldots, x^n)$ be coordinates for M so that $\mathcal{D}^0 = \left\langle dx^{k+1}, \ldots, dx^n \right\rangle_{C^\infty(M)}$. We denote these coordinates symbolically by (x_1, x_2) where $x_1 = (x^1, \ldots, x^k)$ and $x_2 = (x^{k+1}, \ldots, x^n)$.

Now suppose that D is invariant under Y. In the coordinates given above, the representative of Y has the form $(Y_1(x_1, x_2), Y_2(x_2))$. Thus

$$\mathcal{I}_{\Sigma',D} = \left\langle dx^{k+1} - Y^{k+1}(x_2)dt, \ldots, dx^n - Y^n(x_2)dt \right\rangle_{C^\infty(\mathbb{R}\times M)}.$$

To show that $\mathcal{I}_{\Sigma',D}$ is integrable we will show that

$$\boldsymbol{d}(dx^a - Y^a(x_2)dt) = 0 \bmod \mathcal{I}_{\Sigma',D}$$

for $a = k+1, \dots, n$. Indeed

$$\begin{aligned} \boldsymbol{d}(dx^a - Y^a(x_2)dt) &= -\sum_{b=k+1}^{n} \frac{\partial Y^a}{\partial x^b} dx^b \wedge dt \\ &= -\sum_{b=k+1}^{n} \frac{\partial Y^a}{\partial x^b} (dx^b - Y^b(x_2)dt) \wedge dt \bmod \mathcal{I}_{\Sigma',D} \\ &= 0 \bmod \mathcal{I}_{\Sigma',D}. \end{aligned}$$

Now suppose that $\mathcal{I}_{\Sigma',D}$ is integrable. In coordinates given by Frobenius' theorem we have

$$\mathcal{I}_{\Sigma',D} = \left\langle dx^{k+1} - Y^{k+1}(x_1, x_2)dt, \dots, dx^n - Y^n(x_1, x_2)dt \right\rangle_{C^\infty(\mathbb{R}\times M)}.$$

Since $\mathcal{I}_{\Sigma',D}$ is integrable we have

$$\boldsymbol{d}(dx^a - Y^a(x_1, x_2)dt) = 0 \bmod \mathcal{I}_{\Sigma',D}$$

for $a = k+1, \dots, n$. Thus

$$-\frac{\partial Y^a}{\partial x^i} dx^i \wedge dt = 0 \bmod \mathcal{I}_{\Sigma',D}$$

for $a = k+1, \dots, n$. But

$$-\sum_{i=1}^{n} \frac{\partial Y^a}{\partial x^i} dx^i \wedge dt = -\sum_{i=1}^{k} \frac{\partial Y^a}{\partial x^i} dx^i \wedge dt \bmod \mathcal{I}_{\Sigma',D}.$$

Therefore, we must have

$$\sum_{i=1}^{k} \frac{\partial Y^a}{\partial x^i} dx^i \wedge dt = 0 \bmod \mathcal{I}_{\Sigma',D}$$

for $a = k+1, \dots, n$. This implies that

$$\frac{\partial Y^a}{\partial x^i} = 0, \quad a = k+1, \dots, n, \;\; i = 1, \dots, k.$$

In other words, D is invariant under Y. ■

Combining Proposition 3.27 and Lemma 3.23 gives the following result.

Proposition 3.28 *Let D be a distribution on M and let Σ be a control system of the form (3.1). Then D is invariant under Σ if and only if the Pfaffian module, $\mathcal{I}_{\Sigma_u,D}$, on $\mathbb{R} \times M$ is integrable for each $u \in \mathbb{R}^m$.*

3.4 Sufficient Conditions for Small-Time Local Controllability

(Sussmann, 1987) gives a general result concerning so-called small-time local controllability. We are interested in a version of Sussmann's result and so will present only as much background as is necessary to state this result.

The control system (3.1) is said to be *small-time locally controllable* (STLC) from $x_0 \in M$ if it is locally accessible from x_0 and if there exists $T > 0$ so that x_0 is in the interior of $\mathcal{R}^V(x_0, \leq t)$ for each $0 < t \leq T$ and each neighborhood V of x_0. If this holds for any $x_0 \in M$ then the system is called STLC.

Let $\boldsymbol{X} = \{X_0, \dots, X_m\}$. We will need some of the notation from Section 2.2.3 regarding free Lie algebras. In particular, $\mathrm{Br}(\boldsymbol{X})$ is the set of "brackets" of elements from $\boldsymbol{X}$ and $\delta_a(B)$ is the number of occurrences of X_a in $B \in \mathrm{Br}(\boldsymbol{X})$. The reader should also recall the Lie algebra rank condition (LARC) from Section 2.4.2. Note that this is a sufficient condition for local accessibility. With further conditions on the types of brackets that a control system possesses, it may also be STLC.

An element $B \in \mathrm{Br}(\boldsymbol{X})$ is said to be *bad* if $\delta_0(B)$ is odd and $\delta_a(B)$ is even for each $a = 1, \dots, m$. A bracket is *good* if it is not bad. Let S_m denote the permutation group on m symbols. For $\pi \in S_m$ and $B \in \mathrm{Br}(\boldsymbol{X})$, define $\bar{\pi}(B)$ to be the bracket obtained by fixing X_0 and sending X_a to $X_{\pi(a)}$ for $a = 1, \dots, m$. Now define

$$\beta(B) = \sum_{\pi \in S_m} \bar{\pi}(B).$$

We may state sufficient conditions for STLC.

Theorem 3.29 (Sussmann, 1987) *Consider the bijection $\phi\colon \boldsymbol{X} \to \{X, Y_1, \dots, Y_m\}$ which sends X_0 to X and X_a to Y_a for $a = 1, \dots, m$. Suppose that (3.1) is such that every bad bracket $B \in \mathrm{Br}(\boldsymbol{X})$ has the property that*

$$\mathrm{Ev}_x(\phi)(\beta(B)) = \sum_{a=1}^{m} \xi^a \,\mathrm{Ev}_x(\phi)(C_a)$$

where C_a are good brackets in $\mathrm{Br}(\boldsymbol{X})$ of lower degree than B and $\xi_a \in \mathbb{R}$ for $a = 1, \dots, m$. Also suppose that (3.1) satisfies the LARC at x. Then (3.1) is STLC at x.

In (Sussmann, 1987) this result is actually a corollary of a special case originally conjectured in (Hermes, 1978) and proven in (Sussmann, 1983).

Chapter 4

Control Theory for Mechanical Systems

In this chapter we give some control theoretic results for certain classes of mechanical systems. A discussion of general mechanical systems with external forces is deferred to Chapter 5. The goal in this chapter is to adapt the ideas of Chapter 3 to both Lagrangian (Section 4.1) and Hamiltonian (Section 4.2) control systems. When we study the Lagrangian problem, we are primarily interested in obtaining conditions for a refined notion of controllability which is relevant for mechanical control systems. On the Hamiltonian side, if a natural Hamiltonian structure is assumed for the control problem, we are able to give nice descriptions of the locally accessible Hamiltonian dynamics and the strongly locally inaccessible Hamiltonian dynamics (the locally inaccessible dynamics are always trivial). Some simple examples are presented in Section 4.3 to illustrate the ideas put forward by the theory.

4.1 Lagrangian Control Theory for Simple Mechanical Control Systems

In this section we study a specific, but large, class of mechanical control systems. Our presentation is from a Lagrangian point of view since this framework seems best adapted to the computations we do.

The systems studied are the so-called *simple mechanical control systems*. Such systems are characterised by the following data:

1. a Riemannian metric g on the n-dimensional configuration manifold Q,

2. a function V on the configuration manifold, and

3. m linearly independent one-forms, $F^1, \ldots, F^m$, on Q.

The Lagrangian for the control system we consider is defined by

$$L(v) = \frac{1}{2} g(v, v) - V \circ \tau_Q(v). \tag{4.1}$$

Thus we consider the Lagrangian to be "kinetic energy minus potential energy." The control torques take their values in the complete subset of T^*Q (see Section 5.2) defined by

$$\Lambda_q = \langle F^1(q), \ldots, F^m(q) \rangle_{\mathbb{R}}.$$

This means that we will allow the possible directions for application of force to be functions of position only. More generally, one may want these directions to be functions of time and velocity as well.

With this data, the Lagrangian control system in local coordinates has the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = u_a F_i^a. \tag{4.2}$$

For the given Lagrangian, these equations may be expressed in a convenient invariant form. To express this we need the notion of the vertical lift of a vector field. Let X be a vector field on Q. Its *vertical lift* is the vector field on TQ defined by

$$X^{lift}(v) = \frac{d}{dt}\left(X(\tau_Q(v)) + tv\right)|_{t=0}.$$

In local coordinates, if

$$X(q) = X^i(q)\frac{\partial}{\partial q^i}$$

then we have

$$X^{lift}(v_q) = X^i(q)\frac{\partial}{\partial v^i}.$$

The reader may also wish to recall the definition of the geodesic spray, Z_g, from Section 2.7. We shall define

$$X_L = Z_g - \operatorname{grad} V^{lift}.$$

Lemma 4.1 *Let L be the Lagrangian defined by (4.1). Then the equations (4.2) are equivalent to the equations*

$$\dot{v}(t) = X_L(v(t)) + u_a(t) Y_a(\tau_Q(v(t))) \tag{4.3}$$

where $Y_a = (F^a)^\sharp$ for $a = 1, \ldots, m$.

Proof: Let $c\colon [0,T] \to Q$ be an integral curve of X_L. Thus, in local coordinates,

$$\ddot{q}^i + \Gamma^i_{jk}\dot{q}^j\dot{q}^k = -g^{ij}\frac{\partial V}{\partial q^j} + u_a g^{ij} F_j^a$$

where

$$\Gamma^i_{jk} = \frac{1}{2}g^{il}\left(\frac{\partial g_{lk}}{\partial q^j} + \frac{\partial g_{lj}}{\partial q^k} - \frac{\partial g_{jk}}{\partial q^l}\right).$$

Note that

$$\frac{\partial L}{\partial q^k} = \frac{1}{2}\frac{\partial g_{ij}}{\partial q^k} v^i v^j - \frac{\partial V}{\partial q^k}, \qquad \frac{\partial L}{\partial v^k} = g_{kj} v^j.$$

Therefore,

$$\begin{aligned} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} &= g_{ij}\ddot{q}^j + \frac{\partial g_{ij}}{\partial q^k}\dot{q}^j \dot{q}^k - \frac{1}{2}\frac{\partial g_{jk}}{\partial q^i}\dot{q}^j \dot{q}^k + \frac{\partial V}{\partial q^i} \\ &= g_{ij}\ddot{q}^j + \left(\frac{\partial g_{ij}}{\partial q^k} - \frac{1}{2}\frac{\partial g_{jk}}{\partial q^i}\right)\dot{q}^j \dot{q}^k + \frac{\partial V}{\partial q^i}. \end{aligned}$$

Now note that

$$\begin{aligned} \Gamma^l_{jk}\dot{q}^j \dot{q}^k &= \frac{1}{2} g^{li}\left(\frac{\partial g_{ik}}{\partial q^j} + \frac{\partial g_{ij}}{\partial q^k} - \frac{\partial g_{jk}}{\partial q^i}\right)\dot{q}^j \dot{q}^k \\ &= g^{li}\left(\frac{\partial g_{ij}}{\partial q^k} - \frac{1}{2}\frac{\partial g_{jk}}{\partial q^i}\right)\dot{q}^j \dot{q}^k. \end{aligned}$$

The lemma now follows by multiplying Lagrange's equations by the "inverse" of g. ■

Note that we may also write (4.3) as

$$\nabla_{c'(t)} c'(t) = \operatorname{grad} V(c(t)) + u^a(t) Y_a(c(t)).$$

We shall use this form of the equations when we define a solution for a simple mechanical control system in Section 4.1.6.

With systems of this type there are some things that are worth noticing before proceeding to the calculations. In particular, note that all of the data for the problem is defined by quantities on the configuration manifold. Therefore, we would like to be able to compute the answers to interesting questions in terms of these quantities. An example of such an interesting question is the following:

Problem Statement Describe the set of configurations which are reachable from a given configuration when starting at rest. □

It is exactly this question which we are interested in and which we shall answer. Furthermore, as we shall see, our answer is obtainable in terms of quantities defined on Q.

Since some rather detailed calculations are required in this section, let us outline what we plan to do. In Section 4.1.1 we present an example which illustrates what we wish to do and why it is interesting. This example shows that the conventional definitions of controllability given in Chapter 3 are not so well adapted to the mechanical systems we are considering. We also perform a few calculations for this example which foreshadow the general results developed in the succeeding sections. In Section 4.1.2 we do some computations with free Lie algebras. The reader should be warned that the presentation in this section may be difficult to follow, but is very important in understanding the basic premise of the sections which follow. We will also find it useful to know some tangent bundle structure. This is presented in Section 4.1.3. This structure becomes of consequence when we restrict the accessibility distribution to $Z(TQ)$. The distribution computations are performed in Section 4.1.4. With these computations, in Section 4.1.5 we are able to state the form of the accessibility distribution restricted to the zero section of TQ. In Section 4.1.6 we present controllability definitions for systems of the form (4.3). These formalise the problem statement given above. Using the computations from Section 4.1.4, we may obtain conditions for our notions of controllability. These are presented in Section 4.1.7. Finally, in Section 4.1.8 some decomposition results are presented which are analogous to Propositions 3.12 and 3.13.

4.1.1 A Motivating Example

In this section we describe in some detail a simple mechanical control system which illustrates the need to refine the treatment of mechanical systems in nonlinear control theory. In particular, this example demonstrates that the nonlinear control calculations which one often performs do not provide a satisfactory resolution to the controllability problem for all mechanical systems. We propose that a weaker notion of controllability may be useful. We also do some computations with this example which hint at how the general calculations will proceed in the sections to follow.

A Description of the System

The example we consider is a rigid body with inertia J which is pinned to ground at its centre of mass. This example was first presented in (Li *et al.*, 1989).[1] The body has attached to it an extensible massless leg and the leg has a point mass with mass m at its tip. The coordinate θ will describe the angle of the body, and ψ will describe the angle of the leg from an inertial reference frame. The coordinate r will describe the extension of the leg. Thus the configuration space for this problem is $Q = \mathbb{T}^2 \times \mathbb{R}^+$. See Figure 4.1. The Lagrangian is

$$L = \frac{1}{2} J v_\theta^2 + \frac{1}{2} m (v_r^2 + r^2 v_\psi^2).$$

If we consider forces applied in the $(\theta - \psi)$ and r-directions, Lagrange's equations are

$$J\ddot{\theta} = u_1 \tag{4.4a}$$

$$mr^2\ddot{\psi} + 2mr\dot{r}\dot{\psi} = -u_1 \tag{4.4b}$$

$$m\ddot{r} - mr\dot{\psi}^2 = u_2. \tag{4.4c}$$

[1] In this paper the example considered is actually in free flight. We present the robotic leg fixed to a point as this simplifies the analysis, but removes none of the essential structure.

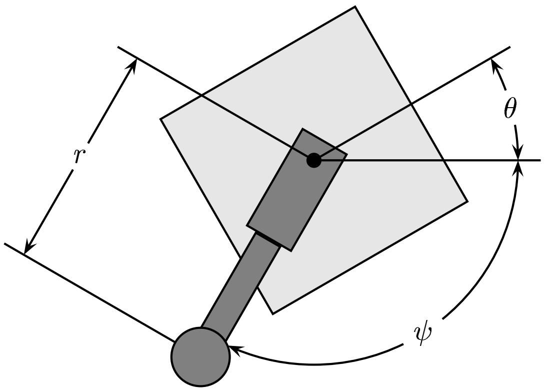

Figure 4.1 The robotic leg

Contradictory Controllability Results

We may rewrite Lagrange's equations in the form (4.3). In this case we compute the Lagrangian vector field as

$$X_L = Z_g = v_\theta \frac{\partial}{\partial \theta} + v_\psi \frac{\partial}{\partial \psi} + v_r \frac{\partial}{\partial r} - \frac{2 v_r v_\psi}{r} \frac{\partial}{\partial v_\psi} + r v_\psi^2 \frac{\partial}{\partial v_r}$$

and the input vector fields as

$$Y_1 = \frac{1}{J} \frac{\partial}{\partial \theta} - \frac{1}{mr^2} \frac{\partial}{\partial \psi}, \qquad Y_2 = \frac{1}{m} \frac{\partial}{\partial r}.$$

The distribution calculations may be performed to obtain the accessibility distribution as

$$C(\theta, \psi, r, v_\theta, v_\psi, v_r) = \left\langle 2mrv_\psi \frac{\partial}{\partial v_\theta} - J \frac{\partial}{\partial r}, mr^2 \frac{\partial}{\partial v_\theta} - J \frac{\partial}{\partial v_\psi}, \frac{\partial}{\partial \theta}, \frac{\partial}{\partial \psi}, \frac{\partial}{\partial v_r} \right\rangle_{\mathbb{R}}.$$

Since this distribution does not span TQ, we conclude that the system is not locally accessible. Nevertheless, it is possible to steer the system from one configuration to another. Indeed we have the following result, some of which was proven in (Murray and Sastry, 1993).

Claim *Select two configurations, $q_1 = (\theta_1, \psi_1, r_1)$ and $q_2 = (\theta_2, \psi_2, r_2)$. Suppose that the system starts at rest in configuration q_1. Then there exists inputs u_1, u_2 which steer the system to rest at q_2.*

Proof: We first note that the inputs leave the total angular momentum,

$$\mu = J\dot{\theta} + mr^2 \dot{\psi},$$

of the system conserved. Thus, when we start at rest at q_1, all consequent motions of the system will have zero angular momentum. This may be thought of as imposing a constraint given by

$$J\dot{\theta} + mr^2\dot{\psi} = 0. \tag{4.5}$$

Let us first answer the question: How many configurations are accessible from q_1 along paths which preserve zero angular momentum? Let D be the distribution defined by (4.5). This distribution has dimension two and the Lie bracket between any two basis vector fields for $\mathcal{D}$ will not lie in $\mathcal{D}$. This shows that D is controllable as discussed in Section 2.4.2. Therefore, by Proposition 2.6, from q_1 it is possible to reach any other configuration while maintaining the constraint of zero angular momentum. To prove the claim, we need to show that all motions of the system which preserve zero angular momentum are realisable using suitable inputs, u_1, u_2. Let c be a path in Q which satisfies the constraint (4.5) and which connects q_1 with q_2. We may suppose that c is reparameterised so that we start at rest at q_1 and end at rest at q_2. From (4.4c) and (4.4a) we immediately have $u_2 = m\ddot{r} - mr\dot{\psi}^2$ and $u_1 = J\ddot{\theta}$. We need only show that, so defined, u_1 satisfies (4.4b). From (4.5) we have

$$J\ddot{\theta} = -mr^2\ddot{\psi} - 2mr\dot{r}\dot{\psi}.$$

Therefore,

$$mr^2\ddot{\psi} + 2mr\dot{r}\dot{\psi} = -u_1$$

which is simply (4.4b). This completes the proof. ■

A Closer Look at the Distribution Calculations

The above claim indicates that we would like to be able to consider this problem controllable in some sense. Let us try to understand how we might do this by taking a closer look at the distribution computations which yield the accessibility distribution. Since we are interested in describing the set of points reachable from initial conditions with zero velocity, we will evaluate all brackets on the zero section of TQ. We may compute

$$\begin{aligned}
[Y_1^{lift}, Y_2^{lift}] &= 0 \\
[Z_g, Y_1^{lift}](0_q) &= -Y_1(q) \\
[Z_g, Y_2^{lift}](0_q) &= -Y_2(q) \\
[Y_1^{lift}, [Z_g, Y_1^{lift}]] &= -\frac{2}{m^2 r^3}\frac{\partial}{\partial v_r} \\
[Y_1^{lift}, [Z_g, Y_2^{lift}]] &= 0 \\
[Y_2^{lift}, [Z_g, Y_2^{lift}]] &= 0 \\
[Z_g, [Z_g, Y_1^{lift}]](0_q) &= 0 \\
[Z_g, [Z_g, Y_2^{lift}]](0_q) &= 0 \\
[Z_g, [Y_1^{lift}, [Z_g, Y_1^{lift}]]](0_q) &= -\frac{2}{m^2 r^3}\frac{\partial}{\partial r} \\
[[Z_g, Y_1^{lift}], [Z_g, Y_2^{lift}]](0_q) &= [Y_1, Y_2](q).
\end{aligned}$$

These turn out to be the only interesting brackets for the robotic leg. If we examine these bracket calculations, we make the following informal observations.

1. The brackets between the input vector fields are zero.

2. The brackets which contain the drift vector field the same number of times as the control vector fields give brackets in the "q-direction" when we evaluate them at zero velocity.

3. The brackets which contain the control vector fields one more time than the drift vector field are vertical lifts of vector fields on Q.

4. The brackets which contain the drift vector field more often than the control vector fields are zero when evaluated at points of zero velocity.

These observations suggest what may happen with general systems of the form (4.3). The sections which follow formally go through the calculations needed to prove the form of the accessibility distribution for these systems when restricted to the zero section of TQ. The reader may wish to refer back to the above bracket calculations at various times during the general exposition.

4.1.2 Computations with Free Lie Algebras

In this section we perform some calculations with a pair of free Lie algebras which are suited to our purposes. The reader should be warned that they may not see what they expect here. Rather than just using a generating set which is in 1–1 correspondence with the set $\{X_L, Y_1^{lift}, \ldots, Y_m^{lift}\}$ of control vector fields and the drift vector field, we also use a generating set which is in 1–1 correspondence with the set $\{Z_g, Y_1^{lift}, \ldots, Y_m^{lift}, \operatorname{grad} V^{lift}\}$. The reason for this will become clear when we perform the distribution calculations in Section 4.1.4.

Let $\boldsymbol{X} = \{X_0, \dots, X_{m+1}\}$ and let $L(\boldsymbol{X})$ be the free Lie algebra generated by the set $\boldsymbol{X}$. We can simplify many of our computations for the controllability analysis of (4.3) by making simplifications to a set of generators for $L(\boldsymbol{X})$.

We first need some notation. Let

$$\begin{aligned}\mathrm{Br}^k(\boldsymbol{X}) &= \{B \in \mathrm{Br}(\boldsymbol{X}) \mid \text{ the degree of } B \text{ is } k\}\,,\\ \mathrm{Br}_k(\boldsymbol{X}) &= \left\{B \in \mathrm{Br}(\boldsymbol{X}) \mid \ \delta_0(B) - \sum_{a=1}^{m+1} \delta_a(B) = k\right\}.\end{aligned}$$

We will also need the concept of a *primitive* bracket.

Definition 4.2 Let $B \in \mathrm{Br}_0(\boldsymbol{X}) \cup \mathrm{Br}_{-1}(\boldsymbol{X})$ and let $B_1, B_2, B_{11}, B_{12}, B_{21}, B_{22}, \dots$ be the decomposition of B into its components. We shall say that B is *primitive* if each of its components is in $\mathrm{Br}_{-1}(\boldsymbol{X}) \cup \mathrm{Br}_0(\boldsymbol{X}) \cup \{X_0\}$. □

The relevant observations that need to be made regarding primitive brackets are:

Prim1. If $B \in \mathrm{Br}_{-1}(\boldsymbol{X})$ is primitive then, up to sign, we may write $B = [B_1, B_2]$ with $B_1 \in \mathrm{Br}_{-1}(\boldsymbol{X})$ and $B_2 \in \mathrm{Br}_0(\boldsymbol{X})$ both primitive.

Prim2. If $B \in \mathrm{Br}_0(\boldsymbol{X})$ is primitive then, up to sign, B may have one of two forms. Either $B = [X_0, B_1]$ with $B_1 \in \mathrm{Br}_{-1}(\boldsymbol{X})$ primitive or $B = [B_1, B_2]$ with $B_1, B_2 \in \mathrm{Br}_0(\boldsymbol{X})$ primitive.

Using these two rules, it is possible to construct primitive brackets of any degree. For example, the primitive brackets of degrees one through four are, up to sign

Degree 1: $\{X_a \mid \ a = 1, \dots, m\}$
Degree 2: $\{[X_0, X_a] \mid \ a = 1, \dots, m\}$
Degree 3: $\{[X_a, [X_0, X_b]] \mid \ a, b = 1, \dots, m\}$
Degree 4: $\{[X_0, [X_a, [X_0, X_b]]] \mid \ a, b = 1, \dots, m\} \cup$
$\{[[X_0, X_a], [X_0, X_b]] \mid \ a, b = 1, \dots, m\}.$

From Proposition 2.1 we know that to generate $L(\boldsymbol{X})$ we need only look at brackets of the form

$$[X_{a_k}, [X_{a_{k-1}}, \dots, [X_{a_2}, X_{a_1}]]] \tag{4.6}$$

where $a_i \in \{0, \dots, m+1\}$ for $i = 1, \dots, k$. We shall see in Section 4.1.4 that brackets from $\mathrm{Br}_j(\boldsymbol{X})$, where $j \geq 1$ or $j \leq -2$, will not be of interest to us. In particular, we shall see that when $j \leq -2$ the brackets evaluate identically to zero. Therefore, in this section we concentrate our attention on brackets in $\mathrm{Br}_0(\boldsymbol{X}) \cup \mathrm{Br}_{-1}(\boldsymbol{X})$ which satisfy certain requirements. We state this in the following lemma.

Lemma 4.3 *Let us impose the condition on elements of* $\mathrm{Br}(\boldsymbol{X})$ *that we shall consider a bracket to be zero if any of its components are in* $\mathrm{Br}_{-j}(\boldsymbol{X})$ *for* $j \geq 2$. *Let* $B \in \mathrm{Br}_0(\boldsymbol{X}) \cup \mathrm{Br}_{-1}(\boldsymbol{X})$. *Then we may write* B *as a finite sum of primitive brackets.*

Proof: It is sufficient to prove the lemma for brackets of the form (4.6). We proceed by induction on k in (4.6). The lemma is true for $k = 1, 2$ by inspection. Now suppose the lemma true for $k = 1, \ldots, l$ and let B be of the form (4.6) for $k = l + 1$. Then we have two cases. Either $B \in \mathrm{Br}_{-1}(\boldsymbol{X})$ or $B \in \mathrm{Br}_0(\boldsymbol{X})$.

We look first at the case where $B \in \mathrm{Br}_{-1}(\boldsymbol{X})$. Since we are considering brackets in $\mathrm{Br}_{-2}(\boldsymbol{X})$ to be zero, we may write $B = [X_a, B']$ with $B' \in \mathrm{Br}_0(\boldsymbol{X})$ of the form (4.6) and $a \in \{1, \ldots, m+1\}$. By the induction hypothesis, B' is a finite sum of primitive brackets and the lemma is proved in this case since B will then also be a finite sum of primitive brackets.

Now we look at the case where $B \in \mathrm{Br}_0(\boldsymbol{X})$. There are two possibilities in this case. The first possibility is that $B = [X_0, B']$ with $B' \in \mathrm{Br}_{-1}(\boldsymbol{X})$. In this case B' is a finite sum of primitive brackets by the induction hypothesis and, therefore, B is also a finite sum of primitive brackets.

The final case is when $B = [X_{a_1}, B']$ with $B' \in \mathrm{Br}_{+1}(\boldsymbol{X})$ of the form (4.6). If $B' = [X_0, B'']$ with $B'' \in \mathrm{Br}_0(\boldsymbol{X})$ then, by Jacobi's identity, we have

$$B = [X_{a_1}, [X_0, B'']] = -[B'', [X_{a_1}, X_0]] - [X_0, [B'', X_{a_1}]].$$

Since $B'' \in \mathrm{Br}_0(\boldsymbol{X})$, by the induction hypotheses it may be written as a finite sum of primitive brackets in $\mathrm{Br}_0(\boldsymbol{X})$. Clearly $[X_{a_1}, X_0]$ is primitive which proves that $[B'', [X_{a_1}, X_0]]$ is a finite sum of primitive brackets. The bracket $[B'', X_{a_1}]$ is in $\mathrm{Br}_{-1}(\boldsymbol{X})$. Therefore, by the induction hypotheses it may be written as a finite sum of primitive brackets. Thus the term $[X_0, [B'', X_{a_1}]]$, and hence B, may be written as a finite sum of primitive brackets.

Now suppose that $B' = [X_{a_2}, B'']$ with $B'' \in \mathrm{Br}_{+2}(\boldsymbol{X})$. First look at the case where $B'' = [X_0, B''']$ with $B''' \in \mathrm{Br}_{+1}(\boldsymbol{X})$. In this case we have

$$\begin{aligned} B =& [X_{a_1}, [X_{a_2}, [X_0, B''']]] = -[X_{a_1}, [B''', [X_{a_2}, X_0]]] - [X_{a_1}, [X_0, [B''', X_{a_2}]]] \\ =& [[X_{a_2}, X_0], [X_{a_1}, B''']] + [B''', [[X_{a_2}, X_0], X_{a_1}]] + \\ & [[B''', X_{a_2}], [X_{a_1}, X_0]] + [X_0, [[B''', X_{a_2}], X_{a_1}]]. \end{aligned}$$

The first, third and fourth terms can be written as finite sums of primitive brackets by the induction hypothesis, and the second term is zero by our condition that brackets in $\mathrm{Br}_{-2}(\boldsymbol{X})$ are taken to be zero.

If $B'' = [X_{a_3}, B''']$ then we keep stripping factors off of B''' until we encounter an X_0. When we do, we repeatedly apply the above procedure. This proves the lemma. ■

An example is useful in illustrating what is behind the lemma.

Example 4.4 Consider the bracket $B = [X_{m+1}, [X_0, [X_0, X_a]]] \in \mathrm{Br}_0(\boldsymbol{X})$. This bracket is in $\mathrm{Br}_0(\boldsymbol{X})$ but is not primitive. However, by Lemma 4.3, we may B as a finite sum of primitive brackets. Indeed, by Jacobi's identity we have

$$\begin{aligned} B =&[X_{m+1},[X_0,[X_0,X_a]]] = -[[X_0,X_a],[X_{m+1},X_0]] - [X_0,[[X_0,X_a],X_{m+1}]] \\ =&[[X_0,X_a],[X_0,X_{m+1}]] + [X_0,[X_{m+1},[X_0,X_a]]]. \end{aligned}$$ □

Now we relate the free Lie algebra $L(\boldsymbol{X})$ with a free Lie algebra which corresponds to the set $\{X_L, Y_1^{lift}, \dots, Y_a^{lift}\}$. Let $\boldsymbol{X}' = \{X_0', \dots, X_m'\}$. We formally set $X_0' = X_0 - X_{m+1}$ and $X_a' = X_a$ for $a = 1, \dots, m$. We may now write brackets in $\mathrm{Br}(\boldsymbol{X}')$ as linear combinations of brackets in $\mathrm{Br}(\boldsymbol{X})$ by $\mathbb{R}$-linearity of the bracket. We may, in fact, be even more precise about this.

Let $B' \in \mathrm{Br}(\boldsymbol{X}')$. We define a subset, $\boldsymbol{S}(B')$, of $\mathrm{Br}(\boldsymbol{X})$ by saying that $B \in \boldsymbol{S}(B')$ if each occurrence of X_a' in B' is replaced with X_a for $a = 1, \dots, m$, and if each occurrence of X_0' in B' is replaced with *either* X_0 *or* X_{m+1}. An example is illustrative. Suppose that

$$B' = [[X_0', X_1'], [X_2', [X_0', X_3']]].$$

Then

$$\begin{aligned} \boldsymbol{S}(B') = \{&[[X_0,X_1],[X_2,[X_0,X_3]]], [[X_0,X_1],[X_2,[X_{m+1},X_3]]], \\ &[[X_{m+1},X_1],[X_2,[X_0,X_3]]], [[X_{m+1},X_1],[X_2,[X_{m+1},X_3]]]\}. \end{aligned}$$

Now we may precisely state how we write brackets in $\mathrm{Br}(\boldsymbol{X}')$.

