Abstract

HICKS, GREGORY PHILIP. Modeling and Control of a Snake-like Serial-link Structure (Under the direction of Kazufumi Ito).

The topic considered is the modeling and control of a snake-like serial-link structure. The system is assumed to have torque controls about the joints, is considered to lie in an isotropic plane, and is assumed to interact with this plane in a manner which adheres to some suitable friction laws. Such a structure is hyper-redundant, making the robotic realization thereof potentially robust with regards to mechanical failure and highly suited for obstacle avoidance tasks and terrain adaptability. It is for these reasons that the structure is studied. Lagrangian mechanics is used to develop a mathematical model for the system. The resulting dynamics possess symmetries which allow them to be placed in a reduced form. Using this form in conjunction with a technique known as feedback linearization, one finds that the dynamics are driven by a three state system describing the evolution of generalized momenta with respect to the device's internal shape progression. The problem is to determine whether or not there is a shape trajectory that can elicit bulk structure movement. In order to determine the appropriate shape for this task a two-pronged approach is taken. One approach is to make a shape selection based on the principle mechanism of undulatory locomotion. The other approach is to set up a variational problem to determine an optimal locomotive shape.

July 21, 2003

MODELING AND CONTROL OF A SNAKE-LIKE SERIAL-LINK STRUCTURE

BY

Gregory P. Hicks

A DISSERTATION SUBMITTED TO THE GRADUATE FACULTY OF

NORTH CAROLINA STATE UNIVERSITY

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

APPLIED MATHEMATICS

RALEIGH, NORTH CAROLINA

July 2003

APPROVED BY:

Kazufumi Ito
CHAIR OF ADVISORY COMMITTEE

H. T. Banks

E. L. Stitzinger

H. T. Tran

D. V. Zenkov

Biography

Gregory P. Hicks was born in Columbus, Ohio, in 1976. At age 5 his education began at Como Elementary School. At approximately age 7, his family moved to a farm located in the rural eastern Kentucky region of Cherokee. In a local township called Blaine, he finished the remainder of grade school and what would be called junior high school. In 1994 he barely graduated from the most disgraceful and pathetic of schools, Lawrence County High School, which is located in Louisa, Kentucky. In that same year he was reintroduced to the idea of learning at Ashland Community College, a local branch of the University of Kentucky. In 1996, he obtained an A.A. degree from this College with high distinction and transferred on scholarship to Morehead State University, which is located in Morehead, Kentucky. In December of 1998, he obtained a B.S. degree in Mathematics with Concentration in Computing with highest honors. Having acquired an associateship at North Carolina State University, he began graduate school in January of 1999. In 2000 he was awarded a M.S. degree in Applied Mathematics.

Acknowledgments

For reasons that are vast in scope, the following people are acknowledged by category. Many names could rightly be placed under multiple headings. However, I shall restrict inclusion to the grouping that indicates the capacity in which I have or had come to know them. I recognize:

Immediate Family Lawrence and Linda Hicks, Rocky Hicks, Rodney Hicks, and Kimberly Davis.

Extended Family Zella Griffith, Laura Stambaugh, and Earl Griffith.

Friends Clayton Camic, Brian Lewis, Joseph Sweeney, Justin Davis, Chris Thompson, Joseph Thompson, Karen Thompson, and Brandi Burke.

Teachers Delores Kazee, Jane Lowe, Brenda Thornberry, Mark Swetnum, Nancy McClelland, Max Jackson, Randy Ross, Lloyd Jaisingh, Dan Seth, Vivian Cyrus, John Boardman, Tom Banks, Dmitry Zenkov, Ernie Stitzinger, Hien Tran, Kazufumi Ito, and Stephen Campbell.

Professional Affiliates Bill Winfree, Madhav Marathe, and Robert Pleasant.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Topic of Study and its Motivation

The primary content of this writing concerns the modeling and locomotion control analysis of a snake-like serial-link structure. This structure is thought to, at least from a primitive morphological point of view, model the structure of a snake or eel. The interest in this structure is to determine how it enables the animal to attain steady and directed bulk locomotion. The majority of this interest is founded in the potentials of the snake to be highly adaptable to its environment and the transfer of these potentials to robots that may carry out tasks that are unsuitable for humans and other standard mechanisms. Some of these tasks are now briefly reviewed for the purpose of providing a legitimate motive for our studies.

We begin with discussion of the latent capability of a snake-like device for delivering non-destructive evaluation (NDE) technologies. It was this potential that initially prompted this research, which has been funded by the NDE group at the *NASA Langley Research Laboratory*. NDE refers to the inspection of areas or objects which is carried out in a manner that is non-degenerative with respect to the structure being inspected and its function. In response to the need for NDE in industry, many indirect, crafty methods have been developed. These methods typically involve the

introduction of some stimuli such as electric current, heat, or vibrations to the system and some measure of the system's response to these inducements via accelerometers, magnometers, etc. The data for both damaged and non-damaged samples are then studied to determine if there are some consistent and fundamental differences between them that would allow one to differentiate the cases and perhaps even different types of damage.

Even though many of these methods have proven to be highly applicable, they tend to have a localization weakness. As an example, consider methods that depend upon disturbances in the magnetic field associated with eddy currents formed by the introduction of electric current to the system. Applying such a technique to an airfoil would produce magnetic fields about the foiling and thus would be applicable to the metal sheathing itself and near-surface interior workings. However, the field may not be strong enough to detect deterioration in the inner framework of the wing. The same could be said for other technologies that invoke, in a similar manner, thermal and radiant energies. Thus, for complete assessment, the sensing medium must be taken closer to the inner workings of some systems. A snake would be ideal for such a task. Thus, the desire of those who develop NDE devices to investigate the possibility of a snake-like device that, like a real snake, would have the capability to navigate the inner workings of convoluted, condensed structures.

Another application of a snake-like robot, though along the same lines, is mine detection and elimination. According to [26], a snake-like robot would be the most appropriate vehicle for carrying sensors of various types into areas for the purpose of mine detection. As pointed out in this article, it is often the case that areas suspected of mine habitation lie dormant for extended periods of time. This results in overgrowth of these locals making them unsafe for most types of approach. Snakes, however, can traverse a variety of media, such as water, grass, and typical overgrowth with ease and without an intrusive presence. Snakes can pass over rough, uneven ground and their long, slender bodies allow them to manipulate winding paths. Fur-

ther, snakes have little difficulty moving over areas where the ground is softened. This is a result of their apparent capability to use the length of their body along with a continuum of contact with the surface to distribute their weight as is fitting for a given situation. Snakes also have a naturally stable form which leaves little concern for difficulties with balance. Finally, since snakes are short and have no appendages, their chances of activating mines via trip wires is significantly reduced. After the authors of [26], "*...they are nearly the perfect mine detectors.*"

A precedent in technologies that permeate the exterior of physical bodies with the notion of probing the interior workings has been in place for some time now in robotics. An example of such a contrivance is the use of large robotic "snakes" to search for human life amongst earthquake rubble and other tragic occurrences that yield massive destruction in populated and developed areas. Additional applications include the inspection and maintenance of narrow sewage pipes [40] and the inspection of feeder cables [53], to name a few.

It is hoped that these examples indicate the need to study snake-like structures. The potential utility of robots that can maneuver and manipulate their surroundings like an actual snake is hard to fathom. With this being said we now turn to the efforts that have been made heretofore so that we may put into perspective the objectives and contributions of our research. We make the note that this review is complete only to the extent that should be expected. Obviously, we did not study every publication in existence. However, we do believe that we have given appropriate attention to those that have significant bearing on our efforts and that we may safely claim that we are aware of the current level of maturation present in the subject area.

1.2 Survey of the Literature

The study of snake locomotion and of structures designed to mimic this form of transport began quite a long time ago. However, as pointed out by S. Hirose [33], the

first investigation of true engineering merit that focused solely on this phenomena was presented in 1946 by experimental biologist J. Gray [28]. In this paper Gray identifies and examines the 4 modes of locomotion observed amongst the genera of snakes. Although some of the conclusions reported in this article had been made previously by zoologist W. Mousour in 1932 [52], Gray provided, from the engineering point-of-view, comparably rigorous explanations of the locomotory capacity of the animal by appealing to the fundamental principles of Newtonian mechanics. A portion of this work followed easily from his earlier research on fish and eels [27]. Further, Gray made considerable efforts to confirm his mechanical insights via experimentation with actual snakes. These efforts were extended by the work of Gray's colleague and collaborator, fellow experimental biologist, H.W. Lissmann. Lissmann independently studied the sidewinding mode of locomotion in 1946. In 1949 he published a study with Gray in which they worked to gather a quantitative understanding of the forces acting on the snake's body during locomotion through experimentation [29].

The efforts of Gray and his peers within the research community seem to have ushered in a new area of study known today as biomechanics. The core of this branch of biology/zoology/morphology is to use the basic principles of mechanics to verify or predict the functionality of animal musculature and/or other morphological or anatomical features. By doing so, inference may be made into evolution of species, etc. After some period of dormancy, the analysis of limbless locomotion was revisited by the biomechanics community. In 1962, Carl Gans presented an essay on the subject [21]. In this essay Gans essentially presents the work of Gray and Lissmann. However, he added discussions on how modifications of the mechanisms of locomotion over differing classes of animals relate to selective advantage, he presented some additional experimental evidence, and did add some commentary on the morphology of the snake's skin/scale patterns and its apparent effect on the differing modes of locomotion. It would seem that this essay was well accepted and became something of a standard. The topic received little further attention until the late 1980's, other than

a few studies such as the one in skin morphology via scanning electron micrographs made by C. Gans and D. Baic in 1977 [23]. In this study the authors extrapolate scale functionality from examination of the rough-tailed snake. It was their conclusion that it was the function of the smooth ventral shields of most snakes to reduce friction and wear, thus aiding the undulatory mode of snake locomotion.

Upon completion of his thesis work in 1985, B.C. Jayne released a series of papers based on this work on snake locomotion. In 1988 Jayne provided his results on the mechanical properties of snake skin [36]. As part of this study he addressed possible correlations between the mechanical properties of snake skin and specializations in locomotion. One of the more interesting observations made was that localized thickenings of snake scales may contribute to the development of anisotropic or directionally dependent friction characteristics in snake skins. Although such a possibility was hinted at by the data collected during the experiments of Gray and Lissmann, they did not comment on the matter. In that same year, Jayne published his work on the muscular activity of two species of colubrid snakes during the lateral undulation mode of locomotion [37]. These EMG results indicated epaxial muscle activity consistent with muscular shorting during lateral undulation and were in agreement with the patterns and shortening that were predicted by Gray using very simple, but powerful, energy arguments.

This work was followed by a similar experiment by J.-P. Gasc et. al., reported in 1989 [24]. These researchers examined the muscle activity, as measure via EMG, of a Python regius when making use of a single rigid peg to obtain leverage for displacement. It would seem that the patterns observed were different from those of Jayne. However, Jayne was making use of a system of pegs in his experiments and it seems relatively clear that one would not expect the muscle activity observed to be consistent after so drastically changing the advantages of the snake's environment.

After a gap of several more years, other works directed at the complete understanding of snake locomotion of some significance have been produced by the biomechanics

community. In 1998, C. Gans and B. Moon published [50]. This research returned to the study of the muscle activity in snakes during locomotion that was produced by contact with one or more rigid pegs. Using EMG analysis along with videography analysis, they attempted to establish some relationship between the observed muscle activity and the axial bending and propulsion observed during locomotion. Their conclusions indicate that, as with Jayne's observations, Gray was probably correct in his evaluation of muscle shortening. However, it is also pointed out that a good deal of variability exists between the magnitude of the EMG bursts observed and the amount of axial bending observed. Thus indicating that muscles other than the epaxial muscles are involved in creation of the winding shape of the animal.

Another relatively recent article that is of interest is [32]. This work was produced by J. Hazel et. al. in 1998 and deals with the surface morphology of snake skin. Being more specific, multi-mode scanning probe microscopy was used to study the structure of the snake skin at the nanoscale level. It was found that the skin of the snake featured many structural properties which would seem to lend themselves well to aiding locomotion. One such feature was a forward-backward friction anisotropy. It is interesting to note that some of the features pointed out in this paper can also be observed from the microscopy results of [23]. However, advancement in the technologies involved allowed a higher resolution microscopy and some of the features, such as the existence of interlocking longitudinal ridges along the borders between two skin cells, were more deeply examined.

