Robotica (2007) volume 25, pp. 765–777. © 2007 Cambridge University Press
doi:10.1017/S0263574707003815 Printed in the United Kingdom

Is it worth learning differential geometric methods for modeling and control of mechanical systems?

Andrew D. Lewis

Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada.

(Received in Final Form: July 27, 2007. First published online: September 26, 2007)

SUMMARY
Evidence is presented to indicate that the answer is, "Yes, sometimes."

1. Introduction

Before I get to the actual writing of this paper, let me say a few words about the manner and spirit in which it is written. Normally when writing technical articles I try to be rigorous and precise, giving clear definitions, stating clear theorems, etc. However, I believe that this should not be the tone of this paper, given the intent of the special issue for which it is being written. Therefore, I make no claims of the paper being precise, rigorous, or even remotely self-contained. Readers looking for a summary of differential geometric methods in mechanics will not find it here. Many such summaries have been written, and many of them are good. Moreover, many books have been written on the topic of differential geometric modeling of mechanical systems, and many of these are good. A reader who has an inkling that they want to learn the differential geometric method will already have ventured into this part of the literature, so it seems pointless to simply reproduce it, or even a small part of it, here.

Instead, I aim to give the reader some insights into *why* I think differential geometric modeling is important. In doing so I will (1) make free use of intuitive "definitions" of concepts and (2) make free use of terminology and notation that I have not defined, but is hopefully either clear or for which the absence of a precise definition is inobtrusive. Thus, the paper should not be read as if it is a source of technical information, but rather as an exposition about technical information that can be found elsewhere.

One of my hopes for this paper is that it will show that there are real advantages to be had from the differential geometric techniques of modeling and control of mechanical systems. Consistent with this, I will also make the point that there are *disadvantages* of not using differential geometric techniques. Moreover, I will illustrate this latter point through particular instances. Therefore, the paper will be, in a very few places I hope, somewhat provocative. In keeping with this, I have decided to write the paper in a less formal and more personal style.

The most common feedback I get from applied researchers not familiar with the mathematics I use is (1) "I do not understand your notation" and (2) "I am not sure you are getting any mileage from all of the mathematics." In this paper, I will attempt to address these concerns.

E-mail: andrew@mast.queensu.ca

First, let me address the notation matter. I am sympathetic to this criticism in that I am certain that some parts of differential geometric notation seem almost intentionally cryptic if one is not used to them. I understand from my own experience that these notational issues can be a serious obstruction to understanding something new. However, my own experience also suggests that sometimes notation can be a placeholder for ideas, and it is the ideas behind the notation that should be focussed upon. I would, therefore, caution against disregarding something as being "mere notation" without taking the time to see whether the notation is actually a manifestation of something real. The affine connection framework for mechanics that I will review in this paper certainly makes use of notation that will be unfamiliar to those with a more classical mechanics background. However, I think that the framework brings with it ideas that can be useful; the notation is merely a consequence of the existence of these ideas.

The other matter—that of whether differential geometric methods actually contribute anything in practice—has real substance, and it is to this matter that this paper is primarily addressed. There is wonderful elegance and clarity in the differential geometric formulation of mechanics. Much of the work done in the area of overlap between differential geometry and mechanics—indeed, almost all of it until, maybe, the past 15 years—has been in understanding the structure of the mathematical models of the mechanics. While such research is certainly justified on intellectual grounds by its outcome, to an engineer looking for solutions to specific problems, it is not clear that the differential geometric formulation contributes to these ends. Thus, any skepticism that might be felt toward differential geometric techniques is justified, at least until it is clearly shown by people like me that these techniques contribute toward the solution of problems that are of interest. What I will try to do in this paper is indicate some of the kinds of problems where differential geometric ideas have made a clear and decisive contribution. I will argue that this contribution might arise in two ways:

1. It is sometimes the case that, merely by formulating a problem in differential geometric language, the right questions reveal themselves. This is often an overlooked contribution of precise mathematical formulations in general. In solving a problem, it is often the question that is the real problem. Sometimes, once one has the right question, the problem either (a) solves itself or (b) reveals itself to be, in a precise way, impossible. For either

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press

eventuality, knowing the right question is much more than half the battle.
2. Sometimes the question one is asking is perfectly clear, and it is the solution that is the real problem. In some such cases (but definitely not all), differential geometric methods are just the thing, and can lead one to unexpectedly simple solutions.

I should make some effort to position what I say in this paper, relative to the growing volume of literature on differential geometric methods for control of mechanical systems. It is impossible to give an accurate overview of the entirety of this research area, and I will not attempt to do so. Instead I will identify research directions that have achieved a somewhat polished state. Probably, the first serious progress in uniting geometric mechanics and geometric control theory was made along Hamiltonian lines. The port-controlled Hamiltonian approach is representative of this, and a huge amount of work has been done here. I refer to the papers of van der Schaft[25]; Dalsmo and van der Schaft[12] as representative of this work. A book that allies this Hamiltonian approach with topics of interest to readers of this journal is that of Stramigioli.[21]

Another significant body of research, which is connected firmly with the origins of geometric mechanics, is that undertaken by researchers including John Baillieul, Anthony Bloch, Peter Crouch, P. S. Krishnaprasad, Naomi Leonard, Jerry Marsden, and Dmitry Zenkov. An account of some of this work can be found in the recent book of Bloch.[4]

The approach to mechanics and control taken in this paper is the one developed in detail in my book[9] coauthored with F. Bullo from the University of California, Santa Barbara. In this book, we develop from "first principles" an approach to the modeling of mechanical systems using Riemannian and affine differential geometry. After this, we outline some of the control problems that can be addressed using the differential geometric modeling techniques. I will not attempt to summarize the content of this book. One of the aims, instead, is to try to give the reader an idea of whether reading this book is worth their effort. For some readers, I believe it *will* be worth the effort, but for others, it may not be.

2. Modeling Mechanical Systems Using Affine Differential Geometry

In this section, I present a quick overview of Chapter 4 of Bullo and Lewis.[9] The intention is to introduce the major mathematical players in the affine differential geometric formulation of mechanics. A novice reader expecting to read this section and walk away understanding affine differential geometry and its use in mechanics will be disappointed. The fact of the matter is that if you are starting with only scant knowledge of differential geometry, there is simply no way one can appreciate a geometric formulation of mechanics in an afternoon. It will take a significant investment of effort and, more importantly, a probable abandoning of certain preconceptions learned during the course of a more classical mechanics education. Thus, what I attempt to do in this section is alert the reader to the "brute facts" of a geometric formulation of mechanics. These are the things that they will have to come to grips with in order to merely get off the ground.

2.1. Configuration manifold

One of the main reasons why differential geometry is an effective tool for formulating mechanics is that mechanical systems evolve naturally on a configuration space that is a manifold. That is to say, the set of configurations of a mechanical system is in natural one-to-one correspondence with a set that has the structure of a differentiable manifold; this manifold is called the configuration space and is typically denoted by Q. This idea is stated clearly by Bullo and Lewis.[9]

At this point I could define what a manifold is, but this seems pointless since this definition can be found in many places, including in Bullo and Lewis.[9] Moreover, the reader will probably have a good idea, or at least some intuition, about what a manifold is. Instead, therefore, let me indicate why the notion of a manifold is important in robotics. The reader is surely familiar with the fact that the configurations of a rigid body fixed at a point in some inertial frame are in one-to-one correspondence with the set SO(3) of 3×3 orthogonal matrices with determinant 1. What is the best way to view SO(3)? Some choose to parameterise SO(3) with Euler angles. This makes it appear as though SO(3) is some subset of three-dimensional Euclidean space, although clearly it is not. Other representations of SO(3) are used (e.g., quaternion representations) to attempt to get a handle on SO(3). But from my point of view, SO(3) is a manifold and that is it. (Actually, SO(3) is a special manifold, since it is also a Lie group, and this Lie group structure is extremely important in robotics. However, I prefer not to get into this aspect of rigid body modeling here, but refer to Bullo and Lewis;[9] Murray *et al.*[19]) A manifold is essentially a set that can locally be parameterised by an open subset Euclidean space (e.g., by Euler angles), and different parameterisations are required to be compatible with one another. A specific choice of parameterisation is sometimes helpful in working on a concrete example. And the existence of parameterisations in general is sometimes useful in that it allows you to believe, at least temporarily, that you are working with familiar calculus in Euclidean space (typically by saying, "Let $(q^1, \ldots, q^n)$ be coordinates for Q"). However, the important thing about manifolds is that, *when used properly*, they force a coordinate independent restriction on the problem formulation. Thus, one is forced to deal with the real structure of the problem and not the one tacked on *ad hoc* by a choice of coordinates.