Lemma 4.5 *Let $B' \in \mathrm{Br}(\boldsymbol{X}')$. Then*

$$B' = \sum_{B \in \boldsymbol{S}(B')} (-1)^{\delta_{m+1}(B)} B.$$

Proof: It suffices to prove the lemma for the case when B' is of the form

$$B' = [X_{a_k}', [X_{a_{k-1}}', [\cdots, [X_{a_2}', X_{a_1}']]]] \tag{4.7}$$

since these brackets generate $L(\boldsymbol{X}')$ by Proposition 2.1. We proceed by induction on k. The lemma is true for $k = 1$. Now suppose the lemma true for $k = 1, \dots, l$ where $l \geq 1$ and let B' be of the form (4.7) with $k = l + 1$. Then either $B' = [X_a', B'']$, $a = 1, \dots, m$ or $B' = [X_0, B'']$ with B'' of the form (4.7) with $k = l$. In the first case, by the induction hypotheses, we have

$$\begin{aligned} B' &= \sum_{B \in \boldsymbol{S}(B'')} [X_a, (-1)^{\delta_{m+1}(B)} B] \\ &= \sum_{B \in \boldsymbol{S}(B')} (-1)^{\delta_{m+1}(B)} B. \end{aligned}$$

In the second case we have

$$
\begin{aligned}
B' &= \sum_{B \in \boldsymbol{S}(B'')} [X_0 - X_{m+1}, (-1)^{\delta_{m+1}(B)} B] \\
&= \sum_{B \in \boldsymbol{S}(B')} (-1)^{\delta_{m+1}(B)} B.
\end{aligned}
$$

This proves the lemma. ■

We shall only be interested in terms in the above decomposition of B' which are in $\mathrm{Br}_0(\boldsymbol{X}) \cup \mathrm{Br}_{-1}(\boldsymbol{X})$ since, as we shall see in Section 4.1.4, these are the only ones which will contribute to $\mathrm{Ev}_{0_q}(\phi')(B')$.

A good understanding of this section is important in any effort to understand the proofs of Proposition 4.11 and Theorem 4.17 which follow. The reader should come back to this section if they are having difficulty with these proofs.

4.1.3 Some Useful Tangent Bundle Structure

Since we are interested in restricting the accessibility distribution to the zero section of TQ, there are some useful properties of the tangent bundle which we shall need.

Since $Z(TQ)$, the zero section of the tangent bundle, is a submanifold of TQ which is canonically diffeomorphic to Q, it is possible to realise T_qQ as a subspace of $T_{0_q}TQ$. At each point $0_q \in Z(TQ)$ we shall call this subspace *horizontal*. Note that this version of horizontal is valid *only* at those points in TQ which are on the zero section. Present as a subspace of $T_{v_q}TQ$ for *any* $v_q \in TQ$ is the *vertical* subspace. Recall that this subspace is the kernel of the map $T_{v_q}\tau_Q$. Also note that at points $0_q \in Z(TQ)$, $T_{0_q}TQ = T_qQ \oplus V_{0_q}Q$. By T_qQ in this decomposition we mean the horizontal subspace of $T_{0_q}TQ$ which is canonically isomorphic to T_qQ. The reader should be aware that this identification will be implicitly made in the sequel.

4.1.4 Distribution Computations for Simple Mechanical Control Systems

In this section we use the simplifications of Section 4.1.2 to get a complete description of the brackets which contribute to the accessibility distribution for (4.3) restricted to $Z(TQ)$. To make the correspondence between the free Lie algebra $L(\boldsymbol{X})$ used in Section 4.1.2 and the accessibility algebra for (4.3), we define a family of vector fields

$$
\mathcal{V} = \{Z_g, Y_1^{lift}, \ldots, Y_m^{lift}, \operatorname{grad} V^{lift}\}
$$

and establish a bijection, ϕ, from $\boldsymbol{X}$ to $\mathcal{V}$ by mapping X_0 to X_L, X_a to Y_a^{lift} for $a = 1, \ldots, m$, and X_{m+1} to $\operatorname{grad} V^{lift}$. Please note that $\mathcal{V}$ is *not* the family of vector fields which generates the accessibility algebra. The accessibility algebra is generated by the family $\mathcal{V}' = \{X_L, Y_1^{lift}, \ldots, Y_m^{lift}\}$. We establish a bijection, ϕ', from $\boldsymbol{X}'$ to $\mathcal{V}'$ by mapping X_0' to X_L and X_a' to Y_a^{lift} for $a = 1, \ldots, m$. By Lemma 4.5, each vector field in $\overline{\mathrm{Lie}}(\mathcal{V}')$ is a $\mathbb{R}$-linear sum of vector fields in $\overline{\mathrm{Lie}}(\mathcal{V})$.

Now we shall show that it is possible to compute the brackets from $\mathrm{Br}(\boldsymbol{X})$ in terms of the problem data. We first present a lemma which gives the basic structure of primitive brackets. In this lemma we see that a large number of brackets are computable in terms of quantities defined on Q. This is worth noting since the vector fields themselves are defined on TQ. Of particular interest in the lemma is the appearance of the covariant derivative which was introduced in Section 2.7.1.

Lemma 4.6 *Suppose that $B \in \mathrm{Br}^k(\boldsymbol{X})$ is primitive.*

i) If $B \in \mathrm{Br}_{-1}(\boldsymbol{X})$ then $\mathrm{Ev}(\phi)(B)$ is the vertical lift of a vector field on Q.

ii) If $B \in \mathrm{Br}_0(\boldsymbol{X})$ then $U = \mathrm{Ev}(\phi)(B)$ has the property that, when expressed in a local chart, the vertical components of U are linear in the fibre coordinates v and the horizontal components are independent of v. In particular, we may define a vector field on Q by $U_Q \colon q \mapsto U(0_q) \in T_qQ \subset T_{0_q}TQ$. There are two cases to consider.

a) $B = [X_0, B_1]$ with $B_1 \in \mathrm{Br}_{-1}(\boldsymbol{X})$: Define U_1 to be the vector field on Q such that $\mathrm{Ev}(\phi)(B_1) = U_1^{lift}$. Then $U(0_q) = -U_1(q)$. Let $U_2 \in \mathfrak{X}(Q)$. Then $[U_2^{lift}, U] = (\nabla_{U_1} U_2 + \nabla_{U_2} U_1)^{lift}$.

b) $B = [B_1, B_2]$ with $B_1, B_2 \in \mathrm{Br}_0(\boldsymbol{X})$: Define $U_{1,Q}, U_{2,Q}$ to be the vector fields on Q corresponding to $\mathrm{Ev}(\phi)(B_1), \mathrm{Ev}(\phi)(B_2)$, respectively. Then $\mathrm{Ev}(\phi)(B)(0_q) = [U_{1,Q}, U_{2,Q}](q)$.

Proof: The proof is by induction on k. The result is true for $k = 1$ trivially. To prove the result for $k = 2$ we introduce some notation which we will find handy for doing the bracket calculations in coordinates. If we have two *general* vector fields

$$X_1 = X^i_{1,h}(q,v)\frac{\partial}{\partial q^i} + X^i_{1,v}(q,v)\frac{\partial}{\partial v^i}, \qquad X_2 = X^i_{2,h}(q,v)\frac{\partial}{\partial q^i} + X^i_{2,v}(q,v)\frac{\partial}{\partial v^i},$$

their Lie bracket will be represented by

$$[X_1, X_2] \sim \begin{bmatrix} \frac{\partial X^i_{2,h}}{\partial q^j} & \frac{\partial X^i_{2,h}}{\partial v^j} \\ \frac{\partial X^i_{2,v}}{\partial q^j} & \frac{\partial X^i_{2,v}}{\partial v^j} \end{bmatrix} \begin{pmatrix} X^j_{1,h} \\ X^j_{1,v} \end{pmatrix} - \begin{bmatrix} \frac{\partial X^i_{1,h}}{\partial q^j} & \frac{\partial X^i_{1,h}}{\partial v^j} \\ \frac{\partial X^i_{1,v}}{\partial q^j} & \frac{\partial X^i_{1,v}}{\partial v^j} \end{bmatrix} \begin{pmatrix} X^j_{2,h} \\ X^j_{2,v} \end{pmatrix}.$$

This is somewhat imprecise, but is convenient notationally.

If X, Y are vector fields on Q we may compute

$$[X^{lift}, Y^{lift}] \sim \begin{bmatrix} 0 & 0 \\ \frac{\partial Y^i}{\partial q^j} & 0 \end{bmatrix} \begin{pmatrix} 0 \\ X^j \end{pmatrix} - \begin{bmatrix} 0 & 0 \\ \frac{\partial X^i}{\partial q^j} & 0 \end{bmatrix} \begin{pmatrix} 0 \\ Y^j \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{4.8}$$

If X is a vector field on Q we compute

$$[Z_g, X^{lift}] \sim \begin{bmatrix} 0 & 0 \\ \frac{\partial X^i}{\partial q^j} & 0 \end{bmatrix} \begin{pmatrix} v^j \\ -\Gamma^j_{kl} v^k v^l \end{pmatrix} - \begin{bmatrix} 0 & \delta^i_j \\ -\frac{\partial \Gamma^i_{kl}}{\partial q^j} & -2\Gamma^i_{jk} v^k \end{bmatrix} \begin{pmatrix} 0 \\ X^j \end{pmatrix}. \tag{4.9}$$

Inspecting (4.9) shows that $[Z_g, X^{lift}](0_q) = -X(q)$. Now let $Y \in \mathfrak{X}(Q)$. We compute

$$[Y^{lift}, [Z_g, X^{lift}]] \sim \begin{bmatrix} -\frac{\partial X^i}{\partial q^j} & 0 \\ \frac{\partial^2 X^i}{q^j q^k} v^k + 2\frac{\partial \Gamma^i_{kl}}{\partial q^j} X^k v^l + 2\Gamma^i_{kl}\frac{\partial X^k}{\partial q^j} v^l & \frac{\partial X^i}{\partial q^j} + 2\Gamma^i_{kj} X^k \end{bmatrix} \begin{pmatrix} 0 \\ Y^j \end{pmatrix} - \begin{bmatrix} 0 & 0 \\ \frac{\partial Y^i}{\partial q^j} & 0 \end{bmatrix} \begin{pmatrix} -X^j \\ \frac{\partial X^j}{\partial q^k} v^k + 2\Gamma^j_{kl} X^k v^l \end{pmatrix}.$$

Reading the coefficients gives

$$[Y^{lift}, [Z_g, X^{lift}]] = \left(\frac{\partial Y^i}{\partial q^j} X^j + \frac{\partial X^i}{\partial q^j} Y^j + 2\Gamma^i_{jk} X^j Y^k \right) \frac{\partial}{\partial v^i} \tag{4.10}$$

which is the coordinate representation of $(\nabla_X Y + \nabla_Y X)^{lift}$. This shows that the lemma is true for $k = 2$.

Now suppose the lemma true for $k = 1, \ldots, l$ for $l \geq 2$ and let $B \in \mathrm{Br}^{l+1}(\boldsymbol{X})$ be primitive.

i: Suppose that $B \in \mathrm{Br}_{-1}(\boldsymbol{X})$. Without loss of generality (by Prim1) we may suppose that $B = [B_1, B_2]$ with $B_1 \in \mathrm{Br}_{-1}(\boldsymbol{X})$ and $B_2 \in \mathrm{Br}_0(\boldsymbol{X})$. Then, by the induction hypotheses, we have

$$\mathrm{Ev}(\phi)(B_1) = \alpha^i(q)\frac{\partial}{\partial v^i}, \qquad \mathrm{Ev}(\phi)(B_2) = \lambda^i(q)\frac{\partial}{\partial q^i} + \mu^i_j(q) v^j \frac{\partial}{\partial v^i}.$$

Now we compute

$$\mathrm{Ev}(\phi)([B_1, B_2]) \sim \begin{bmatrix} \frac{\partial \lambda^i}{\partial q^j} & 0 \\ \frac{\partial \mu^i_k}{\partial q^j} v^k & \mu^i_j \end{bmatrix} \begin{pmatrix} 0 \\ \alpha^j \end{pmatrix} - \begin{bmatrix} 0 & 0 \\ \frac{\partial \alpha^i}{\partial q^j} & 0 \end{bmatrix} \begin{pmatrix} \lambda^j \\ \mu^j_k v^k \end{pmatrix}.$$

Note that the components in the q-direction are zero and the components in the v-direction are only functions of q. This means that this vector field is the vertical lift of a vector field on Q. This proves i.

ii: Suppose that $B \in \mathrm{Br}_0(\boldsymbol{X})$. Without loss of generality (by Prim2) we may suppose that either (a) $B = [X_0, B_1]$ with $B_1 \in \mathrm{Br}_{-1}(\boldsymbol{X})$ or that (b) $B = [B_1, B_2]$ with $B_1, B_2 \in \mathrm{Br}_0(\boldsymbol{X})$. Let us deal with the first case. Equation (4.9) gives $\mathrm{Ev}(B)(\phi)(0_q) = -U_1(q)$ where U_1 is the vector field on Q so that $\mathrm{Ev}(\phi)(B_1) = U_1^{lift}$ (such a vector field exists by i). For every vector field U_2 on Q we have $[U_2^{lift}, [Z_g, U_1^{lift}]] = (\nabla_{U_1} U_2 + \nabla_{U_2} U_1)^{lift}$ by (4.10). This proves ii(a).

Now suppose that we have $B_1, B_2 \in \mathrm{Br}_0(\boldsymbol{X})$. Then, by the induction hypotheses, we have

$$\mathrm{Ev}(\phi)(B_1) = \alpha^i(q)\frac{\partial}{\partial q^i} + \beta^i_j(q) v^j \frac{\partial}{\partial v^i}, \qquad \mathrm{Ev}(\phi)(B_2) = \lambda^i(q)\frac{\partial}{\partial q^i} + \mu^i_j(q) v^j \frac{\partial}{\partial v^i}.$$

We compute

$$\operatorname{Ev}(\phi)([B_1, B_2]) \sim \begin{bmatrix} \frac{\partial \lambda^i}{\partial q^j} & 0 \\ \frac{\partial \mu^i_k}{\partial q^j} v^k & \mu^i_j \end{bmatrix} \begin{pmatrix} \alpha^j \\ \beta^j_k v^k \end{pmatrix} - \begin{bmatrix} \frac{\partial \alpha^i}{\partial q^j} & 0 \\ \frac{\partial \beta^i_k}{\partial q^j} v^k & \beta^i_j \end{bmatrix} \begin{pmatrix} \lambda^j \\ \mu^j_k v^k \end{pmatrix}.$$

The components have the order in v specified by the lemma. Also, it is clear that the vector fields on Q defined by B_1 and B_2 are

$$U_{1,Q} = \alpha^i(q)\frac{\partial}{\partial q^i}, \qquad \text{and} \qquad U_{2,Q} = \lambda^i(q)\frac{\partial}{\partial q^i},$$

respectively. It is easy to see that $\operatorname{Ev}(\phi)(B)(0_q) = [U_{1,Q}, U_{2,Q}](q)$. This completes the proof of the lemma. ■

This lemma provides us with a strong step towards computing the value of all primitive brackets when evaluated using $\operatorname{Ev}(\phi)$. Next we show that these are the *only* types of brackets we need to consider. First we look at brackets in $\operatorname{Br}_l(\boldsymbol{X})$ for $l \geq 1$.

Lemma 4.7 *Let $l \geq 1$ be an integer and let $B \in \operatorname{Br}_l(\boldsymbol{X})$. Then $\operatorname{Ev}(\phi)(B)(0_q) = 0$ for each $q \in Q$.*

Proof: The lemma may be proved by showing that, in a coordinate chart for TQ, the horizontal components of $U = \operatorname{Ev}(\phi)(B)$ are polynomial in the fibre coordinates of degree l, and the vertical components of U are polynomial of degree $l+1$ in the fibre coordinates. This will follow if we can show that bracketing by X_a, $a = 1, \ldots, m$ reduces the polynomial order of the components by one and bracketing by X_0 increases the polynomial order of the components by one. This is a simple calculation which follows along the same lines as the calculations done for Lemma 4.6. ■

Now we look at the remaining brackets, those in $\operatorname{Br}_{-l}(\boldsymbol{X})$ for $l \geq 2$.

Lemma 4.8 *Let $l \geq 2$ be an integer and let $B \in \operatorname{Br}^k(\boldsymbol{X}) \cap \operatorname{Br}_{-l}(\boldsymbol{X})$ for $k \geq 2$. Then $\operatorname{Ev}(\phi)(B) = 0$.*

Proof: We prove the lemma by induction on k for brackets of the form (4.6). The result makes no sense for $k = 1$ and is true for $k = 2$ by (4.8). Now suppose the lemma true for $k = 2, \ldots, j$ and let $B \in \operatorname{Br}^{j+1}(\boldsymbol{X}) \cap \operatorname{Br}_{-l}(\boldsymbol{X})$ for $l \geq 2$ be of the form (4.6). Then either $B = [X_0, B']$ with $B' \in \operatorname{Br}_{-l-1}(\boldsymbol{X})$ or $B = [X_a, B']$ with $B' \in \operatorname{Br}_{-l+1}(\boldsymbol{X})$ and $a = 1, \ldots, m+1$. In either case the result follows immediately from the induction hypotheses and (4.8). ■

Let us summarise what we have done in this section. First we obtained a characterisation of primitive brackets in $\boldsymbol{X}$ when we evaluate them in $\mathcal{V}$ via $\operatorname{Ev}(\phi)$. This characterisation involved Lie brackets and covariant derivatives of the vector fields $Y_1, \ldots, Y_m, \operatorname{grad} V$. Then we showed in Lemmas 4.7 and 4.8 that the primitive brackets are the only ones we need be concerned with if we are evaluating the vector fields on the zero section of TQ.

4.1.5 The Form of the Accessibility Distribution Restricted to $Z(TQ)$ for Simple Mechanical Control Systems

In this section we compute the accessibility distribution for (4.3) when restricted to the zero section of TQ. By Lemma 4.5 we know that we may write the vector fields in the accessibility algebra in terms of vector fields in $\overline{\mathrm{Lie}}(\mathcal{V})$. In Section 4.1.4 we saw some hints that we might be able to write vector fields in $\overline{\mathrm{Lie}}(\mathcal{V})$ in terms of covariant derivatives and Lie brackets of the input vector fields and $\operatorname{grad} V$. First we resolve this issue by saying exactly what the vector fields in $\overline{\mathrm{Lie}}(\mathcal{V})$ look like when we restrict them to $Z(TQ)$. Recall from Section 2.4.2 that $D_{\overline{\mathrm{Lie}}(\mathcal{V})}$ is the distribution defined by

$$D_{\overline{\mathrm{Lie}}(\mathcal{V})}(v) = \left\langle U(v) \mid \quad U \in \overline{\mathrm{Lie}}(\mathcal{V}) \right\rangle_{\mathbb{R}}.$$

The reader will also wish to recall the ideas from symmetric algebras presented in Section 2.7.2. In particular recall that the symmetric product on Q is defined by

$$\langle U_1 : U_2 \rangle = \nabla_{U_1} U_2 + \nabla_{U_2} U_1$$

for $U_1, U_2 \in \mathfrak{X}(Q)$. We denote $\mathcal{Y} = \{Y_1, \ldots, Y_m\}$. Recall from Section 4.1.3 that T_qQ may be canonically included in $T_{0_q}TQ$. Also recall from that section that VTQ is the bundle of vertical vectors on TQ.

Lemma 4.9 *Let $q \in Q$. Then*

$$D_{\overline{\mathrm{Lie}}(\mathcal{V})}(0_q) \cap V_{0_q}TQ = (D_{\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\})}(q))^{lift}$$

and

$$D_{\overline{\mathrm{Lie}}(\mathcal{V})}(0_q) \cap T_qQ = D_{\overline{\mathrm{Lie}}(\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\}))}(q).$$

Proof: From Lemmas 4.7 and 4.8 we know that the only brackets from $\mathrm{Br}(\boldsymbol{X})$ which we need to consider are the primitive brackets. From Lemma 4.6 we know that the brackets which are in $\mathrm{Br}_{-1}(\boldsymbol{X})$ will generate the vertical directions, and the brackets which are in $\mathrm{Br}_0(\boldsymbol{X})$ will generate the horizontal directions.

First we show that $(D_{\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\})}(q))^{lift} \subset D_{\overline{\mathrm{Lie}}(\mathcal{V})}(0_q)$. This may be done inductively. Define $\mathrm{Sym}^{(1)}(\mathcal{Y} \cup \{\operatorname{grad} V\}) = \mathcal{Y} \cup \{\operatorname{grad} V\}$ and inductively define

$$\mathrm{Sym}^{(k)}(\mathcal{Y} \cup \{\operatorname{grad} V\}) = \{\langle U_1 : U_2 \rangle \mid \\ U_i \in \mathrm{Sym}^{(k_i)}(\mathcal{Y} \cup \{\operatorname{grad} V\}),\ k_1 + k_2 = k\}.$$

Clearly

$$\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\}) = \bigcup_{k \in \mathbb{Z}^+} \mathrm{Sym}^{(k)}(\mathcal{Y} \cup \{\operatorname{grad} V\}).$$

It is trivially true that $(\mathrm{Sym}^{(1)}(\mathcal{Y} \cup \{\operatorname{grad} V\}))^{lift} \subset \overline{\mathrm{Lie}}(\mathcal{V})$. Now suppose that $(\mathrm{Sym}^{(k)}(\mathcal{Y} \cup \{\operatorname{grad} V\}))^{lift} \subset \overline{\mathrm{Lie}}(\mathcal{V})$ for $k = 1, \ldots, l$ for $l \geq 1$. We see that $(\mathrm{Sym}^{(l+1)}(\mathcal{Y} \cup \{\operatorname{grad} V\}))^{lift} \subset \overline{\mathrm{Lie}}(\mathcal{V})$ since we may generate all elements of $(\mathrm{Sym}^{(l+1)}(\mathcal{Y} \cup \{\operatorname{grad} V\}))^{lift}$ by considering brackets of the form $[U_1^{lift}, [Z_g, U_2^{lift}]]$ where $U_i \in \mathrm{Sym}^{(l_i)}(\mathcal{Y}, V)$ and $l_1 + l_2 = l + 1$. This follows from (4.10). This shows that $(D_{\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\})}(q))^{lift} \subset D_{\overline{\mathrm{Lie}}(\mathcal{V})}(0_q)$.

Now we show that $D_{\overline{\mathrm{Lie}}(\mathcal{V})}(0_q) \subset (D_{\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\})}(q))^{lift}$. To do this we must show that the image under $\mathrm{Ev}(\phi)$ of all primitive brackets in $\mathrm{Br}_{-1}(\boldsymbol{X})$ may be written as a linear combination of vector fields in $\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\})$. A primitive bracket in $\mathrm{Br}_{-1}(\boldsymbol{X})$ may be written as $B = [B_1, B_2]$ with $B_1 \in \mathrm{Br}_{-1}(\boldsymbol{X})$ and $B_2 \in \mathrm{Br}_0(\boldsymbol{X})$ both being primitive. Therefore, either $B_2 = [X_0, B_2']$ with B_2' primitive and in $\mathrm{Br}_{-1}(\boldsymbol{X})$ or $B_2 = [B_2', B_2'']$ with $B_2', B_2'' \in \mathrm{Br}_0(\boldsymbol{X})$ both primitive. In the first case $\mathrm{Ev}(\phi)(B) \in \mathrm{Sym}^{(k)}(\mathcal{Y} \cup \{\operatorname{grad} V\})$ for some k by (4.10). In the second case we may use Jacobi's identity to obtain

$$B = -[B_2'', [B_1, B_2']] + [B_2', [B_1, B_2'']].$$

We may apply the above argument to the terms $[B_1, B_2']$ and $[B_1, B_2'']$ repeatedly using (4.10) until they are expressed in terms of covariant derivatives. When this is done, $\mathrm{Ev}(\phi)(B)$ will then be a $\mathbb{R}$-linear combination of elements in $\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\})$. This shows that $D_{\overline{\mathrm{Lie}}(\mathcal{V})}(0_q) \subset (D_{\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\})}(q))^{lift}$.

To demonstrate the proposed form of $D_{\overline{\mathrm{Lie}}(\mathcal{V})} \cap T_qQ$, by Lemma 4.6 ii(b) we need only show that $\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\})(q) \subset D_{\overline{\mathrm{Lie}}(\mathcal{V})}(0_q)$. But this is clear from Lemma 4.6 ii(a). This completes the proof of the lemma. ■

Remark 4.10 Notice that the constructions in the above lemma only depend upon $\{Y_1, \ldots, Y_m, \operatorname{grad} V\}$. The effects of the geodesic spray do not appear explicitly. However, its contribution is obviously important in the essential computations performed in Section 4.1.4. □

From Lemma 4.5 we know that the vector fields which contribute to $\overline{\mathrm{Lie}}(\mathcal{V}')$ when we evaluate on $Z(TQ)$ will be $\mathbb{R}$-linear combinations of vector fields from $\overline{\mathrm{Lie}}(\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\}))$. Thus, to compute these vector fields, we need to figure out which vector fields need to be "removed" from $\overline{\mathrm{Lie}}(\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\}))$. We present an algorithm which we shall prove determines exactly which $\mathbb{R}$-linear combinations from $\overline{\mathrm{Lie}}(\overline{\mathrm{Sym}}(\mathcal{Y} \cup \{\operatorname{grad} V\}))$ we need to compute. We define *two* sequences of families of vector fields on Q which we shall denote by $\mathcal{C}_{ver}^{(k)}(\mathcal{Y}, V)$ and $\mathcal{C}_{hor}^{(k)}(\mathcal{Y}, V)$ where $k \in \mathbb{Z}^+$. In Figure 4.2 the algorithm is presented for computing these families. When we have computed these sequences we define

$$\mathcal{C}_{ver}(\mathcal{Y}, V) = \bigcup_{k \in \mathbb{Z}^+} \mathcal{C}_{ver}^{(k)}(\mathcal{Y}, V), \qquad \mathcal{C}_{hor}(\mathcal{Y}, V) = \bigcup_{k \in \mathbb{Z}^+} \mathcal{C}_{hor}^{(k)}(\mathcal{Y}, V).$$

The distributions defined by these families of vector fields shall be denoted $C_{ver}(\mathcal{Y}, V)$ and $C_{hor}(\mathcal{Y}, V)$, respectively.

Algorithm 4.1

For $i \in \mathbb{Z}^+$ do
    For $B \in \mathrm{Br}^{(i)}(\boldsymbol{X})$ primitive do
        If $\delta_{m+1}(B) = 0$ then
            If $B \in \mathrm{Br}_{-1}(\boldsymbol{X})$ then
                $U \in \mathcal{C}_{ver}^{\frac{1}{2}(i+1)}(\mathcal{Y}, V)$ where $\mathrm{Ev}(\phi)(B) = U^{lift}$
            else
                $U \in \mathcal{C}_{hor}^{(i/2)}(\mathcal{Y}, V)$ where $U(q) = \mathrm{Ev}_{0_q}(\phi)(B)$
            end
        else
            If B has no components of the form $[X_0, X_{m+1}]$ then
                Compute $B' \in \mathrm{Br}(\boldsymbol{X})$ by replacing every occurrence of X_0 *and* X_{m+1} in B with X_0' and by replacing every occurrence of X_a in B with X_a' for $a = 1, \dots, m$.
                Let $B'' = 0$.
                For $\tilde{B} \in \boldsymbol{S}(B') \cap (\mathrm{Br}_{-1}(\boldsymbol{X}) \cup \mathrm{Br}_0(\boldsymbol{X}))$ do
                    Write $\tilde{B}$ as a finite sum of primitive brackets in $\mathrm{Br}(\boldsymbol{X})$ by Lemma 4.3.
                    $B'' = B'' + (-1)^{\delta_{m+1}(\tilde{B})}\tilde{B}$
                end
                If $B \in \mathrm{Br}_{-1}(\boldsymbol{X})$ then
                    $U \in \mathcal{C}_{ver}^{\frac{1}{2}(i+1)}(\mathcal{Y}, V)$ where $\mathrm{Ev}(\phi)(B'') = U^{lift}$
                else
                    $U \in \mathcal{C}_{hor}^{(i/2)}(\mathcal{Y}, V)$ where $U(q) = \mathrm{Ev}_{0_q}(\phi)(B'')$
                end
            end
        end
    end
end
end

Figure 4.2 Algorithm for computing $\overline{\mathrm{Lie}}(\mathcal{V}') \mid Z(TQ)$

We may now state the form of the accessibility distribution $\overline{\text{Lie}}(\mathcal{V}')$ for (4.3) when restricted to the zero section of TQ.

Proposition 4.11 *Let $q \in Q$. Then*

$$D_{\overline{\text{Lie}}(\mathcal{V}')}(0_q) \cap V_{0_q}TQ = (C_{ver}(\mathcal{Y}, V)(q))^{lift}$$

and

$$D_{\overline{\text{Lie}}(\mathcal{V}')}(0_q) \cap T_qQ = C_{hor}(\mathcal{Y}, V)(q).$$

Proof: Studying the algorithm that we have used to compute $\mathcal{C}_{ver}(\mathcal{Y}, V)$ and $\mathcal{C}_{hor}(\mathcal{Y}, V)$, the reader will notice that we have exactly taken each primitive bracket $B \in \text{Br}(\boldsymbol{X})$ and computed which $\mathbb{R}$-linear combinations from $\text{Br}(\boldsymbol{X})$ appear along with B in the decomposition of some $B' \in \text{Br}(\boldsymbol{X}')$ given by Lemma 4.5. Since it is only these primitive brackets which appear in $\overline{\text{Lie}}(\mathcal{V}') \mid Z(TQ)$, this will, by construction, generate $D_{\overline{\text{Lie}}(\mathcal{V}')} \mid Z(TQ)$.

We need to prove that, as stated in the first step of the algorithm, if $\delta_{m+1}(B) = 0$, then $\text{Ev}_{0_q}(\phi)(B) \in D_{\overline{\text{Lie}}(\mathcal{V}')}(0_q)$. To show that this is in fact the case, let $B' \in \text{Br}(\boldsymbol{X}')$ be the bracket obtained by replacing X_a with X'_a for $a = 0, \ldots, m$. We claim that the only bracket in $\boldsymbol{S}(B')$ which contributes to $\text{Ev}(\phi')(B')$ is B. This is true since any other brackets in $\boldsymbol{S}(B')$ are obtained by replacing X_0 in B with X_{m+1}. Such a replacement will result in a bracket which has at least one component which is in $\text{Br}_{-l}(\boldsymbol{X})$ for $l \geq 2$. These brackets evaluate to zero by Lemma 4.8.

We also need to show that if B has components of the form $[X_0, X_{m+1}]$ then it will not contribute to $\overline{\text{Lie}}(\mathcal{V}') \mid Z(TQ)$. This is clear since, when constructing B' in the algorithm, the component $[X_0, X_{m+1}]$ will become $[X'_0, X'_0]$ which means that B' will be identically zero. ■

It is perhaps useful to construct a few of the families $\mathcal{C}^{(k)}_{ver}(\mathcal{Y}, V)$ and $\mathcal{C}^{(k)}_{hor}(\mathcal{Y}, V)$ to show how the algorithm works. We shall do this for $k = 1, 2$. Our notation in these calculations follows that in the algorithm.