This brings us to the present with regards to those contributions to the subject of snake locomotion made by, roughly speaking, biologists. However, the face of the subject of snake locomotion was altered by the study of the topic made by S. Hirose. This study was Hirose's Ph.D. thesis work and was submitted to the Tokyo Institute of Technology in 1976. After more than 10 years of lying dormant, Hirose republished the work in the form of a book in 1987, citing an apparent need for such research to emerge [33]. The most impressive part of Hirose's study is its breadth. A single

statement is insufficient to describe the tasks that he performed.

Firstly, Hirose studied the biomechanics of snakes and applied some of his observations to conclude (although incorrectly) that none of the descriptions of the gliding form of snakes up to that point were worthy of merit, being unnatural from the point of view of muscular physiology. He proposed that a sinusoidal pattern for muscle contraction was most natural and upon assuming a direct relation of this activity to the axial bends of the snakes vertebrae structure, he derived a curve to describe the gliding form of the snake that is mathematically expressed using Bessel functions. He called this function the serpenoid curve. Hirose claimed that this curve was a good match to that taken by actual snakes in a certain controlled environment.

Using the continuum point-of-view for the form of the snake and assuming that a serial-link structure model of the snakes vertebra was an approximation to the continuum, Hirose developed kinematic expressions for the continuum form by the limiting process. Using these kinematic expressions in conjunction with his self-named serpenoid curve, Hirose drew the conclusion that decoupled directionally dependent friction was necessary to attain undulatory locomotion. Assuming this necessity to be the case, he went on to derive a variety of expressions for physical quantities related to snake gait, such as the power and motive force required for steady motion. This portion of the study took on an engineering flavor that was beyond the more mild hint of mechanics used by the forefathers of biomechanics.

Impressive was the extent to which Hirose went to confirm the validity of the kinematic expressions that he derived via actual experimentation with live snakes. Amongst these were admittedly less successful experiments used to determine the frictional nature of the ventral portion of the animal and some experimentation to obtain EMG measures of the normal force exerted by snakes during the lateral undulation form of locomotion. Many of the methods that Hirose employed were quite progressive given the extent and nature of previous studies up to that point. Certainly, he provided a fine contribution to the field of biomechanics as it concerns

snakes.

As if this were not enough, Hirose took the validation of the information obtained concerning the locomotion of snakes to a new level when he actually constructed a snake-like robot that performed the lateral undulation form of snake locomotion. This was the first time that a robot was to use a method of propulsion other than standard drive wheel mechanisms. This was revolutionary in the realm of robotics and would anew interest in the topic of snake locomotion and how it may relate to the advancement of technologies.

Since that time, a plethora of research efforts from the robotics and general engineering communities have been accrued and we now provide a survey of some of those that we have come to be aware of. The discussion of these efforts is broken down into two categories: those projects that are associated with affiliates of the California Institute of Technology (Caltech) and those that are not. This may seem like an odd dichotomy, but during the last several years the efforts produced by affiliations with Caltech are distinctive in the way that the subject matter of robotics, and dynamical systems in general, is handled.

We begin with those works that are independent of the Caltech trademarks. These studies tend to be of one of two types, those that deal with both the actual construction of a robot that uses a snake-like mode of locomotion and a control method and those that deal solely with topic of how to mathematically formulate a control method to be utilized by these structures. With regards to the former, our interest lies in the mechanism through which locomotion was achieved (i.e., the control method) and not so much with the technical details of the actual robot's architecture. We have already mentioned that the first snake-like automated locomotion device, to our knowledge, was constructed by S. Hirose in the early part of the 1970's. This robot's name was the ACM III or Active Cord Mechanism Mark III. Hirose has constructed several snake-like robots since that time including descendants of the ACM III, namely the ACM IV, ACM V, and the ACM VI; the Koryu I and Koryo II robots; and the ACM

R1 [33, 20, 34, 18]. Each of these robots was a serial-link structure which made the use of wheel pairs to create a no side-slip condition for each of the links. In the case of all of these robotic realizations the so-called serpenoid curve and variants thereof were used for locomotion control. Being more specific, the differential angles between adjacent links were forced to follow sinusoidal trajectories that were slightly out of phase with one another. The effect is to bring about a wave propagation along the length of the structures body resulting in oriented displacement. For turning, a bias was given to the robots differential angles resulting in an undulatory locomotion along a circular arc whose radius of curvature is proportional to the bias selected. No true mathematical method for selection of the sinusoid parameters and bias was employed. Of the articles cited, [18] provides the most detail in this regard.

In 1996 a team of German researchers from the German National Research Center for Information Technology (GMD) introduced a prototype for a snake-like robot called the GMD-Snake [70]. This robot was different from those that had previously been constructed in that the team, Rainer Worst and Ralf Linnemann, had conceptualized a design that did not include wheels, legs, or any other appendages. As in the case of Hirose's robots, this robot was comprised of series of individual units. However, these units were flexible, being made up of several octagonal pieces of aluminum connected by rubber pieces. The segments were bent by a combination of drives and strings. These men were working toward something that seemed even closer to nature than the robots that had been previously developed. To elicit movement from this structure, the team used trial and error to determine a set of basis motions that were to be used in combination to construct motions satisfying more complex requirements. In other words, they used a table look-up method for actuation patterns. Indeed, it would appear that the researchers were able to demonstrate the capability of a snake-like creeping motion, though the basis patterns used to accomplish this task were not provided in their report. One of the prototype flaws, due to its flexibility, was that it experienced unexpected torsional effects, etc. The team indicated that they would

next envelope the structure in an elastic skin to provide a greater stability to the body. Unfortunately, it does not appear, to our knowledge, that this was followed up.

The GMD-Snake2 was introduced in 1999 [40]. Apparently, the composition of the research team had been altered, to Bernhard Klassen and Karl L. Paap, and so had the design of the GMD-snake, completely. This device was a serial-link structure that was actively driven by wheels and used the clothoid curve as a basis for motion control. This curve has been used for years in the design of highways and is also related strongly to the serpenoid curve proposed by Hirose. Being it so that the entire method of propulsion had been altered, this structure ceased to be snake-like. In fact we only mention this robot at all because a portion of the studies made into how to design the motion are quite intriguing with regards to some of the findings of our current research. In order to determine what the motion of the GMD-Snake2 should be, the group at GMD devised an optimization scheme [43, 60]. They proposed that the structure should follow the path such that the energy dissipation due to the friction experienced would be minimized via a method known as the "Davidson-Fletcher-Powell Method for Minimization". They attributed friction to two causes. The first source being the friction experienced between joints and the second being due to the differential between the direction of motion of the wheels and that of the segment to which the wheels were connected. It was found that the optimal motion would be one such that differential angles of the structure took on constant and equal measures. That is, they determined that the structure should always follow a circular path to move from location to location. For technical purposes, the team at GMD used the clothoid curve.

Only recently, a team of researchers comprised of M. Saito of Hatachi, M. Fukaya of the Tokyo Intitute of Technology, and T. Iwasaki of the University of Virginia, have constructed a snake-like serial link structure [62]. This structure differs from those of Hirose in that rails are used in place of wheels to attain a virtual no side-slip condition on the links. The creators of this robot claim that this alteration leads to higher

terrain adaptability. The control methodology used by the team to steer this robot is worth making note of. Like Hirose, the robot uses phase shifted sinusoidal changes in the differential angles between adjacent links to elicit the lateral-undulation form of snake locomotion. However, unlike Hirose and others, these researchers devised a feedback scheme to update the frequency and bias of these sinusoidal forms in order to alter the robots speed and direction, respectively, that makes use of the system's dynamics. The idea was to make use of feedforward simulation using the sinusoidal shape functions to heuristically determine an optimal value of the phase shift for a given number of links and amplitude for a given ratio of "friction" coefficients. Then PID and H_∞ controllers were used to design feedback control of the frequency and bias of the differential angles. The simulations presented were convincing. This was quite a good effort in beginning to developing a feedback control for the lateral undulation form of locomotion.

An interesting departure away from the robots and locomotion types discussed thus far is the MS-1 and MS-2 robots of Yansong Shan and Yoram Koren of Michigan State University. These researchers developed a serial-link structure whose links are supported by costers and carry a solenoid (MS-1). In the later design (MS-2), the costers were removed. The robots use the concertina type of locomotion. Essentially, a certain portion of the structure is held fixed to the ground via the solenoids while the actuation of the joints is used to reposition the free portion of the robot. A pattern of forward motion was devised using this mechanism. A motion planning method was then implemented to adapt this motion to the task of direction control.

There are several more robots that are typically included in the general category of "snake-like" robots, such as the NTUA Robotic Snake [63, 49]. However, for the most part, these robots are simply articulated and do not actually use a propulsive mechanism observed in snakes. For this reason, they will not be discussed.

We next move on to discuss a few contributions to the literature that are purely focused on the control of actual snake-like serial-link structures. One such effort is pro-

vided by K. Dowling in [17]. His idea was to use a table look-up method, as suggested by the GMD group. However, instead of using trial and error to determine locomotive patterns, he made the suggestion of using Fourier series coefficients as parameters for the functional form of the structure's body, and ideas of learning for selection of those parameters. He suggested the use of specific resistance as a measure of gait success. This non-dimensional quantity is given by $sr = Power/(Weight \times Velocity)$. So, for a fixed weight structure, it is the ratio of Power to Velocity. Then, parameters are sought to minimize this quantity. Little detail is provided concerning specific expressions for this quantity and the techniques used for optimization. However, it is purported that by using this method that 3 of the 4 typical modes of locomotion used by snakes appeared. Again, little tangible information concerning specifics of the resulting Fourier expansions were provided.

In 2000 a group of colleagues from the Tokyo Institute of Technology introduced a control methodology for serial-link structure with a no side-slip condition based on the concept of dynamic manipulability [15, 16]. This concept had first been introduced in the context of manipulator arms, however, it was easily extended to the snake-like articulated structure. The basic idea is to use either constraint equations or some portion of the system's dynamics to solve for position acceleration in terms of applied torques at the structure's joints. The resulting coefficient matrix of the torques is called the manipulability matrix, as it indicates the extent to which the mechanism's acceleration can be affected by application of torque. With a certain amount of algebraic manipulation the image of this mapping can be viewed as an ellipsoid. The lengths of the axes, or certain ratios thereof, are considered to be a measure of manipulability. The control technique then proposed, is one that attempts to strike a compromise between motion in a desired direction and maintaining a suitably high measure of manipulability. i.e., the structure should move toward the target but at the same time should do so in a manner that facilitates its capability to continue to make progress. It would appear that when doing this, the shapes determined often

had a zigzag geometry. To eliminate this, the team added the condition of also trying to minimize the side constraint forces. Upon doing so, a smoothing or regularization effect was obtained and a wave-like geometry developed which produced a locomotion that is strikingly similar to that obtained via sinusoidal differential angles. This is most impressive. To our knowledge, this is the only effort other than our own where a mathematical control construct was used in conjunction with system dynamics to elicit a snake-like motion from the structure that satisfied some sort of optimality criterion *and* made no reference to predetermined forms.

We now turn our focus to the review of those works concerning snake locomotion generated by associates of Caltech. To begin, we consider the work of G. Chirikjian and J. Burdick. Burdick and Chirickjian are well known in several research communities for their persistent contributions to the study of so-called hyper-redundant or high degree-of-freedom robots. Hyper-redundant is a term that they coined to describe robots with structural morphologies similar to that of snakes, elephant trunks, etc. An example of such robots are manipulator arms with high degrees of freedom (HDF) and serial-link snake-like robots. In the respective works [13, 14] of 1991 and 1992, concerning HDF manipulators, the team took the perspective that HDF systems approximate continuum forms, which they would call backbone curves. Using this perspective they were able to describe the geometry of articulated structure via generalized curvature functions written in terms of time and arc-length varying tangent length and angle functions. Modal approximations to these length and angle functions were then used along with the end effector coordinates to determine algorithms for inverse kinematic solutions in the presence of obstacle constraints. In other words, a backbone curve was determined from the constraints and the desired location of the end effector. The backbone curve was then used to determine the appropriate time varying geometry of the articulated system.

Using this same perspective, Burdick and Chirickjian were able to take up the topic of locomotion for such structures with only a few modifications [12, 10]. In this

case, the machinery of the moving frame was added and the modal approximations were selected so as to create stationary-wave, varying-amplitude and travelling-wave, constant-amplitude body forms. Upon the application of some physical constraints and certain no-slip conditions, the problem of locomotion via these modal forms was made a strictly kinematic problem that could be solved for the modal approximation parameters in the same manner as the inverse kinematics problem for the fixed-end manipulator. Although this technique is ill-suited for the analysis and control of the lateral undulation and concertina snake locomotion types, it is well-suited for the production of rectilinear and side-winding forms, which were both illustrated in these efforts.