Let me try to explain this in a concrete context. In Fig. 1, I have depicted two simple systems that are frequently used as benchmark-type systems to implement algorithms, and so on. In the top left is a simple mobile robot. Suppose that the three wheels are actuated, so this is really just a fully actuated planar rigid body. The objective is to steer the body to the $\times$ with a prescribed orientation. The bottom right of Fig. 1 depicts the classical pendulum/cart system where the objective is to balance the pendulum in the upright configuration by the application of force to the cart. One approach to modeling the mobile robot problem is to declare the coordinates describing the configurations of the system to be (x, y, θ), and then proceed as if the configuration

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press

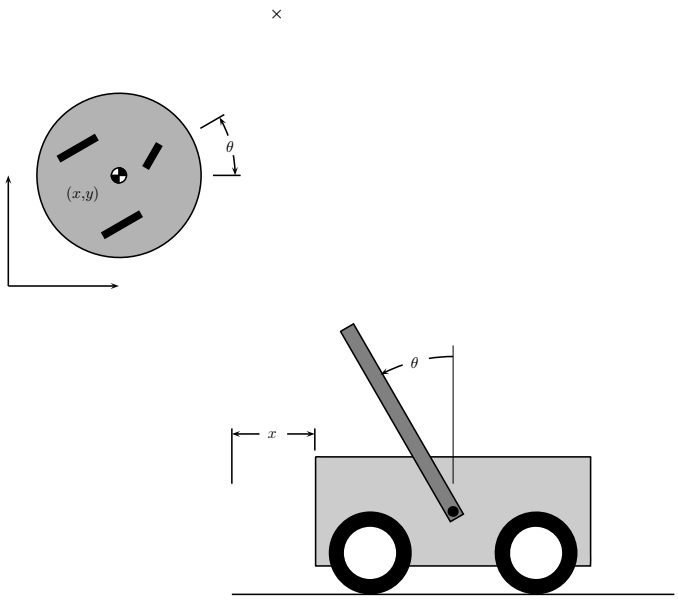


Fig. 1. These systems are not globally stabilisable using continuous feedback.

space is $\mathbb{R}^3$. Similarly, for the pendulum/cart, one could take coordinates (x, θ) and proceed as if the configuration space is $\mathbb{R}^2$. In practice, this is often accentuated by writing the coordinates as $[x\ y\ \theta]^T$ and $[x\ \theta]^T$ to drive home the fact the these are "coordinate vectors." But the configuration spaces are not $\mathbb{R}^3$ or $\mathbb{R}^2$ since, in each case, θ and $\theta + 2k\pi$ describe the same configuration for every integer k. The configuration spaces are, in fact, SE(2) (which is naturally diffeomorphic to $\mathbb{R}^2 \times \mathbb{S}^1$) and $\mathbb{R} \times \mathbb{S}^1$, where $\mathbb{S}^1$ is the unit circle. (Other representations of the configuration spaces are possible, but they will be equivalent to what I have given in the precise way that they are diffeomorphic.)

A good question is, "Is it important to make this distinction?" I believe that it is, and let me give one reason why. It is possible for a mechanical control system with a configuration space of $\mathbb{R}^n$ to possess a globally stable equilibrium point using continuous feedback. However, it is *not* possible for a mechanical control system with a configuration space having rotational degrees of freedom to have a globally stable equilibrium point using continuous feedback. In particular, there cannot be equilibria for the systems depicted in Fig. 1 that are globally stable using continuous feedback. Despite this, the literature is filled with papers where continuous globally stabilizing feedbacks are designed for the systems in Fig. 1 or systems like them. That is to say, merely by believing that the configuration manifold is Euclidean space, many researchers are led to claim something impossible. This may seem as if it is not an important point on which to be fussy, but Bhat and Bernstein [3] illustrate clearly why this matter of rotational degrees of freedom are, in fact, something the control designer ought to be aware of.

2.2. Tangent bundle of the configuration manifold

The configuration manifold describes the set of configurations of a mechanical system. The tangent bundle of the configuration manifold describes the set of configurations and velocities of a mechanical system. Intuitively, the tangent space at a point $q \in Q$, denoted by T_qQ, is the set of all velocities at this configuration. From this intuitive description, I hope it is evident that T_qQ is a vector space of the same dimension as Q. The tangent bundle, denoted by TQ, is the union of all tangent spaces. The tangent bundle itself is a differentiable manifold, and inherits from Q a differentiable structure in a natural way. Coordinates for TQ are often denoted by $((q^1, \dots, q^n), (v^1, \dots, v^n))$. Thus, TQ has twice the dimension of Q. The natural differentiable structure for TQ works out well since it demands that, if one has two compatible sets of coordinates $(q^1, \dots, q^n)$ and $(\tilde{q}^1, \dots, \tilde{q}^n)$ for Q, then the corresponding velocities obey the rather chain rule-like formula

$$\tilde{v}^j = \frac{\partial \tilde{q}^j}{\partial q^k} v^k, \qquad j \in \{1, \dots, n\}.$$

(Here, I use the summation convention commonly employed in differential geometry where summation is implied over indices that are repeated.) The tangent bundle must be thought of as the state space for the system. That is to say, the dynamics of the system are uniquely determined by initial conditions in the tangent bundle, this by virtue of the equations of motion being second-order (as we shall soon see, in case you have forgotten).

One piece of useful notation concerns the natural basis for a tangent space induced by a set of coordinates. Thus, if $(q^1, \dots, q^n)$ are coordinates for Q, there is a natural basis for each tangent space, which is denoted by $\{\frac{\partial}{\partial q^1}, \dots, \frac{\partial}{\partial q^n}\}$. This basis has the property that a tangent vector with components $(v^1, \dots, v^n)$ has the representation $v^j \frac{\partial}{\partial q^j}$. The peculiar partial derivative notation for a tangent vector is motivated by the fact that tangent vectors are, in an appropriate sense, differential operators. But this interpretation need not concern us here.

There is one caveat I would like to make concerning the tangent bundle. The representation of tangent bundle coordinates as $((q^1, \dots, q^n), (v^1, \dots, v^n))$ would seem to imply that there is some sort of "product structure" to the tangent bundle. That is to say, it appears that the tangent bundle is the Cartesian product of the configuration manifold and the set of velocities. *This is not the case.* It is true that TQ is *locally* a product, but even this local product representation is not natural; it arises from a specific choice of coordinates, and different coordinates give different local product representations. Thus, while the coordinates for TQ seem to validate the notation of representing points in TQ as (q, v), this representation is, in a precise way, incorrect. It certainly is the case that one would like to be able to specify the configuration at which a velocity lives. A good way to do this is by writing a point in TQ as v_q, thus declaring by the notation that this is a tangent vector in the tangent space at $q \in Q$.

2.3. External forces

In a geometric setting for mechanics, forces take values in the cotangent bundle. Let me quickly describe the cotangent bundle. The cotangent space at $q \in Q$ is the dual space of

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press

T_qQ (i.e., the set of $\mathbb{R}$-valued linear maps on the vector space T_qQ) and is denoted by T^*_qQ. The cotangent bundle is the union of all the cotangent spaces and is denoted by T^*Q. Often, authors feel compelled to attempt to give an intuitive idea of what cotangent vectors are, typically because it is felt that cotangent vectors are somehow more mysterious than tangent vectors. My own view is that the definition is perfectly adequate, and one just has to accept it and learn that various physical quantities, e.g., forces, naturally are represented as elements of the cotangent bundle.

But why should forces take values in the cotangent bundle? Well, what does a force do? This is actually a fairly involved metaphysical question, but let me sidestep the metaphysics and just answer the question: A force does work on a system as it moves. Precisely, the situation is as follows. Suppose that a mechanical system undergoes a motion described by a curve $t \mapsto \gamma(t)$ in Q. As it undergoes this motion, a force $t \mapsto F(t)$ is applied to the system (being a little vague about just where F lives for the moment). The work done is then

$$W = \int_I F(t) \cdot \gamma'(t)\, \mathrm{d}t,$$

where I is the time interval on which the motion is defined, i.e., the work is the integral of the product of force with velocity. I have just written the product $F(t) \cdot \gamma'(t)$ as if it means something, but let me consider what F must be in order for this expression to have meaning. The expression $F(t) \cdot \gamma'(t)$ should live in $\mathbb{R}$ (work is a scalar quantity) and should be linear in $\gamma'(t)$. Thus, $F(t)$ can live in no place other than $T^*_{\gamma(t)}Q$. This is the explanation I give about why forces are cotangent bundle-valued, and my undergraduate students seem to buy it. Note that the "$\cdot$" in the expression $F(t) \cdot \gamma'(t)$ is not the "dot product." It represents an element of the dual space acting on a vector.

Just as there is a natural basis for tangent vectors induced by a choice of coordinates, there is also a basis for cotangent vectors associated with coordinates $(q^1, \dots, q^n)$. This basis is denoted by $\{\mathrm{d}q^1, \dots, \mathrm{d}q^n\}$ so that a force has the coordinate representation $F_j \mathrm{d}q^j$. It is not entirely obvious how one goes from the representation of forces in Newtonian mechanics to the geometric representation. This is discussed by Bullo and Lewis[9] and Murray *et al.*[19]

Forces in mechanics, at least when one is considering the control theory of mechanical systems, come in two flavours (1) external forces coming from the environment (potential forces, friction, and dissipative forces, etc.) and (2) control forces. Let me denote the total uncontrolled external force by F, noting that F may be a function of position, velocity, and/or time. I write the total control force as a linear combination of forces $F^1, \dots, F^m$. Thus, the control force is

$$\sum_{a=1}^{m} u^a F^a,$$

where I ask that the forces F^a be functions only of position, but allow the controls u^a to be functions of position, velocity, and/or time, as necessitated by the control scheme used.