Let $i = 1$. The only primitive brackets in $\text{Br}^{(1)}(\boldsymbol{X})$ are $X_1, \ldots, X_{m+1}$. For the brackets $B = X_a$, $a = 1, \ldots, m$, $\delta_{m+1}(B) = 0$. Note that $\text{Ev}(\phi)(B) = Y_a^{lift}$ so $Y_a \in \mathcal{C}^{(1)}_{ver}(\mathcal{Y}, V)$ for $a = 1, \ldots, m$. The bracket X_{m+1} has no components of the form $[X_0, X_{m+1}]$ so it is a candidate for providing an element of $\mathcal{C}^{(1)}_{ver}(\mathcal{Y}, V)$. If $B = X_{m+1}$ we compute $B' = X'_0$. Therefore, $\boldsymbol{S}(B') = \{X_0, X_{m+1}\}$. The only element in $\boldsymbol{S}(B')$ which is in $\text{Br}_{-1}(\boldsymbol{X}) \cup \text{Br}_0(\boldsymbol{X})$ is X_{m+1}. Therefore, $B'' = -X_{m+1}$. We then see that $\text{Ev}(\phi)(B'') = -\operatorname{grad} V^{lift}$ from which we conclude that $\operatorname{grad} V \in \mathcal{C}^{(1)}_{ver}(\mathcal{Y}, V)$. In summary,

$$\mathcal{C}^{(1)}_{ver}(\mathcal{Y}, V) = \{Y_1, \ldots, Y_m, \operatorname{grad} V\}.$$

Now we look at the case when $i = 2$. The primitive brackets in $\text{Br}^{(2)}(\boldsymbol{X})$ are $\{[X_0, X_1], \ldots, [X_0, X_{m+1}]\}$. The brackets $B = [X_0, X_a]$, $a = 1, \ldots, m$ have the property that $\delta_{m+1}(B) = 0$. We compute $\text{Ev}_{0_q}(\phi)(B) = -Y_a(q)$ and so conclude that $Y_a \in \mathcal{C}^{(1)}_{hor}(\mathcal{Y}, V)$. The bracket $[X_0, X_{m+1}]$ is not a candidate for providing an element of $\mathcal{C}^{(1)}_{hor}(\mathcal{Y}, V)$ so we have

$$\mathcal{C}^{(1)}_{hor}(\mathcal{Y}, V) = \{Y_1, \ldots, Y_m\}.$$

In a similar manner we may compute

$$\mathcal{C}^{(2)}_{ver}(\mathcal{Y}, V) = \{\langle Y_a : Y_b \rangle \mid \ a, b = 1, \ldots, m\} \cup \{\langle Y_a : \operatorname{grad} V \rangle \mid \ a = 1, \ldots, m\}$$

and

$$\begin{aligned}\mathcal{C}^{(2)}_{hor}(\mathcal{Y}, V) = \mathcal{C}^{(2)}_{ver}(\mathcal{Y}, V) \cup \{[Y_a, Y_b] \mid \ a, b = 1, \ldots, m\} \cup \\ \{2 \langle Y_a : \operatorname{grad} V \rangle + [Y_a, \operatorname{grad} V] \mid \ a = 1, \ldots, m\}.\end{aligned}$$

To compute the terms $2 \langle Y_a : \operatorname{grad} V \rangle + [Y_a, \operatorname{grad} V]$ in $\mathcal{C}^{(2)}_{hor}(\mathcal{Y}, V)$ we have used the computations of Example 4.4.

It would be interesting to be able to derive an inductive formula for computing the families $\mathcal{C}^{(k)}_{ver}(\mathcal{Y}, V)$ and $\mathcal{C}^{(k)}_{hor}(\mathcal{Y}, V)$. However, such an inductive formula appears to be quite complex.

There are some important statements which can easily be made regarding the distributions $C_{hor}(\mathcal{Y}, V)$ and $C_{ver}(\mathcal{Y}, V)$.

Remarks 4.12

1. The generators we have written for $\mathcal{C}^{(k)}_{ver}(\mathcal{Y}, V)$ and $\mathcal{C}^{(k)}_{hor}(\mathcal{Y}, V)$ are not linearly independent. Thus one should be able to generate these families with fewer calculations than are necessary to compute the generators we give. One way to do this is to choose a Philip Hall basis for $L(\boldsymbol{X}')$ and compute the image of these brackets under $\mathrm{Ev}(\phi')$. This will work for any given example. However, we are unable to give the general form for the image of a Philip Hall basis under $\mathrm{Ev}(\phi')$.

2. We claim that $C_{hor}(\mathcal{Y}, V)$ is involutive. Let $B'_1, B'_2 \in \mathrm{Br}(\boldsymbol{X}')$ be brackets which, when evaluated under $\mathrm{Ev}_{0_q}(\phi')$, give vector fields $U_1, U_2 \in \mathcal{C}_{hor}(\mathcal{Y}, V)$. Then the decomposition of B_i given by Lemma 4.5 has the form $B'_i = B_i + \tilde{B}_i$ where $B_i \in \mathrm{Br}_0(\boldsymbol{X})$ and $\tilde{B}_i$ is a sum of brackets in $\mathrm{Br}_j(\boldsymbol{X})$ for $j \geq 2$. Therefore, $[B'_1, B'_2] = [B_1, B_2] + B''$ where B'' is a sum of brackets in $\mathrm{Br}_j(\boldsymbol{X})$ for $j \geq 2$. This shows that $[U_1, U_2] \in \mathcal{C}_{hor}(\mathcal{Y}, V)$. Here we have imposed the condition that brackets in $\mathrm{Br}_{-j}(\boldsymbol{X})$ are taken to be zero for $j \geq 2$ (see Lemma 4.3).

3. An interesting special case, and one that we shall see in the examples in Section 4.3, is that when $V = 0$. In this case we have

$$\mathcal{C}_{ver}(\mathcal{Y}, V) = \overline{\mathrm{Sym}}(\mathcal{Y}), \qquad \mathcal{C}_{hor}(\mathcal{Y}, V) = \overline{\mathrm{Lie}}(\overline{\mathrm{Sym}}(\mathcal{Y})).$$

   This is easily seen in the algorithm by following the path when $\delta_{m+1}(B) = 0$.

4. The calculations of this section and Section 4.1.4 remain valid if we replace $\operatorname{grad} V$ with an arbitrary vector field on Q. □

4.1.6 Controllability Definitions for Simple Mechanical Control Systems

It is possible to simply adopt the controllability definitions given in Section 3.1.1 since our system is of the form (3.1). However, since we are dealing with simple mechanical control systems, it is of more interest to us to know what is happening to the *configurations.* A good example of a question of interest in mechanics is "What is the set of configurations which are reachable from a given configuration if we start at rest?" This is in fact exactly the question we pose.

Definition 4.13 A *solution* of (4.3) is a pair, (c, u), where $c\colon [0, T] \to Q$ is a piecewise smooth curve and $u \in \mathcal{U}$ such that

$$\nabla_{c'(t)} c'(t) = \operatorname{grad} V(c(t)) + u^a(t) Y_a(c(t)).$$ □

Let $q_0 \in Q$ and let U be a neighborhood of q_0. We define

$$\mathcal{R}_Q^U(q_0, T) = \{q \in Q \mid \text{ there exists a solution } (c, u) \text{ of } (4.3) \\ \text{such that } c'(0) = 0_{q_0},\ c(t) \in U \text{ for } t \in [0, T], \text{ and } c'(T) \in T_q Q\}$$

and denote

$$\mathcal{R}_Q^U(q_0, \leq T) = \bigcup_{0 \leq t \leq T} \mathcal{R}_Q^U(q_0, t).$$

Notice that our definitions for reachable configurations do not require us to get to a point in the reachable set at *zero* velocity. They merely ask that we be able to reach that point at *some* velocity. It is, however, required that the initial velocity be zero.

We now introduce our notions of controllability.

Definition 4.14 We shall say that (4.3) is *locally configuration accessible* at $q_0 \in Q$ if there exists $T > 0$ such that $\mathcal{R}_Q^U(q_0, \leq t)$ contains a non-empty open set of Q for all neighborhoods U of q_0 and all $0 < t \leq T$. If this holds for any $q_0 \in Q$ then the system is called *locally configuration accessible.*

We say that (4.3) is *strongly locally configuration accessible* at $q_0 \in Q$ if there exists $T > 0$ such that $\mathcal{R}_Q^U(q_0, t)$ contains a non-empty open set of Q for all neighborhoods U of q_0 and all $0 < t \leq T$. If this holds for any $q_0 \in Q$ then the system is called *strongly locally configuration accessible.*

We say that (4.3) is *small-time locally configuration controllable* (STLCC) at q_0 if it is locally configuration accessible at q_0 and if there exists $T > 0$ such that q_0 is in the interior of $\mathcal{R}_Q^U(q_0, \leq t)$ for every neighborhood U of q_0 and $0 < t \leq T$. If this holds for any $q_0 \in Q$ then the system is called *small-time locally configuration controllable.* □

Note that this definition may be made to apply to any control system which evolves on TQ.

Another definition of controllability may be interesting in some cases. We shall say that $q \in Q$ is an *equilibrium point* for L if $X_L(0_q) = 0$. Let $\mathfrak{E}(L)$ denote the set of equilibrium points for L. We shall say that (4.3) is *equilibrium controllable* if, for $q_1, q_2 \in \mathfrak{E}(L)$, there exists a solution (c, u) of (4.3) where $c\colon [0, T] \to Q$ is such that $c(0) = q_1$, $c(T) = q_2$ and both $c'(0)$ and $c'(T)$ are zero.

4.1.7 Conditions for Controllability of Simple Mechanical Control Systems

In (Lewis and Murray, 1995a) sufficient conditions were presented for local configuration accessibility. Here, since we have a complete description of $\overline{\mathrm{Lie}}(\mathcal{V}') \mid Z(TQ)$, we can give stronger results.

Theorem 4.15 *The control system (4.3) is locally configuration accessible at q if $C_{hor}(\mathcal{Y}, V)(q) = T_qQ$.*

Proof: Following Section 3.1, let C denote the accessibility distribution. Since $C_{hor}(\mathcal{Y}, V)(q) \subset C(0_q)$ by Proposition 4.11, and $C_{hor}(\mathcal{Y}, V)(q) = T_qQ$ by hypothesis, $Z(TQ)$ must be an integral manifold of C. Let Λ be the maximal integral manifold which contains $Z(TQ)$. Since C is the accessibility distribution, Λ must be invariant under the system (4.3) and the system must be locally accessible when restricted to Λ. Thus the set $\mathcal{R}^{\tilde{U}}(0_q, \leq T)$ is open in Λ for every neighborhood $\tilde{U} \subset \Lambda$ of 0_q and for every T sufficiently small. Now let U be a neighborhood of q and define a neighborhood of 0_q in Λ by $\tilde{U} = \tau_Q^{-1}(U) \cap \Lambda$. The set $\tau_Q(\mathcal{R}^{\tilde{U}}(0_q, \leq T))$ is open in Q for T sufficiently small since τ_Q is an open mapping. This proves the theorem. ■

We also have a partial converse to Theorem 4.15 when the potential energy is zero.

Theorem 4.16 *Suppose that $V = 0$ and that (4.3) is locally configuration accessible. Then $C_{hor}(\mathcal{Y}, V)(q) = T_qQ$ for q in an open dense subset of Q.*

Proof: First note that if $C_{hor}(\mathcal{Y}, V)(q_0) = T_{q_0}Q$ then $C_{hor}(\mathcal{Y}, V)(q) = T_qQ$ in a neighborhood of q_0. This proves that the set of points q where $C_{hor}(\mathcal{Y}, V)(q) = T_qQ$ is open. Now suppose that $C_{hor}(\mathcal{Y}, V)(q) \subsetneq T_qQ$ in an open subset U of Q. Then there exists an open subset $\bar{U} \subset U$ so that $\operatorname{rank}(C_{hor}(\mathcal{Y}, V)(q)) = k < n$ for all $q \in \bar{U}$. However, this contradicts local configuration accessibility by Theorem 4.19. Therefore, there can be no open subset of Q on which $C_{hor}(\mathcal{Y}, V)(q) \subsetneq T_qQ$. Thus the set of points q where $C_{hor}(\mathcal{Y}, V)(q) = T_qQ$ is dense. This completes the proof. ■

We may also prove an easy statement about STLCC. We need to say a few things about "good" and "bad" symmetric products. Let $\boldsymbol{Y} = \{X_1, \ldots, X_{m+1}\}$ and establish a bijection $\psi\colon \boldsymbol{Y} \to \mathcal{Y} \cup \{\operatorname{grad} V\}$ by asking that $\psi(X_a) = Y_a$ for $a = 1, \ldots, m$ and $\psi(X_{m+1}) = \operatorname{grad} V$. If $P \in \Pr(\boldsymbol{Y})$ we shall say that P is *bad* if

$\gamma_a(P)$ is even for each $a = 1, \dots, m$. We say that P is *good* if it is not bad. Let S_m denote the permutation group on m symbols. For $\pi \in S_m$ and $P \in \mathrm{Pr}(\boldsymbol{Y})$ define $\bar{\pi}(P)$ to be the bracket obtained by fixing X_{m+1} and sending X_a to $X_{\pi(a)}$ for $a = 1, \dots, m$. Now define

$$\rho(P) = \sum_{\pi \in S_m} \bar{\pi}(P).$$

We may now state the sufficient conditions for STLCC.

Theorem 4.17 *Suppose that* $\mathcal{Y} \cup \{\operatorname{grad} V\}$ *is such that every bad symmetric product in* $\mathrm{Pr}(\boldsymbol{Y})$ *has the property that*

$$\mathrm{Ev}_{0_q}(\psi)(\rho(P)) = \sum_{a=1}^{m} \xi_a \, \mathrm{Ev}_{0_q}(\psi)(C_a)$$

where C_a *are good symmetric products in* $\mathrm{Pr}(\boldsymbol{Y})$ *of lower degree than* P *and* $\xi_a \in \mathbb{R}$ *for* $a = 1, \dots, m$. *Also, suppose that (4.3) is locally configuration accessible at* q. *Then (4.3) is STLCC at* q.

Proof: First recall from the proof of Theorem 4.15 that if (4.3) is locally configuration accessible at q, then $Z(TQ)$ is an integral manifold for the accessibility distribution. We let Λ be the maximal integral manifold for the accessibility distribution which contains $Z(TQ)$. Restricted to Λ, (4.3) is locally accessible. To show that (4.3) is STLCC at q, it clearly suffices to show that (4.3) is STLC at 0_q when restricted to Λ. We do this by showing that (4.3) satisfies the hypotheses of Theorem 3.29 if it satisfies the stated hypotheses on the symmetric products. To do this we shall show that there is a 1–1 correspondence between bad brackets in $\mathrm{Br}(\boldsymbol{X}')$ and bad symmetric products in $\mathrm{Pr}(\boldsymbol{Y})$ and good brackets in $\mathrm{Br}(\boldsymbol{X}')$ and good symmetric products in $\mathrm{Pr}(\boldsymbol{Y})$.

Suppose that $B' \in \mathrm{Br}(\boldsymbol{X}')$ is bad. Thus $\delta_a(B')$ is even for $a = 1, \dots, m$ and $\delta_0(B')$ is odd. When we evaluate $\mathrm{Ev}_{0_q}(\phi')(B')$, the only terms that will remain in the decomposition of $\mathrm{Ev}(\phi')(B')$ given by Lemma 4.5 are the terms obtained from brackets in $\boldsymbol{S}(B')$ which are in $\mathrm{Br}_0(\boldsymbol{X}) \cup \mathrm{Br}_{-1}(\boldsymbol{X})$. Since B' is bad, we must have $\delta_a(B)$ even and $\delta_0(B) + \delta_{m+1}(B)$ odd for each $B \in \boldsymbol{S}(B')$. If $\delta_0(B)$ is odd then $\delta_{m+1}(B)$ must be even. In this case we get $\sum_{a=1}^{m+1} \delta_a(B)$ as even and $\delta_0(B)$ as odd. Thus the only brackets in $\boldsymbol{S}(B')$ which contribute to $\mathrm{Ev}(\phi')(B')$ must be in $\mathrm{Br}_{-1}(\boldsymbol{X})$. This will give us a vector in $V_{0_q}TQ$ which comes from a symmetric product which is bad. Now suppose that $\delta_0(B)$ is even for $B \in \boldsymbol{S}(B')$. Then $\delta_{m+1}(B)$ must be odd. In this case $\sum_{a=1}^{m+1} \delta_a(B)$ is odd and $\delta_0(B)$ is even and again, the only brackets in $\boldsymbol{S}(B')$ which contribute to $\mathrm{Ev}(\phi')(B')$ must be in $\mathrm{Br}_{-1}(\boldsymbol{X})$. We then conclude that $\mathrm{Ev}_{0_q}(\phi')(B')$ must be of the form $(\mathrm{Ev}_q(\psi)(P))^{\mathit{lift}}$ where $P \in \mathrm{Pr}(\boldsymbol{Y})$ is bad.

Now suppose that $B' \in \mathrm{Br}(X')$ is good. It is clear that if $\delta_a(B')$ is odd for any $a = 1, \dots, m$ then B' cannot give rise to a bad symmetric product. Thus we may suppose that $\delta_a(B')$ is even for each $a = 0, \dots, m$. Now let's look at what the brackets look like from $\boldsymbol{S}(B')$ which contribute to $\mathrm{Ev}(\phi')(B')$. Let B be such a bracket. We must have $\delta_a(B)$ even for $a = 1, \ldots, m$ and $\delta_0(B) + \delta_{m+1}(B)$ even. If $\delta_0(B)$ is odd then $\delta_{m+1}(B)$ must be odd. Since B is primitive this means that $\sum_{a=1}^{m+1} \delta_a(B)$ and $\delta_0(B)$ are odd. Therefore, B must be in $\mathrm{Br}_0(\boldsymbol{X})$. Now suppose that $\delta_0(B)$ is even. Then $\delta_{m+1}(B)$ must also be even. Thus $\sum_{a=1}^{m+1} \delta_a(B)$ and $\delta_0(B)$ are even and so $B \in \mathrm{Br}_0(\boldsymbol{X})$. Therefore, good brackets from $\mathrm{Br}(\boldsymbol{X}')$ do not generate any bad symmetric products. ■

We make some observations about the results of this section.

Remarks 4.18

1. Notice that Theorem 4.15 explains the example from Section 4.1.1. More precisely, we have shown that it is not necessary to be able to generate *all* directions on TQ to obtain controllability in the configuration variables. Indeed, the only vertical directions we generate are $C_{ver}(\mathcal{Y}, V)$ which need not span $V_{0_q}TQ$.

2. Note that the result we have proved for STLCC in Theorem 4.17 is stronger than the definition for STLCC. In fact, it is true that, starting from rest at q_0, we may reach a neighborhood of q_0 *at rest*. In particular, if $\mathfrak{C}(L) \neq \emptyset$, then (4.3) is equilibrium controllable if it satisfies the hypotheses of Theorem 4.17. This result may be made even stronger if we allow a point $q \in Q$ to be an equilibrium point if $\operatorname{grad} V(q)$ is in the span of the inputs at q. □

4.1.8 Decompositions for Simple Mechanical Control Systems

Now we give decomposition results which mirror Propositions 3.12 and 3.13. Our first result gives a decomposition which is valid for systems with no potential energy.

Theorem 4.19 *Suppose that $V = 0$ for the control system (4.3) and suppose that $C_{hor}(\mathcal{Y}, V)$ has constant rank k in a neighborhood of $q_0 \in Q$. Then there exists a coordinate chart, (U, ϕ), around q_0 such that the submanifold*

$$S_{q_0} = \{q \in U \mid \; q^i(q) = q^i(q_0), \;\; i = k+1, \ldots, n\}$$

is an integral manifold of $C_{hor}(\mathcal{Y}, V)$. Then, for any neighborhood $W \subset U$ of q_0 and for all $T > 0$ sufficiently small, $\mathcal{R}_Q^W(q_0, T)$ is contained in S_{q_0}. Hence the system restricted to S_{q_0} is locally configuration accessible.

Proof: The coordinate decomposition exists since $C_{hor}(\mathcal{Y}, V)$ is integrable as pointed out in Remark 4.12(2). Since $V = 0$, we have $\mathfrak{C}_{ver}(\mathcal{Y}, V) = \overline{\mathrm{Sym}}(\mathcal{Y})$ and $\mathfrak{C}_{hor}(\mathcal{Y}, V) = \overline{\mathrm{Lie}}(\overline{\mathrm{Sym}}(\mathcal{Y}))$ as in Remark 4.12(3). This implies that $\mathfrak{C}_{ver}(\mathcal{Y}, V) \subset \mathfrak{C}_{hor}(\mathcal{Y}, V)$ and so all solutions of (4.3) which start on S_{q_0} with zero initial velocity will remain on S_{q_0}. Thus $\mathcal{R}_Q^W(q_0, T) \subset S_{q_0}$. It is also clear that the system is locally configuration accessible when restricted to initial conditions in S_{q_0} since $\dim(S_{q_0}) = \operatorname{rank}(C_{hor}(\mathcal{Y}, V) \mid S_{q_0})$. ■

Now we give a result which gives the form of the equations on the integral manifolds of $C_{hor}(\mathcal{Y}, V)$ when the potential energy is non-zero.

Theorem 4.20 *Suppose that $C_{hor}(\mathcal{Y}, V)$ has constant rank k in a neighborhood of $q_0 \in Q$. Then there exists coordinates $(x^1, \ldots, x^k, y^1, \ldots, y^{n-k})$ so that the system (4.3) has the form*

$$\ddot{x}^i + \Gamma^i_{jk}(x,y)\dot{x}^j\dot{x}^k + \Gamma^i_{j\alpha}(x,y)\dot{x}^j\dot{y}^\alpha + \Gamma^i_{\alpha\beta}(x,y)\dot{y}^\alpha\dot{y}^\beta + g^{ij}\frac{\partial V}{\partial x^j} + g^{i\alpha}\frac{\partial V}{\partial y^\alpha} = u^a Y^i_a$$
$$\ddot{y}^\alpha + \Gamma^\alpha_{j\beta}(x,y)\dot{x}^j\dot{y}^\beta + \Gamma^\alpha_{\beta\gamma}(x,y)\dot{y}^\beta\dot{y}^\gamma + g^{\alpha j}\frac{\partial V}{\partial x^j} + g^{\alpha\beta}\frac{\partial V}{\partial y^\beta} = 0.$$

Furthermore, for each fixed value of y, the control system

$$\ddot{x}^i + \Gamma^i_{jk}(x,y)\dot{x}^j\dot{x}^k + g^{ij}\frac{\partial V}{\partial x^j} + g^{i\alpha}\frac{\partial V}{\partial y^\alpha} = u^a Y^i_a$$

is locally configuration accessible.

Proof: Since $C_{hor}(\mathcal{Y}, V)$ has constant rank in a neighborhood of q_0 and $C_{hor}(\mathcal{Y}, V)$ is integrable, by Frobenius' theorem we may find coordinates $(x^1, \ldots, x^k, y^1, \ldots, y^{n-k})$ for Q so that

$$C_{hor}(\mathcal{Y}, V)(q_0) = \left\langle \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^k} \right\rangle_{\mathbb{R}}.$$

In general, the equations (4.3) in these coordinates will have the form

$$\ddot{x}^i + \Gamma^i_{jk}(x,y)\dot{x}^j\dot{x}^k + \Gamma^i_{j\alpha}(x,y)\dot{x}^j\dot{y}^\alpha + \Gamma^i_{\alpha\beta}(x,y)\dot{y}^\alpha\dot{y}^\beta + g^{ij}\frac{\partial V}{\partial x^j} + g^{i\alpha}\frac{\partial V}{\partial y^\alpha} = u^a Y^i_a \tag{4.11a}$$
$$\ddot{y}^\alpha + \Gamma^\alpha_{jk}(x,y)\dot{x}^j\dot{x}^k + \Gamma^\alpha_{j\beta}(x,y)\dot{x}^j\dot{y}^\beta + \Gamma^\alpha_{\beta\gamma}(x,y)\dot{y}^\beta\dot{y}^\gamma + g^{\alpha j}\frac{\partial V}{\partial x^j} + g^{\alpha\beta}\frac{\partial V}{\partial y^\beta} = 0. \tag{4.11b}$$

We claim that the term $\Gamma^\alpha_{jk}(x,y)\dot{x}^j\dot{x}^k$ in (4.11b) must be zero. This follows from Theorem 4.19 proving the given form of the decomposition. That the top system is locally configuration accessible follows from the fact that $\operatorname{rank}(C_{hor}(\mathcal{Y}, V)) = k$. (It makes sense to speak of local configuration accessibility of this system by Remark 4.12(4) and the statement immediately following Definition 4.14.) ■

Remark 4.21 In Theorem 4.19 the act of restricting to S_{q_0} has specific meaning. We may pull-back the Riemannian metric to S_{q_0} since it is a submanifold of Q. Doing so defines a Riemannian metric on S_{q_0}. This defines a simple mechanical control system (with zero potential energy) on S_{q_0} and, as long as we begin with zero initial velocity, the trajectories of this control system will be the same as those of the larger system. □

4.2 Decompositions for Hamiltonian Control Systems

In this section we investigate mechanical control systems in the Hamiltonian framework. We look at Hamiltonian systems which evolve on a general symplectic manifold rather than just on a cotangent bundle. With this more general structure, it is natural to restrict the control problem to one which fully respects the symplectic properties of the phase space. To this end we shall suppose that the control vector fields are Hamiltonian vector fields. In this section, for simplicity, we shall assume that all distributions have constant rank.

First we discuss some relevant reductions of symplectic manifolds to smaller symplectic manifolds and Poisson manifolds. In Section 4.2.1 we present the reduction which will give the locally accessible dynamics in Proposition 4.28. The reduction which leads to the strongly locally inaccessible dynamics of Proposition 4.30 is discussed in Section 4.2.2.

Note In this section, if D is a distribution on a symplectic manifold (P, Ω), then we will denote $D^\perp = \Omega^\perp D$ in order to simplify the notation. □

4.2.1 Reduction of Symplectic Manifolds by Invariant Foliations

The reduction we discuss is from (Abraham and Marsden, 1978) and was originally developed in (Weinstein, 1977).

We suppose that we have an integrable distribution D on P. Denote by $\mathcal{F}_D$ the maximal foliation corresponding to this integrable distribution. From the distribution D we may compute, in a natural way, a smaller distribution which is also integrable.

Lemma 4.22 *The distribution $D \cap D^\perp$ is an integrable distribution on P.*

Proof: Suppose that X_1, X_2 are sections of $D \cap D^\perp$. We have

$$\begin{aligned}\boldsymbol{d}\Omega(X_1, X_2, Y) = X_2(\Omega(X_1, Y)) - X_1(\Omega(X_2, Y)) - Y(\Omega(X_1, X_2)) +\\ \Omega([X_1, X_2], Y) - \Omega([X_1, Y], X_2) + \Omega([X_2, Y], X_1).\end{aligned}$$

for all sections Y of D. Using the facts that Ω is closed and that D is integrable, we get $\Omega([X_1, X_2], Y) = 0$ for all sections Y of D. Thus $[X_1, X_2] \in \mathcal{D}^\perp$. Since D is integrable, this means that $[X_1, X_2] \in \mathcal{D} \cap \mathcal{D}^\perp$ for all $X_1, X_2 \in \mathcal{D} \cap \mathcal{D}^\perp$. Thus $D \cap D^\perp$ is integrable. ■

Now let us fix a leaf Λ of $\mathcal{F}_D$. The distribution $D \cap D^\perp$ restricts to Λ and so, by Frobenius' Theorem, there exists a maximal foliation of Λ for which the tangent spaces to the leaves are elements of $D \cap D^\perp$. We denote this foliation of Λ by $\mathcal{F}_\Lambda$. We shall further assume that $\Lambda/\mathcal{F}_\Lambda$ is a manifold. We can show that this quotient inherits a symplectic structure from P in a natural manner.

Proposition 4.23 *Let Λ be a leaf of the foliation $\mathcal{F}_D$. Suppose that Λ is a submanifold of P and that $\mathcal{F}_\Lambda$ is simple. Then $\Lambda/\mathcal{F}_\Lambda$ inherits a canonical symplectic structure Ω_Λ which satisfies the property*

$$\pi^*\Omega_\Lambda = i^*\Omega.$$

Here $\pi\colon \Lambda \to \Lambda/\mathcal{F}_\Lambda$ is the projection and $i\colon \Lambda \to P$ is the inclusion.

Proof: We first note that, as a vector bundle over Λ, $D/D \cap D^\perp$ is symplectic with the symplectic form on the fibres defined by

$$\tilde{\Omega}(p)([v_1],[v_2]) = \Omega(p)(v_1, v_2)$$

for $v_1, v_2 \in D(p) = T_p\Lambda$ and $p \in \Lambda$. To show that this form is well-defined we must show that it is independent of the choice of representatives v_1, v_2. So let $v_1', v_2' \in D(p)$ be such that $[v_1] = [v_1']$ and $[v_2] = [v_2']$. Then $v_1' = v_1 + u_1$ and $v_2 = v_2' + u_2$ for some $u_1, u_2 \in D(p) \cap D(p)^\perp$. We now have

$$\begin{aligned}\Omega(p)(v_1', v_2') =& \Omega(p)(v_1 + u_1, v_2 + u_2)\\ =& \Omega(p)(v_1, v_2) + \Omega(p)(v_1, u_2) +\\ & \Omega(p)(u_1, v_2) + \Omega(p)(u_1, u_2)\\ =& \Omega(p)(v_1, v_2).\end{aligned}$$

To show that $\tilde{\Omega}$ is nondegenerate, fix $[v_1] \in D(p)/D(p) \cap D(p)^\perp$ and suppose that $\tilde{\Omega}(p)([v_1],[v_2]) = 0$ for all $[v_2] \in D(p)/D(p) \cap D(p)^\perp$. Then $\Omega(p)(v_1, v_2) = 0$ for all $v_2 \in D(p)$. This implies that $v_1 \in D(p)^\perp$ and so $[v_1] = 0$. Thus we have shown that $D/D \cap D^\perp$ is a symplectic vector bundle.

Now suppose that X is a vector field which is tangent to a leaf of $\mathcal{F}_\Lambda$. Then X is a section of $D \cap D^\perp \mid \Lambda$. Therefore,

$$\boldsymbol{\mathcal{L}}_X i^*\Omega = \boldsymbol{d}(X \lrcorner\, i^*\Omega) = 0.$$

Thus $i^*\Omega$ is "constant" on leaves of $\mathcal{F}_\Lambda$. Therefore, it reduces to a two-form Ω_Λ on the quotient $\Lambda/\mathcal{F}_\Lambda$ defined by $\pi^*\Omega_\Lambda = i^*\Omega$. This form is nondegenerate as we have an identification of $T_{[p]}\Lambda/\mathcal{F}_\Lambda$ with $D(p)/D(p) \cap D(p)^\perp$. Also, $\pi^*\boldsymbol{d}\Omega_\Lambda = i^*\boldsymbol{d}\Omega = 0$ and so $\boldsymbol{d}\Omega_\Lambda = 0$ since π is a submersion. ■

Now we look at how we may reduce a Hamiltonian vector field on P to a Hamiltonian vector field on the quotient $\Lambda/\mathcal{F}_\Lambda$. We will suppose that we are given a Hamiltonian H whose Hamiltonian vector field is denoted by X_H.

Recall that a distribution D is *invariant* under a vector field X if $[X, Y] \in \mathcal{D}$ for every $Y \in \mathcal{D}$ (see Section 3.3).

Proposition 4.24 *If the distribution D is invariant under X_H then so is the distribution $D \cap D^\perp$.*

Proof: We must show that $[X_H, Y] \in \mathcal{D} \cap \mathcal{D}^\perp$ for every $Y \in \mathcal{D} \cap \mathcal{D}^\perp$. Let $Z \in \mathcal{D}$.

We compute

$$\begin{aligned}\boldsymbol{d}\Omega(X_H, Y, Z) =& Y(\Omega(X_H, Z)) - X_H(\Omega(Y, Z)) - Z(\Omega(X_H, Y)) + \\ & \Omega([X_H, Y], Z) - \Omega([X_H, Z], Y) + \Omega([Y, Z], X_H) \\ =& -Y(\boldsymbol{d}H \cdot Z) + Z(\boldsymbol{d}H \cdot Y) - \Omega([X_H, Y], Z) - \\ & \Omega([Y, Z], X_H) \\ =& -\Omega([X_H, Y], Z) - [Y, Z] \cdot \boldsymbol{d}H + \Omega(X_H, [Y, Z]) \\ =& -\Omega([X_H, Y], Z).\end{aligned}$$

Thus, since Ω is closed, we have $\Omega([X_H, Y], Z) = 0$ for every $Z \in \mathcal{D}$ and $Y \in \mathcal{D} \cap \mathcal{D}^\perp$. Thus $[X_H, Y] \in \mathcal{D} \cap \mathcal{D}^\perp$ for $Y \in \mathcal{D} \cap \mathcal{D}^\perp$. ■

If we make a stronger assumption on the Hamiltonian vector field, we may drop the dynamics to the quotient manifolds $\Lambda/\mathcal{F}_\Lambda$.