It would seem that once the problem of locomotion was taken up at Caltech, many doors were opened. This institution is one of the few that actively approach the subject of mechanics from a differential-geometric perspective. The foundations of this approach to the subject were laid by the masters themselves: Lagrange, Poincaré, Poisson, Lie, Laplace, Newton, Euler, and many others. Oddly, it would seem that the general treatment of the subject of dynamics from this point-of-view has never been wholly adopted by the scientific community. It is the opinion of the author that this is not too difficult to understand. This perspective is almost always made to be unduly technical and as such, demands a mathematical fluency and knowledge base possessed by, relatively speaking, few. However, in recent times it has seen a resurgence in development, understanding, and use in the explanation of physical phenomena [44]. This is particularly the case in the Caltech community. There, the locomotion problem, including that of snakes, has been placed under the modern analytic mechanics umbrella. This is the distinguishing feature of the studies coming out of that community of researchers.

Inspired by the work of Joel Burdick and Gregory Chirikjian, P.S. Krishnaprasad and D.P. Tsakiris would take the problem of undulatory locomotion of HDF systems to the geometric context in their 1994 paper [41]. In particular, they modelled the

no side-slip constraints of serial-link structure units (i.e, early Hirose type snake-like robots), which are members of the general class of nonholonomic constraints, as a connection on a principle fiber bundle. They then proceeded to show that shape variations (i.e., differential angle variations) would elicit geometric phase (i.e., displacement) in the coordinates of the structures imposed body frame through the connection form (i.e., by using the huge differential between normal and tangent reaction forces from the substratum). One of the important features of this work was the decomposition of the mathematical parameters of the serial-link structure into the translation variables and the shape variables. The translation or body coordinate frame variables would be called the group variables, for their possible values form a group G. The researchers attempted to coin the term $G - snakes$ to refer to such structures.

In 1994, S. Kelly and R. Murray of Caltech released a publication on locomotion which also casted the topic in a geometric context [39]. As did Krishnaprasad and Tsakiris, they modelled kinematic constraints as connections on principle fiber bundles and used the concept of geometric phase to dynamically explain the capability of structures to attain gait. Also similar was the explicit decomposition of the structures phase space into translation or group variables and shape variables. However, this work was different in several respects. To begin, they opened up the problem to all forms of locomotion under nonholonomic constraints which result in a principle kinematic connection and were propulsion is attained through this connection by cyclic changes in the shape variables. Thus, the sidewinding and inchworm gaits of Burdick and Chirikjian fall under this framework as specific cases. Furthermore, they showed how the modeling of constraints on a principle fiber bundle can be used to more easily assess the system's controllability via geometric control methods. Finally, this work is commendable for its attention to application via several complete examples.

Shortly following this work was that of Jim Ostrowski under the advisement of Joel Burdick [57], 1996 (also see [58, 59]). Ostrowski expanded the work of Kelly

and Murray in several ways. To begin, he generalized the work to encompass the modeling of constraints through connections that were not necessarily of the principle kinematic type. This would allow the inclusion of systems whose dynamic effects are crucial to the resulting motion of the system, such as later Hirose type snake-like robots whose constraints do not completely determine the motion of the robot. In fact, he uses his results to explain via the mathematical construct of the connection why the out of phase sinusoidal differential angle trajectores elicit a displacement of the robot. Secondly, he demonstrated how system symmetries could be utilized to perform system reduction in a straight forward manner. As a byproduct of this reduction, which we have found to be the most useful in our present effort, is a partial decoupling that clearly illustrates the relations between a system's control and shape and the system's shape and momentum. Further, he demonstrated that this form of the dynamics could always be attained for locomotion systems, both constrained and unconstrained. Ostrowski also extended the controllability results of Kelly and Murray to this larger class of systems with mixed constraints and illustrated how to use the components of the reduced equations to furnish a quick check of the controllability conditions. Perhaps the nicest feature of Ostrowski's contribution however, is its readability. His presentation of the material risks demystifying the geometric perspective on the locomotion problem to the point that most scientists can understand it with some effort.

Finally, we conclude with review of the very recent work of Ostrowski and his student K. McIsaac [47, 48, 46]. They considered the dynamics of a serial-link structure and the propulsive mechanism of the lateral undulation form of gait. Their studies are quite good. Using the reduction theory and in particular the Euler-Poincaré equations, they developed a model of the structure that exhibits the before mentioned partial decoupling by using Ostrowski's calculation procedures. To model the interaction of the system with the environment, they did not place no side-slip constraints on the links of the structure. Instead, they considered the case of the snake in a fluid and

used a viscous friction model taken from the work of Ekeberg to create a normal to tangent pressure force differential along the centerline axis of the structure. Using the resulting dynamics they demonstrated the propagation that the system can achieve by using a sinusoidal pattern for the structure's differential angles or shape. To aid in their understanding of the effects of the input parameters on the resulting generalized momenta, they assumed a small amplitude perturbation ϵ in the sinusoidal forms and proceeded to express the momenta as an expansion about ϵ. By comparing certain terms, information about the system's response was extracted. In particular, as determined in previous work, they found that an offset or bias in the shape form may be used to steer the system. Using this feature they developed a hierarchical approach to motion planning for the system and showed that this scheme could be used successfully. This differs from the method originally employed by Hirose only in that there is a feedback mechanism that can be employed to make the control autonomous. This work along with that of [62] are the only two efforts known to us that use a feedback mechanism in conjunction with sinusoidal shape variable forms to direct the lateral undulation type of locomotion.

1.3 Statement of Purpose

We now attempt to indicate the purpose of our research, in what way we pursue that purpose, and the success of the directions that we have taken. This project began with a very simple query, "Is a snake-like robot feasible?", and the purpose of our research has been to provide a satisfactory answer. Paraphrasing this question we have, "Can a robot navigate throughout its environment the way that a snake can?" It requires little thought to come to the conclusion that this problem is actually two! The first is, "How does a snake move about its surroundings?" or "What is its locomotory mechanism?". Then there is the secondary question, "Is it possible to build a machine that could utilize that same mechanism?" Both are very interesting

questions. We make progress in answering the first of these, confining our attention to the lateral undulation type of motion.

Our efforts are exerted in two differing, though related, directions. As one of these, we expend a great deal of effort to convey the fact that the general question concerning the mechanism of gait has already been answered but that only a remnant of this answer remains in the conscience of those who are pursuing snake-like locomotion simulators. John Gray completely solved this problem in 1946 [28]. It would seem that there has been a good deal of confusion in this matter. Perhaps the confusion began when S. Hirose claimed that his was the first work that connected the gliding form of the snake with the animal's capability to attain locomotion via lateral undulations. This was not the case, and in fact Hirose proposed the same gliding form that Gray had constructed only as an example of a form that a snake could use to elicit motion. However, Hirose's claim has apparently stuck, and now it is common place to read in publications on the subject that Hirose showed that snakes follow a winding curve, which he named, the serpenoid curve. By assuming this as the beginning and end of snake locomotion, most people miss out on the fact that this form is but one that is acceptable to attain the lateral-undulation type of locomotion and that the underlying principles that Gray used to originally construct this form possess a generality that is far-reaching in utility and potential. It was for this reason that Gray only presented the now so-called clothoid curve or its smoothing, the serpenoid curve, as an example. However, by not making the more singular claim that he had discovered "the" gliding form of snakes, Gray's efforts have been partially obscured. Thus, one of our goals is to revive his conclusions and to place them in their rightful place as "the" explanation of mechanism of locomotion in snakes. In particular, the explanation of the lateral undulation form of locomotion.

We approach this task in the following manner. We begin by developing with sufficient rigor a mathematical model of a serial-link structure. We then return to Gray's conclusions concerning lateral undulation and following his development of the basic

and most general criterion for educing motion via lateral undulations. Based on these criterion we reconstruct Gray's example of a possible gliding form for snakes. Following this, we show how the shape Gray proposed can be expressed mathematically. Further, we describe the shape which leads to the so-called composite clothoid and serpenoid curve gliding forms. In doing so we show that the composite clothoid shape is the shape proposed by Gray. Further, we show via a Fourier series approximation that the serpenoid shape is the proper smoothing of Gray's proposed shape. Next we illustrate through simulation that indeed the shape proposed by Gray evokes gait and after doing so, illustrate that the reason for this is that the shape was constructed to be in accord with Gray's general criterion. We end this exercise by discussing the generality of the criterion and then illustrating that other shapes adhering to them will also educe gait.

Several contributions to the literature are made by this undertaking:

Contribution 1. To begin, our modeling approach is somewhat unique. We apply the geometric mechanics approach to the derivation of the system dynamics. Although this technique is not new in general, it is not yet widely in use and its application in the context of the serial-link structure locomotion problem in the absence of no side-slip constraints is new. Unfortunately for us, after having made this effort, it was discovered that Ostrowski and McIsaac had reported a similar undertaking [47] as part of an IEEE proceedings. This work had initially eluded our literature survey due to the fact that the title makes reference to an underwater eel robot instead of a terrestrial snake robot. All the same, we believe our work to be a contribution as an independent work.

Contribution 2. Our treatment of friction modeling as it is applied to the planar rigid body is also new. In the literature the topic of friction is only discussed in rare instances. Typically the subject is avoided all

together by assuming frictionless environments. We take an elemental approach. By doing so, we avoid the pitfall of applying friction only to a body's center of mass. Additionally, using this approach, we develop and apply a method for describing dry friction forces and directionally dependent dry forces. This allows illustration of the effects of anisotropic snake-skin for the first time. Namely, we properly show that the anisotropy of a snake's skin is not responsible for the locomotive capability of lateral undulations.

Contribution 3. By reviving Gray's research it is believed that we make a profound contribution. Firstly, we properly show that J. Gray was the first to introduce the wave form that now receives such high acclaim as "the" gliding form of snakes. Further, we explain how this form was developed from more generic criterion that contain the freedom in application that has arisen the interest of the scientific community in the animal. For those who truly want to realize the complete functionality of a snake-like locomotor, the understanding of the generality of these criterion and the separation from the notion that there is "a" gliding form for snakes will be essential. We hope to begin this realization.

The second direction of our efforts is toward the determination of optimal gait eliciting shapes. It is interesting that S. Hirose claimed that he was able to show through experimentation that the shape actually taken by real snakes while performing lateral undulation over an essentially uniform surface is very well approximated by the example shape proposed by J. Gray. Thus the question, why is this so? Is this shape truly what is most natural for the animal when placed in the ideal conditions of a uniform planar environment? Of course, there are all sorts of notions of what is natural. But, following the principle of least action, it makes sense that what "should" be natural is a motion that produces the displacement desired by the animal with minimal exertion of effort or loss of energy.

Thus the second of our goals is to establish what the natural or optimal gait generation shape is. We do so by developing a methodology for determining optimal control/shapes. Using this method we determine the natural locomotive tendencies of the serial link structure that we use as a representation of a snake.

By the completion of this task we make these additional contributions to the literature:

Contribution 1. The optimization method introduced is quite unique in several regards and can be applied to provide optimal periodic locomotion shapes for a wide class of systems, including the class of systems with symmetries enveloped by Ostrowski's work.

Contribution 2. We are the first researchers to successfully obtain optimal gait inducing shapes for the planar snake-like serial-link structure in the absence of constraints (or with them for that matter). The desire for such a result has been widely expressed in the current literature.

Contribution 3. Particular features of the optimal snake locomotion patterns illustrate the pertinence of our efforts to emphasize a return to the basic principles set forth by J. Gray. In this regard, we contribute to the merit of the afore mentioned contributions.

1.4 Thesis Layout

The presentation of our efforts follow the ensuing program. In Chapter 2 descriptions of two systems that will be discussed throughout the remainder of the work, and some discussion of these systems with regards to the physical scenarios that they represent, are presented. In Chapter 3 we discuss the use of Lagrangian mechanics for developing the dynamical models of the systems described in Chapter 2 and the use of geometric mechanics in the reduction of these dynamics. Chapter 4 is a continuation of sorts

of Chapter 3, addressing the description of the forcing terms that appear in the system dynamics and finalizing the form of the dynamics that will be used, more or less, for the remainder of the work. These chapters essentially comprise the requisite preparatory work that will be used to study undulatory locomotion in snakes.

In Chapter 5 we take up the resuscitation of the work of J. Gray on undulatory locomotion using the mathematics developed in the foregoing portion of this writing. Finally, the thesis ends with the optimal control formulation and its results being presented in Chapter 6.