2.4. *Kinetic energy metric*

The key to the affine differential geometric approach to mechanics is the fact that the kinetic energy of the system defines a Riemannian metric on the configuration space. Researchers in robotics will know this Riemannian metric as the "inertia matrix," the "mass matrix," or something similar. In the classical way of looking at things, the kinetic energy metric is a symmetric positive-definite matrix-valued function of configuration. In the geometric way of looking at things, the kinetic energy metric is a smooth assignment of an inner product to each tangent space of the configuration space. In differential geometry, such an assignment is called a Riemannian metric. In the particular case where the assignment comes from the kinetic energy for a mechanical system, I shall call the corresponding Riemannian metric the kinetic energy metric.

The notation used in Bullo and Lewis[9] for the kinetic energy metric is $\mathbb{G}$. Thus, the inner product on the tangent space at $q \in Q$ is denoted by $\mathbb{G}(q)$ (in the geometric formulation, the kinetic energy metric is a genuine map, it being a section of a certain tensor bundle), and the inner product of two tangent vectors $u_q, v_q \in T_qQ$ is denoted by $\mathbb{G}(q)(u_q, v_q)$, or simply by $\mathbb{G}(u_q, v_q)$ since the point $q \in Q$ is explicit in the notation already. Therefore, the kinetic energy is the function on TQ given by

$$\mathrm{KE}(v_q) = \frac{1}{2}\mathbb{G}(v_q, v_q).$$

At a point $q \in Q$, $\mathbb{G}(q)$ is an inner product on the tangent space T_qQ. Therefore, it is defined by the n^2 components $\mathbb{G}_{jk} = \mathbb{G}(\frac{\partial}{\partial q^j}, \frac{\partial}{\partial q^k})$, $j, k \in \{1, \dots, n\}$. The way in which one represents the kinetic energy metric in coordinates is then $\mathbb{G} = \mathbb{G}_{jk} \mathrm{d}q^j \otimes \mathrm{d}q^k$.

Once one has at hand a physical model for the system, including useful representations of the configuration space and its tangent bundle, the actual task of writing the kinetic energy metric is actually quite simple, at least in principle. Of course, in practice, there are genuine issues involved with determining the inertial properties of one's system, etc., but there is not much one can say about these matters, within or without a geometric formulation of mechanics.

There is a useful bit of notation associated with a Riemannian metric $\mathbb{G}$. Since I have found that this notation can be off-putting to many readers, let me devote a few moments to an apology for it. Given a Riemannian metric $\mathbb{G}$ on Q, there are naturally defined maps $\mathbb{G}^\flat : TQ \to T^*Q$ and $\mathbb{G}^\sharp : T^*Q \to TQ$ defined by

$$\mathbb{G}^\flat(v_q) \cdot u_q = \mathbb{G}(u_q, v_u), \qquad u_q, v_q \in T_qQ$$
$$\alpha_q \cdot \mathbb{G}^\sharp(\beta_q) = \mathbb{G}^{-1}(\alpha_q, \beta_q), \qquad \alpha_q, \beta_q \in T^*_qQ$$

where $\mathbb{G}^{-1}$ denotes the inner product on T^*_qQ induced by that on T_qQ. If $\mathbb{G}_{jk}$, $j, k \in \{1, \dots, n\}$, are the components of $\mathbb{G}$ in a set of coordinates, then $\mathbb{G}^\flat$ is simply a multiplication by this matrix of components. If $\mathbb{G}^{jk}$, $j, k \in \{1, \dots, n\}$ are the components of the inverse of the matrix with components $\mathbb{G}_{jk}$, then $\mathbb{G}^\sharp$ is multiplication by this inverse matrix. The reason for the funny notation $\flat$ and $\sharp$ is that, in terms of indices, $\mathbb{G}^\flat$ takes an object whose indices are superscripts (a tangent vector) and converts them to an object whose indices are subscripts (a cotangent vector). Thus, $\mathbb{G}^\flat$ "flattens" the indices. In a similar manner, $\mathbb{G}^\sharp$ raises or "sharpens" the indices.

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press

2.5. Levi-Civita connection

The components of the description of a mechanical system described thus far—the configuration manifold, its tangent bundle, forces, and the kinetic energy metric—will be well known to anyone in robotics, although not always necessarily in just the way that I think of these things. However, the Levi-Civita connection that I will describe now is, if you are not familiar with it, really something new. While it can be related to some things that may be familiar, it *is* precisely none of these things. The Levi-Civita affine connection is what it is, and for those unfamiliar with it, this will take a little hard swallowing.

Since I am making the affine connection formulation of mechanics the centerpiece of this paper, I ought to state what an affine connection is. I suppose the reader to be familiar with the basic concept of a vector field on a manifold and some of the notation surrounding the differential geometry of dealing with vector fields. Let me work with a manifold Q (which may or may not be the configuration manifold for a mechanical system). An affine connection[1] on Q is a mapping that assigns to two vector fields X and Y a third vector field denoted $\nabla_X Y$, and this assignment has the following properties:

1. The map $(X, Y) \mapsto \nabla_X Y$ is $\mathbb{R}$-bilinear.
2. For vector fields X and Y and a function f, we have $\nabla_{fX} Y = f \nabla_X Y$.
3. For vector fields X and Y and a function f, we have $\nabla_X(fY) = f \nabla_X Y + (\mathcal{L}_X f)Y$, where $\mathcal{L}_X$ is the Lie derivative with respect to X.

The vector field $\nabla_X Y$ is called the covariant derivative of Y with respect to X. So this is the definition, and how very uninsightful it is! However, the notion of an affine connection comes up naturally in mechanics, and I will explain this as we go along. Let me first say a few general things about affine connections:

1. An affine connection is extra structure on a manifold. That is to say, manifolds do not come equipped with a nice natural affine connection that you can go ahead and start using. You have to introduce your own connection. Thankfully, in mechanics, there are affine connections that come from the mechanical problem data. We shall see how these arise.
2. An affine connection ∇ is described in coordinates $(q^1, \dots, q^n)$ by writing

$$\nabla_{\frac{\partial}{\partial q^j}} \frac{\partial}{\partial q^k} = \Gamma^l_{jk} \frac{\partial}{\partial q^l}, \qquad j, k \in \{1, \dots, n\},$$

which is always possible for n^3 functions Γ^l_{jk}, $j, k, l \in \{1, \dots, n\}$. These are called the Christoffel symbols for the affine connection, and they clearly depend on a choice of coordinates. Using these, the covariant derivative of two general vectors fields is easily written as

$$\nabla_X Y = \left(\frac{\partial Y^l}{\partial q^j} X^j + \Gamma^l_{jk} X^j Y^k \right) \frac{\partial}{\partial q^l}.$$

3. Associated with every Riemannian metric $\mathbb{G}$ is a unique affine connection called the Levi-Civita connection and denoted by me as $\overset{\mathbb{G}}{\nabla}$. There is a nice intrinsic characterisation of $\overset{\mathbb{G}}{\nabla}$ and for this I refer the reader to Bullo and Lewis.[9] Here, let me give a hands-on, but rather grimy characterisation by giving the formula for its Christoffel symbols:

$$\overset{\mathbb{G}}{\Gamma}{}^l_{jk} = \frac{1}{2} \mathbb{G}^{lr} \left(\frac{\partial \mathbb{G}_{rj}}{\partial q^k} + \frac{\partial \mathbb{G}_{rk}}{\partial q^j} - \frac{\partial \mathbb{G}_{jk}}{\partial q^r} \right), \quad j, k, l, \in \{1, \dots, n\}.$$

We shall see in Section 2.7 where this affine connection arises from.
4. Associated with an affine connection ∇ are special curves on Q called the geodesics of ∇. These admit a nice intrinsic description too, but let me simply give the differential equation in coordinates that these curves must satisfy:

$$\ddot{q}^l(t) + \Gamma^l_{jk} \dot{q}^j(t) \dot{q}^k(t) = 0, \qquad l \in \{1, \dots, n\}. \tag{1}$$

In coordinate-free language, Eq. (1) can be written as $\nabla_{\gamma'(t)} \gamma'(t) = 0$. Thus, $\nabla_{\gamma'(t)} \gamma'(t)$ is the slick way of writing what is on the left-hand side of Eq. (1), where $t \mapsto \gamma(t)$ is the curve that in coordinates has the form $t \mapsto (q^1(t), \dots, q^n(t))$.