Corollary 4.25 *Suppose that $X_H \in \mathcal{D}$. Then, for each leaf Λ of $\mathcal{F}_D$, the vector field X_H gives rise to a Hamiltonian vector field on the quotient $\Lambda/\mathcal{F}_\Lambda$. The Hamiltonian, H_Λ, for the reduced vector field is defined by $\pi^* H_\Lambda = i^* H$.*

Proof: This simply amounts to observing that if $X_H \in \mathcal{D}$ then X_H restricts to each leaf of $\mathcal{F}_D$. Now, since this restricted vector field leaves the distribution $D \cap D^\perp$ invariant by Proposition 4.24, it drops to the quotient by Proposition 3.20. To show that the vector field is Hamiltonian with the Hamiltonian H_Λ we need only show that H projects. However this is clear since X_H is tangent to the leaves of $D \cap D^\perp$. ■

4.2.2 Decompositions of Symplectic Manifolds into Poisson Manifolds

We saw in Section 2.9 that a Poisson manifold is foliated by symplectic leaves. In this section we show that distributions with certain properties may be used to form Poisson manifolds from symplectic ones. This discussion is extracted from (Libermann and Marle, 1987).

Proposition 4.26 *Let (P, Σ) be a Poisson manifold and let $\phi\colon P \to N$ be a surjective submersion. Then the following are equivalent:*

i) The restriction of $\{f \circ \phi, g \circ \phi\}$ to $\phi^{-1}(x)$ is constant for each $x \in N$ and for each $f, g \in C^\infty(N)$, and

ii) there exists a Poisson structure on N so that ϕ is a Poisson mapping.

Proof: Since ϕ is a surjective submersion, the map $f \mapsto f \circ \phi$ is an isomorphism of $C^\infty(N)$ with the subspace of $C^\infty(M)$ consisting of functions which are constant on $\phi^{-1}(x)$ for each $x \in N$. Therefore, if N has a Poisson structure so that ϕ is a Poisson mapping, the bracket (on P) of two functions which are constant on $\phi^{-1}(x)$ must also be constant on $\phi^{-1}(x)$. Thus ii implies i.

Now suppose that i is true. We may then define a Poisson structure on N by making it the unique function which satisfies the property

$$\{f \circ \phi, g \circ \phi\} = \{f, g\} \circ \phi.$$

This definition makes sense since i is true, and it clearly makes ϕ a Poisson mapping. ∎

This result is especially interesting when the Poisson manifold is symplectic.

Proposition 4.27 *Let (P, Ω) be a symplectic manifold and let D be an integrable distribution on P.*

i) The distribution $D^\perp$ is integrable if and only if the Poisson bracket of every pair of integrals of D is an integral of D.

ii) Suppose there is a submersion $\phi: P \to N$ so that $D = \ker(T\phi)$. If there exists a Poisson structure on N so that ϕ is a Poisson mapping, then $D^\perp$ is integrable. Conversely, if $D^\perp$ is integrable and if $\phi^{-1}(x)$ is connected for each $x \in N$, then there exists a unique Poisson structure on N so that ϕ is a Poisson mapping.

Proof: From Lemma 2.17 i we have have

$$[\Omega^\sharp \boldsymbol{d}f, \Omega^\sharp \boldsymbol{d}g] = \Omega^\sharp \boldsymbol{d}\{f, g\}$$

for all $f, g \in C^\infty(P)$. This holds in particular if f, g are integrals of D. Note that $\Omega^\sharp \boldsymbol{d}f$ and $\Omega^\sharp \boldsymbol{d}g$ are sections of $D^\perp$ if f and g are integrals of D (see Lemma 2.17 ii). Therefore, $[X_f, X_g]$ is a section of $D^\perp$ if and only if $\{f, g\}$ is an integral of $D^\perp$. If the rank of D is k and the dimension of P is $2n$, then around every point $p \in P$ we may find $2n - k$ linearly independent functions, $f_1, \ldots, f_{2n-k}$, which are integrals of D. The resulting Hamiltonian vector fields span $D^\perp$. This proves i by Frobenius' theorem.

Note that if $D = \ker(T\phi)$, then the integrals of D are the functions which are constant on $\phi^{-1}(x)$ for each $x \in N$. Now apply Proposition 4.26 and i to get the first part of ii.

Now suppose that $\ker(T\phi)^\perp$ is integrable. By i, for every two functions $f, g \in C^\infty(N)$, $\{f \circ \phi, g \circ \phi\}$ is constant on $\phi^{-1}(x)$ for each $x \in N$. But the sets $\phi^{-1}(x)$ are the leaves of the maximal foliation of D since we have assumed these sets to be connected. The second part of ii now follows from Proposition 4.26. ∎

4.2.3 Applications of Decompositions to Hamiltonian Control Systems

The reductions of Sections 4.2.1 and 4.2.2 may be applied to Hamiltonian control systems. We shall define a *Hamiltonian control system* on a symplectic manifold

(P, Ω) to be an affine control system whose drift and control vector fields are Hamiltonian. We shall write such a system as

$$\dot{p} = X_H(p) + u^a X_a(p) \tag{4.12}$$

where the vector fields X_a are assumed to be Hamiltonian with Hamiltonian H_a for $a = 1, \ldots, m$. This type of control system is actually quite common in mechanics, and is discussed in (Nijmeijer and van der Schaft, 1990) and the references cited therein. Examples of systems which are (at least locally) Hamiltonian control systems are those which evolve on the symplectic manifold T^*Q and where the control Hamiltonians are simply coordinate functions on Q.

Following Section 3.1, we denote by C the accessibility distribution and by C_0 the strong accessibility distribution for the control system (4.12). The reduction of Section 4.2.1 may be applied to the case where the integrable distribution is the accessibility distribution since in this case the leaves of the foliation will be invariant under both the control vector fields and the drift vector field. By Corollary 4.25, the drift and control vector fields drop to the quotient $\Lambda/\mathcal{F}_\Lambda$ for each leaf $\Lambda \in \mathcal{F}_C$. In this way we obtain a family of Hamiltonian control systems, one for each leaf of $\mathcal{F}_C$. We have the following result.

Proposition 4.28 *Let Λ be a leaf of $\mathcal{F}_C$ and let $\mathcal{F}_\Lambda$ be the foliation induced on Λ by the distribution $C \cap C^\perp$. The control system (4.12) drops to a locally accessible Hamiltonian control system on the symplectic manifold $\Lambda/\mathcal{F}_\Lambda$.*

Proof: By Proposition 4.23 the manifold $\Lambda/\mathcal{F}_\Lambda$ is symplectic. By Corollary 4.25 the control system (4.12) drops to a Hamiltonian control system on the quotient. It remains to be shown that this reduced control system is locally accessible. Suppose it is not. Then the accessibility distribution on the quotient is not maximal. This immediately implies that the accessibility distribution on Λ must not be maximal since if it were, it would project to the maximal distribution on Λ. Thus we contradict local accessibility on the leaf Λ. ■

Now we study the uncontrollable dynamics. We shall use the decomposition results of Section 4.2.2. To do so we shall need some properties of the strong accessibility distribution C_0 associated with (4.12).

Lemma 4.29 *The distribution $C_0^\perp$ is integrable.*

Proof: By Proposition 4.27 i it suffices to show that the Poisson bracket of every pair of integrals of C_0 is also an integral of C_0. It is clear that C_0 is the distribution generated by Hamiltonian vector fields whose Hamiltonians are of the form

$$\{F_k, \{F_{k-1}, \{\cdots \{F_1, H_a\} \cdots\}\}$$

for $a = 1, \ldots, m$ and where $F_i \in \{H, H_1, \ldots, H_m\}$ for $i = 1, \ldots, k$. Let us denote this collection of functions by $\mathcal{O}$. The functions on P which commute under Poisson bracket with $\mathcal{O}$ will be denoted $\mathcal{O}^\perp$. The integrals of C_0 are all functions which are annihilated by the vector fields which generate C_0. These functions are then also those functions in $\mathcal{O}^\perp$. Let $G_1, G_2 \in \mathcal{O}^\perp$. We claim that the Poisson bracket of $\{G_1, G_2\}$ with any function in $\mathcal{O}$ is zero. Indeed, let $F \in \mathcal{O}$. Then, by Jacobi's identity

$$0 = \{F, \{G_1, G_2\}\} + \{G_2, \{F, G_1\}\} + \{G_1, \{G_2, F\}\} = \{F, \{G_1, G_2\}\}.$$

This proves the lemma by Proposition 4.27. ■

With this result we immediately have the following characterisation of the strongly inaccessible dynamics.

Proposition 4.30 *Suppose that the foliation $\mathcal{F}_{C_0}$ is simple. Then the manifold $P/\mathcal{F}_{C_0}$ has a Poisson structure so that the projection from P to the quotient is a Poisson mapping. Furthermore, the vector field X_H induces a vector field on $P/\mathcal{F}_{C_0}$ which is Poisson.*

Proof: The first part of the proposition follows from Lemma 4.29 and Proposition 4.27 ii. That the vector field X_H drops to the quotient is a consequence of C_0 being invariant under X_H and of Proposition 3.20. To show that the projected vector field, X, on the quotient is Poisson, let $f, g \in C^\infty(P/\mathcal{F}_{c_0})$. Then, by Lemma 2.19 and since π is a Poisson mapping, we have

$$\begin{aligned}\mathfrak{L}_X\{f,g\} =& \mathfrak{L}_{X_H}\pi^*\{f,g\} = \mathfrak{L}_{X_H}\{\pi^* f, \pi^* g\} = \{\mathfrak{L}_{X_H}\pi^* f, \pi^* g\} + \{\pi^* f, \mathfrak{L}_{X_H}\pi^* g\} \\ &= \{\mathfrak{L}_X f, g\} + \{f, \mathfrak{L}_X g\}.\end{aligned}$$

This shows that X is a Poisson vector field. ■

4.3 Examples of Mechanical Control Systems

In this section we present some examples. The examples are rather simple and are intended to illustrate the concepts put forward by the theory. One of the advantages of the conditions for local configuration accessibility given in Theorem 4.15 is that it lends itself to symbolic computation. Indeed, a *Mathematica* package was written to facilitate the computations in this section.

4.3.1 The Robotic Leg

In this section we return to the example discussed in Section 4.1.1. This example, although simple, exhibits much of the subtle behaviour that makes the study of mechanical systems interesting. This system may be cast both as a simple mechanical control system as in Section 4.1 and a Hamiltonian control system as in Section 4.2. The reader should be aware that the Hamiltonian representation is only valid locally since one of the input vector fields is only locally Hamiltonian and not Hamiltonian.

In the calculations in this section we shall avoid any problems which arise in the computations when the velocities or momenta are zero.

The Robotic Leg as a Simple Mechanical Control System

In the coordinates (θ, ψ, r) presented in Section 4.1.1, the Riemannian metric for the robotic leg is

$$g = J d\theta \otimes d\theta + mr^2 d\psi \otimes d\psi + m dr \otimes dr,$$

the input one-forms are

$$F^1 = d\theta - d\psi, \qquad F^2 = dr,$$

and the potential energy function is zero. In Section 4.1.1 we computed the input vector fields to be

$$Y_1 = \frac{1}{J}\frac{\partial}{\partial\theta} - \frac{1}{mr^2}\frac{\partial}{\partial\psi}, \qquad Y_2 = \frac{1}{m}\frac{\partial}{\partial r}.$$

Since there is no potential energy present, the distribution $C_{hor}(\mathcal{Y}, V)$ is simply generated by the vector fields $\overline{\mathrm{Lie}}(\overline{\mathrm{Sym}}(\mathcal{Y}))$.

We will find the following computations to be sufficient:

$$\begin{aligned}
\langle Y_1 : Y_1 \rangle &= -\frac{2}{m^2 r^3}\frac{\partial}{\partial r}, \\
\langle Y_1 : Y_2 \rangle &= 0, \\
\langle Y_2 : Y_2 \rangle &= 0, \\
[Y_1, Y_2] &= -\frac{2}{m^2 r^3}\frac{\partial}{\partial\psi}, \\
[Y_1, \langle Y_1 : Y_1 \rangle] &= \frac{4}{m^3 r^6}\frac{\partial}{\partial\psi}.
\end{aligned}$$

The reader may wish to compare these calculations with the bracket calculations of Section 4.1.1.

We may now go ahead and determine the configuration controllability of the robotic leg for the following three combinations of inputs:

RL1. Inputs Y_1 and Y_2: In this case it is clear that the system is locally configuration accessible by Theorem 4.15 as the input vector fields and their Lie bracket generates the maximal distribution on Q. Also, the bad symmetric product $\langle Y_1 : Y_1 \rangle$ is a multiple of Y_2 so the system is STLCC by Theorem 4.17.

RL2. Input Y_1: In this case the system is again locally configuration accessible since the vector fields $\{Y_1, \langle Y_1 : Y_1 \rangle, [Y_1, \langle Y_1 : Y_1 \rangle]\}$ generate the maximal distribution on Q. Note that the bad symmetric product $\langle Y_1 : Y_1 \rangle$ does not lie in the span of the inputs. Therefore, with this input, the robotic leg violates the sufficient conditions of Theorem 4.17 for STLCC.

RL3. Input Y_2: In this case we only generate the direction Y_2 and so the system is not locally configuration accessible. Indeed, starting from rest and only applying force in the r-direction, the only behaviour that can be observed is motion back and forth of the mass on the end of the leg. The decomposition of Theorem 4.20 in this case is given by

$$\ddot{r} - r\dot{\psi}^2 = \frac{1}{m}u_1$$

$$\ddot{\theta} = 0$$
$$\ddot{\psi} + \frac{2}{r}\dot{r}\dot{\psi} = 0.$$

The top system is obviously locally configuration accessible and also STLCC.

RL4. The linearisation of this system around points of zero velocity is not controllable so the cases where the system is STLCC do not follow from the linear calculations.

Remarks 4.31

1. The fact that the system is STLCC with both inputs (RL1) is not surprising given the discussion of Section 4.1.1. Here we have just verified the claim in that section using the formalism developed in Section 4.1.
2. Observe that the decomposition in RL3 is just as specified in Theorem 4.20. No inputs appear in the bottom two equations, and no terms which are quadratic in $\dot{r}$ appear in the bottom two equations.
3. Although the system only violates the *sufficient* conditions for STLCC in RL2, one may easily see by looking at the r-component of Lagrange's equations that the system is, in fact, not STLCC. The reason for this is that, since $\ddot{r} \geq 0$, r will always increase no matter what happens to the other variables. Thus our initial configuration will never be in the interior of the set of reachable configurations. □

The Robotic Leg as a Hamiltonian Control System

Now we look at the robotic leg as a Hamiltonian control system. The symplectic manifold is the cotangent bundle of the configuration manifold $Q = \mathbb{T}^2 \times \mathbb{R}^+$. As coordinates for T^*Q we shall use $(\theta, \psi, r, p_\theta, p_\psi, p_r)$ where (θ, ψ, r) are as explained in Section 4.1.1. The symplectic structure we consider is the canonical one for T^*Q. Thus

$$\Omega = d\theta \wedge dp_\theta + d\psi \wedge dp_\psi + dr \wedge dp_r.$$

Note In this section, if D is a distribution on the symplectic manifold (T^*Q, Ω), then we will denote $D^\perp = \Omega^\perp D$ in order to simplify the notation. □

The Hamiltonian is

$$H = \frac{1}{2J}p_\theta^2 + \frac{1}{2m}\left(p_r^2 + r^{-2}p_\psi^2\right).$$

We may then compute the drift vector field to be

$$X_H = \frac{p_\theta}{J}\frac{\partial}{\partial\theta} + \frac{p_\psi}{mr^2}\frac{\partial}{\partial\psi} + \frac{p_r}{m}\frac{\partial}{\partial r} + \frac{p_\psi^2}{mr^3}\frac{\partial}{\partial p_r}.$$

The control vector fields we shall consider are computed to be

$$X_1 = \frac{\partial}{\partial p_\theta} - \frac{\partial}{\partial p_\psi}, \qquad X_2 = \frac{\partial}{\partial p_r}.$$

The reader should be aware that X_1 is only locally Hamiltonian.

As we did when we considered the robotic leg as a simple mechanical control system, we consider three combinations of inputs.

Inputs X_1 and X_2:

The accessibility distribution may be computed to be

$$C = \left\langle \frac{\partial}{\partial r}, \frac{\partial}{\partial\theta}, \frac{\partial}{\partial\psi}, \frac{\partial}{\partial p_r}, \frac{\partial}{\partial p_\theta} - \frac{\partial}{\partial p_\psi} \right\rangle.$$

The leaves of the maximal foliation of this distribution are easily seen to be defined by $j \triangleq p_\theta + p_\psi = \textit{constant}$. As coordinates for $j^{-1}(\mu)$ we shall use $(\theta, \psi, r, p_\phi = p_\theta - p_\psi, p_r)$. The inclusion of $j^{-1}(\mu)$ into T^*Q is given by

$$(\theta, \psi, r, p_\phi, p_r) \mapsto \left(\theta, \psi, r, \tfrac{1}{2}(\mu - p_\phi), \tfrac{1}{2}(\mu + p_\phi), p_r\right).$$

We readily compute

$$C^\perp = \left\langle \frac{\partial}{\partial\theta} + \frac{\partial}{\partial\psi} \right\rangle$$

and so we see that $C \cap C^\perp = C^\perp$ (thus C is coisotropic). As the theory says, $C \cap C^\perp$ is integrable and its integral manifolds in this case are defined by

$$\theta - \psi = r = p_\theta = p_\psi = p_r = \textit{constant}.$$

As coordinates for $j^{-1}(\mu)/\mathcal{F}_{C\cap C^\perp}$ we will use $(\phi \triangleq \theta - \psi, r, p_\phi, p_r)$. Thus $j^{-1}(\mu)/\mathcal{F}_{C\cap C^\perp}$ is symplectomorphic to $T^*(\mathbb{S}^1 \times \mathbb{R}^+)$. The drift vector field drops to this manifold and is the Hamiltonian vector field on $T^*(\mathbb{S}^1 \times \mathbb{R}^+)$ with the canonical symplectic structure and Hamiltonian

$$H_\mu = \frac{1}{2m}p_r^2 + \left(\frac{1}{8J} + \frac{1}{8mr^2}\right)p_\phi^2 + \left(\frac{1}{4mr^2} - \frac{1}{4J}\right)\mu p_\phi + \left(\frac{1}{8J} + \frac{1}{8mr^2}\right)\mu^2.$$

The control vector fields drop to $T^*(\mathbb{S}^1 \times \mathbb{R}^+)$ and are given by

$$\tilde{X}_1 = \frac{\partial}{\partial p_\phi}, \qquad \tilde{X}_2 = 0$$

corresponding to X_1 and X_2, respectively. This defines for us the locally accessible Hamiltonian control system corresponding to the leaf $j^{-1}(\mu)$ as specified by Proposition 4.28.

Remark 4.32 These reduced dynamics are the same as would be obtained by performing standard symplectic reduction (see (Abraham and Marsden, 1978)) corresponding to the group action of $G = \mathbb{S}^1$ on Q given by

$$(\alpha, (\theta, \psi, r)) \mapsto (\theta + \alpha, \psi + \alpha, r). \qquad \square$$

We may also compute the strongly inaccessible dynamics in this case as in Proposition 4.30. When both inputs are present we compute $C_0 = C$. Thus $T^*Q/\mathcal{F}_{C_0}$ is a Poisson manifold which, in this case, may be coordinatised by $\mu = p_\theta + p_\psi$. The Poisson tensor is identically zero and there are no dynamics on this reduced manifold. This is a consequence of the fact that the system is strongly locally accessible on each leaf of $\mathcal{F}_{c_0}$.

Input X_1:

In this case the accessibility distribution and the strong accessibility distribution are the same as those computed with both inputs, so the reductions are the same.

Input X_2:

In this case the accessibility distribution may be computed to be

$$C = \left\langle \frac{\partial}{\partial \theta}, \frac{\partial}{\partial \psi}, \frac{\partial}{\partial r}, \frac{\partial}{\partial p_r} \right\rangle.$$

The leaves are defined by $p_\theta = constant$ and $p_\psi = constant$. As coordinates for any leaf we shall use the coordinates (θ, ψ, r, p_r). The injection of the leaf defined by $p_\theta = \mu$ and $p_\psi = \nu$ is given by

$$(\theta, \psi, r, p_r) \mapsto (\theta, \psi, r, \mu, \nu, p_r).$$

We may readily compute

$$C^\perp = \left\langle \frac{\partial}{\partial \theta}, \frac{\partial}{\partial \psi} \right\rangle$$

and so

$$C \cap C^{\perp} = \left\langle \frac{\partial}{\partial \theta}, \frac{\partial}{\partial \psi} \right\rangle.$$

This distribution is integrable, as it must be, and its integral manifolds are defined by

$$r = p_\theta = p_\psi = p_r = \textit{constant}.$$

As coordinates for the quotient of a leaf by $\mathcal{F}_{C\cap C^{\perp}}$ we shall use (r, p_r). Thus the reduced manifolds are symplectomorphic to $T^*\mathbb{R}^+$. The drift vector field drops to the quotient and is the Hamiltonian vector field on $T^*\mathbb{R}^+$ with the canonical symplectic structure and Hamiltonian

$$H_{\mu,\nu} = \frac{1}{2m} p_r^2 + \frac{1}{2mr^2}\nu^2 + \frac{1}{2J}\mu^2.$$

The drift vector field also factors through the quotient to yield the Hamiltonian vector field

$$\tilde{X}_2 = \frac{\partial}{\partial p_r}$$

on $T^*\mathbb{R}^+$. In this way we obtain the locally accessible Hamiltonian control system corresponding to the leaf $p_\theta = \mu$, $p_\psi = \nu$ as specified by Proposition 4.28.

Remark 4.33 The reduced dynamics that one obtains in this manner are those corresponding to symplectic reduction by the group action of $G = \mathbb{S}^1 \times \mathbb{S}^1$ on Q given by

$$((\alpha, \beta), (\theta, \psi, r)) \mapsto (\theta + \alpha + \beta, \psi + \alpha, r). \qquad \square$$

We may also compute the strong accessibility distribution as

$$C_0 = \left\langle \frac{\partial}{\partial \psi}, \frac{\partial}{\partial r}, \frac{\partial}{\partial p_r} \right\rangle.$$

This integrable distribution defines a foliation whose leaves are described by

$$\theta = p_\theta = p_\psi = \textit{constant}.$$

We may use $(\theta, p_\theta, p_\psi)$ as coordinates for the Poisson manifold $T^*Q/\mathcal{F}_{C_0}$. In these coordinates the Poisson tensor is given by

$$\Sigma = \frac{\partial}{\partial \theta} \wedge \frac{\partial}{\partial p_\theta}.$$

The vector field X_H drops to a vector field on the quotient which may be easily computed to be

$$\tilde{X} = \frac{p_\theta}{J}\frac{\partial}{\partial\theta}.$$

This is Hamiltonian on the quotient with the Hamiltonian

$$\tilde{H} = \frac{1}{2J}p_\theta^2.$$

Therefore, the reduced vector field is indeed Poisson on the quotient as in Proposition 4.30. These dynamics describe the locally strongly inaccessible dynamics for the robotic leg with the input in the r-direction. Physically these dynamics are a manifestation of the fact that, with this input, the rigid body part of the system will rotate completely unaffected by what is happening with the actuated leg.

4.3.2 The Forced Planar Rigid Body

In this section we study the planar rigid body with various combinations of forces and torques. The configuration space for the system is the Lie group $SE(2)$. To establish the correspondence between the configuration of the body and $SE(2)$, fix a point $O \in \mathbb{R}^2$ and let $\{\boldsymbol{e}_1 = \frac{\partial}{\partial x}, \boldsymbol{e}_2 = \frac{\partial}{\partial y}\}$ be the standard orthonormal frame at that point. Let $\{\boldsymbol{f}_1, \boldsymbol{f}_2\}$ be an orthonormal frame attached to the body at its centre of mass. The configuration of the body is determined by the element $g \in SE(2)$ which maps the point O with its frame $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ to the position, P, of the centre of mass of the body with its frame $\{\boldsymbol{f}_1, \boldsymbol{f}_2\}$. See Figure 4.3. The inputs for this problem consist of forces applied at an arbitrary point and a torque about the centre of mass. Without loss of generality (by redefining our body reference frame $\{\boldsymbol{f}_1, \boldsymbol{f}_2\}$) we may suppose that the point of application of the force is a distance h along the $\boldsymbol{f}_1$ body-axis from the centre of mass. The situation is illustrated in Figure 4.4.

With this convention fixed, we shall use coordinates (x, y, θ) for the planar rigid body where (x, y) describe the position of the center of mass and θ describes the orientation of the frame $\{\boldsymbol{f}_1, \boldsymbol{f}_2\}$ with respect to the frame $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$. In these coordinates, the Riemannian metric for the system is

$$g = m dx \otimes dx + m dy \otimes dy + J d\theta \otimes d\theta.$$

Here m is the mass of the body and J is its moment of inertia about the centre of mass. The inputs are described by the one-forms

$$F^1 = \cos\theta dx + \sin\theta dy, \qquad F^2 = -\sin\theta dx + \cos\theta dy - h d\theta, \qquad F^3 = d\theta$$

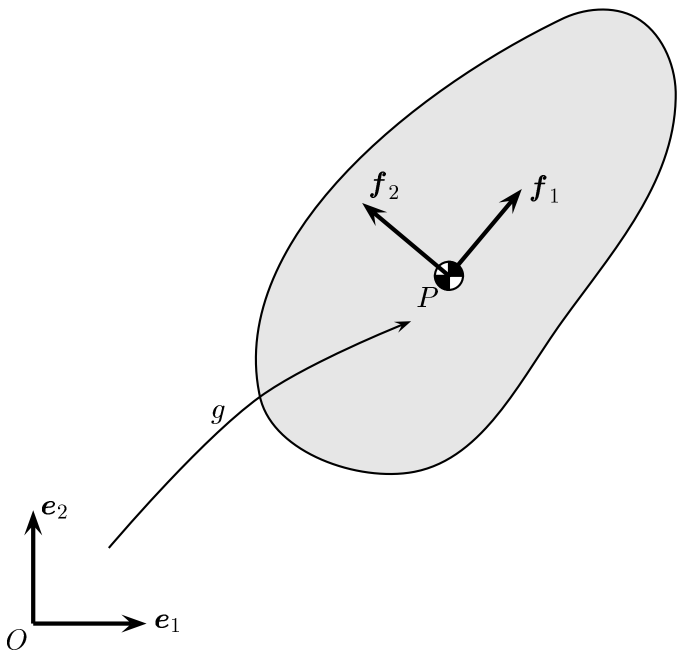

Figure 4.3 The configuration of a planar body as an element of $SE(2)$

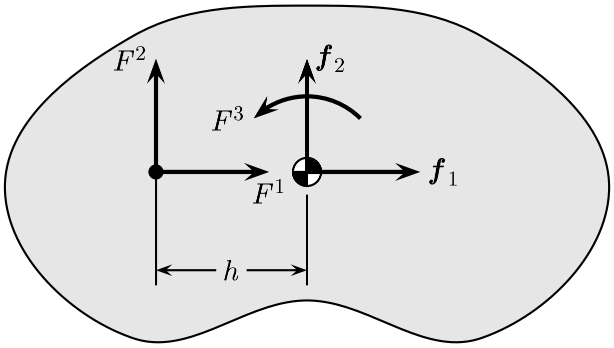

Figure 4.4 Positions for application of forces on a planar rigid body after simplifying assumptions

from which we compute the input vector fields to be

$$Y_1 = \frac{\cos\theta}{m}\frac{\partial}{\partial x} + \frac{\sin\theta}{m}\frac{\partial}{\partial x},$$
$$Y_2 = -\frac{\sin\theta}{m}\frac{\partial}{\partial x} + \frac{\cos\theta}{m}\frac{\partial}{\partial y} - \frac{h}{J}\frac{\partial}{\partial\theta}, \qquad Y_3 = \frac{1}{J}\frac{\partial}{\partial\theta}.$$

Again, as with the robotic leg, there is no potential energy so the distribution $C_{hor}(\mathcal{Y}, V)$ may be computed by calculating $\overline{\mathrm{Lie}}(\overline{\mathrm{Sym}}(\mathcal{Y}))$.

The following computations are sufficient to obtain the results we desire:

$$\begin{aligned}
\langle Y_1 : Y_1\rangle &= 0,\\
\langle Y_1 : Y_2\rangle &= \frac{h\sin\theta}{mJ}\frac{\partial}{\partial x} - \frac{h\cos\theta}{mJ}\frac{\partial}{\partial y},\\
\langle Y_1 : Y_3\rangle &= -\frac{\sin\theta}{mJ}\frac{\partial}{\partial x} + \frac{\cos\theta}{mJ}\frac{\partial}{\partial y},\\
\langle Y_2 : Y_2\rangle &= \frac{2h\cos\theta}{mJ}\frac{\partial}{\partial x} + \frac{2h\sin\theta}{mJ}\frac{\partial}{\partial y},\\
\langle Y_2 : Y_3\rangle &= -\frac{\cos\theta}{mJ}\frac{\partial}{\partial x} - \frac{\sin\theta}{mJ}\frac{\partial}{\partial y},\\
\langle Y_3 : Y_3\rangle &= 0,\\
[Y_1, Y_2] &= -\frac{h\sin\theta}{mJ}\frac{\partial}{\partial x} + \frac{h\cos\theta}{mJ}\frac{\partial}{\partial y},\\
[Y_1, Y_3] &= \frac{\sin\theta}{mJ}\frac{\partial}{\partial x} - \frac{\cos\theta}{mJ}\frac{\partial}{\partial y},\\
[Y_2, Y_3] &= \frac{\cos\theta}{mJ}\frac{\partial}{\partial x} + \frac{\sin\theta}{mJ}\frac{\partial}{\partial y},\\
[Y_2, \langle Y_2 : Y_2\rangle] &= \frac{2h^2\sin\theta}{mJ^2}\frac{\partial}{\partial x} - \frac{2h^2\cos\theta}{mJ^2}\frac{\partial}{\partial y}.
\end{aligned}$$

With the computations done, we may proceed to determine configuration controllability for the planar rigid body with various combinations of inputs. Since the case where all inputs are present is trivial from the point of view of controllability, we do not present it.

PB1. Inputs Y_1 and Y_2: In this case the maximal distribution on Q is generated by the inputs and their Lie bracket. Therefore, the system in locally configuration accessible with these inputs by Theorem 4.15. Also, the bad symmetric product $\langle Y_2 : Y_2\rangle$, is a multiple of Y_1 so the system is STLCC by Theorem 4.17.

PB2. Inputs Y_1 and Y_3: It is easy to see that the vector fields $\{Y_1, Y_3, [Y_1, Y_3]\}$ generate the maximal distribution on Q and so the system is locally configuration accessible with these inputs. All bad symmetric products vanish so the system is also STLCC.

PB3. Input Y_1: The only direction generated by all symmetric products and Lie brackets is Y_1 itself. Thus the system is not locally configuration accessible. To use the decomposition of Theorem 4.20 we must make a change of coordinates. In the coordinates $(\xi, \eta, \psi) = (x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta, \theta)$ the equations have the form

$$\ddot{\xi} + 2\left(\frac{m\eta^2}{J} - \frac{J + m\eta^2}{J}\right)\dot{\eta}\dot{\psi} + \left(\frac{m\xi\eta^2}{J} - \frac{\xi J + m\xi\eta^2}{J}\right)\dot{\psi}^2 = \left(\frac{J + m\eta^2}{J} - \frac{\eta^2}{J}\right)u_1$$

$$\ddot{\eta} + 2\left(\frac{J + m\xi^2}{J} - \frac{m\xi^2}{J}\right)\dot{\xi}\dot{\psi} + \left(\frac{m\eta\xi^2}{J} - \frac{\eta J + m\eta\xi^2}{J}\right)\dot{\psi}^2 = 0$$

$$\ddot{\psi} = 0.$$

The top system is locally configuration accessible and STLCC.