Should one be perusing this work and find this outline too vague, it is suggested that the introductions to the individual chapters be consulted. A good deal of effort has been made to provide detailed chapter outlines as part of these prefaces. It was felt that providing the more detailed delineation at those junctures rather than here served the greater function of structuring and directing the focus of the chapter contents and the document as a whole.

Chapter 2

The System Descriptions

2.1 Introduction

In this chapter we shall describe the two systems that will serve as the objects on which we shall experiment mathematically. The first of these objects, a mass-spring system, will primarily be of interest only as an example. This system's description will be simplistic enough to allow intuition on the qualitative level and ease of complete calculations on the quantitative side of things. The second object will be a serial-link structure used to model the locomotive morphology of snakes. This system is a bit more complex and as a result can elude intuition and provide computational challenges. Our descriptions are comprised of the following components. In each case we provide qualitative descriptions of the systems, identify generalized coordinate sets and related coordinate transformations, provide some basic kinematic expressions, and include some qualitative description of the forces that are to act on the systems. By doing so, we set up the potential to describe the systems' dynamics.

2.2 The Mass-Spring System

We consider a system made up of a large mass, a smaller mass, and a spring. The masses are viewed as blocks, a large one and a small one. Let the large block be denoted B_1 and the small block B_2. B_1 will lie on a line. B_2 is then attached to the top of the larger one via the spring. This is, qualitatively, the system. Refer to Figure 2.1 for an illustration of the mass-spring system.

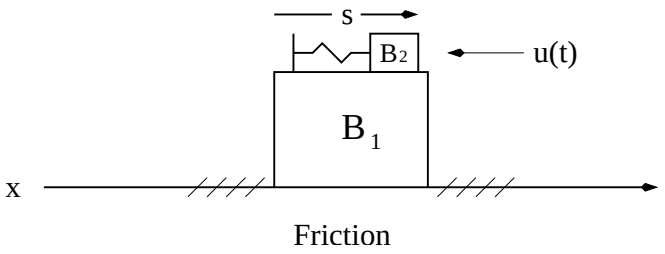


Figure 2.1: The simple mass-spring system.

An inertial reference frame is imposed on the line that supports B_1. The line is assigned the typical euclidian frame. We denote the position along this axis by x. Let x_1 denote the position of B_1 and x_2 the position of B_2. These two coordinates completely specify the configuration of the system with respect to the imposed frame and thus comprise a generalized coordinate set or vector. As an alternative to this, we may select x_1 as before, renaming it g, and select as a second coordinate the position of B_2 relative to B_1, say s. Then g and s completely specify the state of the system and as such, also comprise a set or vector of generalized coordinates. Being it the case that there are no geometric constraints for this system (technically there are, but none that we wish to formally impose), its degrees of freedom are $df = 2 - 0 = 2$. Thus, both generalized coordinate sets are independent ones. However, the second of these is of the type that is desired most for future purposes. For g indicates the position of the system as a unit and s its internal configuration or shape at that particular position.

We shall denote the coordinate vector $(x_1, x_2)^t$ by z and the coordinate vector $(g, s)^t$ by q. The vector z may be expressed in terms q via the the coordinate transformation

$$z(q) = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_1 + (x_2 - x_1) \end{bmatrix} = \begin{bmatrix} g \\ g + s \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} q \ . \tag{2.1}$$

From this relationship, the relation between the kinematics of the system in z coordinates are easily related to the kinematics of the system in q coordinates and visa versa. Although there is no utility in doing so, we state the kinematic relations between the coordinates sets. The coordinate velocities are related by the equation

$$\dot{z} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \dot{q} \ ,$$

and the coordinate accelerations by the equation

$$\ddot{z} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \ddot{q} \ .$$

The final component in the description of this system is a statement of the applied forces that are assumed to act on it. Firstly, we envision friction acting on B_1 as a result of its tangential motion over the surface of the line. Secondly, it is assumed that B_2 experiences some friction or damping, such as air resistance or a dashpot effect. Finally, we allow ourselves a way to contribute or act on the system. That is, we introduce a control. It shall be assumed that a translational force may be placed in the direction of the variable s. This force is denoted by u.

Before leaving this system, a differing qualitative description is expressed, along with the purpose of introducing it. We see the mass-spring system as a crude model of the following physical situation. Imagine a person on top of a crate, facing a particular direction. The person is B_2, the crate is B_1, and the direction is the positive x-axis.

Further, imagine that there is no object that may be used for leverage and thus the only way to apply a force to the system is to rock it back and forth by the shifting of body weight. This shifting is the control u and of course there is some muscle action, and this is represented by the spring. When the person on the crate begins to rock back and forth, the crate will have a tendency to displace along the direction that the individual is facing. This tendency is opposed by a reaction force due to the surface interaction between the crate and the substratum on which it rests. This reaction force is called friction whence its inclusion in the modeling as an applied force. Why would the person rock the crate? To try to get it to slide along the ground...to attain locomotion! Now come the questions. Is it possible for this individual to make any progress? If so, why? In order to try to answer these questions, we will use the system description provided to formulate the systems dynamics and then ask the questions at a later juncture in the language of mathematics.

2.3 The Snake-Like Serial-Link Structure

Consider a series of absolutely rigid rods joined together in a fashion similar to that of a snake's vertebrae structure. This system is placed in a plane onto which an inertial reference frame is imposed. This frame is selected to be the typical Euclidian type and we denote the position along the horizontal and vertical axes by x and y, respectively. In this plane, a rod is considered to be a line segment which has a linear mass density function ρ and half-length l attributed to it. To a given rod another is attached at one of its endpoints via a revolute joint and in this fashion one may continue connecting rods one to another at free endpoints (previously unconnected) until n rods are joined one with another. That is, we construct a continuous piecewise linear function in the plane. Each piece of this function is to be considered a rod or vertebra and these rods may rotate relative to one another. Each of the n vertebrae are labelled in the following manner. Choose one of the two vertebra that have a free

end and call it B_1. This segment will be known as the head. Label the rod adjacent to B_1, B_2 and continue in the logical fashion to label each of the segments in the system. The final segment in the labelling will be B_n and is to be considered the tail. So we have a sequence of n rigid bodies $\{B_i\}_{i=1}^n$ which comprises the system, denoted $\mathcal{S}$. Denote the half-length of body B_i, l_i, and the mass density function for that body, ρ_i. For us, the length and mass density of the rods will be uniform along the system. However, for the sake of generality, the model is developed with the potential to allow rods of differing mass and length.

For each rigid body B_i, three coordinates completely describe the state thereof: Two translational coordinates, denoted x_i and y_i, and one orientation coordinate, denoted θ_i. The translational coordinates can be those of any particle of B_i. We choose to select the coordinates of the particle located at the body's center of mass, which we denote in vector form by $R_i = (x_i, y_i)^t$. Thus we may completely describe the state of the system $\mathcal{S}$ by specifying the coordinate vector $z \doteq (R_1^t, \theta_1, \ldots, R_n^t, \theta_n)^t$. See Figure 2.2. Due to this fact, z is a generalized coordinate vector.

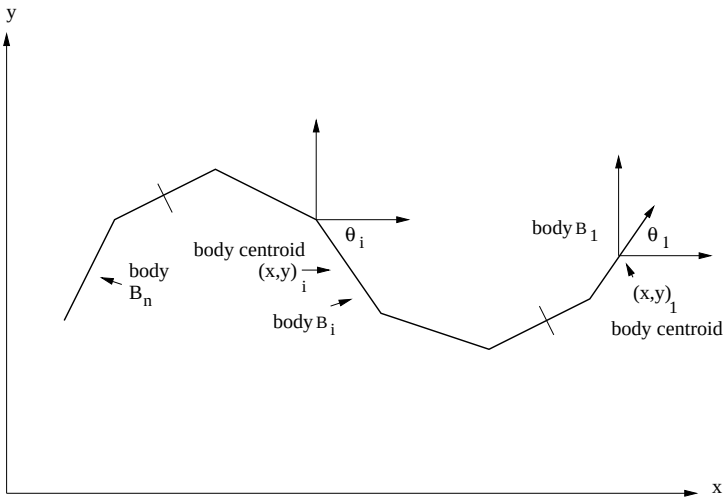


Figure 2.2: The snake-like serial-link structure described in terms of the generalized coordinate vector z_d.

It is noted that the positions of the bodies comprising this system are not independent. In fact, since the connected endpoints of adjacent bodies must translate in identical fashion, it is clear that at each such joint there are two constraints, one corresponding to horizontal translation in the plane of motion and another to vertical translation in the plane of motion. Since there are 3 needed parameters per body and 2 constraints per joint, the system degrees of freedom are $df = 3n - 2(n-1) = n+2$. i.e., there are precisely $n+2$ parameters in an independent generalized coordinate set. In this case, it is clear by inspection that $\tilde{z} = (R_1^t, \theta_1, \ldots, \theta_n)^t$ is an independent generalized coordinate vector.

Expressions for the body centroids; $R_2, \ldots, R_n$; in terms of the independent coordinates are now provided. These expressions will be of later use. To begin, we recognize θ_i as the angle coordinate of a particle P of body B_i whose location in a frame located at R_i is given in polar coordinates by $(d(P), \theta_i)$. Of course, we can locate any other point along the rod by simply assigning the correct value of $d(P)$, $-l_i \leq d(P) \leq l_i$, as θ_i is fixed over the range of the body. Using the polar coordinates of a particle P of body B_i, the particle's coordinates in the global/inertial xy-frame may be attained by first determining the euclidian coordinates in the frame imposed at R_i from the polar coordinates $(d(P), \theta_i)$ and then adding the result to the coordinates R_i. This is a conversion from a body fixed coordinate system to a global one. The idea of the body coordinate system is a very important one and a more formal discussion of this will be given at a later juncture. We shall call the global coordinates of point P, $r_i^{d(P)}$. As described,

$$r_i^{d(P)} = R_i + d(P) \begin{bmatrix} \cos(\theta_i) \\ \sin(\theta_i) \end{bmatrix}. \tag{2.2}$$

The joint constraints may now be expressed mathematically via equation (2.2) as $r_i^{-l_i} = r_{i+1}^{l_{i+1}}$, $i = 1, \ldots, n-1$. From these relations we arrive in an inductive manner

at the expression

$$R_i = R_1 - l_1 \begin{bmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{bmatrix} - \sum_{j=2}^{i-1} 2l_j \begin{bmatrix} \cos(\theta_j) \\ \sin(\theta_j) \end{bmatrix} - l_i \begin{bmatrix} \cos(\theta_i) \\ \sin(\theta_i) \end{bmatrix}, \tag{2.3}$$

which is also clear from the geometry. This relation defines a coordinate transformation from the independent generalized coordinate vector $\tilde{z}$ to the dependent generalized coordinate vector z. The remaining kinematic relationships between the coordinate vectors $\tilde{z}$ and z are now stated. These relations are

$$\dot{R}_i(z) = \dot{R}_1 - l_1 \begin{bmatrix} -\sin(\theta_1) \\ \cos(\theta_1) \end{bmatrix} \dot{\theta}_1 - \sum_{j=2}^{i-1} 2l_j \begin{bmatrix} -\sin(\theta_j) \\ \cos(\theta_j) \end{bmatrix} \dot{\theta}_j - l_i \begin{bmatrix} -\sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix} \dot{\theta}_i \ , \tag{2.4}$$

and

$$\begin{aligned} \ddot{R}_i(z) = \ddot{R}_1 - l_1 \begin{bmatrix} -\sin(\theta_1) \\ \cos(\theta_1) \end{bmatrix} \ddot{\theta}_1 - \sum_{j=2}^{i-1} 2l_j \begin{bmatrix} -\sin(\theta_j) \\ \cos(\theta_j) \end{bmatrix} \ddot{\theta}_j - l_i \begin{bmatrix} -\sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix} \ddot{\theta}_i \\ + l_1 \begin{bmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{bmatrix} \dot{\theta}_1^2 + \sum_{j=2}^{i-1} 2l_j \begin{bmatrix} \cos(\theta_j) \\ \sin(\theta_j) \end{bmatrix} \dot{\theta}_j^2 + l_i \begin{bmatrix} \cos(\theta_i) \\ \sin(\theta_i) \end{bmatrix} \dot{\theta}_i^2 \ . \end{aligned} \tag{2.5}$$

Unlike the additional kinematic relations that were obtained for the coordinate relations of the mass-spring system, these relations, at least the first, will reappear at a later portion of this writing and will actually be a critical element in the proof of a primary theorem.