It is through geodesics that one can most easily begin to understand what an affine connection "means." So let me say a few words about this:

a. If $\nabla = \overset{\mathbb{G}}{\nabla}$ is the Levi-Civita connection, then geodesics are those curves which locally minimize length.
b. Note that the coordinate expression in Eq. (1) for $\nabla_{\gamma'(t)} \gamma'(t)$ is a sum of "$\ddot{q}^l$" and "$\Gamma^l_{jk} \dot{q}^j(t) \dot{q}^k(t)$." *It is an act of unspeakable violence to break these terms apart.* These terms always belong together, and together they mean "$\nabla_{\gamma'(t)} \gamma'(t)$." Separately they mean absolutely positively nothing.[2] One consequence of this is that the expression "$\nabla_{\gamma'(t)} \gamma'(t)$" gives a way of understanding the acceleration along the curve γ. But note that this expression of acceleration is not intrinsic as it depends on the choice of affine connection. Some lengths are taken in Section 4.3.5 of Bullo and Lewis[9] to describe this.

The primary message of the preceding discussion is as follows:

> *Associated in a unique way with the kinetic energy metric $\mathbb{G}$ for a mechanical system, is an affine connection called the Levi-Civita connection.*

(Its importance will be made clear in Section 2.7.)

[1] What I call an "affine connection" is called a "linear connection" by some authors, and these authors have something else (something that is related, however) in mind when they say "affine connection." However, the use of "affine connection" as I use is fairly widespread.

[2] The expression $\ddot{q}^i$ makes sense as a local coordinate for the second jet bundle. But I do not wish to go in that direction, so let me stick to my expression of the unspeakable violence of rendering these terms as under, since you cannot be led wrong in accepting it.

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press

2.6. Nonholonomic constraints

The final ingredient in my description of mechanics before I write the equations of motion is nonholonomic constraints. Sometimes, it seems to me as if there are as many techniques for deriving the equations of motion in the presence of nonholonomic constraints as there are people who have thought about doing this. These methods are, at least the correct ones, all distinguished by one simple fact: *they are all the same!* That is to say, Method X for writing equations of motion for a system subject to nonholonomic constraints is correct if and only if it agrees with Newton's and Euler's equations in the intersection where Method X and Newton's and Euler's equations apply. Many authors claim mystical advantages of their correct technique over other (necessarily equivalent) correct techniques. The mystical advantages are too often described in some metaphysical form, and Newton and Euler do not even get mentioned. My take on this is: If your technique is equivalent to that of Newton and Euler, then you are correct and so is everyone else whose method is equivalent to that of Newton and Euler. There can be real issues in practice, such as coming up with a technique that is easy to implement. By no means do I wish to assert that these practical considerations do not merit consideration. However, there are no "in principle" advantages of any one correct method over any other correct method.

That all being said, let me just clarify here what I mean when I say "nonholonomic constraint," and how these constraints get accounted for in the equations of motion. A nonholonomic constraint on a configuration manifold Q is a distribution $\mathcal{D}$ on Q. That is, we select a subspace $\mathcal{D}_q$ of each tangent space T_qQ that describes the set of velocities admissible at that configuration. Generalizations are possible, e.g., to time-dependent or nonlinear constraints. But this complication adds little to the conceptual picture.

The way in which nonholonomic constraints get incorporated into the equations of motion is by adding a force, called the constraint force, to the equations of motion. A constraint force is, by definition, one that does no work on admissible velocities. That is, at a configuration $q \in Q$, a force $\lambda_q \in T_q^*Q$ is a constraint force if $\lambda_q \cdot v_q = 0$ for every $v_q \in \mathcal{D}_q$. The idea, as we shall see clearly in the next section, is that a constraint force is introduced into the problem as an unknown to be solved for, a little like, but not exactly like, a Lagrange multiplier.

2.7. Equations of motion

As I suggested in the preceding section, the equations of motion for a mechanical system are given by Newton's and Euler's equations that are based on simple laws of force and moment balance. The precise variant of Newton's and Euler's equations that one wishes to use—for example, the standard Euler–Lagrange equations, the Gibbs–Appell equations, Gauss's Principle of Least Action, the Principle of Virtual Work—is subject to taste and possibly other considerations. For my approach, the key is the following result.

Theorem 1 (General equations of motion). *Consider a mechanical system with configuration manifold Q, kinetic energy metric $\mathbb{G}$, external force F, control force $\sum_{a=1}^m u^a F^a$, and nonholonomic constraint $\mathcal{D}$. For a curve $t \mapsto \gamma(t)$, the following are equivalent:*

1. *The force/moment balance equations of Newton/Euler hold along γ.*
2. *γ satisfies the equations*

$$\mathbb{G}^\flat\big(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\gamma'(t)\big) = F(t,\gamma'(t)) + \sum_{a=1}^m u^a(t)F^a(\gamma(t)) + \lambda(t),$$
$$\gamma'(t) \in \mathcal{D}_{\gamma(t)},$$

where $t \mapsto \lambda(t)$ is a constraint force along γ.

For unconstrained systems, i.e., where the nonholonomic constraint is $\mathcal{D} = TQ$, this immediately gives the following corollary.

Corollary 2 (Equations of motion for unconstrained systems). *Consider an unconstrained mechanical system with configuration manifold Q, kinetic energy metric $\mathbb{G}$, external force F, and control force $\sum_{a=1}^m u^a F^a$. For a curve $t \mapsto \gamma(t)$, the following are equivalent:*

1. *The force/moment balance equations of Newton/Euler hold along γ.*
2. *γ satisfies the equation*

$$\mathbb{G}^\flat\big(\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\gamma'(t)\big) = F(t,\gamma'(t)) + \sum_{a=1}^m u^a(t)F^a(\gamma(t)).$$

It is perhaps worth comparing these equations with the equations of motion in the usual form seen in the robotics and mechanics literature. The usual form of the equations is something like

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{N}(t,\boldsymbol{q},\dot{\boldsymbol{q}}) = \boldsymbol{B}\boldsymbol{u}.$$

In components, the equations of Corollary 2 are

$$\mathbb{G}_{rl}\ddot{q}^l + \overset{\mathbb{G}}{\Gamma}{}^l_{jk}\dot{q}^j\dot{q}^k = F_r(t,q,\dot{q}) + \sum_{a=1}^m u^a F_r^a(q),$$
$$r \in \{1,\dots,n\}.$$

Thus, we can directly compare the terms in the two equations as per Table I. Note that I *do not* separate $\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}}$, and neither should you, given my comments in item 4b in Section 2.5.

Table I. Comparison of terms in "classical," "geometric component," and "intrinsic" formulation of equations of motion.

Classical form	Geometric coordinate form	Intrinsic form
$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}}$	$\mathbb{G}_{rl}\ddot{q}^l + \mathbb{G}_{rl}\overset{\mathbb{G}}{\Gamma}{}^l_{jk}\dot{q}^j\dot{q}^k$	$\overset{\mathbb{G}}{\nabla}_{\gamma'(t)}\gamma'(t)$
$\boldsymbol{N}(t,\boldsymbol{q},\dot{\boldsymbol{q}})$	$-F_r(t,q,\dot{q})$	$-F(t,\gamma'(t))$
$\boldsymbol{B}\boldsymbol{u}$	$\sum_{a=1}^m u^a F_r^a(q)$	$\sum_{a=1}^m u^a F^a(\gamma(t))$

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press

For systems with nonholonomic constraints, the usual way to proceed is to eliminate the constraint force in some way. In terms of the affine connection picture I am painting, the magical thing is that after elimination of the constraint force, one still gets equations involving an affine connection, although it is not generally the Levi-Civita connection for some Riemannian metric. I will not explain the way this works here since the most direct explanation involves manipulations with affine connections that are one step beyond elementary, and I do not want to get into too much intricate stuff. The elimination of the constraint force also results in a modification of the external and control forces. Let me just summarize descriptively the final tagline, and refer to Bullo and Lewis[9] for a thorough discussion of how this works.

For a mechanical system with configuration manifold Q, kinetic energy metric $\mathbb{G}$, external force F, control force $\sum_{a=1}^{m} u^a F^a$, and nonholonomic constraint $\mathcal{D}$, there exists (1) an affine connection $\overset{\mathcal{D}}{\nabla}$ depending on $\mathbb{G}$ and $\mathcal{D}$, (2) a TQ-valued map Y depending on F, $\mathbb{G}$, and $\mathcal{D}$, and (3) vector fields $Y_1, \ldots, Y_m$, depending on $F^1, \ldots, F^m$, $\mathbb{G}$, and $\mathcal{D}$, such that for a curve $t \mapsto \gamma(t)$, the following are equivalent:

1. *The force/moment balance equations of Newton/Euler hold along γ.*
2. *γ satisfies the equations*

$$\overset{\mathcal{D}}{\nabla}_{\gamma'(t)}\gamma'(t)) = Y(t, \gamma'(t)) + \sum_{a=1}^{m} u^a(t) Y_a(\gamma(t)),$$
$$\gamma'(0) \in \mathcal{D}_{\gamma(0)}. \tag{2}$$

Therefore, the point is that the equations of motion are given by Eq. (2), and a fundamental rôle in the equations is played by an affine connection.