PB4. Inputs Y_2 and Y_3: With these inputs the maximal distribution on Q is generated by the input vector fields and their Lie bracket. Thus the system is locally configuration accessible. However, the bad symmetric product $\langle Y_2 : Y_2\rangle$ does not lie in the span of the inputs so the sufficient conditions of Theorem 4.17 are violated and the system may not be STLCC.

PB5. Input Y_2: With this input the maximal distribution on Q is generated by the vector fields $\{Y_2, \langle Y_2 : Y_2\rangle, [Y_2, \langle Y_2 : Y_2\rangle]\}$. Thus the system is locally configuration accessible by Theorem 4.15. The bad symmetric product $\langle Y_2 : Y_2\rangle$, is not a multiple of Y_2 so the system does not satisfy the sufficient conditions for STLCC.

PB6. Input Y_3: In this final case all symmetric products and Lie brackets are in the direction Y_3. Thus the system is not locally configuration accessible. We may use the coordinates (θ, x, y) to render the system in the form specified by Theorem 4.20. We obtain

$$\ddot{\theta} = \frac{1}{J}u_3$$

$$\ddot{x} = 0$$

$$\ddot{y} = 0.$$

The top system is clearly locally configuration accessible and STLCC.

Remarks 4.34

1. In this example, in the cases when the system fails to satisfy the sufficient conditions for STLCC of Theorem 4.17, we are not able to say whether the system is, in fact, not STLCC. In fact, in PB4, even though the system does

not satisfy the sufficient conditions of Theorem 4.17, it is easy to see that it *is* STLCC.

2. On a related note, in the robotic leg we saw that it was "Coriolis forces" which caused the loss of STLCC in RL2. In this example the metric is flat so the same explanation does not work. It would be interesting to ascertain why STLCC may be lost in the cases where the metric is flat.

3. The reader should verify that the decompositions given in PB3 and PB6 are in fact of the form guaranteed by Theorem 4.20.

4. The linearisation of this system around points of zero velocity is not controllable so the cases where the system is STLCC do not follow from the linear calculations.

5. The planar rigid body we presented in this section is an example of a class of systems whose configuration manifold is a Lie group, and the Riemannian metric and the input one-forms are left-invariant. In this case the control vector fields will also be left-invariant. We may choose a basis, $\{\xi_1, \dots, \xi_n\}$, for the Lie algebra of the group. Corresponding to this basis will be a basis of left-invariant vector fields, $\{X_1, \dots, X_n\}$, obtained by left translating the Lie algebra basis to each point in the group. The covariant derivative $\nabla_{X_i} X_j$ will also be a left-invariant vector field and so we may write $\langle X_i : X_j \rangle = \gamma_{ij}^k X_k$ for some set of constants γ_{ij}^k. Similarly we may write $[X_i, X_j] = c_{ij}^k X_k$ where the constants c_{ij}^k are the *structure constants* for the Lie algebra relative to the given basis. The conditions for local configuration accessibility and STLCC may then be expressed in terms of the constants γ_{ij}^k and c_{ij}^k. □

4.3.3 The Pendulum on a Cart

In this section we study the problem of a pendulum suspended from a cart. The configuration manifold for the system is $Q = \mathbb{R} \times \mathbb{S}^1$. As coordinates we shall use (x, θ) as shown in Figure 4.5. In this case the Riemannian metric for the system is

$$g = (M + m)dx \otimes dx + ml \cos\theta dx \otimes d\theta + ml \cos\theta d\theta \otimes dx + ml^2 d\theta \otimes d\theta.$$

Here M is the mass of the cart and m is the mass of the pendulum. The potential energy is

$$V = ma_g l(1 - \cos\theta)$$

where a_g is the acceleration due to gravity. The input is given by the one-form

$$F^1 = dx.$$

The input vector field is then readily computed to be

$$Y_1 = \frac{ml^2}{m^2l^2 + Mml^2 - m^2l^2\cos^2\theta} \frac{\partial}{\partial x} + \frac{ml\cos\theta}{m^2l^2 + Mml^2 - m^2l^2\cos^2\theta} \frac{\partial}{\partial \theta}.$$

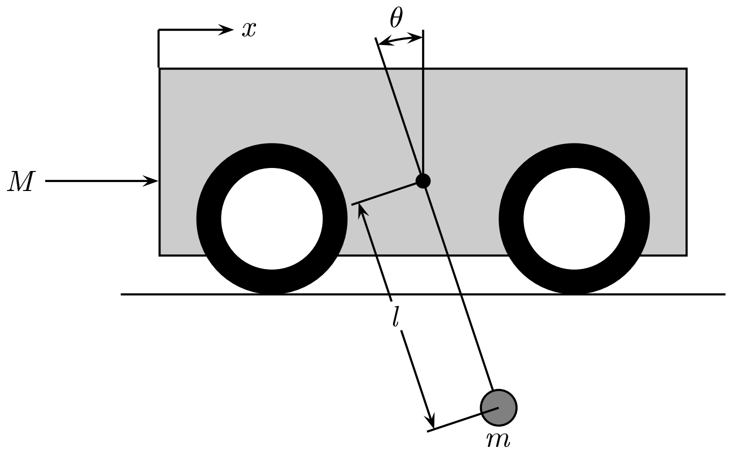

Figure 4.5 Pendulum suspended from a cart

To compute $C_{hor}(\mathcal{Y}, V)$ we need the following computations:

$$\langle Y_1 : Y_1 \rangle = \frac{16m\cos^2\theta\sin\theta}{l(m+2M-m\cos 2\theta)^3}\frac{\partial}{\partial x} + \frac{8(M+m)\sin\theta}{l^2(m\cos 2\theta - m - 2M)^3}\frac{\partial}{\partial\theta},$$

$$\langle Y_1 : \operatorname{grad} V \rangle = \frac{4a_g m\cos\theta(m - m\cos 2\theta - 2M\cos 2\theta)}{l(m\cos 2\theta - m - 2M)^3}\frac{\partial}{\partial x} +$$
$$\frac{4a_g(2M^2\cos 2\theta + 3Mm\cos 2\theta + m^2\cos 2\theta - Mm - m^2)}{l^2(m\cos 2\theta - m - 2M)^3}\frac{\partial}{\partial\theta}.$$

Note that at all points $q \in Q$ except those where $\theta \in \{0, \pi\}$, the vector fields $\{Y_1, \langle Y_1 : Y_1\rangle\}$ generate the tangent space at q. This means that the system is locally configuration accessible at these points. Also, at these points the bad symmetric product $\langle Y_1 : Y_1\rangle$ is not a multiple of Y_1 so the system may not be STLCC at these points. At points where $\theta \in \{0, \pi\}$ the vector fields $\{Y_1, \langle Y_1 : \operatorname{grad} V\rangle\}$ span T_qQ and so the system is also locally configuration accessible at these points. Most importantly, however, the bad symmetric product vanishes at these two points so the system is STLCC at these equilibria. This must be so as, at these two points, the linearised system is controllable.

Remark 4.35 This example may also be regarded as a Hamiltonian control system. However, it is uninteresting since the resulting Hamiltonian control system is strongly locally accessible. Therefore, there is only one leaf of the accessibility distribution: all of T^*Q. This means that the locally accessible dynamics are just the original dynamics. Also, $T^*Q/\mathcal{F}_{C_0}$ is trivial and so has no dynamics. □

Chapter 5

Formulations of General Mechanical Systems with External Forces

In Section 4.1 we studied a class of Lagrangian systems which was specified by having a particular Lagrangian and a particular set of inputs. Since we would eventually like to be able to consider more general Lagrangians and more general inputs, it is worth formulating, in a precise way, a formulation of mechanics which lends itself to this task. In the majority of modern geometric descriptions of mechanics, whether from the Lagrangian or Hamiltonian point of view, the representation of external inputs has been neglected. An example of some work which *does* incorporate inputs is the dissertation of (Yang, 1992). In that work, the geometry of the tangent bundle is used to describe forces in the Lagrangian setting. In this chapter we present an intrinsic description of mechanics which takes into account the presence of external forces by modifying the time-dependent Hamiltonian point of view. Our formulations use the jet bundles $J^1(\mathbb{R}, Q)$ in the Lagrangian case and $J^1(Q, \mathbb{R})$ in the Hamiltonian case. See Section 2.10 for a discussion of jet bundles.

We begin in Section 5.1 by introducing the basic objects of mechanics: the Lagrangian and the Hamiltonian. We also introduce the Legendre transformation which is used to go from one formulation to the other. In Section 5.2 we present the objects which we shall use to model inputs in mechanical systems. In Section 5.3 we present the Hamiltonian formulation of mechanics with external inputs. We present the Hamiltonian representation first since it fits most naturally with our use of two-forms to describe the equations of motion. Next we adapt our description of Hamiltonian mechanics to arrive at a Lagrangian formulation in Section 5.4. In the case when the Lagrangian is hyperregular, the formulations of Lagrangian and Hamiltonian mechanics are equivalent via the Legendre transformation. We briefly present these results in Section 5.5. Most of what we say in this section is a direct consequence of what was done in Section 5.1. In Section 5.6 we introduce an object which we call the *Lagrange force field*. With this we are able to write Lagrange's equations in a manner which is reminiscent of Newton's equations. This also serves the purpose of realising Lagrange's equations as the coefficients of a geometric object. Using the notions from Section 2.6, we may construct a Pfaffian module which describes the equations of motion. This approach is described for the Lagrangian point of view in Section 5.7.

Throughout this chapter, Q will denote the configuration manifold which we assume to have dimension n.

5.1 Lagrangians, Hamiltonians, and the Legendre Transformation

To discuss mechanics we need to say what we mean by the very basic concepts of a Lagrangian and a Hamiltonian. We will also introduce the Legendre transformation which is used to relate the Lagrangian and Hamiltonian formulations in the hyperregular case. Our definitions are generalisations to the time-dependent case of those given in (Abraham and Marsden, 1978) for the time-independent case.

Definition 5.1 A *Lagrangian* on Q is a function, L, on $J^1(\mathbb{R}, Q)$. Let L be a Lagrangian on Q. The *Legendre transformation* of L is the map, $\boldsymbol{F}L\colon J^1(\mathbb{R}, Q) \to J^1(Q, \mathbb{R})$, defined as follows: Let $[c]_1 \in J^1(\mathbb{R}, Q)_{t,q}$. Let $L_{t,q}$ denote the restriction of L to $J^1(\mathbb{R}, Q)_{t,q}$. Then

$$\boldsymbol{F}L([c]_1) = \{[f] \in J^1(Q, \mathbb{R})_{q,t} \mid \boldsymbol{d}f(q) = DL_{t,q}([c]_1)\}.$$

We say that L is *regular* if $\boldsymbol{F}L$ is a local diffeomorphism and *hyperregular* if $\boldsymbol{F}L$ is a diffeomorphism. □

In natural coordinates for $J^1(\mathbb{R}, Q)$ and $J^1(Q, \mathbb{R})$ we have

$$\boldsymbol{F}L(t, q^1, \ldots, q^n, v^1, \ldots, v^n) = \left(q^1, \ldots, q^n, \frac{\partial L}{\partial v^1}, \ldots, \frac{\partial L}{\partial v^n}, t\right).$$

It is clear from this coordinate expression that L is regular if and only if the matrix with components

$$\frac{\partial^2 L}{\partial v^i \partial v^j}([c]_1)$$

is nondegenerate for each $[c]_1 \in J^1(\mathbb{R}, Q)$. Note that we may consider $\boldsymbol{F}L$ to be fibre preserving in the sense that the following diagram commutes.

$$\begin{tikzcd} J^1(\mathbb{R}, Q) \arrow[rr, "\boldsymbol{F}L"] \arrow[dr, "\rho_1"'] & & J^1(Q, \mathbb{R}) \arrow[dl, "\rho_1^*"] \\ & Q & \end{tikzcd}$$

Now we may define the action corresponding to a function on $J^1(\mathbb{R}, Q)$.

Definition 5.2 Let L be a Lagrangian on Q. The *action* corresponding to L is the function, $\boldsymbol{A}L$, on $J^1(\mathbb{R}, Q)$ defined by

$$\boldsymbol{A}L([c]_1) = \boldsymbol{d}f(q) \cdot c'(t)$$

where $[c]_1 \in J^1(\mathbb{R}, Q)_{t,q}$ and f is a function on Q defined by $[f] = \boldsymbol{F}L([c]_1)$. $\square$

In natural coordinates for $J^1(\mathbb{R}, Q)$ the function $\boldsymbol{A}L$ looks like

$$(t, q^1, \ldots, q^n, v^1, \ldots, v^n) \mapsto \frac{\partial L}{\partial v^i} v^i.$$

It is also possible to define the Legendre transformation for a function on $J^1(Q, \mathbb{R})$.

Definition 5.3 A *Hamiltonian* on Q is a function, H, on $J^1(Q, \mathbb{R})$. Let H be a Hamiltonian on Q. The *Legendre transformation* of H is the map, $\boldsymbol{F}H \colon J^1(Q, \mathbb{R}) \to J^1(\mathbb{R}, Q)$, defined as follows: Let $[f] \in J^1(Q, \mathbb{R})_{q,t}$. Let $H_{q,t}$ denote the restriction of H to $J^1(Q, \mathbb{R})_{q,t}$. Then

$$\boldsymbol{F}H([f]) = \{[c]_1 \in J^1(\mathbb{R}, Q)_{t,q} \mid \ c'(t) = DH_{q,t}([f])\}.$$

We say H is *regular* if $\boldsymbol{F}H$ is a local diffeomorphism and *hyperregular* if $\boldsymbol{F}H$ is a diffeomorphism. $\square$

Here we have made the canonical identification of $(T_q^*Q)^*$ with T_qQ. In natural coordinates we have

$$\boldsymbol{F}H(q^1, \ldots, q^n, p_1, \ldots, p_n, t) = \left(t, q^1, \ldots, q^n, \frac{\partial H}{\partial p_1}, \ldots, \frac{\partial H}{\partial p_n}\right).$$

It is clear from this coordinate expression that H is regular if and only if the matrix with components

$$\frac{\partial^2 H}{\partial p_i \partial p_j}([f])$$

is nondegenerate for each $[f] \in J^1(Q, \mathbb{R})$. Like $\boldsymbol{F}L$, $\boldsymbol{F}H$ is fibre preserving in that the following diagram commutes.

$$\begin{tikzcd} J^1(Q, \mathbb{R}) \arrow[rr, "\boldsymbol{F}H"] \arrow[dr, "\rho_1^*"'] & & J^1(\mathbb{R}, Q) \arrow[dl, "\rho_1"] \\ & Q & \end{tikzcd}$$

Now we define the action corresponding to a Hamiltonian.

Definition 5.4 Let H be a Hamiltonian on Q. The *action* corresponding to H is the function, $\boldsymbol{A}H$, on $J^1(Q, \mathbb{R})$ defined by

$$\boldsymbol{A}H([f]) = \boldsymbol{d}f(q) \cdot c'(t)$$

where $[f] \in J^1(Q, \mathbb{R})_{q,t}$ and c is a curve on Q defined by $[c]_1 = \boldsymbol{F}H([f])$. $\square$

In natural coordinates for $J^1(Q,\mathbb{R})$, $\boldsymbol{A}H$ looks like

$$(q^1,\ldots,q^n,p_1,\ldots,p_n,t)\mapsto \frac{\partial H}{\partial p_i}p_i.$$

The two mappings $\boldsymbol{F}L$ and $\boldsymbol{F}H$ turn out to be inverses of each other in the hyperregular case.

Proposition 5.5 *Let L be a hyperregular Lagrangian and define $H=(\boldsymbol{A}L-L)\circ\boldsymbol{F}L^{-1}$. Then H is a hyperregular Hamiltonian and $\boldsymbol{F}H=\boldsymbol{F}L^{-1}$.*

Proof: We use natural coordinates for $J^1(\mathbb{R},Q)$ and $J^1(Q,\mathbb{R})$. The Legendre transformation for L looks like

$$\boldsymbol{F}L(t,q^1,\ldots,q^n,v^1,\ldots,v^n)=\left(q^1,\ldots,q^n,p_1=\frac{\partial L}{\partial v^1},\ldots,p_n=\frac{\partial L}{\partial v^n},t\right)$$

and the Hamiltonian looks like

$$H(q,p,t)=\left(\frac{\partial L}{\partial v^i}v^i-L\right)\circ\boldsymbol{F}L^{-1}(q,p,t)$$

which, using the Legendre transformation, we write as

$$H(q,p,t)=p_iv^i-L.$$

We are thinking of the v's as functions of the q's and p's via the Legendre transformation. Thus, using the chain rule, we have

$$\begin{aligned}\frac{\partial H}{\partial p_i}&=v^i+p_j\frac{\partial v^j}{\partial p_i}-\frac{\partial L}{\partial v^j}\frac{\partial v^j}{\partial p_i}\\&=v^i+p_j\frac{\partial v^j}{\partial p_i}-p_j\frac{\partial v^j}{\partial p_i}=v^i.\end{aligned}$$

Thus $\boldsymbol{F}H\circ\boldsymbol{F}L=\mathrm{id}_{J^1(\mathbb{R},Q)}$ and so $\boldsymbol{F}H$ is the inverse of $\boldsymbol{F}L$ and so is a diffeomorphism. Thus H is hyperregular. ■

Now we have the dual of this.

Proposition 5.6 *Let H be a hyperregular Hamiltonian and define $L=(H-\boldsymbol{A}H)\circ\boldsymbol{F}H^{-1}$. Then L is a hyperregular Lagrangian and $\boldsymbol{F}L=\boldsymbol{F}H^{-1}$.*

Proof: We use natural coordinates for $J^1(\mathbb{R},Q)$ and $J^1(Q,\mathbb{R})$. The Legendre transformation for H looks like

$$\boldsymbol{F}H(q^1,\ldots,q^n,p_1,\ldots,p_n,t)=\left(t,q^1,\ldots,q^n,v^1=\frac{\partial H}{\partial p_1},\ldots,v^n=\frac{\partial H}{\partial p_n}\right)$$

and the Lagrangian looks like

$$L(t,q,v)=\left(H-\frac{\partial H}{\partial p_i}p_i\right)\circ\boldsymbol{F}H^{-1}(t,q,v)$$

which, using the Legendre transformation, we write as

$$L(t, q, v) = H - v^i p_i.$$

We are thinking of the p's as functions of the q's and v's via the Legendre transformation. Thus, using the chain rule, we have

$$\begin{aligned}\frac{\partial L}{\partial v^i} &= p_i + v^j \frac{\partial p_j}{\partial v^i} - \frac{\partial H}{\partial p_j}\frac{\partial p_j}{\partial v^i} \\ &= p_i + v^j \frac{\partial p_j}{\partial v^i} - v^j \frac{\partial p_j}{\partial v^i} = p_i.\end{aligned}$$

Thus $\boldsymbol{FL} \circ \boldsymbol{FH} = \mathrm{id}_{J^1(Q,\mathbb{R})}$ and so $\boldsymbol{FL}$ is the inverse of $\boldsymbol{FH}$ and so is a diffeomorphism. Thus L is hyperregular. ■

5.2 External Forces for Mechanical Systems

In this section we say what we shall mean by an external force for a mechanical system. We shall be quite general and allow the forces to depend on time and on both position and higher derivatives of position with respect to time.

Definition 5.7 A subset Λ of T^*Q is called *complete* if $\pi_Q(\Lambda) = Q$. If $\Lambda \subset T^*Q$ is complete, we denote $\Lambda_q = \Lambda \cap T_q^*Q$.

Let $q \in Q$. An *m-force at q* is a map from $\rho_m^{-1}(q)$ to T_q^*Q. If Λ is a complete subset of T^*Q, we say that an m-force at q is *Λ-compatible* if it lies in Λ_q. The set of Λ-compatible m-forces at q is denoted by $\mathcal{F}_q^m(\Lambda)$. We will denote the totality of Λ-compatible m-forces on Q by

$$\mathcal{F}^m(\Lambda) = \mathring{\bigcup_{q \in Q}} \mathcal{F}_q^m(\Lambda).$$

We formally regard $\mathcal{F}^m(\Lambda)$ as a fibre space over Q and denote the projection from $\mathcal{F}^m(\Lambda)$ to Q by σ_Λ^m. If $\Lambda = T^*Q$ we write $\mathcal{F}^m(\Lambda) = \mathcal{F}^m(Q)$ and $\sigma_\Lambda^m = \sigma_Q^m$.

A *Λ-compatible m-force field* on Q is a section of $\mathcal{F}^m(\Lambda)$. If $\Lambda = T^*Q$ we will simply call a Λ-compatible m-force field an *m-force field.* We will also think of a Λ-compatible m-force field as a map, $F\colon J^m(\mathbb{R}, Q) \to \Lambda$, such that the following diagram commutes.

$$\begin{tikzcd} J^m(\mathbb{R}, Q) \arrow[rr, "F"] \arrow[dr, "\rho_m"'] & & \Lambda \arrow[dl, "\pi_Q"] \\ & Q & \end{tikzcd}$$

If $c\colon [a,b] \to Q$ is a curve on Q, an *m-force field along c* is a mapping, $F_c\colon [a,b] \to \mathcal{F}^m(Q)$, such that the following diagram commutes.

$$\begin{tikzcd} {[a,b]} \arrow[rr, "F_c"] \arrow[dr, "c"'] & & \mathcal{F}^m(Q) \arrow[dl, "\sigma_Q^m"] \\ & Q & \end{tikzcd}$$

If Λ is a complete subset of T^*Q, an m-force field along c, F_c, is Λ-*compatible* if $F_c(t) \in \mathcal{F}^m_{c(t)}(\Lambda)$ for each $t \in [a,b]$. □

In Section 4.1 we were interested in the case where Λ is a subbundle of T^*Q and where the force fields are only allowed to depend on position (thus they are 0-force fields in our terminology). The most generality one needs for external forces is probably only that of a 1-force. However, in Section 5.6 we introduce an interesting 2-force field associated to a curve, so the extra generality is maintained.

Note that since an m-force field takes its values in T^*Q, we may write it in natural coordinates for T^*Q as

$$F([c]_m) = F_i([c]_m)dq^i.$$

Thus we can think of an m-force field as a one-form on Q whose coefficients are functions on $J^m(\mathbb{R}, Q)$. It is also clear from this representation that one may regard an m-force field as a horizontal one-form on the bundle $\rho_m\colon J^m(\mathbb{R}, Q) \to Q$. Note that an l-force field, F, may be regarded as an m-force field, $\tilde{F}$, for $m > l$ by

$$\tilde{F}([c]_m) = F(\tau_{m,l}([c]_l)).$$

Now we discuss forces which are allowed to depend upon momenta for use in the Hamiltonian formulation.

Definition 5.8 Let $q \in Q$. A *coforce at q* is a map from $\pi_{1,0}^{-1}(q)$ to T_q^*Q. If Λ is a complete subset of T^*Q, we say that a coforce at q is Λ-*compatible* if it lies in Λ_q. The set of Λ-compatible coforces at q is denoted by $\mathcal{F}_q^*(\Lambda)$. We will denote the totality of Λ-compatible coforces on Q by

$$\mathcal{F}^*(\Lambda) = \overset{\circ}{\bigcup_{q\in Q}} \mathcal{F}_q^*(\Lambda).$$

We formally regard $\mathcal{F}^*(\Lambda)$ as a fibre space over Q and denote the projection from $\mathcal{F}^*(\Lambda)$ to Q by σ_Λ^*. If $\Lambda = T^*Q$ we write $\mathcal{F}^*(\Lambda) = \mathcal{F}^*(Q)$ and $\sigma_\Lambda^* = \sigma_Q^*$.

A Λ-*compatible coforce field* on Q is a section of $\mathcal{F}^*(\Lambda)$. If $\Lambda = T^*Q$ we will simply call a Λ-compatible coforce field a *coforce field*. We will also think of a Λ-compatible coforce field as a map, $F^*\colon J^1(Q, \mathbb{R}) \to T^*Q$, such that the following diagram commutes.

$$\begin{tikzcd} J^1(Q,\mathbb{R}) \arrow[rr, "F^*"] \arrow[dr, "\rho_1^*"'] & & T^*Q \arrow[dl, "\pi_Q"] \\ & Q & \end{tikzcd}$$

If $c\colon [a,b] \to Q$ is a curve on Q, a *coforce field along* c is a mapping, $F_c^*\colon [a,b] \to \mathcal{F}^*(Q)$, such that the following diagram commutes.

$$\begin{tikzcd} {[a,b]} \arrow[rr, "F_c^*"] \arrow[dr, "c"'] & & \mathcal{F}^*(Q) \arrow[dl, "\sigma_Q^*"] \\ & Q & \end{tikzcd}$$

If Λ is a complete subset of T^*Q, a coforce field along c, F_c^*, is Λ-*compatible* if $F_c^*(t) \in \mathcal{F}^*_{c(t)}(\Lambda)$ for each $t \in [a,b]$. □

As with m-forces, we have a convenient representation of a coforce field in terms of coordinates. We may write

$$F^*([f]) = F_i([f])dq^i$$

and so regard a coforce field as a differential form on Q whose coefficients are functions on $J^1(Q,\mathbb{R})$. As with m-forces, we may also regard a coforce field as a horizontal one-form on the bundle $\rho_1^*\colon J^1(Q,\mathbb{R}) \to Q$.

5.3 The Hamiltonian Formulation

The development of the Hamiltonian formalism with external inputs which we present here is based on that developed in (Hermann, 1982). Our development is a generalisation of the time-dependent contact formulation for Hamiltonian systems which is presented, for example, in (Abraham and Marsden, 1978).

As the basic data for our presentation we take the configuration manifold Q, the Hamiltonian H on Q, and a coforce field F^*. On $J^1(Q,\mathbb{R})$ we define a two-form by

$$\Omega(H,F^*) = \Omega_0 + \boldsymbol{d}H \wedge \boldsymbol{d}t + F^* \wedge \boldsymbol{d}t. \tag{5.1}$$

Here Ω_0 is the canonical symplectic structure on T^*Q pulled back to $J^1(Q,\mathbb{R})$ via the projection p_Q. We now use this two-form to define what is meant by the equations of motion for the system with Hamiltonian H and coforce field F^*.

Definition 5.9 Let H be a Hamiltonian on Q and let F^* be a coforce field on Q. A curve $c\colon [a,b] \to J^1(Q,\mathbb{R})$ is said to be a *solution of Hamilton's equations* with Hamiltonian H and coforce field F^* if $\boldsymbol{d}t \cdot c'(t) = 1$ and if $c'(t) \lrcorner\, \Omega(H,F^*) = 0$ for each $t \in [a,b]$. □

Let us see what a solution looks like in natural coordinates for $J^1(Q,\mathbb{R})$. Since $\boldsymbol{dt} \cdot c'(t) = 1$ we may suppose that the curve has the form

$$s \mapsto (q^1(s), \ldots, q^n(s), p_1(s), \ldots, p_n(s), s).$$

In coordinates we have

$$\Omega(H, F^*) = dq^i \wedge dp_i + \left(\frac{\partial H}{\partial q^i} dq^i + \frac{\partial H}{\partial p_i} dp_i + F_i^* dq^i \right) \wedge dt.$$

Thus we compute

$$c'(t) \lrcorner\, \Omega(H, F^*) = \left(\dot{q}^i - \frac{\partial H}{\partial p_i} \right) dp_i - \left(\dot{p}_i + \frac{\partial H}{\partial q^i} + F_i^* \right) dq^i + \left(\frac{\partial H}{\partial q^i} \dot{q}^i + \frac{\partial H}{\partial p_i} \dot{p}_i + F_i^* \dot{q}^i \right) dt.$$

This gives us the standard form of Hamilton's equations for the systems we are considering:

$$\begin{aligned} \dot{q}^i &= \frac{\partial H}{\partial p_i} \\ \dot{p}_i &= -\frac{\partial H}{\partial q^i} - F_i^*. \end{aligned}$$

The equation

$$\frac{\partial H}{\partial q^i} \dot{q}^i + \frac{\partial H}{\partial p_i} \dot{p}_i + F_i^* \dot{q}^i = 0$$

describes how the value of the Hamiltonian changes along solutions of Hamilton's equations. When $F^* = 0$ it is a statement of conservation of energy.

Remarks 5.10

1. When there are no forces present (i.e., when $F^* = 0$), the two-form $\Omega(H, F^*)$ is a contact form on $J^1(Q,\mathbb{R})$. In this case the Hamiltonian vector field is the characteristic vector field for the contact system. See (Abraham and Marsden, 1978) for a discussion of this.

2. When $F^* \neq 0$, the two form $\Omega(H, F^*)$ is not in general closed. However, when $F^* = 0$ it is even exact since Ω_0 is exact. □

5.4 The Lagrangian Formulation

Now we develop the Lagrangian formulation. This goes much like the Hamiltonian formulation with the obvious modifications to put one on $J^1(\mathbb{R}, Q)$ rather than on $J^1(Q,\mathbb{R})$.

The basic data here is the configuration manifold Q, the Lagrangian L on Q, and a 1-force field F. Associated to the Lagrangian is the action $\boldsymbol{A}L$ which was defined in Section 5.1. The *energy* is the function $\boldsymbol{E}L$ defined by

$$\boldsymbol{E}L = \boldsymbol{A}L - L.$$

Now define a two-form on $J^1(\mathbb{R}, Q)$ by

$$\Omega(L, F) = \boldsymbol{F}L^*\Omega_0 + \boldsymbol{dE}L \wedge \boldsymbol{d}t + F \wedge \boldsymbol{d}t.$$

We state what is meant by a solution of Lagrange's equations.

Definition 5.11 Let L be a Lagrangian on Q and let F be a 1-force field on Q. A curve $c\colon [a,b] \to Q$ is said to be a *solution of Lagrange's equations* with Lagrangian L and 1-force field F if $j^1c'(t) \lrcorner \Omega(L,F) = 0$ for each $t \in [a,b]$. □

It is helpful for the following calculations to manipulate the coordinate expression for $\Omega(L, F)$.

$$\begin{aligned}
\Omega(L,F) =& dq^i \wedge \boldsymbol{d}\left(\frac{\partial L}{\partial v^i}\right) + \boldsymbol{d}\left(\frac{\partial L}{\partial v^i}v^i - L\right) \wedge dt + F_i dq^i \wedge dt \\
=& -\boldsymbol{d}\left(\frac{\partial L}{\partial v^i}\right) \wedge (dq^i - v^i dt) - \boldsymbol{d}L \wedge dt + \frac{\partial L}{\partial v^i} dv^i \wedge dt + F_i dq^i \wedge dt \\
=& -\boldsymbol{d}\left(\frac{\partial L}{\partial v^i}\right) \wedge (dq^i - v^i dt) + \frac{\partial L}{\partial q^i} dq^i \wedge dt + F_i dq^i \wedge dt \\
=& -\left(\boldsymbol{d}\left(\frac{\partial L}{\partial v^i}\right) - \frac{\partial L}{\partial q^i} dt - F_i dt\right) \wedge (dq^i - v^i dt).
\end{aligned} \tag{5.2}$$

Now it is a simple matter to verify that our definition of a solution of Lagrange's equations agrees with the usual coordinate expression. By definition, a solution to Lagrange's equations has the form

$$s \mapsto (s, q^1(s), \ldots, q^n(s), \dot{q}^1(s), \ldots, \dot{q}^n(s)).$$

Therefore, the expression $j^1c(t) \lrcorner \Omega(L,F) = 0$ is simply equivalent to

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = F_i \tag{5.3}$$

which is the usual expression of Lagrange's equations with external forces.

Remarks 5.12

1. Unlike in the Hamiltonian case, the form $\Omega(L,F)$ may not be a contact form on $J^1(\mathbb{R}, Q)$ even when $F = 0$. This is a consequence of the fact that the Lagrangian may not be regular which may cause the form $\boldsymbol{F}L^*\Omega_0$ to lose rank.

2. In spite of 1, the equations (5.3) always make sense. □

5.5 The Equivalence of the Lagrangian and Hamiltonian Formulations

As might be gleaned by looking at the respective formulations of Hamiltonian and Lagrangian mechanics in Sections 5.3 and 5.4, when the Lagrangian or Hamiltonian is hyperregular, the two formulations agree. In this section we merely state this precisely.

First we state how to go from the Lagrangian to the Hamiltonian formulation.