Although an independent generalized coordinate set has been identified for this system, it must be observed that this set is not of the type that will be desired as our development progresses. That is, the coordinate vector $\tilde{z}$ does not naturally decompose into coordinates describing the position of the system and those that describe its internal configuration. The selection is half satisfactory. For it is the case that x_1, y_1, and θ_1 can be used to identify the position of the system. So let

us make the assignment $g \doteq (x_1, y_1, \theta_1)^t$. However, the remainder of the coordinates, $\theta_2, \ldots, \theta_n$, suffer from the fact that they are determined by making reference to the global coordinate frame. Hence, the values of these coordinates are not internal to the system. However, it is geometrically clear that the values of the relative angles

$$\phi_i \doteq \theta_i - \theta_{i+1} \ , \tag{2.6}$$

$i = 1, \ldots, n-1$, depend only on the relative positioning of the links and are thus internal to $\mathcal{S}$. See Figure 2.3. Let us denote the vector of these relative coordinates $(\phi_1, \ldots, \phi_n)^t$ by s. Then $q \doteq (g^t, s^t)^t$ completely determines the state of $\mathcal{S}$ and naturally decomposes in the manner desired. Furthermore, the cardinality of q is $n+2$, which equals the system degrees of freedom. Therefore, q is an independent generalized coordinate vector. The definitions of equation (2.6) define the coordinate transformation $\tilde{z}(q)$ and consequently $z(q)$. We now state the kinematic relation involving the velocities $\dot{q}$ and $\dot{z}$ which comes about through the kinematic relation (2.4). The relation

$$\begin{aligned} \dot{R}_i(q) = \dot{R}_1 - l_1 \begin{bmatrix} -\sin(\theta_1) \\ \cos(\theta_1) \end{bmatrix} \dot{\theta}_1 \\ - \sum_{j=2}^{i-1} 2l_j \begin{bmatrix} -\sin(\theta_1 - \sum_{k=1}^{j-1} \phi_k) \\ \cos(\theta_1 - \sum_{k=1}^{j-1} \phi_k) \end{bmatrix} \left(\dot{\theta}_1 - \sum_{k=1}^{j-1} \dot{\phi}_k \right) \\ - l_i \begin{bmatrix} -\sin(\theta_1 - \sum_{k=1}^{i-1} \phi_k) \\ \cos(\theta_1 - \sum_{k=1}^{i-1} \phi_k) \end{bmatrix} \left(\dot{\theta}_1 - \sum_{k=1}^{i-1} \dot{\phi}_k \right) \end{aligned} \tag{2.7}$$

describes the translational velocity of the dependent bodies $B_2, \ldots, B_n$. The rotational velocity of these bodies is given by

$$\dot{\theta}_i(q) = \left(\dot{\theta}_1 - \sum_{k=1}^{i-1} \dot{\phi}_k \right) \ . \tag{2.8}$$

Lastly, we have the relations described by the equation

$$(\dot{R}_1^t, \dot{\theta}_1)^t = \dot{g} \ . \tag{2.9}$$

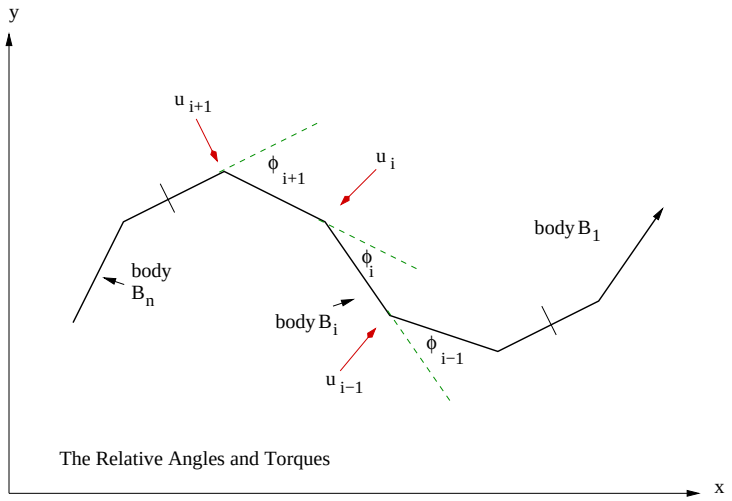


Figure 2.3: The snake-like serial-link structure described in terms of the independent generalized coordinate vector q. ϕ_i represents the relative angle between adjacent bodies B_i and B_{i+1} and u_i represents the torque applied to the joint connecting these bodies via the muscle-like actuator.

Still there is the matter of the applied forces that will act on $\mathcal{S}$. These forces are now described beginning with those forces that we shall assume to be controls. To each body B_i, a massless segment of length $2l_r$ is attached at length l_r along itself and perpendicular to the body at R_i. These segments are to be thought of as the ribs of a snake. We consider the endpoints of these massless segments to be attached one to another in a longitudinal manner via a massless muscle-like actuator. This actuator may contract or relax, and by doing so, apply a torque around the joint connecting the adjacent bodies to which the actuator is attached. In addition to the actuator forces, friction is also assumed to occur due to the tangential contact between $\mathcal{S}$ and

the surface on which it lies. It is assumed that the surface of contact is uniform.

The external forces that induce torque and friction forces lie in the plane of motion. As a result, all moments are perpendicular to the plane of motion. That is, the system does not experience torsion.

As indicated already, our intent is to liken this system's physical circumstances to that of a snake on a relatively homogeneous surface. Now we may ask the locomotion question and those questions related to it. Can the serial-link structure attain significant locomotive capability via its interaction with the substratum by use of its actuators? If so, what should the muscle contractions or control forces be? Why does the motion happen? What is the most efficient way to move from one position to another? Is this motion like a snake's?

A disclaimer of sorts is in order at this point. This caveat concerns the extent to which this system's locomotive capability should be comparable to that of a snake's. We only expect to get out what we put in, which may or may not be much. The following notes on basic snake morphology are taken from the well-known bio-mechanics text [22].

Feature 1. Limbless life forms such as snakes tend to be elongated. It is postulated that such elongation makes two of the primary motions utilized by snakes more readily attainable and more effective.

Feature 2. In animals such as snakes one finds a high vertebra count in the range of 100 to 300.

Feature 3. There is a pair of ribs to each vertebra present.

Feature 4. In a snake there are at least 20 discrete muscles on each side of each vertebra. These muscles are used to connect vertebra to vertebra, vertebra to rib, vertebra to skin, rib to rib, and rib to skin. Muscles are also found that attach to longitudinal tendons whose vertebrae span may be quite large.

Commenting on this list it is first noted that our system description is flexible enough to allow for elongation and a high vertebra count. However, the modeling and analysis of such a system, although not impossible, is complicated by this feature. It is easier to handle $\mathcal{S}$ with a lower body count. This could make it seem as though we should concede failure from the beginning with respect to comparing our system to a snake. However, note that the claim is that the elongation makes certain snake motions efficient. This is not to say that it is a primary factor in the underlining mechanisms of locomotion.

Also, we have included in $\mathcal{S}$ the existence of a pair of ribs per vertebra. However, we have only included in our description 2 muscle-like actuators per joint which extend from rib tip to rib tip. Thus, from the outset, we have excluded the effects of 19 or more other elements that may affect a snake's potentials for movement. It is purported in [28] that

> "*From a mechanical point of view, the axial skeleton of a snake can be regarded as a series of rigid rods hinged together to form a chain, whilst the axial musculature can be regarded as a series of elastic elements operating, laterally to the hinges, between adjacent rods.*"

No in-depth explanation of this statement is given. However, it would seem that this description has been used to embody the snake system, as it concerns its locomotive capacity, in every study up to the present. We shall do the same. However, upon viewing musculature diagrams such as those found in [52] and [21], it is fairly clear that the positioning of many of the muscles connecting rib to rib are not parallel to the ground and thus more than likely have the potential to collectively produce significant spacial moments. How much such moments may affect locomotion is still unclear in the opinion of the author. However, being it the case that we are not experimental biologists, we will for the course of this writing, follow this description, as the discussion thus far should indicate.

In the case of the mass-spring description provided in section 2.2 we were able to

give some qualitative description of the control actuation that should take place for a person to scoot a crate in a particular direction. It makes sense to simply state that a periodic shifting of the individual's body weight is the natural thing to do. It is natural because every person has performed this experiment at some point, albeit informally, as a child (or perhaps adult) while playing. However, not too many people have slithered across a floor like a snake and thus there is not a whole lot of reliable intuition that can be utilized in a qualitative description of motion. Thus it makes sense that we must rely on observations of snake motions as a source for what "types" of motion can be expected to bring about meaningful displacement of the serial-link structure. Three primary forms of motion have been observed amongst most snake genera. These forms are *lateral undulation*, *concertina motion*, and *rectilinear motion*. The first of these motion types is the most common and is the only one observed when a snake moves at a steady velocity in a specified direction. The other two forms rely on manipulation of static frictional forces and they operate in a sort of discrete fashion and tend to be observed only in very particular scenarios [22]. We now describe all three of these motions qualitatively, closely following the descriptions found in [22].

The essence of lateral undulation is as follows. A series of alternating contractions occur throughout the trunk of the snake inducing alternating "planes" along a wave-form facing left and right, being it assumed that we view the snake along its longitudinal axis. These planes apply force against the surface (water, sand, etc.), which provides a returning force that is used to obtain leverage.

The mechanism and description of the concertina type motion is in fact simple. Consider a snake whose body is split into two states, one that is in motion and one that is not. Along the portion of the body that is not in motion, the weight of the snake applies a ventral force against the substratum. If the surface has a non-negligible friction coefficient, this weight force will allow a horizontal force of a bounded magnitude to be applied to stationary points of contact without slippage. This horizontal force merely needs to be less than the maximum static friction force.

In this case, no net force and thus no acceleration can take place. Thus, these points act as anchors to be used for leverage by the remainder of the animal's trunk, which is in motion. The snake may then extend the remainder or a portion of its trunk (here is where the horizontal component of an extension or muscle force is applied in the direction of motion) into a new position. In this new position the anchor mechanism is induced by the portion of the body previously in motion and then the portion following behind, previously at rest, may slide forward and the process is repeated. Of course, there is a continuous transition along the body of the animal from the state of rest to that of motion. The ventral portion of the animal acts as the legs and the upper portion moves steadily along like the human torso.

The third type of motion to be described is the rectilinear motion. This motion is actually very similar to human running. While running or jogging, the human torso moves at a relatively constant velocity. However, the actuating members do not move in such a continuous fashion. One leg is removed from the substratum and is propelled forward with sharp acceleration so as to overtake, with respect to displacement, the torso. The other member is planted on the ground and anchors via static friction to obtain leverage for horizontal or forward propulsion. The anchoring is aided with muscle work. Near the end of the forward thrust of the free leg, it decelerates and instantaneously comes to a stop. At or near that moment, a horizontal force is applied to the planted member so as to invoke slippage so that both legs are temporarily removed from the substrate. The anchor leg is accelerated and then decelerated into the kick position as the other member is positioned to then act as the anchor. This is iterated to obtain forward propulsion. A snake can mimic this sort of motion along the length of its body in order to creep forward in a straight-line fashion.

We are not really interested rectilinear motion, where alternating points of contact are used to attain locomotion. This phenomena has been extensively studied in the case of bipeds and quadrupeds. We are interested in the case of motion in the presence of continuous contact along the length of the body, as in the case of the mass-spring

system. Secondly, we do not intend to give attention to the concertina motion, where static friction is utilized. Our interest is in the kinetic, continuous motion case. This is the type most commonly observed amongst all limbless life forms. Thus, undulatory locomotion serves as our intuition concerning the nature of the controls to be used to attain locomotion from the serial-link structure.

Chapter 3

Modeling and System Reduction

3.1 Introduction

In this chapter a framework is established from which mathematical models of system dynamics can be extracted. This framework, and thus this chapter, is composed of two primary parts. The first of these components deals with the topic of Lagrangian dynamics. The second of the components is an exposition on system reduction via symmetries. This development loosely follows the work of Ostrowski [58].

Lagrangian dynamics makes use of a well-known variational principle, known as as Hamilton's principle, to extract system dynamics from a function known as the system's Lagrangian. This Lagrangian is defined to be the difference between a system's kinetic and potential energies. Thus, the following program is followed with regards to this component of the chapter. We begin by describing, in a very concise fashion, the idea of the body coordinate frame for a planar rigid body. We then develop an expression for the velocity of a particle of a body in terms of the body coordinate frame. This velocity expression is used to define the concept of kinetic energy. Immediately thereafter, the concept of potential energy is defined and the Lagrangian is formulated. Upon completing the description of the Lagrangian, we appeal to Hamilton's principle and extract the well-known Euler-Lagrange equations.