2.8. An example

In this section, I will quickly go through an example to illustrate how the geometric concepts arise in a concrete context. The example I use is the so-called "snakeboard" system, which is a model for a commercially available skateboard-like product (http://www.snakeboard.com/). The mathematical model I present here originated in the paper of Lewis *et al.*[17] Here, I shall present an affine differential geometric formulation of the equations of motion; further details on this model can be found in Bullo and Lewis.[7, 9]

The model for the snakeboard is shown in Fig. 2. The idea of the system is that there is a central coupler, at each end of which are wheels whose angle relative to the coupler is controlled. On top of the coupler, between the wheels, is an inertial rotor that is also controlled. Thus, the system has five degrees of freedom and two controls. This is an example of a system with nonholonomic constraints, since one places on the velocities of the system the constraint that the wheels roll without slipping.

2.8.1. Configuration manifold. In the model considered here, the angle of the front and back wheels relative to the coupler are fixed as being equal. This means that a configuration of the system is specified by (1) the position of the coupler, (2) the angle of the wheels, and (3) the angle of the rotor. The coupler is essentially a planar rigid body, and it can take any configuration of a planar rigid body, i.e., a point in $\mathrm{SE}(2) \simeq \mathbb{R}^2 \times \mathbb{S}^1$. The wheels are then specified by an angle, i.e., by a point in $\mathbb{S}^1$. Similarly, the rotor angle is determined by a point in $\mathbb{S}^1$. The configuration space is then $Q = \mathrm{SE}(2) \times \mathbb{R}^1 \times \mathbb{S}^1$. I use coordinates $(x, y, \theta, \psi, \phi)$ for Q, where the physical meaning of these coordinates is as indicated in Fig. 2 (and with (x, y) being the position of the center of mass of the coupler).

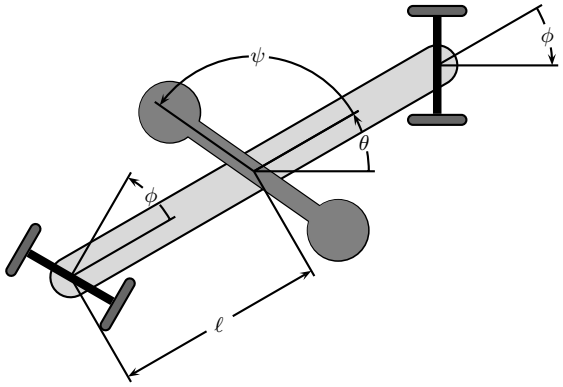


Fig. 2. Snakeboard.

Note that in simple planar systems like the snakeboard, it is often easy to "by hand" determine the configuration manifold and useful coordinates for it. However, for complicated interconnected systems, this process can often be a challenge. Some guidance for doing this in a systematic way can be found in Bullo and Lewis[9] and Murray *et al.*[19] Moreover, the modeling techniques outlined in Bullo and Lewis[9] make this step of determining the configuration manifold the step of paramount importance. That is to say, once this step is done properly, the other facets of the model will often fall into place fairly easily.

2.8.2. Tangent bundle to the configuration manifold. The coordinates $(x, y, \theta, \psi, \phi)$ induce natural coordinates $((x, y, \theta, \psi, \phi), (v_x, v_y, v_\theta, v_\psi, v_\phi))$ for the tangent bundle TQ. A point in the tangent bundle can be written as

$$v_x \frac{\partial}{\partial x} + v_y \frac{\partial}{\partial y} + v_\theta \frac{\partial}{\partial \theta} + v_\psi \frac{\partial}{\partial \psi} + v_\phi \frac{\partial}{\partial \phi}$$

relative to the basis of vector fields $\{\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial \theta}, \frac{\partial}{\partial \psi}, \frac{\partial}{\partial \phi}\}$ induced by the coordinate chart.

2.8.3. External forces. In the simple model considered here, the only external forces are the control forces. These consist of torques applied to the wheels and the rotor. A general force F is then written in coordinates as

$$F = F_x \mathrm{d}x + F_y \mathrm{d}y + F_\theta \mathrm{d}\theta + F_\psi \mathrm{d}\psi + F_\phi \mathrm{d}\phi.$$

In the particular case of our control forces, a torque applied to the rotor will have the form $F^1 = u_\psi \mathrm{d}\psi$, and a torque applied to the wheels will have the form $F^2 = u_\phi \mathrm{d}\phi$. Here, u_ψ and u_ϕ are to be thought of as the controls, and so are

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press

functions of time, position, and/or velocity, depending on the control task that one is carrying out.

2.8.4. Kinetic energy metric. I use the following physical parameters:

m_c mass of coupler;
m_r mass of rotor;
m_w mass of each wheel assembly;
J_c inertia of coupler about center of mass;
J_r inertia of rotor about center of mass;
J_w inertia of wheel assembly about center of mass;
ℓ distance from coupler center of mass to wheel assembly.

The kinetic energy is easily determined by "elementary methods" to be

$$\begin{aligned}\mathrm{KE} = {} & \frac{1}{2}(m_c + m_r + 2m_w)\left(v_x^2 + v_y^2\right) \\ & + \frac{1}{2}(J_c + J_r + 2(J_w + m_w\ell^2))v_\theta \\ & + \frac{1}{2}J_r v_\psi^2 + J_w v_\phi^2 + J_r v_\theta v_\psi.\end{aligned}$$

In differential geometric notation, the kinetic energy metric is then given by

$$\begin{aligned}\mathbb{G} = {} & (m_c + m_r + 2m_w)(\mathrm{d}x \otimes \mathrm{d}x + \mathrm{d}y \otimes \mathrm{d}y) \\ & + (J_c + J_r + 2(J_w + m_w\ell^2))\mathrm{d}\theta \otimes \mathrm{d}\theta \\ & + J_r\mathrm{d}\psi \otimes \mathrm{d}\psi + 2J_w\mathrm{d}\phi \otimes \mathrm{d}\phi \\ & + J_r(\mathrm{d}\theta \otimes \mathrm{d}\psi + \mathrm{d}\psi \otimes \mathrm{d}\theta).\end{aligned}$$

2.8.5. Levi-Civita connection. Since the components of the kinetic energy metric are constants, independent of the coordinates we are using, the Christoffel symbols of the Levi-Civita connection associated to $\mathbb{G}$ are all identically zero. However, since this is a system with nonholonomic constraints, the Levi-Civita connection, *per se*, is not of much interest.

2.8.6. Nonholonomic constraints. The nonholonomic constraints for the snakeboard are prescribed by no slip conditions on the wheels. These conditions are given by the two constraints

$$\begin{aligned}-\sin(\theta + \phi)v_x + \cos(\theta + \phi)v_y - \ell\cos\phi v_\theta = 0, \\ -\sin(\theta - \phi)v_x + \cos(\theta - \phi)v_y + \ell\cos\phi v_\theta = 0.\end{aligned}$$

To write a basis of vector fields that satisfy the constraints, let me introduce the vector field

$$V = \cos\theta\frac{\partial}{\partial x} + \sin\theta\frac{\partial}{\partial y}.$$

One can then check by direct computation that the three vector fields

$$X_1 = \ell\cos\phi V - \sin\phi\frac{\partial}{\partial\theta}, \quad X_2' = \frac{\partial}{\partial\psi}, \quad X_3 = \frac{\partial}{\partial\phi}$$

form a basis for the set of velocities satisfying the constraints (after noting that one must exclude points in Q where $\phi = \pm\frac{\pi}{2}$, since at these points, the dimension of the subspace of admissible velocities changes from three to four). Below we will see in the following that it is useful to use a basis for the admissible velocities that are orthogonal with respect to $\mathbb{G}$. One can check that if we define

$$X_2 = \frac{J_r\ell\cos\phi\sin\phi}{c_1(\phi)}V - \frac{J_r\sin^2\phi}{c_1(\phi)}\frac{\partial}{\partial\theta} + \frac{\partial}{\partial\psi},$$

where

$$\begin{aligned}c_1(\phi) = {} & (m_c + m_r + 2m_w)\ell^2\cos^2\phi \\ & + (J_c + J_r + 2(J_w + m_w\ell^2))\sin^2\phi,\end{aligned}$$

then the vector fields $\{X_1, X_2, X_3\}$ form a $\mathbb{G}$-orthogonal basis for the set of velocities satisfying the constraints.

Constraint forces are those which annihilate the velocities satisfying the constraints. Thus, a typical constraint force has the form

$$\begin{aligned}&\lambda_1(-\sin(\theta + \phi)\,\mathrm{d}x + \cos(\theta + \phi)\,\mathrm{d}y - \ell\cos\phi\,\mathrm{d}\theta) \\ &\quad + \lambda_2(-\sin(\theta - \phi)\,\mathrm{d}x + \cos(\theta - \phi)\,\mathrm{d}y + \ell\cos\phi\,\mathrm{d}\theta)\end{aligned}$$

for some coefficients λ_1 and λ_2.