Proposition 5.13 *Let L be a hyperregular Lagrangian on Q and define a Hamiltonian on Q by $H = \boldsymbol{F}L_*(\boldsymbol{A}L - L)$. Let F be a 1-force field on Q and define a coforce field on Q by $F^* = \boldsymbol{F}L_*F$. Then $c\colon [a,b] \to Q$ is a solution of Lagrange's equations with Lagrangian L and 1-force field F if and only if $\boldsymbol{F}L \circ j^1c\colon [a,b] \to J^1(Q,\mathbb{R})$ is a solution of Hamilton's equations with Hamiltonian H and coforce field F^*.*

Proof: By Proposition 5.5 we have $\Omega(H,F^*) = \boldsymbol{F}L^*\Omega(L,F)$. It is also clear that $\boldsymbol{dt} \cdot \boldsymbol{F}L \circ j^1c(t) = 1$. The proposition now follows since a characteristic vector field of $\Omega(L,F)$ will be mapped to a characteristic vector field of $\Omega(H,F^*)$ under the diffeomorphism $\boldsymbol{F}L$. ■

Now we go from Hamiltonian to Lagrangian.

Proposition 5.14 *Let H be a hyperregular Hamiltonian on Q and define a Lagrangian on Q by $L = \boldsymbol{F}H_*(H - \boldsymbol{A}H)$. Let F^* be a coforce field on Q and define a 1-force field on Q by $F = \boldsymbol{F}H_*F^*$. Then $c\colon [a,b] \to J^1(Q,\mathbb{R})$ is a solution of Hamilton's equations with Hamiltonian H and coforce field F^* if and only if $\rho_1^* \circ c\colon [a,b] \to Q$ is a solution of Lagrange's equations with Lagrangian L and 1-force field F.*

Proof: We need only show that the curve $\boldsymbol{F}H \circ c$ is of the form $j^1\sigma$ for some curve σ on Q. We know, as in Proposition 5.13, that $FH \circ c'(t) \lrcorner\, \Omega(L,F) = 0$. If we refer to (5.2) we can see that this implies that $\boldsymbol{F}H \circ c$ is indeed the lift of a curve on Q since the forms $dq^i - v^i dt$ must annihilate $FH \circ c'$. This completes the proof of the proposition. ■

5.6 Lagrange's Equations with the Lagrange Force Field

In this section we give an alternate formulation of Lagrangian mechanics by introducing a geometric object which we call the *Lagrange force field.* This object is the 2-force field on Q which has the form

$$F_L = \left(\frac{\partial^2 L}{\partial v^i \partial s} + \frac{\partial^2 L}{\partial v^i \partial q^j} v^j + \frac{\partial^2 L}{\partial v^i \partial v^j} a^j - \frac{\partial L}{\partial q^i} \right) dq^i$$

in natural coordinates for $J^2(\mathbb{R},Q)$. Although we define F_L in coordinates, it is, in fact, independent of the choice of natural coordinates. We present the calculation here for completeness. Suppose that $(Q^1,\dots,Q^n)$ are coordinates in another chart for Q. The corresponding natural coordinates for $J^2(\mathbb{R},Q)$ are given by

$$V^i = \frac{\partial Q^i}{\partial q^j} v^j$$

and

$$A^i = \frac{\partial Q^i}{\partial q^j} a^j + \frac{\partial^2 Q^i}{\partial q^j \partial q^k} v^j v^k.$$

Now we compute F_L in these new coordinates.

$$\begin{aligned}
F_L &= \left(\frac{\partial^2 L}{\partial v^i \partial s} + \frac{\partial^2 L}{\partial v^i \partial q^j} v^j + \frac{\partial^2 L}{\partial v^i \partial v^j} a^j - \frac{\partial L}{\partial q^i}\right) dq^i \\
&= \Bigg(\frac{\partial^2 L}{\partial V^k \partial s}\frac{\partial Q^k}{\partial q^i} + \frac{\partial^2 L}{\partial V^k \partial Q^l}\frac{\partial Q^k}{\partial q^i}\frac{\partial Q^l}{\partial q^j} v^j + \\
&\quad \frac{\partial^2 L}{\partial V^k \partial V^l}\frac{\partial Q^k}{\partial q^i}\frac{\partial^2 Q^l}{\partial q^j \partial q^m} v^j v^m + \frac{\partial L}{\partial V^k}\frac{\partial^2 Q^k}{\partial q^i \partial q^j} v^j + \\
&\quad \frac{\partial^2 L}{\partial V^k \partial V^l}\frac{\partial Q^k}{\partial q^i}\frac{\partial Q^l}{\partial q^j} a^j - \frac{\partial L}{\partial Q^k}\frac{\partial Q^k}{\partial q^i} - \frac{\partial L}{\partial V^k}\frac{\partial^2 Q^k}{\partial q^i \partial q^j} v^j\Bigg) dq^i \\
&= \frac{\partial Q^k}{\partial q^i}\left(\frac{\partial^2 L}{\partial V^k \partial s} + \frac{\partial^2 L}{\partial V^k \partial Q^l} V^l + \frac{\partial^2 L}{\partial V^k \partial V^l} A^l - \frac{\partial L}{\partial Q^k}\right) dq^i \\
&= \left(\frac{\partial^2 L}{\partial V^k \partial s} + \frac{\partial^2 L}{\partial V^k \partial Q^l} V^l + \frac{\partial^2 L}{\partial V^k \partial V^l} A^l - \frac{\partial L}{\partial Q^k}\right) dQ^k.
\end{aligned}$$

If $c\colon [a,b] \to Q$ is a curve on Q we may define a 2-force field along c, which we will also denote by F_L, by $F_L(t) = F_L(j^2c(t))$. If the curve c has the form

$$s \mapsto (q^1(s),\dots,q^n(s))$$

in coordinates, then the corresponding 2-force field along c is given by

$$F_L(t) = \left(\frac{\partial^2 L}{\partial v^i \partial s} + \frac{\partial^2 L}{\partial v^i \partial q^j}\dot q^j + \frac{\partial^2 L}{\partial v^i \partial v^j}\ddot q^j - \frac{\partial L}{\partial q^i}\right) dq^i = \left(\frac{d}{ds}\left(\frac{\partial L}{\partial \dot q^i}\right) - \frac{\partial L}{\partial q^i}\right) dq^i.$$

With this representation we easily see that the following result is true.

Proposition 5.15 *A curve $c\colon [a,b] \to Q$ is a solution of Lagrange's equations with Lagrangian L and force field F if and only if*

$$F_L(j^2c(t)) = F(j^1c(t))$$

for each $t \in [a,b]$.

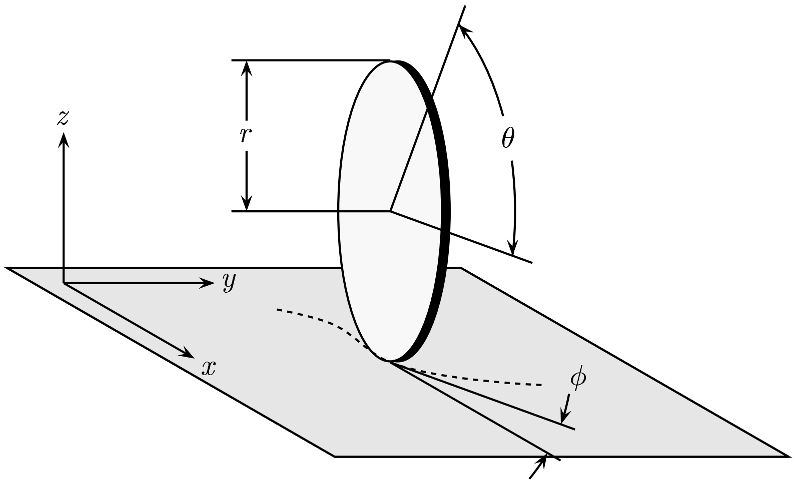

Figure 5.1 The rolling penny

A simple example helps to illustrate the concept of the Lagrange force field.

Example 5.16 We will consider the example of a rolling penny as a simple illustration of the concepts we have introduced. As coordinates for the penny we will use (x, y, θ, ϕ) as shown in Figure 5.1. Thus $Q = \mathbb{R}^2 \times \mathbb{T}^2$. The Lagrangian we will consider is

$$L = \frac{1}{2}m\left(v_x^2 + v_y^2\right) + \frac{1}{2}Iv_\theta^2 + \frac{1}{2}Jv_\phi^2.$$

Here m is the mass of the wheel, and I and J are the moments of inertia and so are strictly positive constants.

Let's compute F_L. Let $c\colon \mathbb{R} \to Q$ be a curve defined by

$$s \mapsto (x(s), y(s), \theta(s), \phi(s))$$

and let j^2c be the corresponding curve in $J^2(\mathbb{R}, Q)$. Let $t \in \mathbb{R}$ and let $q = c(t)$. Note that $[c]_2 \in J^2(\mathbb{R}, Q)_{t,q}$ is defined in the given coordinates by

$$\begin{aligned}
[c]_2 &= j^2c(s)\,|_{s=t} \\
&= (s, x(s), y(s), \theta(s), \phi(s), \dot{x}(s), \dot{y}(s), \dot{\theta}(s), \dot{\phi}(s), \ddot{x}(s), \ddot{y}(s), \ddot{\theta}(s), \ddot{\phi}(s))\,|_{s=t} \\
&= (t, x(t), y(t), \theta(t), \phi(t), \dot{x}(t), \dot{y}(t), \dot{\theta}(t), \dot{\phi}(t), \ddot{x}(t), \ddot{y}(t), \ddot{\theta}(t), \ddot{\phi}(t)).
\end{aligned}$$

It is straightforward to compute

$$F_L([c]_2) = m\ddot{x}(t)dx + m\ddot{y}(t)dy + I\ddot{\theta}(t)d\theta + J\ddot{\phi}(t)d\phi.$$

If we use natural coordinates for $J^2(\mathbb{R}, Q)$ we get: If

$$[c]_2 = (t, x, y, \theta, \phi, v_x, v_y, v_\theta, v_\phi, a_x, a_y, a_\theta, a_\phi)$$

then we have

$$F_L([c]_2) = ma_x dx + ma_y dy + Ia_\theta d\theta + Ja_\phi d\phi. \qquad \square$$

5.7 An Exterior Differential Systems Formulation for Lagrangian Mechanics

In this section we shall study the Pfaffian module on $J^1(\mathbb{R}, Q)$ specified by the Cartan system (see Section 2.6) of $\Omega(L, F)$ where L is a Lagrangian on Q and F is a 1-force field on Q. For the sake of notation, let us denote this system by $\mathcal{I}(L, F)$. It is interesting to see what this Pfaffian module looks like in natural coordinates for $J^1(\mathbb{R}, Q)$. By (5.2) and Lemma 2.14, if the Lagrangian is regular, a local basis for $\mathcal{I}(L, F)$ is generated by

$$\Big\{ dq^1 - v^1 dt, \dots, dq^n - v^n dt, \\ \boldsymbol{d}\left(\frac{\partial L}{\partial v^1}\right) - \frac{\partial L}{\partial q^1} dt - F_1 dt, \boldsymbol{d}\left(\frac{\partial L}{\partial v^n}\right) - \frac{\partial L}{\partial q^n} dt - F_n dt \Big\}.$$

It will be useful to introduce the following notation:

$$\begin{aligned} \alpha^i &= dq^i - v^i dt, \qquad i = 1, \dots, n, \\ \Theta_i(L, F) &= \boldsymbol{d}\left(\frac{\partial L}{\partial v^i}\right) - \frac{\partial L}{\partial q^i} dt - F_i dt, \qquad i = 1, \dots, n. \end{aligned}$$

To make precise the relationship between between solutions of Lagrange's equations and $\mathcal{I}(L, F)$ we prove the following result.

Lemma 5.17 *Let L be a Lagrangian on Q and let F be a 1-force field on Q. Suppose that c is a solution to Lagrange's equations with Lagrangian L and 1-force field F. Then the curve j^1c is an integral curve of $(\mathcal{I}(L, F), [\boldsymbol{d}t])$.*

Conversely, suppose that N is an integral manifold of $(\mathcal{I}(L, F), [\boldsymbol{d}t])$. Then there exists a curve, $c\colon [a, b] \to Q$, on Q so that N is locally the image of the curve j^1c. Furthermore, c is a solution of Lagrange's equations with Lagrangian L and 1-force field F.

Proof: If c is a solution to Lagrange's equations, the curve j^1c is such that $\boldsymbol{d}t \cdot j^1c' = 1$. It is a simple calculation to verify that $\alpha^i \cdot j^1c' = 0$ and $\Theta_i(L, F) \cdot j^1c' = 0$ for $i = 1, \dots, n$. Thus σ is an integral curve of $(\mathcal{I}(L, F), [\boldsymbol{d}t])$.

Now suppose that N is an integral manifold of $(\mathcal{I}(L, F), [\boldsymbol{d}t])$. Since $\boldsymbol{d}t \neq 0$ on N, we may locally use t as a coordinate for N. Thus we regard N as a graph over

t. Therefore, in coordinates, N is of the form

$$(s, q^1(s), \ldots, q^n(s), v^1(s), \ldots, v^n(s)).$$

Since N is an integral manifold for $\mathcal{I}(L, F)$, α^i, $i = 1, \ldots, n$, is zero when restricted to N. Therefore,

$$dq^i(s) - v^i(s)dt = \left(\frac{dq^i}{ds} - v^i(s)\right) dt = 0$$

for $i = 1, \ldots, n$. This means that on integral manifolds of $(\mathcal{I}(L, F), [\boldsymbol{dt}])$ we have

$$v^i = \frac{dq^i}{ds}.$$

Therefore, N locally has the form

$$(s, q^1(s), \ldots, q^n(s), \dot{q}^1(s), \ldots, \dot{q}^n(s)).$$

Thus N is locally the image of a curve on $J^1(\mathbb{R}, Q)$ of the form $j^1 c$ for some curve c on Q. It remains to check that c is a solution to Lagrange's equations. This follows from expanding the expression $\Theta_i(L, F) \cdot \sigma'(s) = 0$ for $i = 1, \ldots, n$ and $s \in [a, b]$. This completes the proof. ■

Remark 5.18 (Hermann, 1982) gives a formulation of Lagrangian mechanics with external forces which is somewhat similar to what we have presented. However, he treats the one-forms $\alpha^1, \ldots, \alpha^n, \Theta_1, \ldots, \Theta_n$ as the basic objects from which to derive the equations of motion. The problem with this approach is that these one-forms are not canonically defined. However, the two-form $\alpha^i \wedge \Theta_i$ *is* canonically defined and this is the object which we regard as basic in our formulation. □

Chapter 6

Mechanical Systems with Constraints

In Chapter 5 we developed Lagrangian and Hamiltonian formalisms for dealing with mechanical systems with external forces. Along with forces, another of the victims of the geometrisation of mechanics has been the inclusion of constraints in the formulation. A fairly modern treatment of constraints from a Hamiltonian point of view may be found in (Weber, 1986). In (Koiller, 1992) some systems with constraints are put in the framework of geometric reduction. A fairly comprehensive statement of the state of the art knowledge of reduction for systems with constraints is contained in (Bloch *et al.*, 1994). A thorough overview of variational methods for systems with constraints is presented in (Lewis and Murray, 1995b). In that work, a simple experiment was performed in an attempt to settle the debate over which of the *nonholonomic* or *vakonomic* variational methods is correct. In Section 6.1 we give an overview of the theoretical results of this paper. In particular, we give a careful formulation of the two variational problems, and show that they are equivalent when the constraints are holonomic. A more general type of constraint is introduced in Section 6.2. With this type of constraint it is possible to prove some natural controllability results for constrained systems.

The reader should be aware that the two sections in this chapter are not related except that they both deal with constrained mechanical systems. Each may be read, and should be interpreted, independently. To keep with the notation of Chapter 5, the presentation is on the jet bundle $J^1(\mathbb{R}, Q)$.

6.1 Variational Methods for Systems with Constraints

In this section we present the nonholonomic and vakonomic methods for deriving the equations of motion for a mechanical system with constraints. We shall try to be somewhat precise without overly burdening the presentation with technicalities.

We shall need to be clear about the type of constraints we consider.

Definition 6.1 An *affine constraint* on Q is a pair, (D, γ), where D is a distribution on Q and γ is a vector field on Q. A curve $c\colon [a,b] \to Q$ will be said to *satisfy* the affine constraint (D, γ) if $c'(t) - \gamma(c(t)) \in D(c(t))$ for all $t \in [a,b]$. □

We shall assume that D has constant rank k for simplicity. We will use this fact to suppose, at least locally, the existence of $n-k$ linearly independent one-forms, $\omega^1, \ldots, \omega^{n-k}$, which annihilate the distribution. That is to say we have

$$D(q) = \ker\{\omega^1(q), \ldots, \omega^{n-k}(q)\}.$$

All solutions of the constrained system are required to satisfy the conditions

$$\omega^a(c'(t)) = \omega^a(\gamma(c(t))), \qquad a = 1, \ldots, n-k.$$

Now we are ready to give precise definitions of the quantities involved in performing the variational calculations in this section.

6.1.1 Variations and Hamilton's Principle

In this section we introduce the basic tools for studying variational principles in mechanics. The main purpose of the discussion is to get the reader acquainted with the techniques we shall be using to pose and solve the variational problems considered. In particular, we introduce the notion of a *variation* and an *infinitesimal variation* of a curve c. The classical functional, J, is defined here as well.

The calculus of variations in its own right is a large subject. A good introduction which addresses some of the same issues we do is (Yan, 1995).

Unconstrained Variations

We will typically be considering curves, $c\colon [a,b] \to Q$, which connect two points, q_1 and q_2, in the configuration manifold Q. These curves may be subject to some constraints, but let us initially deal with the unconstrained case for the sake of concreteness. The set of all such curves which are C^2 will be denoted by $C^2(q_1, q_2, [a,b])$. It may be demonstrated that this set is a smooth infinite-dimensional manifold (see (Klingenberg, 1982)). The tangent space at a point $c \in C^2(q_1, q_2, [a,b])$ may be shown to be given by

$$\begin{aligned} T_cC^2(q_1, q_2, [a,b]) = \{u\colon [a,b] \to TQ \mid \ & u \text{ is } C^2, \\ & \tau_Q \circ u = c,\ u(a) = 0, \text{ and } u(b) = 0\}. \end{aligned}$$

We may think of a tangent vector u at c as being a vector field along c which vanishes at the endpoints (see Figure 6.1). Since u is a tangent vector, we may write it as the tangent vector to a curve which passes through c. A curve in $C^2(q_1, q_2, [a,b])$ will be written as

$$\mathbb{R} \ni s \mapsto c_s \in C^2(q_1, q_2, [a,b]).$$

For any $u \in T_cC^2(q_1, q_2, [a,b])$ we may write

$$u = \left.\frac{dc_s}{ds}\right|_{s=0}$$

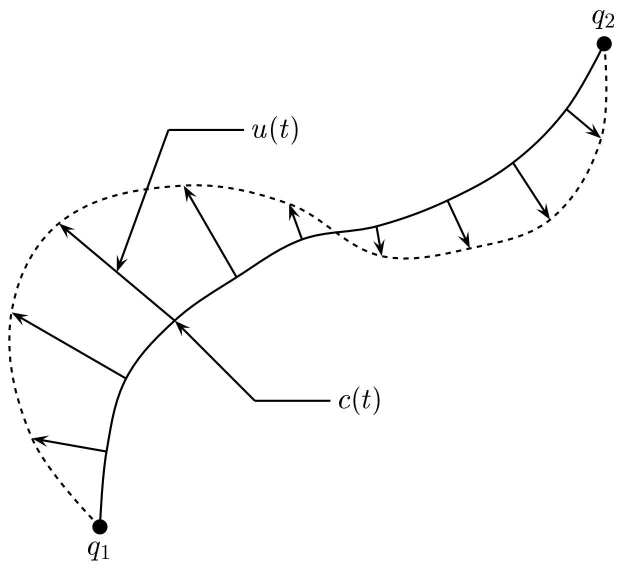

Figure 6.1 An infinitesimal variation

for some curve c_s in $C^2(q_1, q_2, [a, b])$. We shall refer to the curve c_s as a *variation* of $c = c_0$ and we shall refer to u as an *infinitesimal variation* of c.

Constrained Variations

Now we place an affine constraint (D, γ) on Q. For $q_1, q_2 \in Q$ we define

$$C^2(q_1, q_2, [a, b], D, \gamma) = \{c\colon [a, b] \to Q \mid \; c \text{ is a } C^2 \text{ curve}, \\ c(a) = q_1,\ c(b) = q_2, \text{ and } c'(t) - \gamma(c(t)) \in D(c(t)) \text{ for } t \in [a, b]\}.$$

It is possible that this subset of $C^2(q_1, q_2, [a, b])$ is empty, but let us suppose that it is not.

We will now define, in the presence of affine constraints, a special class of infinitesimal variations. In the classical literature these are commonly referred to as *virtual displacements*. Let $c \in C^2(q_1, q_2, [a, b], D, \gamma)$. Define

$$X_c(q_1, q_2, [a, b], D) = \{u \in T_c C^2(q_1, q_2, [a, b]) \mid \; c'(t) + u(t) - \gamma(c(t)) \in D(c(t))\}.$$

In words, $X_c(q_1, q_2, [a, b], D)$ is the set of infinitesimal variations which, when added to c', still satisfy the affine constraints. Clearly, since $c \in C^2(q_1, q_2, [a, b], D, \gamma)$, $u \in X_c(q_1, q_2, [a, b], D)$ if and only if $u(t) \in D(c(t))$, i.e., if u satisfies the non-affine constraints. This is why no reference to γ appears in the name of $X_c(q_1, q_2, [a, b], D)$.

The Functional J

Since we are on the manifold $C^2(q_1, q_2, [a,b])$, we may speak of smooth functions which may be differentiated. We therefore know what it means for a function to have a critical point. We will only define the functional for unconstrained systems. It is given by

$$\begin{aligned} J\colon\ C^2(q_1, q_2, [a,b]) &\to \mathbb{R} \\ c &\mapsto \int_a^b L(j^1 c(t))dt \end{aligned} \tag{6.1}$$

where L is a Lagrangian on Q. Note that $\mathbf{d}J(c) = 0$ if and only if $\mathbf{d}J(c)\cdot u = 0$ for every $u \in T_c C^2(q_1, q_2, [a,b])$. It is convenient to write

$$\mathbf{d}J(c)\cdot u = \frac{d}{ds}J(c_s)\bigg|_{s=0}.$$

With J as given by (6.1) we have

$$\mathbf{d}J(c)\cdot u = \frac{d}{ds}\int_a^b L(j^1 c_s(t))dt\bigg|_{s=0} = \int_a^b \frac{d}{ds}L(j^1 c_s(t))\bigg|_{s=0} dt.$$

We wish to evaluate this expression in local coordinates for Q. By the chain rule we have

$$\mathbf{d}J(c)\cdot u = \int_a^b \left(\frac{\partial L}{\partial q^i}\frac{\partial q^i}{\partial s} + \frac{\partial L}{\partial \dot{q}^i}\frac{\partial \dot{q}^i}{\partial s}\right)\bigg|_{s=0} dt.$$

Hamilton's Principle

As an example of how to apply the above concepts, we present *Hamilton's Principle*. This establishes a correspondence between solutions of Lagrange's equations and the solutions of a variational problem. We present this as a proposition whose proof goes much like the one in (Abraham and Marsden, 1978).

Proposition 6.2 (Hamilton's Principle) *Let L be a Lagrangian on Q. A curve $c\colon [a,b] \to Q$ joining q_1 with q_2 in Q is a solution to Lagrange's equations,*

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = 0, \qquad i = 1, \ldots, n,$$

if and only if $\mathbf{d}J(c) = 0$.

Proof: We need to show that c is a solution to Lagrange's equations if and only if $\mathbf{d}J(c)\cdot u = 0$ for every $u \in T_c C^2(q_1, q_2, [a,b])$. For any $u \in T_c C^2(q_1, q_2, [a,b])$ we may then write

$$u = \frac{dc_s}{ds}\bigg|_{s=0}$$

for some variation c_s of c. Then we have

$$\begin{aligned} \boldsymbol{d}J(c)\cdot u &= \left.\frac{d}{ds}J(c_s)\right|_{s=0} \\ &= \left.\frac{d}{ds}\int_a^b L(j^1c_s(t))dt\right|_{s=0}. \end{aligned}$$

The differentiation may be moved under the integral sign and in coordinates we have

$$\begin{aligned} \boldsymbol{d}J(c)\cdot u &= \int_a^b \left.\frac{d}{ds}L(q(t,s),\dot{q}(t,s),t)\right|_{s=0} dt \\ &= \int_a^b \left.\left(\frac{\partial L}{\partial q^i}\frac{\partial q^i}{\partial s} + \frac{\partial L}{\partial \dot{q}^i}\frac{\partial \dot{q}^i}{\partial s}\right)\right|_{s=0} dt. \end{aligned}$$

For the variation given we have

$$\left.\frac{\partial q^i(t,s)}{\partial s}\right|_{s=0} = u^i(t), \quad \text{and} \quad \left.\frac{\partial \dot{q}^i(t,s)}{\partial s}\right|_{s=0} = \left.\frac{d}{dt}\frac{\partial q^i(t,s)}{\partial s}\right|_{s=0} = \dot{u}^i(t).$$

We thus have, using integration by parts,

$$\begin{aligned} \boldsymbol{d}J(c)\cdot u &= \int_a^b \left(\frac{\partial L}{\partial q^i}u^i + \frac{\partial L}{\partial \dot{q}^i}\dot{u}^i\right) dt \\ &= \int_a^b \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right) u^i dt + \left.\frac{\partial L}{\partial \dot{q}^i}u^i\right|_a^b \\ &= \int_a^b \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right) u^i dt. \end{aligned}$$

Clearly then, $\boldsymbol{d}J(c)\cdot u = 0$ for every u if and only if Lagrange's equations are satisfied. This completes the proof. ■

Now we may apply the basic ideas of this section to the formulation of variational principles in the presence of constraints.

6.1.2 The Nonholonomic Method

In this variational method[1] one applies the constraints *after* making the functional J stationary. Let us formulate this problem more precisely. Let (D,γ) be an affine constraint on Q. Recall from Section 6.1.1 the definition of $C^2(q_1,q_2,[a,b],D,\gamma)$. From now on we shall tacitly assume that $C^2(q_1,q_2,[a,b],D,\gamma)$ is not empty. That is to say, we suppose that there are C^2 curves which connect q_1 and q_2 and which satisfy the affine constraint. We shall regard $C^2(q_1,q_2,[a,b],D,\gamma)$

[1] We call the nonholonomic method a variational method even though, in the strictest sense, it really is not. However, since we *do* use variations in discussing this method, our nomenclature is not entirely inappropriate.

as a subset of $C^2(q_1, q_2, [a, b])$. Also recall from Section 6.1.1 that at a point $c \in C^2(q_1, q_2, [a, b], D, \gamma)$, we defined X_c as the subset of the $T_c C^2(q_1, q_2, [a, b])$ consisting of *virtual displacements*.

The nonholonomic variational problem may now be stated as a definition.

Definition 6.3 A curve $c \in C^2(q_1, q_2, [a, b], D, \gamma)$ will be called a *solution to the nonholonomic constrained variational problem* if $\boldsymbol{d}J(c) \cdot u = 0$ for every $u \in X_c(q_1, q_2, [a, b], D)$. □

The following result is natural given our definition of the problem.

Proposition 6.4 *Let L be a regular Lagrangian on Q and let (D, γ) be an affine constraint on Q. Then $c \in C^2(q_1, q_2, [a, b], D, \gamma)$ is a solution of the nonholonomic constrained variational principle if and only if*

$$\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i}\right] u^i(t) = 0$$

for every $u \in X_c(q_1, q_2, [a, b], D)$.

Proof: Let c_s be a variation whose infinitesimal variation is $u \in X_c(q_1, q_2, [a, b], D)$. Then, as in the proof of Proposition 6.2, we have

$$\boldsymbol{d}J(c) \cdot u = \int_a^b \left(\frac{\partial L}{\partial q^i}\frac{\partial q^i}{\partial s} + \frac{\partial L}{\partial \dot{q}^i}\frac{\partial \dot{q}^i}{\partial s}\right)\bigg|_{s=0} dt.$$

In this case we simply have

$$\left.\frac{\partial q^i(t,s)}{\partial s}\right|_{s=0} = u^i(t), \quad \text{and} \quad \left.\frac{\partial \dot{q}^i(t,s)}{\partial s}\right|_{s=0} = \dot{u}^i(t).$$

If we do the usual integration by parts we have

$$\boldsymbol{d}J(c) \cdot u = \int_a^b \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}\right) u^i dt$$

from which the proposition follows. ■

Remarks 6.5

1. Note that we do not require c_s to be in $C^2(q_1, q_2, [a, b], D, \gamma)$ for $s \neq 0$. Thus we do not require our variations to satisfy the constraints. We only require the infinitesimal variations to satisfy the non-affine constraints. For a discussion of this see Section 6.1.1. The fact that the variations themselves do not necessarily satisfy the constraints allows us to interchange the order of differentiation with respect to s and t in determining $\partial \dot{q}^i / \partial s$. In classical terms, this allows us to interchange the "operators" δ and d/dt.

2. Observe that, unlike Hamilton's Principle, the nonholonomic constrained variational problem does not immediately give the equations of motion. This task is taken up when we discuss the Principle of Virtual Work in Section 6.1.4. There we will show that the equations of motion for the nonholonomic method are

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = \lambda_a \omega_i^a, \qquad i = 1, \ldots, n \tag{6.2}$$

along with the constraint equations

$$\omega_i^a \dot{q}^i = \omega_i^a \gamma^i, \qquad a = 1, \ldots, n-k.$$

There are other forms of the equations of motion for the nonholonomic method. An example of another form is the so-called *Lagrange-d'Alembert* equations. See (Bloch *et al.*, 1994) for a discussion of this along with other forms of the equations of motion using Ehresmann connections on fibre bundles.

3. See Figure 6.2 for a visual representation of the nonholonomic constrained variational problem. Observe how it differs from the representation of the vakonomic problem next to it. In particular, observe that we allow the variations to leave $C^2(q_1, q_2, [a,b], D, \gamma)$ in the nonholonomic method. □

6.1.3 The Vakonomic Method

In this variational technique one makes the functional J stationary *after* asking that the solutions satisfy the constraints. Thus this is a classical constrained minimisation problem, and may be solved with techniques from the calculus of variations with constraints. To make this method precise we must introduce some involved notation.

We begin with the definition of the solution to the vakonomic problem.

Definition 6.6 A curve $c \in C^2(q_1, q_2, [a,b], D, \gamma)$ will be called a *solution to the vakonomic constrained variational problem* if c is a critical point of $\boldsymbol{d}J \mid C^2(q_1, q_2, [a,b], D, \gamma)$. □

We now give a rough derivation of the equations of motion for the vakonomic constrained variational problem. We shall be somewhat informal here for the sake of clarity.

Since the vakonomic method is simply a constrained minimisation problem, we need some results from that field. The main result we shall use is the Lagrange Multiplier Theorem, the version which we use being taken from (Abraham *et al.*, 1988).

Lemma 6.7 (The Lagrange Multiplier Theorem) *Let M be a smooth manifold and let F be a Banach space with $g\colon M \to F$ a smooth submersion so that $N = g^{-1}(0)$ is a submanifold of M. Let $f\colon M \to \mathbb{R}$ be a smooth function. Then*

$n \in N$ is a critical point of $f \mid N$ if and only if there exists $\lambda \in F^*$ such that n is a critical point of $f - \lambda \circ g$.