The Euler-Lagrange equations are the system dynamics.

Reduction is the process of eliminating particular groups of coordinates involved in the mathematical expression of a system's dynamics by taking advantage of symmetries (cyclic coordinates) or certain properties of invariance present therein. The program for describing this process is as follows. To begin, a good deal of mathematical language borrowed from differential geometry is described. This language is then used to put the idea of symmetry or invariance on more solid ground mathematically. The resulting mathematical definition of invariance is used to develop the reduced Lagrangian, which is the Lagrangian under the coordinate change that eliminates the cyclic coordinates. We then acquire a bit more technical background, taking up the subject of Lie-algebras. This background will allow us give form to some of the reduction calculations that follow its introduction. We then revisit the Euler-Lagrange equations developed during the first part of the chapter. These equations are rewritten in terms of the reduced Lagrangian. Upon doing so we arrive at partially decoupled system dynamics that are free of cyclic coordinates and highlight the natural separation of the configuration space $Q = G \times S$.

3.2 Part I: Lagrangian Dynamics

3.2.1 The Body Frame and Associated Kinematic Expressions

Here, some kinematics of the planar rigid body B_i, a member of a multi-body system $\{B_i\}_{i=1}^n$, are described in terms of a body coordinate system. Consider the standard xyz inertial coordinate frame on R^3 with standard basis $\mathcal{B} = \{e^{(1)}, e^{(2)}, e^{(3)}\}$. The body B_i is assumed to act in the x-y plane of this coordinate system. A body fixed coordinate system is set up for B_i in the following way. Consider the coordinate transformation that rotates $e^{(1)}$ and $e^{(2)}$ through the angle θ_i, which is used to indicate

the orientation of the body. Namely,

$$A_i(\theta_i) = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) & 0 \\ \sin(\theta_i) & \cos(\theta_i) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let $\mathfrak{e}_i^{(j)} = A_i(\theta_i)e^{(j)}$, $j = 1, 2, 3$, and let $R_i = (x_i, y_i, 0)^t$ denote the coordinate vector of the center of mass of body B_i in the inertial frame. Then affixing $\mathfrak{e}_i^{(j)}$, $j = 1, 2, 3$, to R_i establishes a local body-fixed coordinate system for B_i. Given a point or particle P of body B_i, it may be identified in the body coordinate system by means of a coordinate vector, say $\mathfrak{r}_i$. This same point may be identified by means of another coordinate vector with respect to the global coordinate system. This global identification is given by $r_i = R_i + A_i\mathfrak{r}_i$. As an example of such a body coordinate frame see Figure 3.1, which is an illustration of the body coordinate frame for a body of the serial-link structure $\mathcal{S}$.

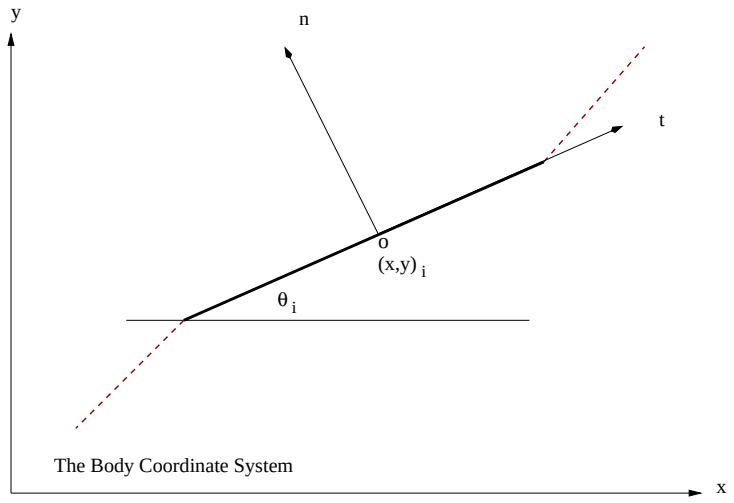


Figure 3.1: The body coordinate frame for the body B_i of the serial-link structure. In this case, the fixed body frame is comprised of normal and tangential axes n and t, respectively.

The time derivative of this expression for the global position of the particle P will provide an expression for the particle's velocity. The resulting velocity expression is

$$\dot{r}_i = \dot{R}_i + \dot{A}_i \mathfrak{r}_i, \tag{3.1}$$

where

$$\dot{A}_i(\theta_i, \dot{\theta}_i) = \begin{bmatrix} -\sin(\theta_i) & -\cos(\theta_i) & 0 \\ \cos(\theta_i) & -\sin(\theta_i) & 0 \\ 0 & 0 & 0 \end{bmatrix} \dot{\theta}_i.$$

Let $\omega_i \doteq \dot{\theta}_i e^3$. Then it can be shown that $A_i(\omega_i \times \mathfrak{r}_i) = \dot{A}_i \mathfrak{r}_i$. Note that $\omega_i \times \mathfrak{r}_i = \tilde{\omega}_i \mathfrak{r}_i = -\tilde{\mathfrak{r}}_i \omega_i$, where the $\tilde{\cdot}$ symbol indicates the matrix defined on the vector $\nu = [\nu^1, \nu^2, \nu^3]^t$ as follows:

$$\tilde{\nu} = \begin{bmatrix} 0 & -\nu^3 & \nu^2 \\ \nu^3 & 0 & -\nu^1 \\ -\nu^2 & \nu^1 & 0 \end{bmatrix}.$$

So, equation (3.1) becomes

$$\dot{r}_i = \dot{R}_i - \dot{\theta}_i A_i \tilde{\mathfrak{r}}_i e^3 = \begin{bmatrix} I^{3\times 3} & -A_i \tilde{\mathfrak{r}}_i e^3 \end{bmatrix} \begin{bmatrix} \dot{R}_i \\ \dot{\theta}_i \end{bmatrix}. \tag{3.2}$$

3.2.2 System Kinetic and Potential Energies

We now derive expressions for the kinetic and potential energies of the planar rigid body. The following definition is required:

Definition 3.1 (Kinetic Energy). *Suppose that the body B_i has uniform mass density ρ_i, and generalized volume V_i. Then, if r_i denotes the position vector of an element or particle of this body, the* kinetic energy *of this body is defined to be*

$$T_i = \frac{1}{2} \int_{V_i} \rho_i (\dot{r}_i)^2 .$$

Using equation (3.2) provides us with the expression

$$\begin{aligned}(\dot{r}_i^\tau)^2 &= \dot{r}_i^\tau \cdot \dot{r}_i^\tau \\ &= \begin{bmatrix}\dot{R}_i^t & \dot{\theta}_i\end{bmatrix} \begin{bmatrix} I^{3\times 3} \\ -(A_i\tilde{\mathfrak{r}}_i e^3)^t \end{bmatrix} \begin{bmatrix} I^{3\times 3} & -A_i\tilde{\mathfrak{r}}_i e^3 \end{bmatrix} \begin{bmatrix}\dot{R}_i \\ \dot{\theta}_i\end{bmatrix} \\ &= \begin{bmatrix}\dot{R}_i^t & \dot{\theta}_i\end{bmatrix} \begin{bmatrix} I^{3\times 3} & -A_i\tilde{\mathfrak{r}}_i e^3 \\ -(e^3)^t\tilde{\mathfrak{r}}_i^t A_i^t & (e^3)^t\tilde{\mathfrak{r}}_i^t\tilde{\mathfrak{r}}_i e^3 \end{bmatrix} \begin{bmatrix}\dot{R}_i \\ \dot{\theta}_i\end{bmatrix} ,\end{aligned}$$

where we have observed that A_i is an orthogonal matrix. Hence, we have

$$T_i = \frac{1}{2}\begin{bmatrix}\dot{R}_i^t & \dot{\theta}_i\end{bmatrix} \left(\int_{V_i} \rho_i \begin{bmatrix} I^{3\times 3} & -A_i\tilde{\mathfrak{r}}_i e^3 \\ -(e^3)^t\tilde{\mathfrak{r}}_i^t A_i^t & (e^3)^t\tilde{\mathfrak{r}}_i^t\tilde{\mathfrak{r}}_i e^3 \end{bmatrix} \right) \begin{bmatrix}\dot{R}_i \\ \dot{\theta}_i\end{bmatrix} .$$

The development of this expression is continued by carrying out the integration in a block-wise fashion. Firstly note that since the body is planar,

$$\tilde{\mathfrak{r}}_i = \begin{bmatrix} 0 & 0 & \mathfrak{r}_i^2 \\ 0 & 0 & -\mathfrak{r}_i^1 \\ -\mathfrak{r}_i^2 & \mathfrak{r}_i^1 & 0 \end{bmatrix} .$$

From this it is easily seen that $(e^3)^t\tilde{\mathfrak{r}}_i^t\tilde{\mathfrak{r}}_i e^3 = (\mathfrak{r}_i^1)^2 + (\mathfrak{r}_i^2)^2$. Hence, the block integrations that follow.

$$m_i^{RR} \doteq \int_{V_i} \rho_i I^{3\times 3} = \rho_i V_i I^{3\times 3} ,$$

$$m_i^{R\theta} \doteq -A_i \left(\int_{V_i} \rho_i \tilde{\mathfrak{r}}_i \right) e^3 = 0 ,$$

as the origin of the body coordinate system is located at the center of mass and thus $\int_{V_i} \rho_i \mathfrak{r}_i = 0$ necessarily, and

$$m_i^{\theta\theta} \doteq \int_{V_i} \rho_i ((\mathfrak{r}_i^1)^2 + (\mathfrak{r}_i^2)^2) .$$

Let

$$\mathcal{M}_i \doteq \begin{bmatrix} m_i^{RR} & m_i^{R\theta} \\ (m_i^{R\theta})^t & m_i^{\theta\theta} \end{bmatrix} = \begin{bmatrix} \rho_i V_i & 0 & 0 & 0 \\ 0 & \rho_i V_i & 0 & 0 \\ 0 & 0 & \rho_i V_i & 0 \\ 0 & 0 & 0 & m_i^{\theta\theta} \end{bmatrix}.$$

The matrix $\mathcal{M}_i$ is know as the *mass matrix* for the system body B_i. So, it is seen that the body's kinetic energy can be expressed in terms of the body's mass matrix as

$$T_i = \frac{1}{2} \begin{bmatrix} \dot{R}_i^t & \dot{\theta}_i \end{bmatrix} \mathcal{M}_i \begin{bmatrix} \dot{R}_i \\ \dot{\theta}_i \end{bmatrix} = \frac{1}{2} m_i (\dot{x}_i)^2 + \frac{1}{2} m_i (\dot{y}_i)^2 + \frac{1}{2} m_i^{\theta\theta} (\dot{\theta}_i)^2 \ , \tag{3.3}$$

where $m_i \doteq \rho_i V_i$ is the total mass of the body. The system kinetic energy is defined by the expression $T = \sum_{i=1}^{n} T_i$.

To solidify the concept described here, attention is turned to the particular systems described in Chapter 2. The kinetic energy of these two systems is now calculated from equation (3.3).

Example 3.1 (The Mass-Spring System). *To begin, we note that since the masses are treated as particles, they have no orientation. Thus, $\theta_i = 0$. Further, it is noted that due to the assumption that the particles lie on the x-axis, $y_i = 0$. Referring to Figure 2.1 we have body B_1 of mass m_1 with position x_1 and body B_2 of mass m_2 with position x_2. Thus, by equation* (3.3)*,*

$$T_1 = \frac{1}{2} m_1 \dot{x}_1^2 \ ,$$
$$T_2 = \frac{1}{2} m_2 \dot{x}_2^2 \ .$$

This, of course, illustrates the consistency of the extension of the definition of kinetic energy, for when the body is considered to be a particle, the definition reduces the familiar expression for the kinetic energy of a particle. The kinetic energy of the system is $T = T_1 + T_2 = \frac{1}{2} m_1 \dot{x}_1^2 + \frac{1}{2} m_2 \dot{x}_2^2$.