2.8.7. Equations of motion. As mentioned earlier, there are many ways to write the equations of motion for a mechanical system subject to nonholonomic constraints. To represent the equations of motion using an affine connection as indicated in Eq. (2), I follow the approach of Bullo and Žefran,[11] which is explained in detail in Bullo and Lewis.[9] The idea of this approach is that one uses coordinates for TQ that are *not* the natural coordinates, but rather coordinates adapted to the constraints. Thus, one chooses a basis $\{X_1, \ldots, X_n\}$ of vector fields, the first r of which are a basis for the admissible velocities allowed by the constraints. In the particular version of this approach presented by Bullo and Žefran,[11] one uses a $\mathbb{G}$-orthogonal basis of vector fields. For the snakeboard, we thus use the orthogonal basis $\{X_1, X_2, X_3\}$ of vector fields for the admissible velocities, and write a typical admissible velocity as

$$\dot{q}^j(t)\frac{\partial}{\partial q^j} = v^\alpha(t)X_\alpha(\gamma(t))$$

hence defining v^1, v^2, and v^3. The functions v^1, v^2, and v^3 are often known in the literature as "pseudo-velocities." One can verify that

$$\begin{aligned}v^1 &= \cos\theta\cos\phi\dot{x} + \sin\theta\cos\phi\dot{y} - l\sin\phi\dot{\theta} - \frac{J_r\ell\sin\phi}{c_1(\phi)}\dot{\psi}, \\ v^2 &= \dot{\psi}, \\ v^3 &= \dot{\phi},\end{aligned}$$

provided that $(\dot{x}, \dot{y}, \dot{\theta}, \dot{\psi}, \dot{\psi})$ satisfy the constraints. Moreover, using the methodology outlined in Bullo and Lewis,[9] one can compute the equations of motion to be

$$\dot{x} = \ell\cos\phi\cos\theta v^1 + \frac{J_r\ell\cos\phi\sin\phi\cos\theta}{c_1(\phi)}\dot{\psi},$$

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press

$$\dot{y} = \ell \cos\phi \sin\theta v^1 + \frac{J_r \ell \cos\phi \sin\phi \sin\theta}{c_1(\phi)}\dot{\psi},$$

$$\dot{\theta} = -\sin\phi v^1 - \frac{J_r \sin^2\phi}{c_1(\phi)}\dot{\psi},$$

$$\dot{v}^1 = -\frac{(J_c + J_r + 2J_w - (m_c + m_r)\ell^2)\cos\phi \sin\phi}{c_1(\phi)} v^1\dot{\phi} - \frac{J_r(m_c + m_r + 2m_w)\ell^2 \cos\phi}{c_1(\phi)^2}\dot{\psi}\dot{\phi},$$

$$\ddot{\psi} = \frac{(m_c + m_r + 2m_w)\ell^2 \cos\phi}{c_2(\phi)} v^1\dot{\phi} + \frac{J_r(m_c + m_r + 2m_w)\ell^2 \cos\phi \sin\phi}{c_1(\phi)c_2(\phi)}\dot{\psi}\dot{\phi} + \frac{c_1(\phi)}{J_r c_2(\phi)}u_\psi\ddot{\phi} = \frac{1}{2J_w}u_\phi,$$

where

$$c_2(\phi) = (m_c + m_r + 2m_w)\ell^2 \cos^2\phi + (J_c + 2(J_w + m_w\ell^2))\sin^2\phi.$$

These equations may not appear to be derived from the geodesic Eq. (1) since the first three equations are first-order in configuration. This, however, is simply a manifestation of the technique of using pseudo-velocities.

3. Some Problems Whose Solution is Facilitated by Using Differential Geometry

The bottom line of the preceding section is Eq. (2), which includes the equations of motion for a large class of mechanical systems. These equations are compact, general, and reveal the geometric structure of the problem. However, the fact of the matter is that, if this is the end of what one can do with the affine connection formalism, then serious questions can be raised about whether there is a payoff in learning the mathematics necessary to understand the formulae. Thankfully, there are problems beyond the mere formulation of the equations of motion that are effectively addressed within the differential geometric framework. In this section, I investigate some of these. Since everything I will talk about here has already appeared in the literature, and since a detailed presentation of the results would necessitate reproducing huge chunks of the papers in question, I instead *describe* the results in the papers in question rather than reproduce them. I think that this approach is more in keeping with the objectives of the paper.

3.1. Controllability

One of the fundamental problems in the control of mechanical systems is, "Can a system be steered from Point A to Point B?" This is a controllability question. For systems that are fully actuated, i.e., when the control forces $F^1, \ldots, F^m$ span T^*_qQ for each $q \in Q$, these controllability questions are trivial since one can always apply a force to make the system go wherever you want. However, for underactuated systems, controllability questions are generally extremely difficult. Moreover, if the system is underactuated *and* if there are no uncontrolled external forces (e.g., no potential forces, so ruling out the multitude of pendulum systems that are often considered), then controllability becomes a difficult problem. The approach considered here is to give a brief general discussion of the sort of controllability problems I will consider, and then look at an example where I outline the controllability that can be deduced based on some more or less recent work that is heavily steeped in techniques of differential geometry.

The controllability problem that I consider is purely a local one. That is, I consider a mechanical system with configuration manifold Q and whose equations of motion are

$$\nabla_{\gamma'(t)}\gamma'(t) = \sum_{a=1}^{m} u^a(t) Y_a(\gamma(t)), \tag{3}$$

recalling from Section 2.7 that this equation models mechanical systems, possibly with nonholonomic constraints, that are subject to no uncontrolled external forces. I let $q_0 \in Q$ and denote by $\mathcal{R}_Q(q_0, \leq T)$ the set of configurations reachable from q_0 in time at most T. The system is small-time locally configuration controllable (STLCC) from q_0 if q_0 is an interior point of $\mathcal{R}_Q(q_0, \leq T)$. This definition is a bit pared down from the most general definition, and I refer to Bullo and Lewis[9] for full details. The results referred to in this section are from the works of Lewis and Murray,[15] Lewis,[14] Bullo and Lewis,[8] Lewis and Tyner.[16] All of the results in these papers rely heavily on the affine connection structure in Eq. (3).

As an example of the sort of system whose controllability can be described by the results I refer to, I consider the planar rigid body system depicted in Fig. 3. This is a system moving on a planar surface orthogonal to the direction of the gravitational field, and I assume that there are no friction forces. One might want to think of this system as an idealized model of a hovercraft whose motive force is supplied by a single thrust fan for which the direction and magnitude of the force can be varied. One problem of interest for this system is to steer the system from rest in a configuration q_A to rest in a configuration q_B. In order to do this, one would require that q_B be in the set of configurations reachable from q_A. Thus, the controllability problem is fundamental to the motion planning problem. In Lewis and Murray,[15] it is shown that this system is, indeed, STLCC from every

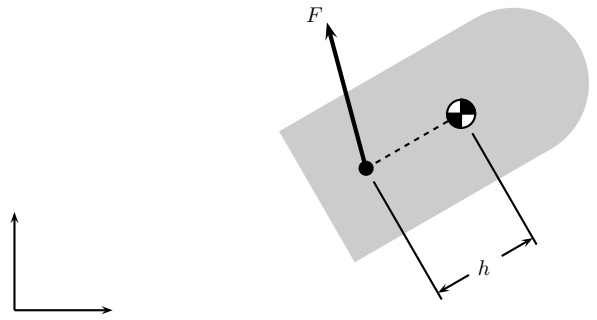


Fig. 3. A controlled planar body.

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press

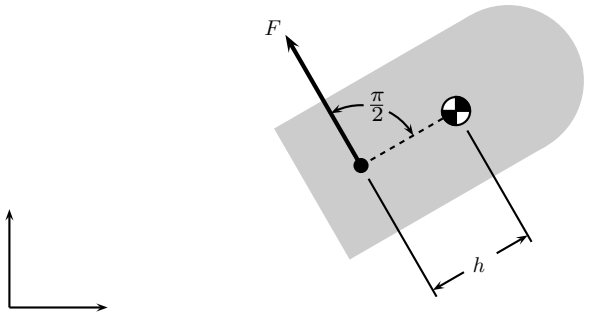


Fig. 4. A controlled planar body with a single control input.

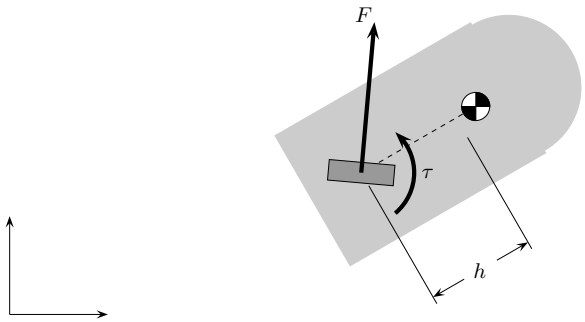


Fig. 5. A controlled planar body with the fan dynamics modeled.

configuration. The results in Lewis and Murray [15] are an affine differential geometric formulation of general accessibility results of Sussmann and Jurdjevic [23] and local controllability results of Sussmann. [22]

Let me now consider a slight alteration of the system by supposing that the direction of the force is now fixed, say at an angle of $\pi/2$ as shown in Fig. 4. Effectively, this system now has a single-input. It turns out that the controllability of this system cannot be determined using the results of Lewis and Murray. [15] However, these results were refined in a certain direction in Bullo and Lewis [8] where it is shown that this system, and indeed any single-input system of the form given in Eq. (3), is controllable from at most a "thin" set[3] of points.