To utilise this lemma, we must further examine the structure of $C^2(q_1, q_2, [a,b], D, \gamma)$, which was defined in Section 6.1.1. If E is a real Banach space, we denote by $\mathcal{F}([a,b], E)$ the Banach space of C^2, E-valued functions on the interval $[a,b]$. Suppose that the distribution D is annihilated by $n-k$ one-forms, $\omega^1, \dots, \omega^{n-k}$. We define a function $g\colon C^2(q_1, q_2, [a,b]) \to \mathcal{F}([a,b], \mathbb{R}^{n-k})$ by

$$g(c) = \left\{ t \mapsto \left(\omega^1(c'(t)) - \omega^1(\gamma(c(t))), \dots, \omega^{n-k}(c'(t)) - \omega^{n-k}(\gamma(c(t))) \right) \right\}. \tag{6.3}$$

We shall assume that g is a smooth submersion. Note that

$$C^2(q_1, q_2, [a,b], D, \gamma) = g^{-1}(0, \dots, 0)$$

is a smooth submanifold with this assumption.

We shall need to have some idea of what elements of $\mathcal{F}([a,b], \mathbb{R}^{n-k})^*$ look like. We shall be purposefully formal here. Note that $\mathcal{F}([a,b], \mathbb{R}^{n-k})$ is naturally isomorphic to the $(n-k)$-fold direct sum of $\mathcal{F}([a,b], \mathbb{R})$ with itself. Therefore, $\mathcal{F}([a,b], \mathbb{R}^{n-k})^*$ will be naturally isomorphic to the $(n-k)$-fold direct sum of $\mathcal{F}([a,b], \mathbb{R})^*$ with itself. Recall that elements of $\mathcal{F}([a,b], \mathbb{R})^*$ are (functional analytic) *distributions* on $[a,b]$. We shall not depart from the tradition of denoting the pairing of elements of $\mathcal{F}([a,b], \mathbb{R})^*$ with elements of $\mathcal{F}([a,b], \mathbb{R})$ by

$$\langle \alpha; f \rangle = \int_a^b \alpha \cdot f(t) dt.$$

We will at times regard elements of $\mathcal{F}([a,b], \mathbb{R})^*$ as elements of $\mathcal{F}([a,b], \mathbb{R})$ via the integral. The reader should be aware of what is taking place, and that it is not wholly precise.

The following result gives the equations of motion for the vakonomic constrained variational problem.

Proposition 6.8 *Let L be a Lagrangian on Q, let (D, γ) be an affine constraint on Q, and let $\omega^1, \dots, \omega^{n-k}$ be $n-k$ linearly independent differential one-forms on Q which annihilate D. Then $c \in C^2(q_1, q_2, [a,b], D, \gamma)$ is a solution of the vakonomic constrained variational problem if and only if there exists $(\lambda_1, \dots, \lambda_{n-k}) \in \mathcal{F}([a,b], \mathbb{R}^{n-k})^*$ such that*

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}^i}\right) - \frac{\partial \mathcal{L}}{\partial q^i} = 0, \qquad i = 1, \dots, n$$

where $\mathcal{L}\colon J^1(\mathbb{R}, Q) \to \mathbb{R}$ is defined along c by

$$\mathcal{L}(j^1 c(t)) = L(j^1 c(t)) - \lambda_a(t)[\omega^a(c'(t)) - \omega^a(\gamma(c(t)))].$$

Proof: Let $(g^1(c), \ldots, g^{n-k}(c))$ denote the components of $g(c)$ under the identification of $\mathcal{F}([a,b], \mathbb{R}^{n-k})$ with $\mathcal{F}([a,b], \mathbb{R}) \oplus \cdots \oplus \mathcal{F}([a,b], \mathbb{R})$. By (6.3) we have

$$g^a(c) = \{t \mapsto \omega^a(c'(t)) - \omega^a(\gamma(c(t)))\}, \qquad a = 1, \ldots, n-k.$$

From the Lagrange Multiplier Theorem we know that c is a solution to the vakonomic constrained variational problem if and only if there exists $(\lambda_1, \ldots, \lambda_{n-k}) \in \mathcal{F}([a,b], \mathbb{R}^{n-k})^*$ such that c is a critical point of the function J_D on $C^2(q_1, q_2, [a,b], D, \gamma)$ defined by

$$J_D(c) = \int_a^b L(j^1 c(t))dt - \lambda_a \cdot g^a(c).$$

Note that c is a critical point of J_D if and only if

$$\left.\frac{dJ_D(c_s)}{ds}\right|_{s=0} = \frac{d}{ds} \int_a^b L(j^1 c_s(t))dt\bigg|_{s=0} - \frac{d}{ds}\lambda_a \cdot g^a(c_s)\bigg|_{s=0} = 0$$

for every variation c_s of c. Now we use the integral notation for the pairing of the distribution λ_a with the element $g^a(c_s)$ of $\mathcal{F}([a,b], \mathbb{R})$. This gives

$$\left.\frac{J_D(c_s)}{ds}\right|_{s=0} = \int_a^b \frac{d}{ds} \left(L(j^1 c_s(t)) - \lambda_a \cdot (\omega^a(c_s'(t)) - \omega^a(\gamma(c_s(t)))\right)\bigg|_{s=0} dt.$$

The result now follows by the arguments used in the proof of Hamilton's Principle, Proposition 6.2. ■

Let us further examine the equations of motion for the vakonomic problem. In coordinates we have

$$\mathcal{L}(q, \dot{q}, t) = L(q, \dot{q}, t) - \lambda_a \omega_i^a \dot{q}^i + \lambda_a \omega_i^a \gamma^i.$$

Lagrange's equations for the Lagrangian $\mathcal{L}$ then read

$$\begin{aligned}\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}^i}\right) - \frac{\partial \mathcal{L}}{\partial q^i} &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i} - \lambda_a \omega_i^a\right) - \frac{\partial L}{\partial q^i} + \lambda_a \frac{\partial \omega_j^a}{\partial q^i}\dot{q}^j - \lambda_a \frac{\partial \omega_j^a}{\partial q^i}\gamma^j - \lambda_a \omega_j^a \frac{\partial \gamma^j}{\partial q^i} \\ &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} - \dot{\lambda}_a \omega_i^a - \lambda_a \frac{\partial \omega_j^a}{\partial q^i}\gamma^j - \lambda_a \omega_j^a \frac{\partial \gamma^j}{\partial q^i} = 0, \\ & \qquad i = 1, \ldots, n.\end{aligned} \tag{6.4a}$$

Appended to these are the constraint equations which are simply the "λ-part" of Lagrange's equations:

$$\omega_i^a \dot{q}^i = \omega_i^a \gamma^i, \qquad a = 1, \ldots, n-k. \tag{6.4b}$$

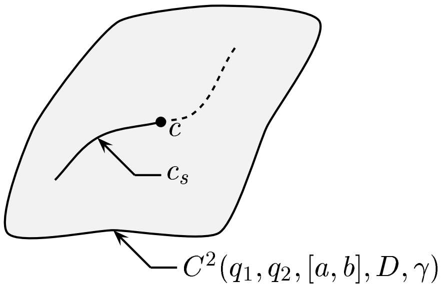

Figure 6.2 A representation of the nonholonomic constrained variational problem

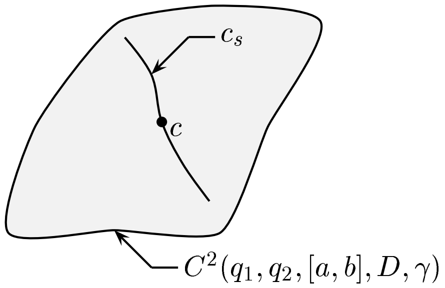

Figure 6.3 A representation of the vakonomic constrained variational problem

Remarks 6.9

1. Observe that, in practice, the equations (6.4a) and (6.4b) constitute a set of implicit first order ordinary differential equations in the variables $(q, \dot{q}, \lambda)$. This means that one must specify initial conditions for the Lagrange multipliers for the vakonomic problem.

2. In the case when $\gamma = 0$, the equations of motion for the vakonomic problem look like the equations of motion for the nonholonomic problem except there is now a $\dot{\lambda}_a$ in place of λ_a.

3. See Figure 6.3 for a visual representation of the vakonomic constrained variational problem. Observe how, unlike in the nonholonomic method, the variations for the vakonomic problem are not allowed to leave $C^2(q_1, q_2, [a,b], D, \gamma)$. $\square$

6.1.4 The Principle of Virtual Work

This principle is classically presented as an axiom of mechanics which is not derivable from the other basic axioms. It is typically stated as follows:

The Principle of Virtual Work *The work done by the forces of constraint is zero on virtual displacements.*

When we say that a force does no work on virtual displacements, we mean that, regarded as a differential one–form, the force annihilates tangent vectors in D. Thus the constraint force annihilates all vectors annihilated by the forms $\omega^1, \ldots, \omega^{n-k}$. We shall say that the Principle of Virtual Work is satisfied by a curve c if there exists external forces, F_i^c, which do no work on the constraints and are such that

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = F_i^c(t)$$

along c. In other words, regarded as a differential form, $F_i^c(t)dq^i$ must lie in the span of $\omega^1(c(t)), \dots, \omega^{n-k}(c(t))$. Thus, for each $t \in \mathbb{R}$ which is in the domain of definition of c, there must exist constants $\lambda_1(t), \dots, \lambda_{n-k}(t)$ such that

$$F_i^c(t)dq^i = \lambda_a(t)\omega^a(c(t)) = \lambda_a(t)\omega_i^a(c(t))dq^i$$

which means that $F_i^c(t) = \lambda_a(t)\omega_i^a(c(t))$ for some constants $\lambda_1(t), \dots, \lambda_{n-k}(t)$. Thus Lagrange's equations may be written as

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = \lambda_a \omega_i^a, \qquad i = 1, \dots, n$$

and we are to solve for the *Lagrange multipliers*, $\lambda_1, \dots, \lambda_{n-k}$, as part of the solution. To get the right number of equations for the number of unknowns, we append the constraint equations

$$\omega_i^a \dot{q}^i = \omega_i^a \gamma^i, \qquad a = 1, \dots, n-k.$$

We have the following easy result which relates the Principle of Virtual Work to the nonholonomic constrained variational problem discussed in Section 6.1.2.

Proposition 6.10 *A curve $c \in C^2(q_1, q_2, [a,b], D, \gamma)$ is a solution of the nonholonomic constrained variational problem if and only if the Principal of Virtual Work is satisfied by c.*

Proof: We must show that

$$\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i}\right] u^i(t) = 0$$

for every $u \in X_c(q_1, q_2, [a,b], D)$ if and only if

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = F_i^c(t)$$

along c, where the forces F_i^c do no work on virtual displacements. By definition, the forces F_i^c do no work on virtual displacements if and only if

$$F_i^c(t)u^i(t) = 0$$

for every $u \in X_c(q_1, q_2, [a,b], D)$ and $t \in [a,b]$. Thus the proposition is proved. ■

This gives a way of determining equations of motion for the nonholonomic constrained variational problem. Existence and uniqueness of solutions of these equations is not something we shall take up here.

6.1.5 The Nonholonomic and Vakonomic Methods when the Constraints are Holonomic

It turns out that when the constraints are holonomic, the nonholonomic and vakonomic problems are equivalent. We shall say that an affine constraint, (D,γ), is *holonomic* if D is integrable and if γ is a section in D. Notice that this is a modest generalisation of what we would denote as an holonomic constraint for systems with no affine part. In that case the constraint is simply the distribution D and is holonomic if D is integrable.

Remark 6.11 Note that if (D,γ) is an holonomic affine constraint, then $C^2(q_1,q_2,[a,b],D,\gamma)$ is non-empty if and only if q_1 and q_2 lie in the same leaf of $\mathcal{F}_D$. Also, any curve that is in a leaf of $\mathcal{F}_D$ will automatically satisfy the constraints. Thus our definition is only a mild generalisation of the usual notion of integrability of a distribution. □

Let Λ be a leaf of $\mathcal{F}_D$. Given a Lagrangian on Q, we may define a Lagrangian L_Λ on Λ by restriction of L to $J^1(\mathbb{R},\Lambda)\subset J^1(\mathbb{R},Q)$. With this Lagrangian we may define a function on $C^2(q_1,q_2,[a,b],D,\gamma)$ by

$$\begin{aligned} J_\Lambda\colon\ & C^2(q_1,q_2,[a,b],D,\gamma)\to\mathbb{R}\\ & c\mapsto\int_a^b L_\Lambda(j^1c(t))dt. \end{aligned}\tag{6.5}$$

The result is thus stated.

Proposition 6.12 *Let L be a Lagrangian on Q and let (D,γ) be an holonomic affine constraint on Q. Let $c\in C^2(q_1,q_2,[a,b],D,\gamma)$ where q_1 and q_2 lie in a leaf, Λ, of $\mathcal{F}_D$. Let J_Λ be the function defined by (6.5). Then the following are equivalent:*

i) c is a solution of the nonholonomic constrained variational problem,

ii) c is a solution of the vakonomic constrained variational problem,

iii) c is a critical point of J_Λ, and

iv) c is a solution of Lagrange's equations on Λ with Lagrangian L_Λ.

Proof: By Frobenius' theorem, we may choose coordinates, $(x^1,\ldots,x^k,y^1,\ldots,y^{n-k})$, around any point $q\in\Lambda$ which have the properties:

1. $(x^1,\ldots,x^k)$ are coordinates for Λ,
2. the injection of Λ into Q looks like $(x^1,\ldots,x^k)\mapsto(x^1,\ldots,x^k,0,\ldots,0)$, and
3. $D=\ker\{dy^1,\ldots,dy^{n-k}\}=\left\langle\frac{\partial}{\partial x^1},\ldots,\frac{\partial}{\partial x^k}\right\rangle$.

We first look at the equations of motion for the nonholonomic problem. By (6.2) we know that $c\in C^2(q_1,q_2,[a,b],D,\gamma)$ is a solution of the nonholonomic constrained variational problem if and only if

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\dot q^i}\right)-\frac{\partial L}{\partial q^i}=\lambda_a\omega_i^a,\qquad i=1,\ldots,n\tag{6.6}$$

for some $\lambda_1, \ldots, \lambda_{n-k}$ defined on $[a, b]$. In the coordinates $(x^1, \ldots, x^k, y^1, \ldots, y^{n-k})$, the curve c looks like

$$t \mapsto (x^1(t), \ldots, x^k(t), 0, \ldots, 0).$$

The equations (6.6) in the coordinates $(x^1, \ldots, x^k, y^1, \ldots, y^{n-k})$ are thus

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^\sigma}\right) - \frac{\partial L}{\partial x^\sigma} = 0, \qquad \sigma = 1, \ldots, k, \tag{6.7a}$$

$$\frac{\partial^2 L}{\partial \dot{y}^a \partial t} - \frac{\partial L}{\partial y^a} = \lambda_a, \qquad a = 1, \ldots, n-k. \tag{6.7b}$$

Note that (6.7b) simply specifies the Lagrange multipliers and has no effect on the solution in Q since all the time evolution there is specified by (6.7a).

Now we turn to the vakonomic problem. The appended Lagrangian to be used in the coordinates coordinates $(x^1, \ldots, x^k, y^1, \ldots, y^{n-k})$ is

$$\mathfrak{L} = L - \lambda_a \dot{y}^a.$$

We may easily determine that the equations (6.4a) appear in these coordinates as

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^\sigma}\right) - \frac{\partial L}{\partial x^\sigma} = 0, \qquad \sigma = 1, \ldots, k \tag{6.8a}$$

$$\frac{\partial^2 L}{\partial \dot{y}^a \partial t} - \frac{\partial L}{\partial y^a} = \dot{\lambda}_a, \qquad a = 1, \ldots, n-k. \tag{6.8b}$$

Here again we have used the fact that $y^1 = \cdots = y^{n-k} = 0$ along c. As with the nonholonomic equations, (6.8b) serves to determine the Lagrange multipliers and does not affect the time evolution of the coordinates $(x^1, \ldots, x^k)$.

In both the nonholonomic and vakonomic equations, the constraint equations are null since γ is a section of D.

Lagrange's equations on Λ for the Lagrangian L_Λ are

$$\frac{d}{dt}\left(\frac{\partial L_\Lambda}{\partial \dot{x}^\sigma}\right) - \frac{\partial L_\Lambda}{\partial x^\sigma} = 0, \qquad \sigma = 1, \ldots, k. \tag{6.9}$$

Note that since $y^1 = \cdots = y^{n-k} = 0$ along c we have

$$\frac{\partial L_\Lambda}{\partial \dot{x}^\sigma} = \frac{\partial L}{\partial \dot{x}^\sigma}, \quad \text{and} \quad \frac{\partial L_\Lambda}{\partial x^\sigma} = \frac{\partial L}{\partial x^\sigma}, \qquad \sigma = 1, \ldots, k. \tag{6.10}$$

From (6.7a) and (6.8a) we see that the components $(x^1, \ldots, x^k)$ evolve according to the same equations of motion in the nonholonomic and vakonomic problems. This proves that i is equivalent to ii. Using (6.9) and (6.10) we also see that iv is equivalent to both i and ii. Hamilton's Principle implies that iii is equivalent to iv. This completes the proof. ■

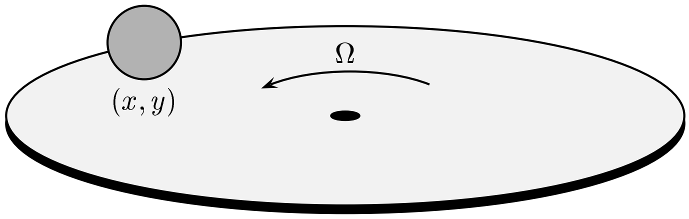

Figure 6.4 The rolling ball

6.1.6 The Nonholonomic and Vakonomic Methods Compared

In Section 6.1.5 we saw that the nonholonomic and vakonomic methods are equivalent when the constraints are holonomic. However, this is not true in general when the constraints are not holonomic. For certain systems, even though their constraints are not holonomic, it is possible to choose the initial conditions for the Lagrange multipliers in the vakonomic equations in such a way that the resulting solution is exactly that determined by the nonholonomic method. This occurs, for example, in the rolling penny considered in Example 5.16 (see (Bloch and Crouch, 1995)).

However, it is not true that it is *always* possible to select the initial conditions for the Lagrange multipliers in the vakonomic method so that the solutions are those of the nonholonomic equations. In this section we quickly review the example presented in (Lewis and Murray, 1995b) which illustrates that the nonholonomic and vakonomic methods are fundamentally different. The system is a ball rolling on a uniformly rotating table with no sliding (see Figure 6.4). Here (x, y) denotes the position of the point of contact of the ball with respect to the center of rotation of the table. The z–axis will be perpendicular to the plane of the table. The ball is assumed to be spherical and to have uniform mass density. The parameters in the problem are:

m	: mass of the ball
r	: radius of the ball
I	: moment of inertia of the ball
Ω	: rotational velocity of the table

The configuration space for the system is $Q = \mathbb{R}^2 \times SO(3)$. We shall use (x, y, R) to represent a typical point in Q. The constraints for the system are given by

$$\dot{x} - r\boldsymbol{e}_1^T \dot{R}R^T \boldsymbol{e}_3 = -\Omega y$$
$$\dot{y} + r\boldsymbol{e}_3^T \dot{R}R^T \boldsymbol{e}_2 = \Omega x$$

where $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ is the standard basis for $\mathbb{R}^3$. Since the matrix $\dot{R}R^T$ is skew symmetric (it represents the angular velocity of the ball in spatial coordinates), we may write

$$\dot{R}R^T = \begin{bmatrix} 0 & -\xi^3 & \xi^2 \\ \xi^3 & 0 & -\xi^1 \\ -\xi^2 & \xi^1 & 0 \end{bmatrix} \tag{6.11}$$

where ξ^1, ξ^2, ξ^3 are the rotational velocities about the x, y, z axes, respectively. With this notation, the constraints assume a more recognisable form:

$$\begin{aligned} \dot{x} - r\xi^2 &= -\Omega y \\ \dot{y} + r\xi^1 &= \Omega x. \end{aligned}$$

The Lagrangian for the rolling ball is

$$L = -\frac{1}{4} I \operatorname{tr}(\dot{R}R^T\dot{R}R^T) + \frac{1}{2}m(\dot{x}^2 + \dot{y}^2).$$

In (Lewis and Murray, 1995b) the equations for the nonholonomic method are shown to be equivalent to the equations derivable from Newton's equations. Therefore, we readily obtain the equations

$$m\ddot{x} = \lambda_1 \tag{6.12a}$$
$$m\ddot{y} = \lambda_2 \tag{6.12b}$$
$$I\dot{\xi}^1 = r\lambda_2 \tag{6.12c}$$
$$I\dot{\xi}^2 = -r\lambda_1 \tag{6.12d}$$
$$I\dot{\xi}^3 = 0. \tag{6.12e}$$

Here λ_1, λ_2 are the Lagrange multipliers which are to be determined from the constraint equations.

The vakonomic equations take a bit more work to derive, but are determined in (Lewis and Murray, 1995b) to be

$$m\ddot{x} - \dot{\lambda}_1 - \Omega\lambda_2 = 0 \tag{6.13a}$$
$$m\ddot{y} - \dot{\lambda}_2 + \Omega\lambda_1 = 0 \tag{6.13b}$$
$$\boldsymbol{I}\dot{\boldsymbol{\xi}} + \hat{\boldsymbol{\xi}}(\lambda_2 r\boldsymbol{e}_1 - \lambda_1 r\boldsymbol{e}_2) + \dot{\lambda}_1 r\boldsymbol{e}_2 - \dot{\lambda}_2 r\boldsymbol{e}_1 = 0 \tag{6.13c}$$
$$\begin{bmatrix} 1 + \frac{I}{mr^2} & 0 \\ 0 & 1 + \frac{I}{mr^2} \end{bmatrix} \begin{pmatrix} \dot{\lambda}_1 \\ \dot{\lambda}_2 \end{pmatrix} = \begin{pmatrix} -\frac{I\Omega}{mr^2}\lambda_2 - \frac{I\Omega}{r^2}\dot{y} - \lambda_2\xi^3 \\ \frac{I\Omega}{mr^2}\lambda_1 + \frac{I\Omega}{r^2}\dot{x} + \lambda_1\xi^3 \end{pmatrix}. \tag{6.13d}$$

Here $\boldsymbol{\xi} = (\xi^1, \xi^2, \xi^3)$ and $\hat{\boldsymbol{\xi}}$ is the skew-symmetric matrix on the right-hand-side of (6.11).

With the two sets of equations, we may now prove a lemma which states that their solutions will be fundamentally different.

Lemma 6.13 *Let* $q_0 = (x_0, y_0, \xi_0^1, \xi_0^2, \xi_0^3) \in \mathbb{R}^2 \times \mathbb{R}^3$ *and let*

$$c_{q_0} : t \mapsto (x(t), y(t), \xi^1(t), \xi^2(t), \xi^3(t))$$

be an integral curve for the nonholonomic equations of motion through q_0 *at* $t = 0$. *Then we may choose* q_0 *so that* c_{q_0} *is not a solution of the vakonomic equations of motion for any choice of initial conditions for the Lagrange multipliers.*

Proof: Substituting (6.13d) into (6.13a) and (6.13b) we get

$$\begin{aligned} m\ddot{x} + \frac{mI\Omega}{I + mr^2}\dot{y} + \Omega\left(\frac{I}{I + mr^2} - 1\right)\lambda_2 + \frac{mr^2}{I + mr^2}\lambda_2\xi^3 &= 0 \\ m\ddot{y} - \frac{mI\Omega}{I + mr^2}\dot{x} + \Omega\left(1 - \frac{I}{I + mr^2}\right)\lambda_1 - \frac{mr^2}{I + mr^2}\lambda_1\xi^3 &= 0. \end{aligned}$$

The nonholonomic equations for x, y may be written as

$$\begin{aligned} m\ddot{x} + \frac{mI\Omega}{I + mr^2}\dot{y} &= 0 \\ m\ddot{y} - \frac{mI\Omega}{I + mr^2}\dot{x} &= 0. \end{aligned}$$

We may easily see that these equations will give the same motions in x and y only if

$$\lambda_2(\xi^3 - \Omega) = 0 \qquad \text{and} \qquad \lambda_1(\xi^3 - \Omega) = 0.$$

Let us choose q_0 so that $\xi_0^3 \neq \Omega$. This means that we must have $\xi^3(t) \neq \Omega$ for all t since $\dot{\xi}^3 = 0$ in the nonholonomic equations. Therefore we must have $\lambda_1(t) = \lambda_2(t)$ for all t. From equations (6.13d) this means that we must have $\dot{x}(t) = \dot{y}(t) = 0$ for all t if a vakonomic solution is to agree with the nonholonomic solution. To prove the lemma we then choose initial conditions so that $\dot{x}(0)^2 + \dot{y}(0)^2 \neq 0$. ■

Remark 6.14 It is worth noting that the ball rolling on a rotating table is a system whose constraint has an affine part. It would be interesting to find an example whose constraint is non-affine, but whose nonholonomic and vakonomic equations are fundamentally different in the manner demonstrated in Lemma 6.13 for the rolling ball. It may be the case that no such example exists and that the nonholonomic and vakonomic methods may be taken to be equivalent in the case where the constraints have no affine part. However, we cannot make a strong statement in either direction at this point. □

Let us wrap up this section with a presentation of the pros and cons of the nonholonomic and vakonomic methods.

VM1. The vakonomic method has the advantage that it is a mathematically clean variational problem. This makes it the more appealing method, at least at first glance, to those who feel that nature seeks to act through a variational principle.

VM2. The nonholonomic method has the profound advantage of agreeing with Newton's equations in the cases where both techniques are applicable. This must certainly be held up as the most philosophically convincing argument in favour of the nonholonomic method.

VM3. In the vakonomic method, one is faced with having to make choices for the initial conditions for the Lagrange multipliers. As we saw in Section 6.1.5, when the constraints are holonomic, the initial conditions for the Lagrange multipliers are inconsequential since all choices of such initial conditions lead to the same physical motions as those specified by the nonholonomic method.

VM4. Finally, in (Lewis and Murray, 1995b) a series of experiments were performed for the ball rolling on a rotating table. If relevant friction effects are added to the nonholonomic model, the simulated equations give reasonable agreement with the experimental observations.

To summarise, the nonholonomic method has two strong arguments in its favour, one philosophical and one experimental. On the philosophical side, since the nonholonomic method agrees with Newton's equations in the cases where both are applicable, one must surely feel that the nonholonomic method is preferable. This philosophical argument is born out by experimental observation as well. While there are still some issues remaining unresolved in the nonholonomic versus vakonomic debate (see Remark 6.14), we feel that embracing the nonholonomic method is correct. For the reader interested in the debate over these two methods, we refer to the papers (Kharlomov, 1992) and (Kozlov, 1992).

6.1.7 Realising Constraints

As a final word in our presentation of the nonholonomic and vakonomic methods, we say a few things about "realising constraints." One may think of constraints as being a limiting process where certain dynamical properties become large and so limit the motion to the unconstrained directions. This may be made precise in the vakonomic and nonholonomic models. These notions are given in their precise forms in (Arnol'd, 1988), but we shall give rough descriptions of these limits here.

The vakonomic solutions may be regarded as a limit as an inertial term becomes large. The inertial term is a degenerate one which supplies no inertial forces to motions allowed by the constraints. When this term goes to infinity, the solutions of Lagrange's equations approach a solution for the vakonomic problem.

The nonholonomic solutions may be regarded as a limit as viscosity becomes large. To be more precise, we add Rayleigh dissipation to the mechanical system which does no work on motions allowed by the constraints (thus the dissipation function is degenerate). Then, as we make the magnitude of the dissipation function go to infinity, the corresponding solutions to Lagrange's equations approach the solutions to the nonholonomic equations.

As a simple example of using these limits to obtain constraints, consider the system in Figure 6.5. We wish to impose the (holonomic, non-affine) constraint $x = 0$. There are several ways to do this. One way would be to let the mass M get

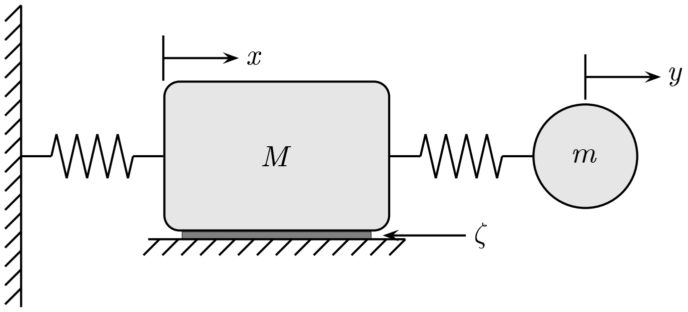

Figure 6.5 An example of realising constraints

very large. This would correspond to the vakonomic limit. Another way to impose the constraint $x = 0$ would be to let the damping coefficient ζ tend to infinity. This would correspond to the nonholonomic limit.

In each case care must be taken in the limit, and the convergence to the vakonomic and nonholonomic solutions in each case is not uniform in time. Note that in the example, since the constraint is holonomic, the limiting processes will produce the same motions by Proposition 6.12.

6.2 General Constraints

In this section we step away from the traditional constraints discussed in Section 6.1 and make more general definitions. The results in this section may be viewed as an extension of what we did in Chapter 5. We are able to use the language from that chapter to give an easy proof of an intuitive controllability result.

6.2.1 Definitions of General Constraints

In this section we define what we mean by a constraint in our general setting. In words we want a constraint to be an assignment of admissible directions which may depend on position, velocity and higher derivatives, and on time. To be completely general, we shall allow affine constraints. We first need to define what we mean by an affine subspace of a vector space.

Let V be a vector space and let $U \subset V$. We shall say that U is an *affine subspace* of V if there exists $v_0 \in V$ and a subspace $s(U)$ of V so that $U = v_0 + s(U)$. We define the *dimension* of an affine subspace U to be the dimension of the corresponding subspace $s(U)$. Note that $s(U)$ is well-defined given the affine subspace U, but the translation vector v_0 is defined only up to addition in $s(U)$. We denote by $\mathrm{Aff}^k(V)$ the set of k-dimensional affine subspaces of V and by $\mathrm{Aff}(V)$ the set of all affine subspaces of V.

Now let us introduce the affine bundle of TQ. We define

$$\mathrm{Aff}(TQ) = \overset{\circ}{\bigcup_{q\in Q}} \mathrm{Aff}(T_qQ).$$

This is a fibre bundle over Q and we denote the projection by ρ_Q.

We will also need the Grassmann bundle of T^*Q. Let $G^k(V)$ denote the set of k-dimensional subspaces of the vector space V and let $G(V)$ denote the set of all subspaces of V. We shall call

$$G(T^*Q) = \overset{\circ}{\bigcup_{q\in Q}} G(T_q^*Q)$$

the *Grassmann bundle* of T^*Q. This is a fibre bundle over Q and we denote the projection by ρ_Q^*.

Now we define constraints.

Definition 6.15 An *m-constraint* on Q is a smooth mapping, $\mathfrak{C}\colon J^m(\mathbb{R},Q) \to \mathrm{Aff}(Q)$, such that the following diagram commutes.

$$\begin{tikzcd} J^m(\mathbb{R},Q) \arrow[rr, "\mathfrak{C}"] \arrow[dr, "\rho_m"'] & & \mathrm{Aff}(Q) \arrow[dl, "\rho_Q"] \\ & Q & \end{tikzcd}$$

□

When we formulate the equations of motion below, we will see that there are connections between constraints and forces. When constraints are present, forces may be thought of as falling into two categories: those which act against the constraints, and those which act in directions complementary to the constrained directions. We make these notions clear with definitions.