Example 3.2 (The Serial-link Structure). *For this system, we recognize that the generalized volume V_i is simply the length of the link or body B_i, which is $2l_i$. So, m_i, the total mass of body B_i, is $2\rho_i l_i$. It remains to specify precisely what the rotational mass $m^{\theta\theta}$ is. We start by noting that $\mathfrak{r}_i^2 = 0$ for the body. Hence,*

$$\begin{aligned} m_i^{\theta\theta} &= \int_{-l_i}^{l_i} \rho_i \sigma^2 \, d\sigma \\ &= m_i \frac{l^2}{3} \ , \end{aligned}$$

and it is seen through equation (3.3) *that the kinetic energy of link B_i is*

$$T_i = \frac{1}{2} m_i \dot{x}_i^2 + \frac{1}{2} m_i \dot{y}_i^2 + \frac{m_i}{2} \frac{l^2}{3} \dot{\theta}_i^2 \ .$$

Of course $T = \sum_{i=1}^n T_i$ for the n-link structure $\mathcal{S}$.

We now wish to describe how one may formulate an expression for the kinetic energy of a planar system in terms of an independent generalized coordinate vector q. So, consider the planar multi-body system $\{B_i\}_{i=1}^n$. It has already been explained that the configuration of each of these bodies may be completely described by assigning values to the position coordinates (x_i, y_i) of the centroid of each body and the orientation θ_i of each body. Label these coordinates $(z_{3i-2}, z_{3i-1}, z_{3i})$. Then $z \doteq (z_1, \ldots, z_{3n})$ is a generalized coordinate vector for the system. It was shown that the body kinetic energy is given by $T_i = \frac{1}{2} m_{3i-2} \dot{z}_{3i-2}^2 + \frac{1}{2} m_{3i-1} \dot{z}_{3i-1}^2 + \frac{1}{2} m_{3i} \dot{z}_{3i}^2$, where m_{3i-2} and m_{3i-1} are the total body mass m_i and m_{3i} is its rotational mass $m_i^{\theta\theta}$. This creates an inconsistency in labelling. However, in an attempt to keep the index i associated with body B_i we will temporarily make the abuse and continue with the new meaning of the subscripted m whose index now runs from $1, \ldots, 3n$. We arrive at T by summing these energies over all n bodies. Thus,

$$T = \frac{1}{2} \sum_{j=1}^{3n} m_j \dot{z}_j^2 \ . \tag{3.4}$$

Now suppose that $z_j = z_j(q)$. Then $\dot{z}_j = \frac{\partial z_j}{\partial q}\dot{q}$ and $\dot{z}_j^2 = \sum_k^m \sum_l^m \frac{\partial z_j}{\partial q_k}\frac{\partial z_j}{\partial q_l}\dot{q}_k\dot{q}_l$. Going back to equation (3.4) and gathering terms corresponding to the product $\dot{q}_k\dot{q}_l$ we arrive at the expression $T(q,\dot{q}) = \frac{1}{2}\sum_k^m \sum_l^m m_{kl}\dot{q}_k\dot{q}_l$, where $m_{kl} = \sum_{j=1}^{3n} m_j \frac{\partial z_j}{\partial q_k}\frac{\partial z_j}{\partial q_l}$. Let $[M_j]_{kl} \doteq [m_j \frac{\partial z_j}{\partial q_k}\frac{\partial z_j}{\partial q_l}]$. Then the matrix M_j is given by $m_j \nabla z_j (\nabla z_j)^t$ and $T(q,\dot{q}) = \frac{1}{2}\dot{q}^t M \dot{q}$, where $M = \sum_{j=1}^{3n} M_j$. M is known as the system's *generalized mass matrix*. Note that since $z_j(q)$, ∇z_j is a function of q and thus $M = M(q)$.

We shall illustrate the calculations just detailed via the mass-spring system and $\mathcal{S}$. Determining the kinetic energies of each system in terms of the generalized coordinate vector q identified for them in Chapter 2.

Example 3.3 (The Mass-Spring System Kinetic Energy). *As seen in Chapter 2, we may define the so-called shape variable s to be $x_2 - x_1$ and upon doing so, the system my be described in terms of $q = (g, s)^t$. Doing so, we arrive at the expressions*

$$z_1(q) = g \ ,$$
$$z_2(q) = x_2 = g + s \ .$$

With these equations at hand, the process of calculating the kinetic energy in terms of the coordinate set q becomes algorithmic. We have

$$M_1 = m_1 \nabla z_1 \nabla z_1^t = m_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} , \quad \textit{and} \quad M_2 = m_2 \nabla z_2 \nabla z_2^t = m_2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} .$$

So,

$$M = M_1 + M_2 = \begin{bmatrix} m_1 + m_2 & m_2 \\ m_2 & m_2 \end{bmatrix}$$

and thus

$$\begin{aligned} T(q,\dot{q}) &= \frac{1}{2}\dot{q}^t M \dot{q} \\ &= \frac{1}{2}((m_1+m_2)\dot{g}^2 + 2m_2\dot{g}\dot{s} + m_2\dot{s}^2) \\ &= \frac{1}{2}(m_1\dot{g}^2 + m_2(\dot{g}+\dot{s})^2) \ . \end{aligned}$$

By noting that $\dot{g}+\dot{s} = \dot{x}_2$*, it is seen that one arrives at the same expression of kinetic energy as found in Example 3.1.*

Example 3.4 (The Serial-link Structure Kinetic Energy). *It is clear that the primary element of calculating the kinetic energy with respect to generalized coordinates is the gradient of* z_j*. Hence, to initiate this example, the appropriate derivatives are now specified. Consider first* $x_i = z_{3i-2}$*,*

$$\frac{\partial x_i}{\partial x} = 1 \ , \qquad\qquad \frac{\partial x_i}{\partial y} = 0 \ ,$$

$$\frac{\partial x_i}{\partial \theta} = l_1 \sin(\theta) + \sum_{j=2}^{i-1} 2l_j \sin(\theta_j) + l_i \sin(\theta_i) \ ,$$

and

$$\frac{\partial x_i}{\partial \phi_p} = \begin{cases} -\sum_{j=2}^{i-1} 2l_j \sin(\theta_j) - l_i \sin(\theta_i) \ , & \text{if } p < i-1; \\ -l_i \sin(\theta_i) \ , & \text{if } p = i-1; \\ 0 \ , & \text{otherwise.} \end{cases}$$

Likewise, for $y_i = z_{3i-1}$*,*

$$\frac{\partial y_i}{\partial x} = 0 \ , \qquad\qquad \frac{\partial y_i}{\partial y} = 1 \ ,$$

$$\frac{\partial y_i}{\partial \theta} = -l_1 \cos(\theta) - \sum_{j=2}^{i-1} 2l_j \cos(\theta_j) - l_i \cos(\theta_i) \ ,$$

and

$$\frac{\partial x_i}{\partial \phi_p} = \begin{cases} \sum_{j=2}^{i-1} 2l_j \cos(\theta_j) + l_i \cos(\theta_i) \ , & \text{if } p < i-1; \\ l_i \cos(\theta_i) \ , & \text{if } p = i-1; \\ 0 \ , & \text{otherwise.} \end{cases}$$

Finally, considering $\theta_i = z_{3i}$,

$$\frac{\partial \theta_i}{\partial x} = 0 \ , \qquad \frac{\partial \theta_i}{\partial y} = 0 \ , \qquad \frac{\partial \theta_i}{\partial \theta} = 1 \ ,$$

and

$$\frac{\partial \theta_i}{\partial \phi_p} = \begin{cases} 0 \ , & \text{if } p \geq i; \\ -1 \ , & \text{otherwise.} \end{cases}$$

Referring to Example 3.2 it is seen that $m_{3i} = \frac{m(i)l_i^2}{3}$ *is the rotational mass, where as a consequence of our notation abuse we are forced to make the definition* $m(i) = m_{3i-2} = m_{3i-1}$. *Now it is simply the process of calculating* $M = \sum_{j=1}^{3n} m_j \nabla z_j \nabla z_j^t$ *and performing the matrix multiplication* $T = \frac{1}{2}\dot{q}^t M \dot{q}$. *This is, admittedly, easier said than done. We may carry out the calculations with computer algebra software such as* Maple, *however, for large* n, *the resulting expressions are intractable. Hence, this system needs to be handled computationally.*

Despite this note, the resultant mass matrix M *is provided here for the case of* $\mathcal{S}$ *with* $n = 2$ *by* Maple *for the sake of completion:*

```
> restart;
> with(linalg);
> z[1] := x;
```

$$z_1 := x$$

```
> z[2] := y;
```

$$z_2 := y$$

```
> z[3] := theta;
```

$$z_3 := \theta$$

```
> z[4] := x - l*cos(theta)-l*cos(theta - phi[1]);
```

$$z_4 := x - l\cos(\theta) - l\cos(-\theta + \phi_1)$$

```
> z[5] := y - l*sin(theta)-l*sin(theta - phi[1]);
```

$$z_5 := y - l\sin(\theta) + l\sin(-\theta + \phi_1)$$

```
> z[6] := theta - phi[1];
```

$$z_6 := \theta - \phi_1$$

```
> for i from 1 by 1 to 6 do
> gz[i]:= grad(z[i],[x,y,theta,phi[1]]);
> end do;
```

$$gz_1 := [1,\ 0,\ 0,\ 0]$$

$$gz_2 := [0,\ 1,\ 0,\ 0]$$

$$gz_3 := [0,\ 0,\ 1,\ 0]$$

$$gz_4 := [1,\ 0,\ l\sin(\theta) - l\sin(-\theta + \phi_1),\ l\sin(-\theta + \phi_1)]$$

$$gz_5 := [0,\ 1,\ -l\cos(\theta) - l\cos(-\theta + \phi_1),\ l\cos(-\theta + \phi_1)]$$

$$gz_6 := [0,\ 0,\ 1,\ -1]$$

```
> for i from 1 to 6 do
> if (i mod 3 = 0) then
> m[i]:=(1/3)*m*l^2;
> else
> m[i]:=m;
> end if;
> end do;
```

```
> M := evalm(m[1]*(gz[1]*transpose(gz[1])));
```

$$M := \begin{bmatrix} m & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

```
> for i from 2 to 6 do
> M := evalm(M + evalm(m[i]*(gz[i]*transpose(gz[i]))));
> end do:
> print(M);
```

$$\begin{bmatrix} 2\,m\,,\ 0\,,\ m\,\%4\,,\ m\,l\,\%2 \\ 0\,,\ 2\,m\,,\ m\,\%3\,,\ m\,l\,\%1 \\ m\,\%4\,,\ m\,\%3\,,\ \frac{2}{3}\,m\,l^2 + m\,\%4^2 + m\,\%3^2\,,\ m\,\%4\,l\,\%2 + m\,\%3\,l\,\%1 - \frac{1}{3}\,m\,l^2 \\ m\,l\,\%2\,,\ m\,l\,\%1\,,\ m\,\%4\,l\,\%2 + m\,\%3\,l\,\%1 - \frac{1}{3}\,m\,l^2\,,\ m\,l^2\,\%2^2 + m\,l^2\,\%1^2 + \frac{1}{3}\,m\,l^2 \end{bmatrix}$$

$$\%1 := \cos(-\theta + \phi_1)$$
$$\%2 := \sin(-\theta + \phi_1)$$
$$\%3 := -l\cos(\theta) - l\,\%1$$
$$\%4 := l\sin(\theta) - l\,\%2$$

The concept of potential energy for a body B_i is now defined.

Definition 3.2 (Potential Energy). *Suppose that a force F acts on the body B_i and that there exists a differentiable function $P(q)$ such that $-\nabla P = F$. Then the potential energy of body B_i due to the force F is said to be $P(q)$.*

A force which has a potential function is said to be *conservative*. The potential energies related to the mass-spring system and $\mathcal{S}$ are now discussed.

Example 3.5 (The Mass-Spring System Potential Energy). *It is assumed that the spring used in the mass-spring example responds to compression or elongation with the returning force $F = -k(1 + a^2 s^2)s$, where k is the so-called spring constant and a is a stiffness parameter. Note that if $a = 0$, the Hookean or ideal spring results. Other values of a correspond to so-called soft or hard springs. This force may be obtained via the potential function $P(s) = \frac{1}{2}ks^2 + \frac{1}{4}ka^2s^4$, as $-\nabla P = F$. All other*

forces acting on the mass-spring system, namely the control and the friction, are non-conservative and do not contribute to the system's potential energy. Thus, P is the potential for the mass-spring system.

The only forces acting on the serial-link structure are control and frictional forces. These forces are non-conservative and due to this fact, $\mathcal{S}$ has no potential energy expression P.

3.2.3 The Lagrangian, Virtual Work, Hamilton's Principle, and the Euler-Lagrange Equations

With the ideas of kinetic and potential energy set forth, the definition of the mechanical Lagrangian can now be stated.