Finally, let me consider another alteration of the model as depicted in Fig. 5. For this model, the fan is assumed to have inertia, so this adds an extra degree of freedom to the model. Once again, this system was the one whose controllability I was unable to analyse when I first encountered it. However, after some work by Lewis and Tyner [16] it was shown that this system was only STLCC from the configurations shown in Fig. 6 and SE(2)-translations of these configurations.

The point of the presentation in this section might be said to be the following.

1. Underactuated mechanical systems often have subtle local controllability properties. This is clearly evidenced in this section by the difficulty of determining the controllability of a simple planar body system with some simple, natural variations in the model.
2. By understanding the geometric structure of the model of the system, particularly the rôle played by the affine connection, it is possible to develop general results that allow the analysis of some nontrivial examples.

[3] Precisely, a set that is locally the set of zeros of a finite collection of analytic functions.

Moreover, the controllability results referred to in this section lead naturally to some techniques in motion planning that are described as follows.

3.2. Trajectory planning using decoupling vector fields

In this section, I shall describe what can be achieved in terms of converting the controllability results used in the preceding section to motion planning results. The idea described here was motivated by explorations of Arai *et al.*, [1] and Lynch *et al.* [18] The initial geometric presentation is due to Bullo and Lynch [10] and the most polished form of the results appear in Bullo and Lewis. [8] The idea is fairly simple. For a given mechanical system whose equations of motion are given by

$$\nabla_{\gamma'(t)}\gamma'(t) = \sum_{a=1}^{m} u^a(t) Y_a(\gamma(t)) \tag{4}$$

one seeks vector fields $X_1, \ldots, X_k$, called decoupling vector fields, that have the property that their integral curves, and any reparameterisations of these integral curves, can be followed by a trajectory of the system given in Eq. (4). One then hopes to find sufficiently many such vector fields that any motion planning problem can be solved by concatenating their integral curves.

Again, rather than restating the results which are given in the language of affine differential geometry, let me give a problem for which the ideas lead to a solution of the motion planning problem. The system considered here is the snakeboard system discussed in Section 2.8. The motion planning strategy presented here is described by Bullo and Lewis. [7]

The objective is to steer the snakeboard from an initial configuration at rest to a desired configuration at rest. Bullo and Lewis [7] show that the snakeboard has two decoupling vector fields and indicate how to use these to explicitly solve the motion planning problem. In Fig. 7, I show the trace of a solution of the motion planning problem where the snakeboard is steered from the position at the origin to a specified final position. The motion planning controller that generates Fig. 7 is very simple, but is directly derived from the controllability results of Bullo and Lewis. [8]

Other systems where the motion planning problem can be solved are given in Chapter 13 of Bullo and Lewis. [9]

The point of what I say in this section includes the following.

1. The idea of motion planning using decoupling vector fields is an excellent example of one where the right question is suggested by the differential geometric formulation.
2. Controllability is often seen as being a somewhat esoteric undertaking. Very often, systems are controllable in a simple way, and in such cases, sophisticated controllability techniques are more than what is needed. However,

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press

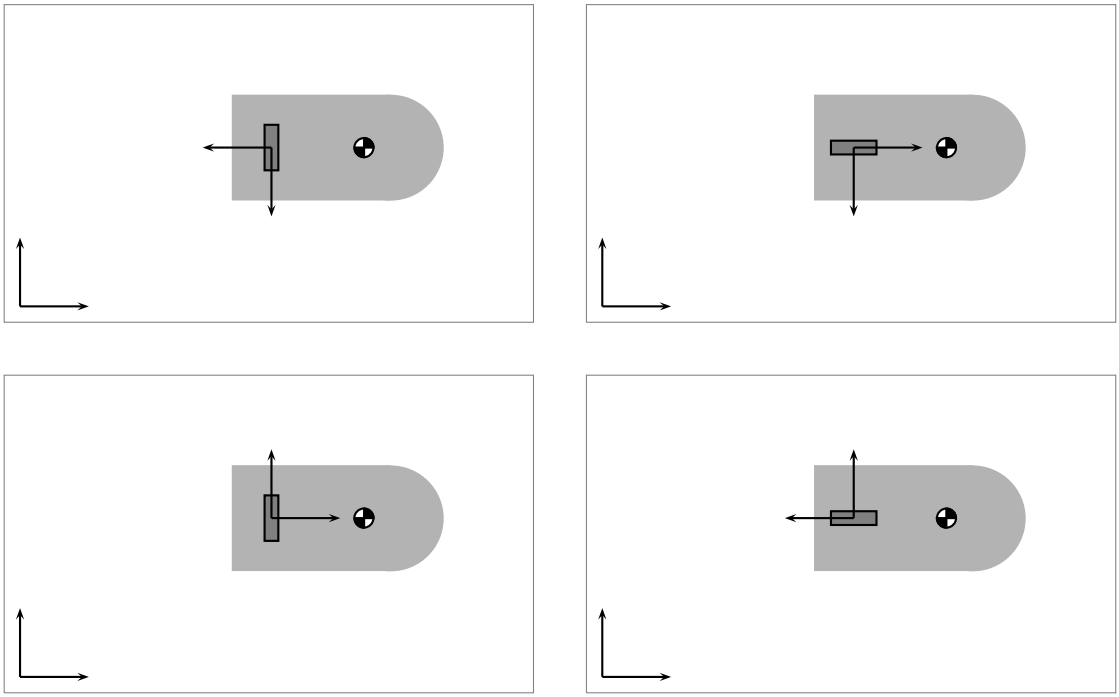


Fig. 6. Configurations from which the planar body with fan dynamics is controllable.

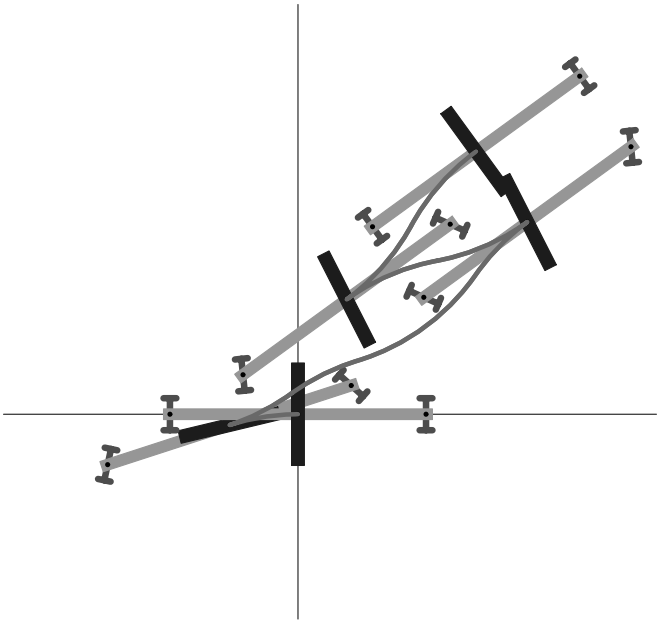


Fig. 7. A solution of the snakeboard motion planning problem.

mechanical systems provide a rich class of systems where controllability is often not easy to characterize. Moreover, by understanding the controllability properties of a system, it is possible that one might be able to derive explicit solutions to the motion planning problem.

3. While the affine connection formulation of nonholonomic systems is fairly complicated, the application of this formulation to the solution of the motion planning problem for the snakeboard justifies the complication.