Definition 6.16 Let $\mathfrak{C}$ be an m-constraint on Q and let F be an m-force field on Q. We say that F is a *$\mathfrak{C}$-constraint force* if $F([c]_m)\cdot v = 0$ for every $v \in s(\mathfrak{C}([c]_m))$ and $[c]_m \in J^m(\mathbb{R},Q)$. A *$\mathfrak{C}$-complementary force distribution* is a map, $\mathfrak{W}\colon J^m(\mathbb{R},Q) \to G^*(Q)$, such that the following diagram commutes

$$\begin{tikzcd} J^m(\mathbb{R},Q) \arrow[rr, "\mathfrak{W}"] \arrow[dr, "\rho_m"'] & & G^*(Q) \arrow[dl, "\rho_Q^*"] \\ & Q & \end{tikzcd}$$

and such that $T_q^*Q = \mathfrak{W}([c]_m) \oplus s(\mathfrak{C}([c]_m))^0$ for each $[c]_m \in J^m(\mathbb{R},Q)_{t,q}$, $q \in Q$, and $t \in \mathbb{R}$. An m-force field F is *$\mathfrak{W}$-admissible* if $F([c]_m) \in \mathfrak{W}([c]_m)$. □

Intuitively we think of $\mathfrak{C}$-constraint forces as those forces which annihilate the admissible directions and so do no work on motions allowed by the constraints. A

$\mathfrak{W}$-admissible force field will apply forces in directions complimentary to the constrained directions and so will contribute to the net motion of the system. Given an m-constraint $\mathfrak{C}$ and a $\mathfrak{C}$-complimentary force distribution $\mathfrak{W}$, we define the following complete subsets of T^*Q.

$$\mathfrak{C}^0 = \{\alpha = F([c]_m) \mid \ [c]_m \in J^m(\mathbb{R}, Q) \text{ and } F \text{ a } \mathfrak{C}\text{-constraint force}\}$$
$$\overline{\mathfrak{W}} = \{\alpha = F([c]_m) \mid \ [c]_m \in J^m(\mathbb{R}, Q) \text{ and } F \text{ a } \mathfrak{W}\text{-admissible force}\}.$$

Thus $\mathfrak{C}^0$ is the set of all $\mathfrak{C}$-constraint forces and $\overline{\mathfrak{W}}$ is the set of all $\mathfrak{W}$-admissible forces.

Now we define what we mean by a solution of a constrained and forced system. Recall the definition of the Lagrange force field F_L from Section 5.6.

Definition 6.17 Let F be a 1-force on Q, let $\mathfrak{C}$ be a 1-constraint on Q, and let L be a Lagrangian on Q. A curve $c\colon [a,b] \to Q$ is called a *solution to Lagrange's equations with force field F and constraint $\mathfrak{C}$* if $c'(t) \in \mathfrak{C}(j^1c(t))$ and if there exists a $\mathfrak{C}$-constraint force, λ, along c such that

$$F_L(j^2c(t)) = \lambda(j^1c(t)) + F(j^1c(t)). \qquad \square$$

A simple example illustrates the concepts. We return to the example of the rolling penny initially presented in Section 5.6.

Example 6.18 (5.16 cont'd) Recall that the configuration space was $Q = \mathbb{R}^2 \times \mathbb{T}^2$ and the Lagrangian was given by

$$L = \frac{1}{2}m(v_x^2 + v_y^2) + \frac{1}{2}Iv_\theta^2 + \frac{1}{2}Jv_\phi^2.$$

The Lagrange force field was computed in Example 5.16. There are velocity constraints on the system determined by the differential forms

$$dx - r\cos\phi d\theta, \quad dy - r\sin\phi d\theta$$

where r is the radius of the penny. We wish to present these as a 1-constraint on Q. Let D be the kernel of these two differential one-forms. i.e.,

$$D = \ker\{dx - r\cos\phi d\theta, dy - r\sin\phi d\theta\}.$$

We define the 1-constraint $\mathfrak{C}$ by

$$\mathfrak{C}([c]_1) = D(\rho_1([c]_1)).$$

Note that the constraint in this case depends only on position and not on velocity or time. A $\mathfrak{C}$-constraint force on Q is given by

$$\lambda([c]_1) = \lambda_1([c]_1)(dx - r\cos\phi d\theta) + \lambda_2([c]_1)(dy - r\sin\phi d\theta)$$

since this is the most general force which will annihilate the constrained directions. A 1-force on Q looks like

$$F([c]_1) = F_x([c]_1)dx + F_y([c]_1)dy + F_\theta([c]_1)d\theta + F_\phi([c]_1)d\phi.$$

Let $c\colon [a,b] \to Q$ be a curve on Q. We then have

$$\begin{gathered}F_L(j^2c(t)) - F(j^1c(t)) - \lambda(j^1c(t)) = (m\ddot{x}(t) - F_x(j^1c(t)) - \lambda_1(j^1c(t)))dx + \\ (m\ddot{y}(t) - F_y(j^1c(t)) - \lambda_2(j^1c(t)))dy + \\ (I\ddot{\theta}(t) - F_\theta(j^1c(t)) + \lambda_1(j^1c(t))r\cos\phi(t) + \lambda_2(j^1c(t))r\sin\phi(t))d\theta + \\ (J\ddot{\phi}(t) - F_\phi(j^1c(t)))d\phi.\end{gathered}$$

Therefore, omitting arguments, c is a solution to Lagrange's equations with force field F and constraint $\mathfrak{C}$ if and only if

$$\begin{aligned}\dot{x} - r\cos\phi\dot{\theta} &= 0\\ \dot{y} - r\sin\phi\dot{\theta} &= 0\end{aligned}$$

and

$$\begin{aligned}&m\ddot{x} - F_x - \lambda_1 = 0\\ &m\ddot{y} - F_y - \lambda_2 = 0\\ &I\ddot{\theta} - F_\theta + \lambda_1 r\cos\phi + \lambda_2 r\sin\phi = 0\\ &J\ddot{\phi} - F_\phi = 0.\end{aligned}$$

The coefficients of F and λ will, in general, depend on position, velocity, and time. □

6.2.2 Controllability Results for General Constraints

Let us first give some definitions of controllability which suit our general notion of a constrained system.

Definition 6.19 Let $\mathfrak{C}$ be a 1-constraint on Q. We say that $\mathfrak{C}$ is *controllable* if, for every two points $q_1, q_2 \in Q$, there exists a piecewise smooth curve $c\colon [a,b] \to Q$ such that c connects q_1 with q_2 and $c'(t) \in \mathfrak{C}(j^1c(t))$ for each $t \in [a,b]$.

Let L be a Lagrangian on Q and let Λ be a complete subset of T^*Q. We say the triple $(L, \mathfrak{C}, \Lambda)$ is *controllable* if, for every two points $q_1, q_2 \in Q$, there exists:

i) a piecewise smooth curve $c\colon [a,b] \to Q$,

ii) a Λ-compatible 1-force field, F, along c, and

iii) a $\mathfrak{C}$-constraint force, λ, along c

such that

i) c connects q_1 with q_2,

ii) $c'(t) \in \mathfrak{C}(j^1c(t))$ for each $t \in [a,b]$,

iii) $F_L(j^2c(t)) = F(j^1c(t)) + \lambda(j^1c(t))$ for each $t \in [a,b]$, and

iv) $c'(a) = 0$ and $c'(b) = 0$. □

Remark 6.20 If the constraint $\mathfrak{C}$ defines a distribution D (thus $\mathfrak{C}([c]_1)$ depends only on $\rho_1([c]_1)$), then controllability of $\mathfrak{C}$ is determined by Proposition 2.6. This problem has also been studied in the nonlinear controls literature. For example see (Hermann and Krenner, 1977). □

First we will show that if $\mathfrak{C}$ is controllable then there exists a complete subset Λ of T^*Q such that, for every Lagrangian L on Q, the triple $(L, \mathfrak{C}, \Lambda)$ is controllable. We also show that, given a complete subset Λ of T^*Q, it is possible to make it smaller by removing those forces which do work against the constraints. These results are presented as a means of demonstrating that our formulation agrees with what we expect from physical arguments.

If $\mathfrak{C}$ is a 1-constraint on Q and $\mathfrak{W}$ is a $\mathfrak{C}$-complementary force distribution, then in Section 6.2.1 we had defined $\mathfrak{C}^0$ to be the image of $J^1(\mathbb{R}, Q)$ under all $\mathfrak{C}$-constraint force fields and $\overline{\mathfrak{W}}$ to be the image of $J^1(\mathbb{R}, Q)$ under all $\mathfrak{W}$-admissible force fields.

Proposition 6.21 *Let $\mathfrak{C}$ be a 1-constraint on Q, let $\mathfrak{W}$ be a $\mathfrak{C}$-complementary force distribution, and let L be a Lagrangian on Q. Then the triple $(L, \mathfrak{C}, \overline{\mathfrak{W}})$ is controllable if and only if $\mathfrak{C}$ is controllable.*

Proof: It is clear that if $(L, \mathfrak{C}, \overline{\mathfrak{W}})$ is controllable then $\mathfrak{C}$ is controllable.

Now suppose that $\mathfrak{C}$ is controllable. Then, for any two points $q_1, q_2 \in Q$, there is a curve $c\colon [a,b] \to Q$ connecting q_1 and q_2 such that $c'(t) \in \mathfrak{C}(j^1c(t))$ for each $t \in [a,b]$. We will show the existence of a $\overline{\mathfrak{W}}$-compatible 1-force field, F, and a $\mathfrak{C}$-constraint force, λ, along c such that the conditions of Definition 6.19 are met.

We reparameterise c with a mapping $\tau\colon [a,b] \to [a,b]$ which has the following properties:

i) τ is a bijection,

ii) $\tau|(a,b)$ is a diffeomorphism, and

iii) $\tau'(a) = \tau'(b) = 0$.

Define $\tilde{c} = c \circ \tau$ as a new curve on Q which connects q_1 with q_2. Clearly $\tilde{c}$ satisfies the constraint $\mathfrak{C}$.

By definition we have

$$T_q^*Q = \mathfrak{W}([\tilde{c}]_1) \oplus (\mathfrak{C}([\tilde{c}]_1))^0$$

where $q = \rho_1([\tilde{c}]_1)$. Note that a 1-force at q is $\overline{\mathfrak{W}}$-compatible if it lies in $\mathfrak{W}([\tilde{c}]_1)$ and a 1-force at q is a $\mathfrak{C}$-constraint force if it lies in $(\mathfrak{C}([\tilde{c}]_1))^0$. We may now define $F(j^1\tilde{c}(t))$ and $\lambda(j^1\tilde{c}(t))$ uniquely by requiring that

$$F_L(j^2\tilde{c}(t)) = F(j^1\tilde{c}(t)) + \lambda(j^1\tilde{c}(t))$$

and asking that $F(j^1\tilde{c}(t)) \in \mathfrak{W}([\tilde{c}]_1)$ and $\lambda(j^1\tilde{c}(t)) \in (\mathfrak{C}([\tilde{c}]_1))^0$. Therefore, we see that $\tilde{c}$ is a solution to Lagrange's equations with force field F and constraint

$\mathfrak{C}$. Furthermore, λ is a $\mathfrak{C}$-constraint force, and F is $\overline{\mathfrak{W}}$-compatible. Finally, by construction, $\tilde{c}'(a)$ and $\tilde{c}'(b)$ are both zero. This proves the proposition. ■

Note that, although we are able to steer the system from a point at rest to another point at rest, it may not remain at rest after time b if the final configuration is not an equilibrium point for the Lagrangian force field.

We present these results for the rolling penny.

Example 6.22 (5.16 cont'd) Recall that the constraint for the rolling penny is

$$\mathfrak{C}([c]_1) = D(q)$$

where $q = \rho_1([c]_1)$ and where the distribution D is defined by

$$D = \ker\{dx - r\cos\phi d\theta, dy - r\sin\phi d\theta\}.$$

We have already seen that a $\mathfrak{C}$-constraint force is of the form

$$\lambda([c]_1) = \lambda_1([c]_1)(dx - r\cos\phi d\theta) + \lambda_2([c]_1)(dy - r\sin\phi d\theta).$$

Note that we may define a $\mathfrak{C}$-complementary force distribution by

$$\mathfrak{W}([c]_1) = \langle d\theta, d\phi\rangle_{\mathbb{R}}.$$

That this is a $\mathfrak{C}$-complementary force distribution follows since

$$\{dx - r\cos\phi d\theta, dy - r\sin\phi d\theta, d\theta, d\phi\}$$

is linearly independent. Thus we may write any $\mathfrak{W}$-admissible force field as

$$F([c]_1) = F_1([c]_1)d\theta + F_2([c]_1)d\phi.$$

We may think of F_1 as forces which contribute to "pure rolling" and F_2 as forces which contribute to "pure spinning." Proposition 6.21 says that we may steer between any two configurations by applying $\mathfrak{W}$-admissible forces (i.e., by applying "rolling" and "spinning" torques). Furthermore, Proposition 6.21 gives a way to compute the torques if we can determine a path on Q which satisfies the constraints. □

Now we show that if we have a Λ-compatible force field with constraint $\mathfrak{C}$, we may always make Λ "smaller" by removing from it the forces which do work against the constraints. Let $\mathfrak{C}$ be a 1-constraint and let $\mathfrak{W}$ be a $\mathfrak{C}$-complementary force distribution. For each $q \in Q$ we define

$$\Lambda_{a,q} = \begin{cases} \Lambda_q \cap \overline{\mathfrak{W}}, & \Lambda_q \cap \overline{\mathfrak{W}} \neq \emptyset \\ 0 \in T_q^*Q, & \Lambda_q \cap \overline{\mathfrak{W}} = \emptyset \end{cases}$$

$$\Lambda_{c,q} = \begin{cases} \Lambda_q \cap \mathfrak{C}^0, & \Lambda_q \cap \mathfrak{C}^0 \neq \emptyset \\ 0 \in T_q^*Q, & \Lambda_q \cap \mathfrak{C}^0 = \emptyset. \end{cases}$$

Thus the subsets

$$\Lambda_a = \overset{\circ}{\bigcup_{q\in Q}} \Lambda_{a,q}, \qquad \Lambda_c = \overset{\circ}{\bigcup_{q\in Q}} \Lambda_{c,q}$$

are complete subsets of T^*Q.

We now have the following result.

Lemma 6.23 *Let Λ be a complete subset of T^*Q, let $\mathfrak{C}$ be a 1-constraint on Q, and let $\mathfrak{W}$ be a $\mathfrak{C}$-complementary force distribution. Then any Λ-compatible 1-force field F may be uniquely decomposed as $F = F_a + F_c$ where F_a is Λ_a-compatible and F_c is Λ_c-compatible.*

Proof: Let $[c]_1 \in J^1(\mathbb{R}, Q)$. By definition we have the unique decomposition

$$F([c]_1) = F_a([c]_1) + F_c([c]_1)$$

where $F_a([c]_1) \in \mathfrak{W}([c]_1)$ and $F_a([c]_1) \in (\mathfrak{C}([c]_1))^0$. It is clear that F_a and F_c so defined are Λ_a and Λ_c-compatible, respectively. ■

Now we have the following result which says that we may effectively consider only Λ_a-compatible force fields out of those which are Λ-compatible.

Proposition 6.24 *Let Λ be a complete subset of T^*Q, let $\mathfrak{C}$ be a 1-constraint on Q, let $\mathfrak{W}$ be a $\mathfrak{C}$-complementary force distribution, and let F be a Λ-compatible 1-force field. Let L be a Lagrangian on Q. Then $c\colon [a,b] \to Q$ is a solution to Lagrange's equations with force field F and constraint $\mathfrak{C}$ if and only if c is a solution to Lagrange's equations with force field F_a and constraint $\mathfrak{C}$. In particular, $(L, \mathfrak{C}, \Lambda)$ is controllable if and only if $(L, \mathfrak{C}, \Lambda_a)$ is controllable.*

Proof: By definition, c is a solution to Lagrange's equations with force field F and constraint $\mathfrak{C}$ if and only if $c'(t) \in \mathfrak{C}(j^1c(t))$ and there exists a $\mathfrak{C}$-constraint force λ along c such that

$$F_L(j^2c(t)) = F(j^1c(t)) + \lambda(j^1c(t))$$

for each $t \in [a,b]$. We may write

$$F(j^1c(t)) = F_a(j^1c(t)) + F_c(j^1c(t))$$

where $F_a(j^1c(t)) \in \mathfrak{W}(j^1c(t))$ and $F_c(j^1c(t)) \in (\mathfrak{C}(j^1c(t)))^0$ by Lemma 6.23. If we define

$$\tilde{\lambda}(j^1c(t)) = F_c(j^1c(t)) + \lambda(j^1c(t))$$

we see that

$$F_L(j^2c(t)) = F_a(j^1c(t)) + \tilde{\lambda}(j^1c(t))$$

which gives the result, since F_a is Λ_a-admissible and $\tilde{\lambda}$ is a $\mathfrak{C}$-constraint force by definition. ■

This may be presented for the rolling penny.

Example 6.25 (5.16 cont'd) If we have a force field which we apply to the rolling penny of the form

$$F([c]_1) = F_x([c]_1)dx + F_y([c]_1)dy + F_\theta([c]_1)d\theta + F_\phi([c]_1)d\phi,$$

Proposition 6.24 says that any path we can follow with these forces may also be followed by a force field which has $F_x([c]_1) = F_y([c]_1) = 0$. □

Remark 6.26 In practice we may wish to choose the $\mathfrak{C}$-complementary force distribution $\mathfrak{W}$ in a particular manner, perhaps to "minimise" the forces wasted doing work against the constraints. However, in the general mathematical formulation, this is not reflected. □

Chapter 7

Conclusions and Future Work

In this dissertation we have developed some aspects of mechanics and control of mechanical systems. Since mechanical systems form a large and interesting class of control systems, and since this class of systems has not received much fundamental attention in the literature, we have tried to establish a solid foundation for analysis of mechanical control systems. It is hoped that this will be merely the first step on a road to developing a complete set of tools for analysis *and* synthesis of controllers for these systems.

7.1 Conclusions

When classical mechanics was in its infancy, the concepts of external forces and constraints were always considered an integral part of any formulation of mechanics. These two notions were, in large part, lost with the recent geometrisation of mechanics, and only recently has there been an attempt to revive constraints and inputs to put them in a proper geometric framework. In this dissertation we have made some additions to this effort. In Sections 5.4 and 5.3 we gave an intrinsic formulation of mechanics with external forces in the Lagrangian and Hamiltonian settings, respectively. The presentation here is loosely based upon that of (Hermann, 1982). However, we use differential two-forms for our formulation. Locally this determines a Pfaffian module which is the Cartan system of the two-form. This is explained for the Lagrangian case in Section 5.7.

Another intrinsic approach to formulating equations of motion in the Lagrangian framework employs a new geometric object which we call the Lagrange force field. This terminology is explained in Section 5.6. With this approach one may allow very general notions of forces and constraints. This is used to advantage in Section 6.2.2 to obtain some preliminary results for control of mechanical systems with constraints.

The other part of the dissertation deals with proper control theory for certain classes of mechanical systems. We may break up the main results in this area into two parts: Lagrangian (Section 4.1) and Hamiltonian (Section 4.2). The former is the more original and interesting of the two. In each case we attempt to apply the formalism of the basic nonlinear control theory presented in Chapter 3 to the class of mechanical systems we are studying.

In the Lagrangian formalism we studied a class of mechanical control systems which we call "simple mechanical control systems." These systems are characterised by their Lagrangian being "kinetic energy minus potential energy." With this class of systems, it is most meaningful to be able to define controllability in terms of the configuration variables. We are then able to use the structure of the system to develop algebraic conditions for testing this configuration controllability. As we show with examples, the distinction between determining controllability in the configuration variables and determining controllability in the configurations *and* velocities is important. In many problems it is the controllability in the configurations which is more useful to us. Our results in Section 4.1.7 may be summarised as follows:

1. We have provided *new* definitions of controllability for mechanical systems. These new versions of controllability are made in terms of the configuration variables for the mechanical system, as this is often what is more interesting. This new version of controllability is a natural one to consider for simple mechanical control systems.

2. We have reduced the number of computations which need to be performed to answer the configuration controllability question. The computations (covariant differentiation and Lie bracket) in the controllability tests we derive are performed on vector fields on Q rather than on TQ. Also, the number of operations (covariant differentiation and Lie bracket) which need to be performed is half that which need to be performed in computing Lie brackets of vector fields on TQ. For example, the bracket

   $$[[X_L, Y_a^{lift}], [X_L, [Y_b^{lift}, [X_L, Y_c^{lift}]]]]$$

   is represented by the expression

   $$[Y_a, \langle Y_b : Y_c \rangle]$$

   in our controllability test.

3. In computing the distributions on Q which determine configuration controllability, we see how the system geometry enters into the problem. Of particular interest is the appearance of the symmetric product. This is something that we would not have guessed before we started working on this problem.

On the Hamiltonian side, we complete the analysis which is presented in (Nijmeijer and van der Schaft, 1990) for Hamiltonian control systems. In particular, we state very precisely the form of the locally accessible dynamics and the strongly locally inaccessible dynamics. With the assumed structure of the Hamiltonian control system, we see that the dynamics in each case is Hamiltonian. In the example of the robotic leg, we see how the decompositions from Hamiltonian control theory are related to the classical reductions by group actions which may be performed for this problem.

The examples presented in Section 4.3, while simple, provide valuable insight into the usefulness of the techniques we have introduced.

7.2 Future Work

The work presented in this dissertation on the topic of mechanical control systems is only a beginning of what *can* be done. Here is a list of possible directions for future work which follow naturally from our analysis of simple mechanical control systems in Section 4.1.

1. Extend the analysis of Section 4.1 to determine the structure of the reachable sets when the initial velocity is non-zero. This, we feel, is connected with the problem of determining the structure of the strongly reachable sets.

2. Allow for more general Lagrangians and inputs. Although simple mechanical systems make up a large number of mechanical control systems, the generalisation to more general Lagrangians and more general inputs may offer more insight into the mechanisms at work in controlling Lagrangian systems. It would be particularly interesting if such analyses could be presented utilising the framework for Lagrangian mechanics presented in Sections 5.4 and 5.6.

3. Generalise the computations of Sections 4.1 to show what happens in the situation when the configuration manifold is acted upon by a Lie group which leaves the problem data invariant. Some work of this type is seen in (Bloch and Crouch, 1992). Since systems with symmetry have received a lot of attention in the recent literature (see (Marsden and Ratiu, 1994) and the references contained therein), this would serve to connect our analysis with some existing analysis, hopefully to the benefit of both.

4. Apply the analysis tools of Section 4.1 to systems with constraints. In (Bloch *et al.*, 1992a) some analysis of this type is performed. However, the hypotheses in their work are restrictive in that a large set of possible forces is assumed. In fact, it is assumed that it is possible to apply forces in all directions complementary to the constraint forces. As we saw in Proposition 6.21, in this situation it is quite natural to suppose that the system will be controllable.

5. The "Holy Grail" in this type of analysis would be the assimilation of the above steps into a complete control theory for systems with constraints *and* symmetries as presented in (Bloch *et al.*, 1994). Some interesting, but preliminary, results may be found in (Ostrowski and Burdick, 1995). In particular the "Snakeboard" example is presented in this paper and is shown to be STLCC.

The directions stated above are along the lines of developing tools for *analysis* of mechanical systems. Just as important, however, is the development of *synthesis* tools. It would be very interesting to develop algorithms for controlling simple mechanical control systems which may be shown to be STLCC. A possible example of existing work which may prove valuable is the work of (M'Closkey, 1995). In this dissertation, stabilisation algorithms are presented for nonlinear systems without a drift vector field. Moreover, some results are presented which are intended to model the "dynamic extension" of the kinematic results. However, this extension must be used with care as it may not properly capture the subtle dynamic effects as seen in the examples presented in Section 4.3.

The picture of our work on the Hamiltonian side is more complete. However, it would still be useful to make connections between the Hamiltonian and Lagrangian presentations in Chapter 4 of this dissertation. Perhaps some of the clean results of the Hamiltonian theory could be combined with the inherently more useful Lagrangian results to yield a deeper understanding of each in the cases where they agree.

Another very interesting avenue of future research may involve using the techniques of Chapter 5 to study fairly general mechanical control systems. The idea here would be to use the two-form $\Omega(L, F)$ (or $\Omega(H, F^*)$ in the Hamiltonian case) as the basic object which describes the control system. The results obtained in this way would have much more of an exterior differential systems flavour, and the ideas and results in Sections 3.2 and 3.3.3 may be helpful here.

Bibliography

Abraham, R., and Marsden, J. E. 1978. *Foundations of Mechanics.* Second edn. Reading, MA: Addison-Wesley.

Abraham, R., Marsden, J. E., and Ratiu, T. S. 1988. *Manifolds, Tensor Analysis, and Applications.* Second edn. New York-Heidelberg-Berlin: Springer-Verlag.

Aeyels, D., and Szafranski, M. 1988. Comments on the stabilizability of the angular velocity of a rigid body. *Systems & Control Letters*, **10**, 35–39.

Arnol'd, V. I. (ed). 1988. *Dynamical Systems.* Vol. III. New York-Heidelberg-Berlin: Springer-Verlag.

Bloch, A. M., and Crouch, P. E. 1992. Kinematics and dynamics of nonholonomic control systems on Riemannian manifolds. *Pages 1–5 of: Proceedings of the 32nd IEEE Conference on Decision and Control.* Tucson, AZ: IEEE.

Bloch, A. M., and Crouch, P. E. 1995. Nonholonomic and vakonomic control systems on Riemannian manifolds. *SIAM Journal on Control and Optimization*, **33**(January), 126–148.

Bloch, A. M., Reyhanoglu, M., and McClamroch, N. H. 1992a. Control and stabilization of nonholonomic dynamic systems. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, **37**, 1746–1757.

Bloch, A. M., Krishnaprasad, P. S., Marsden, J. E., and Sánchez de Alvarez, G. 1992b. Stabilization of rigid body dynamics by internal and external torques. *Automatica. The Journal of IFAC. The International Federation of Automatic Control*, **28**, 745–756.

Bloch, A. M., Krishnaprasad, P. S., Marsden, J. E., and Murray, R. M. 1994. *Nonholonomic Mechanical Systems with Symmetry.* Tech. rept. CDS 94-013. Caltech Department of Control and Dynamical Systems. Submitted to Archives for Rational Mechanics and Analysis. Available electronically from `http://avalon.caltech.edu/cds`.

Brockett, R. W. 1983. Asymptotic stability and feedback stabilization. *Pages 181–191 of:* Brockett, R. W., Millman, R. S., and Sussmann, H. J. (eds), *Differential Geometric Control Theory.* Boston/Basel/Stuttgart: Birkhäuser.

Bryant, R. L., Chern, S. S., Gardner, R. B., Goldschmidt, H. L., and Griffiths, P. A. 1991. *Exterior Differential Systems.* New York-Heidelberg-Berlin: Springer-Verlag.

Goldschmidt, H. L. 1967. Integrability criteria for systems of nonlinear partial differential equations. *Journal of Differential Geometry*, **1**, 269–307.

Golubitsky, M., and Guillemin, V. 1973. *Stable Mappings and Their Singularities.* New York-Heidelberg-Berlin: Springer-Verlag.

Hermann, R. 1982. The differential geometric structure of general mechanical systems from the Lagrangian point of view. *Journal of Mathematical Physics*, **23**, 2077–2089.

Hermann, R., and Krenner, A. J. 1977. Nonlinear controllability and observability. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, **22**(5), 728–740.

Hermes, H. 1978. On local controllability. *SIAM Journal on Control and Optimization*, **20**, 211–220.

Jurdjevic, V., and Sussmann, H. J. 1972. Control systems on Lie groups. *Journal of Differential Equations*, **12**, 313–329.

Kelly, S. D., and Murray, R. M. 1994. Geometric phases and locomotion. *The Journal of Robotic Systems.* To appear, available electronically via `http://avalon.caltech.edu/cds`.

Kharlomov, P. V. 1992. A critique of some mathematical models of mechanical systems with differential constraints. *Journal of Applied Mathematics and Mechanics. Translation of the Soviet journal Prikladnaya Matematika i Mekhanika*, **56**, 584–594.

Klingenberg, W. 1982. *Riemannian Geometry.* Berlin-New York: Walter de Gruyter.

Koiller, J. 1992. Reduction of some classical nonholonomic systems with symmetry. *Archive for Rational Mechanics and Analysis*, **118**, 113–148.

Kozlov, V. V. 1992. The problem of realizing constraints in dynamics. *Journal of Applied Mathematics and Mechanics. Translation of the Soviet journal Prikladnaya Matematika i Mekhanika*, **56**, 594–600.

Krishnaprasad, P. S. 1985. Lie-Poisson structures, dual-spin spacecraft, and asymptotic stability. *Nonlinear Analysis. Theory, Methods, and Applications*, **9**, 1011–1035.

Lang, S. 1984. *Algebra.* Second edn. Reading, MA: Addison Wesley.

Lewis, A. D., and Murray, R. M. 1995a. *Equilibrium Controllability of a Class of Mechanical Systems.* Submitted to the 34th IEEE Conference on Decision and Control. Available electronically via `http://avalon.caltech.edu`.

Lewis, A. D., and Murray, R. M. 1995b. Variational principles for constrained systems: Theory and experiment. *International Journal of Nonlinear Mechanics.* To appear.

Lewis, A. D., Ostrowski, J. P., Murray, R. M., and Burdick, J. W. 1994 (May). Nonholonomic mechanics and locomotion: The Snakeboard example. *Pages 2391–2400 of: Proceedings of the 1994 IEEE International Conference on Robotics and Automation.* IEEE, San Diego.

Li, Z., Montgomery, R., and Raibert, M. 1989. Dynamics and control of a legged robot in flight phase. *In: Proceedings of the 1989 IEEE International Conference on Robotics and Automation.* IEEE.

Libermann, P., and Marle, C.-M. 1987. *Symplectic Geometry and Analytical Mechanics.* Dordrecht/Boston/Lancaster/Tokyo: D. Reidel Publishing Company.

Marmo, G., Saletan, E. J., Simoni, A., and Vitale, B. 1985. *Dynamical Systems: A Differential Geometric Approach to Symmetry and Reduction.* New York, New York: John Wiley and Sons.

Marsden, J. E., and Ratiu, T. S. 1994. *Introduction to Mechanics and Symmetry.* New York-Heidelberg-Berlin: Springer-Verlag.

M'Closkey, R. T. 1995. *Exponential Stabilization of Driftless Nonlinear Control Systems.* Ph.D. thesis, California Institute of Technology.

Meyer, G. 1971. *Design and Global Analysis of Spacecraft Attitude Control Systems.* Tech. rept. TR R-361. NASA, Ames Research Center.

Murray, R. M., and Sastry, S. S. 1993. Nonholonomic motion planning: Steering using sinusoids. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, **38**(5), 700–716.

Nijmeijer, H., and van der Schaft, A. J. 1990. *Nonlinear Dynamical Control Systems.* New York-Heidelberg-Berlin: Springer-Verlag.

Ong, C. P. 1975. Curvature and mechanics. *Advances in Mathematics*, **15**, 269–311.

Ostrowski, J. P., and Burdick, J. W. 1995. *Control of Dynamical Systems with Symmetries and Nonholonomic Constraints.* Submitted to the 1995 IEEE Conference on Decision and Control. Available electronically via `http://robby.caltech.edu/papers`.

San Martin, L., and Crouch, P. E. 1984. Controllability on principal fibre bundles with compact structure group. *Systems & Control Letters*, **5**, 35–40.

Serre, J.-P. 1992. *Lie Algebras and Lie Groups.* Lecture Notes in Mathematics, vol. 1500. New York-Heidelberg-Berlin: Springer-Verlag.

Sontag, E. D. 1990. *Mathematical Control Theory: Deterministic Finite Dimensional Systems.* New York-Heidelberg-Berlin: Springer-Verlag.

Sussmann, H. J. 1983. Lie brackets and local controllability: A sufficient condition. *SIAM Journal on Control and Optimization*, **21**, 686–713.

Sussmann, H. J. 1987. A general theorem on local controllability. *SIAM Journal on Control and Optimization*, **25**(January), 158–194.

Wang, L. S., and Krishnaprasad, P. S. 1992. Gyroscopic control and stabilization. *Journal of Nonlinear Science*, **2**, 367–415.

Weber, R. W. 1986. Hamiltonian systems with constraints and their meaning in mechanics. *Archive for Rational Mechanics and Analysis*, **91**, 309–335.

Weinstein, A. 1977. *Lectures on Symplectic Manifolds.* Conference Board of the Mathematical Sciences, Regional Conference Series in Mathematics, vol. 29. Providence, Rhode Island: American Mathematical Society.

Yan, F. Y. M. 1995. *Introduction to the Calculus of Variations and its Applications.* New York/London: Chapman & Hall.

Yang, R. 1992. *Nonholonomic Geometry, Mechanics and Control.* Ph.D. thesis, University of Maryland, College Park.

Author Index

Index

Note: *Italicised numbers indicate where the indexed quantity is defined.*