3.3. Energy shaping

The last idea discussed in this section is something that is actually extremely complicated, and is still the topic of ongoing research. Suppose one is given a mechanical system, let me call it the "open-loop system," with kinetic energy metric $\mathbb{G}_{\text{ol}}$ and with the uncontrolled external forces being those coming from a potential function V_{ol}. Using feedback, one wishes to transform the system into a mechanical system, let me call it the "closed-loop system," with kinetic energy metric $\mathbb{G}_{\text{cl}}$ and external forces being those coming from a potential function V_{cl}. The value of this idea is that much can be said about the behavior of a mechanical system by knowing the properties of the potential function. In particular, equilibria and their stability are easily determined from the properties of the potential function. Thus, this technique of "energy shaping," i.e., changing the energy of the system using feedback, is a useful tool for stabilization. The original idea of potential shaping for fully actuated systems dates to Takegaki and Arimoto.[24] Potential shaping in the underactuated case was worked out by van der Schaft.[26] The possibility of shaping the kinetic energy was considered first in a Lagrangian setting by Bloch *et al.*[5,6] and in a Hamiltonian setting by Ortega *et al.*[20]

In our affine differential geometric formulation, energy shaping amounts to finding a state feedback u^a_{shp}, $a \in \{1, \dots, m\}$, that translates the open-loop system to the closed-loop system. In equations this reads

$$\overset{\mathbb{G}_{\text{ol}}}{\nabla}_{\gamma'(t)}\gamma'(t) + \mathbb{G}^{\sharp}_{\text{ol}} \circ \boldsymbol{d}V_{\text{ol}}(\gamma(t)) - \sum_{a=1}^{m} u^a_{\text{shp}}(v_q)\mathbb{G}^{\sharp}_{\text{ol}} \circ F^a(q)$$

$$= \overset{\mathbb{G}_{\text{cl}}}{\nabla}_{\gamma'(t)}\gamma'(t) + \mathbb{G}^{\sharp}_{\text{cl}} \circ \boldsymbol{d}V_{\text{cl}}(\gamma(t)).$$

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press

The process of energy shaping can be broken into two steps, kinetic energy shaping and potential energy shaping. For kinetic energy shaping, one seeks feedback $u^a_{\rm kin}$, $a \in \{1, \ldots, m\}$, such that

$$\sum_{a=1}^{m} u^a_{\rm kin}(v_q)\mathbb{G}^\sharp_{\rm ol} \circ F^a(q) = \underbrace{(\overset{\mathbb{G}_{\rm cl}}{\nabla} - \overset{\mathbb{G}_{\rm ol}}{\nabla})}_{D_{\rm ol,cl}}(v_q, v_q) \qquad (5)$$

where we use the fact that the difference of affine connections [$D_{\rm ol,cl}$ in Eq. (5)] is a tensor field. For potential energy shaping, one seeks feedback $u^a_{\rm pot}$, $a \in \{1, \ldots, m\}$, such that

$$\sum_{a=1}^{m} u^a_{\rm pot}(q)\mathbb{G}^\sharp_{\rm ol} \circ F^a(q) = \mathbb{G}^\flat_{\rm ol} \circ \mathbb{G}^\sharp_{\rm cl} \circ \boldsymbol{d}V_{\rm cl}(q) - \boldsymbol{d}V_{\rm ol}(q). \quad (6)$$

Equation (5) is a quasilinear first-order partial differential equation for the closed-loop metric $\mathbb{G}_{\rm cl}$. Equation (6) is a linear first-order partial differential equation for the closed-loop potential function $V_{\rm cl}$.

Thus, energy shaping leads to partial differential equations and, particularly when kinetic energy shaping is considered, these partial differential equations are extremely complicated. It is difficult to even give conditions ensuring that these partial differential equations have interesting solutions (there always exists the trivial solution $\mathbb{G}_{\rm cl} = \mathbb{G}_{\rm ol}$ and $V_{\rm cl} = V_{\rm ol}$). A rather coarse general result is given by Auckly and Kapitanski,[2] and they make use of some Riemannian geometry in their development. A fairly refined result concerning part of the problem is given by Lewis.[13] In both cases, the results are too complicated to reproduce here.

From a practical point of view, the jury is still out on the value of energy shaping as a means of stabilization. Because of the complexity of the partial differential equations involved, the method has actually been implemented only in simple examples (see refs. [2], [13]). Myself, I am dubious about the merits of the technique being widely applicable. However, the mathematical questions arising from energy shaping are deep and interesting, and may possibly shed some light on more practical matters. However, at this point, one cannot really be certain about any of this.

But the point is: Sometimes natural engineering questions can lead to extremely challenging problems that require sophisticated mathematics for their solution.

4. Conclusion

In this paper, I have attempted to illustrate that there is value in the differential geometric approach to modeling and control of mechanical systems. In particular, in Section 3, I have attempted to make it clear that this value extends beyond a mere elegant formulation of the equations of motion. I hope that uninitiated readers interested in the differential geometric approach might now attack the daunting literature with a sense of purpose, knowing that there are possible rewards awaiting them at the end of their journey.

References

1. H. Arai, K. Tanie and N. Shiroma, "Nonholonomic control of a three-DOF planar underactuated manipulator," *IEEE Trans. Robotics Autom.* **14**(5), 681–695 (1998).
2. D. R. Auckly and L. V. Kapitanski, "On the λ-equations for matching control laws," *SIAM J. Control Optim.* **41**(5), 1372–1388 (2002).
3. S. P. Bhat and D. S. Bernstein, "A topological obstruction to continuous global stabilization of rotational motion and the unwinding phenomenon," *Syst. Control Lett.* **39**, 63–70 (2000).
4. A. M. Bloch, *Nonholonomic Mechanics and Control* of *Interdisciplinary Applied Mathematics* **24** (Springer-Verlag, New York-Heidelberg-Berlin, 2003).
5. A. M. Bloch, N. E. Leonard and J. E. Marsden, "Controlled Lagrangians and the stabilization of mechanical systems. I. The first matching theorem," *IEEE Trans. Autom. Control* **45**(12), 2253–2270 (2000).
6. A. M. Bloch, D. E. Chang, N. E. Leonard and J. E. Marsden, "Controlled Lagrangians and the stabilization of mechanical systems. II. Potential shaping," *IEEE Trans. Autom. Control* **46**(10), 1556–1571 (2001).
7. F. Bullo and A. D. Lewis, "Kinematic controllability and motion planning for the snakeboard," *IEEE Trans. Robot. Autom.* **19**(3), 494–498 (2003).
8. F. Bullo and A. D. Lewis, "Low-order controllability and kinematic reductions for affine connection control systems," *SIAM J. Control Optim.* **44**(3), 885–908 (2005).
9. F. Bullo and A. D. Lewis. *Geometric Control of Mechanical Systems: Modeling, Analysis, and Design for Simple Mechanical Systems*. *Texts in Applied Mathematics* **49**. (Springer-Verlag, New York-Heidelberg-Berlin, 2004).
10. F. Bullo and K. M. Lynch, "Kinematic controllability and decoupled trajectory planning for underactuated mechanical systems," *IEEE Trans. Robot. Autom.* **17**(4), 402–412 (2001).
11. F. Bullo and M. Žefran, "On mechanical systems with nonholonomic constraints and symmetries," *Syst. Control Lett.* **45**(2), 133–143 (2002).
12. M. Dalsmo and A. J. van der Schaft, "On representations and integrability of mathematical structures in energy-conserving physical systems," *SIAM J. Control Optim.* **37**(1), 54–91 (1998).
13. A. D. Lewis, "Potential Energy Shaping After Kinetic Energy Shaping," *Proceedings of the 45th IEEE Conference on Decision and Control*, San Diego, CA (Dec. 2006) pp. 3339–3344.
14. A. D. Lewis, "Simple mechanical control systems with constraints," *IEEE Trans. Autom. Control* **45**(8), 1420–1436 (2000).
15. A. D. Lewis and R. M. Murray, "Controllability of simple mechanical control systems," *SIAM J. Control Optim.* **35**(3), 766–790 (1997).
16. A. D. Lewis and D. R. Tyner, "Controllability of a Hovercraft Model (and Two General Results)," *Proceedings of the 43rd IEEE Conference on Decision and Control*, Paradise Island, Bahamas, (Dec. 2004) pp. 1204–1209.
17. A. D. Lewis, J. P. Ostrowski, R. M. Murray and J. W. Burdick, "Nonholonomic Mechanics and Locomotion: The Snakeboard Example," *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*, San Diego, CA, (May 1994) pp. 2391–2400.
18. K. M. Lynch, N. Shiroma, H. Arai and K. Tanie, "Collision-free trajectory planning for a 3-DOF robot with a passive joint," *Int. J. Robot Res.* **19**(12), 1171–1184 (2000).
19. R. M. Murray, Z. X. Li and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation* (CRC Press, Boca Raton, Florida, 1994).
20. R. Ortega, M. W. Spong, F. Gómez-Estern and G. Blankenstein, "Stabilization of a class of underactuated mechanical systems via interconnection and damping assignment," *IEEE Trans. Autom. Control* **47**(8), 1218–1233 (2002).

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press

21. S. Stramigioli, *Modeling and IPC Control of Interactive Mechanical Systems—A Coordinate-Free Approach*. *Lecture Notes in Control and Information Sciences* **266** (Springer-Verlag, New York-Heidelberg-Berlin, 2001).
22. H. J. Sussmann, "A general theorem on local controllability," *SIAM J. Control Optim.* **25**(1), 158–194 (1987).
23. H. J. Sussmann and V. Jurdjevic, "Controllability of nonlinear systems," *J. Differential Equations* **12**, 95–116 (1972).
24. M. Takegaki and S. Arimoto, "A new feedback method for dynamic control of manipulators," *Trans. ASME Ser. G J. Dynam. Syst. Meas. Control* **103**(2), 119–125 (1981).
25. A. J. van der Schaft, "Port-controlled Hamiltonian systems: Towards a theory for control and design of nonlinear physical systems," *J. Soc. Instrum. Control Eng.* **39**(2), 91–98 (2000).
26. A. J. van der Schaft, "Stabilization of Hamiltonian systems," *Nonlinear Anal. TMA* **10**(10), 1021–1035 (1986).

https://doi.org/10.1017/S0263574707003815 Published online by Cambridge University Press