Definition 3.3 (The Mechanical Lagrangian). *Consider a system described with generalized system coordinates q that has total kinetic energy $T(q, \dot{q})$ and total potential energy $P(q)$. The mechanical Lagrangian for that system is defined to be*

$$L(q, \dot{q}) = T(q, \dot{q}) - P(q) \ .$$

Example 3.6 (The Mass-Spring System Lagrangian). *No calculation needs to be done here. Simply referring to the expressions obtained in Examples 3.3 and 3.5 we obtain*

$$\begin{aligned} L(q, \dot{q}) &= T(q, \dot{q}) - P(q) \\ &= \frac{1}{2}(m_1 \dot{x}_1^2 + m_2(\dot{x}_1 + \dot{s})^2) - \left(\frac{1}{2}ks^2 + \frac{1}{4}ka^2 s^4\right) \ . \end{aligned}$$

As for the Lagrangian of the serial-link structure, $P = 0$ and thus the system Lagrangian is simply the system's kinetic energy. i.e., $L = T$.

The concepts of virtual work and the generalized force are now introduced for the planar rigid body. Consider a force F_i acting on the body $B_i \in \{B_i\}_{i=1}^n$ at the point P

with global coordinates $r_i^P = R_i + A_i \mathfrak{r}_i$. Consider a virtual displacement or variation of this particle's position, δr_i^P. The *virtual work* done by the force F_i is then defined to be $\delta W_i = F_i^t \delta r_i^P$. This is similar to the ordinary definition of work. However, it is labelled virtual since the variation in the position is not actually due to the dynamics of the body. Calculating the variation in r_i^P due to variations in R_i and θ_i provides the expression

$$\begin{aligned} \delta r_i^P &= \delta R_i + A_i^\theta \mathfrak{r}_i \delta\theta \\ &= \begin{bmatrix} I^{(2\times 2)} & A_i^\theta \mathfrak{r}_i \end{bmatrix} \begin{bmatrix} \delta x_i \\ \delta y_i \\ \delta\theta_i \end{bmatrix} , \end{aligned}$$

where A_i^θ is the entry-wise partial of the matrix A with respect to θ. Thus,

$$\delta W_i = \begin{bmatrix} F_i^t & F_i^t A_i^\theta \mathfrak{r}_i \end{bmatrix} \begin{bmatrix} \delta x_i \\ \delta y_i \\ \delta\theta_i \end{bmatrix} . \tag{3.5}$$

Noting that $F_i^t A_i^\theta \mathfrak{r}_i = \pm \| A_i \mathfrak{r}_i \times F_i \| = M_{F_i}$, the moment of force F_i about the center of mass, this expression implies a very well-known result. For a rigid body, we may translate a force from anywhere on the body to the center of mass so long as the corresponding moment of the force is accounted for.

Now, suppose that the body coordinates $\vec{z}_i \doteq (z_{3i-2}, z_{3i-1}, z_{3i})^t = (x_i, y_i, \theta_i)^t$, as defined in previous discussion, may be expressed in terms of some other generalized coordinate vector q. Then

$$\delta \vec{z}_i(q) = \frac{\partial \vec{z}_i}{\partial q_k} \delta q_k \ .$$

Upon substitution of this equation into Equation (3.5) we arrive at the equation

$$\delta W_i = \begin{bmatrix} F_i^t & M F_i \end{bmatrix} \frac{\partial \vec{z}_i}{\partial q_k} \delta q_k \ . \tag{3.6}$$

Indeed it may be the case that more than one force acts on a particular body. However, all of the quantities discussed follow some type of superposition principle and thus we look at F_i as a composite force with the knowledge that these same calculations may be carried out force-by-force. The total virtual work of the system due to the forces $\{F_i\}_{i=1}^n$ is $\delta W = \sum_i \delta W_i$. Using equation (3.6) we obtain the following expression:

$$\begin{aligned}
\delta W &= \sum_{i=1}^{n} \sum_{k=1}^{m} \begin{bmatrix} F_i^t & MF_i \end{bmatrix} \frac{\partial \vec{z}_i}{\partial q_k} \delta q_k \\
&= \sum_{k=1}^{m} \left(\sum_{i=1}^{n} \begin{bmatrix} F_i^t & MF_i \end{bmatrix} \frac{\partial \vec{z}_i}{\partial q_k} \right) \delta q_k \\
&= \sum_{k=1}^{m} Q_k \delta q_k \ ,
\end{aligned}$$

where

$$Q_k = \sum_{i=1}^{n} \begin{bmatrix} F_i^t & M_i \end{bmatrix} \frac{\partial \vec{z}_i}{\partial q_k} \ .$$

The Q_k, $k = 1, \ldots, m$, are known as the system's *generalized forces*. Given that F_i and M_i represent a cumulative force and corresponding cumulative moment on body i, it is seen that we obtain a single generalized force for each generalized system coordinate. Explicit expressions for the non-conservative generalized forces acting on the mass-spring and serial-link structure systems, namely controls and friction, will be discussed in detail in Chapter 4.

The progression has brought the discussion to the point where *Hamilton's Principle* may be introduced. Hamilton's principle is

$$\delta \int_{t_1}^{t_2} L(q(t), \dot{q}(t)) \ dt + \int_{t_1}^{t_2} \delta W_{nc} \ dt = 0 \ . \tag{3.7}$$

Here, $q(t)$ is the trajectory of a system's generalized coordinates, $\dot{q}(t) = \frac{d}{dt} q(t)$, $q(t_1)$ and $q(t_2)$ are fixed, and δW_{nc} is the virtual work of the system due to non-conservative forces. An account of the derivation of this principle may be found in texts on

mechanics and the calculus of variations [19].

At this juncture a brief discussion on those aspects of the calculus of variations that will be useful in extrapolating system dynamics from Hamilton's principle will be presented. Consider the integral functional $J : \mathcal{C}^1[t_1, t_2] \mapsto \mathbb{R}$ defined by

$$J(x) = \int_{t_1}^{t_2} g(t, x(t), \dot{x}(t))\ dt\ ,$$

for $x \in \mathcal{C}^1[t_1, t_2]$. Here, t_1 and t_2 are fixed, the values of $x(t)$ are fixed at these values, and it is assumed that g is smooth enough to allow the calculations that are to follow. An expression for the *first variation* of this functional due to a variation in x is sought. What we intend by variation is to be made clear as we progress with the calculation thereof. To begin, consider the one parameter family of *variations* defined by

$$x_\alpha = x(t) + \alpha X(t)\ ,$$

where α is a sufficiently small scalar quantity and X is a smooth function on $[t_1, t_2]$ satisfying the boundary conditions

$$X(t_1) = X(t_2) = 0\ . \tag{3.8}$$

In this way, it is assured that x_α will belong to a particular class of functions, typically referred to as the admissible set or the set of admissible variations. Define the functional $\Phi : \mathbb{R} \to \mathbb{R}$ through J by

$$\Phi(\alpha) = J(x_\alpha) = \int_{t_1}^{t_2} g(t, x_\alpha(t), \dot{x}_\alpha(t))\ dt\ .$$

Since precautions have been made to ensure that Φ is sufficiently smooth, Taylor's theorem may be applied to Φ to arrive at the expression

$$\Phi(\alpha) = \Phi(0) + \alpha \frac{d\Phi}{d\alpha}(0) + R(\alpha)\ ,$$

where R is the familiar remainder term.

$$\delta J = \alpha \frac{d\Phi}{d\alpha}(0)$$

is known as the first variation of J due to the variation in x, $\delta x \doteq \alpha X$.

The calculations are continued by stating

$$\int_{t_1}^{t_2} (g_x(t, x, \dot{x})X + g_{\dot{x}}(t, x, \dot{x})\dot{X})\ dt$$

as the expression of $\frac{d\Phi}{d\alpha}(0)$ upon application of the chain rule. Performing integration by parts on the second term of this integration results in the expression

$$\begin{aligned}\int_{t_1}^{t_2} g_{\dot{x}}\dot{X}\ dt &= g_{\dot{x}}X\Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{d}{dt}(g_{\dot{x}})X\ dt \\ &= -\int_{t_1}^{t_2} \frac{d}{dt}(g_{\dot{x}})X\ dt\ ,\end{aligned}$$

where the boundary conditions of equation (3.8) have been observed. Thus one has

$$\delta J = \int_{t_1}^{t_2} \left(g_x - \frac{d}{dt}g_{\dot{x}}\right) \delta x\ dt\ . \tag{3.9}$$

This is precisely the expression needed to acquire what is desired from Hamilton's principle and no more talk of the calculus of variations will ensue.

Returning to Hamilton's principle with Equation (3.9) in hand we may restate the principle as

$$\int_{t_1}^{t_2} \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} - Q_{nc}\right) \delta q\ dt = 0\ ,$$

where Q_{nc} is the vector of generalized non-conservative forces. Given sufficient smoothness of Q_{nc}, it is relatively clear that the only way this principle is to hold is if

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = Q_{nc}\ . \tag{3.10}$$

These equations are the so-called *Lagrange-D'Alembert* equations in mechanics and the *Euler-Lagrange* equations in the realm of general optimization mathematics. They provide a system of ODEs in the independent generalized coordinates that describe the dynamics of the system under consideration.

3.3 Part II: Symmetry and Reduction

3.3.1 Geometric Mechanics: Some Basic Language

This section begins with a very quick introduction to some definitions and nomenclature used quite often in the field of geometric mechanics. This exposition is provided primarily to allow the use of the language and to provide some intuition. The mathematics involved is quite standard and we wish to emphasize this point, as the geometric perspective can burden the understanding of some end results by mystifying things in a quagmire of terminology. There will be no talk of connections, forms, and other such terms that usually accompany this type of discussion.

Firstly, the definition of a Lie group is given. This definition is constructed on the object known as a *manifold*. We do not wish to get into very mathematically rigorous descriptions of manifolds. Instead, suffice it to say that an n-dimensional manifold is a space that has a natural correspondence at the local level with $\mathbb{R}^n$ via objects known as charts . This includes $\mathbb{R}^n$ itself but also includes surfaces such as the generalized spherical surface $\mathbb{S}^n$. The point is that, if there is a natural correspondence of a space with $\mathbb{R}^n$, then we can have a calculus on this space defined through this mapping. We feel confidant that the manifolds discussed herein leave little to the imagination and thus there is no problem of thinking in $\mathbb{R}^n$. Also, the concept of the algebraic construct known as a group is used. We assume that this concept is well understood.

Definition 3.4 (Lie Group). *A manifold G with a group structure is said to be a* Lie group *if the product mapping $hg : G \times G \to G$ as well as the inverse mapping $g^{-1} : G \to G$ are C^∞ mappings.*

The best interpretation of the Lie group for the purposes here is to consider these objects as sets of coordinate changes. This interpretation is facilitated by providing appropriate examples and examining the resultant objects closely. This is done directly.

Example 3.7 (The Lie Group SE(1)). *Consider the smooth manifold $G = \mathbb{R}$ and the group structure $G = (\mathbb{R}, +)$. It is very obvious that both the product mapping, which is addition, and the inverse mapping, which is negation, are infinitely smooth. Thus, G is a Lie group. This group is known as the special Euclidean group of one dimension and is denoted $SE(1)$. This group is quite simply the translation group on the real line.*

It is noted that this Lie group can be identified with a matrix Lie group (a group of matrices under ordinary matrix multiplication). Let $\tilde{G}$ be the manifold whose elements have the form

$$\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix},$$

where $t \in \mathbb{R}$. Clearly $\tilde{G}$ and G are identical as manifolds and in fact share charts. $\tilde{G}$ has a group structure under matrix multiplication.

$$\varphi(t) \doteq \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

provides the group isomorphism.

Example 3.8 (The Lie Group SE(2)). *As a second example, consider the smooth manifold $G = \mathbb{R}^2 \times \mathbb{S}$ and two representative elements thereof, $g = (x, y, \theta)$ and $h = (a, b, \alpha)$. Group multiplication on this manifold is defined by $hg = (a + x\cos(\alpha) - y\sin(\alpha), b + x\sin(\alpha) + y\cos(\alpha), \theta + \alpha)$ and as such, $g^{-1} = (-x\cos(\theta) - y\sin(\theta), x\sin(\theta) - y\cos(\theta), -\theta)$. It is clear by these expressions that both the product and inverse mappings are smooth. Thus G is a Lie group. This group also carries a name. It is*

known as the special Euclidean group of two dimensions or SE(2). In simple terms, this is the group of translations and rotations in the plane.

Like SE(1), SE(2) can also be identified with a matrix Lie group. The isomorphism is given by the mapping

$$\varphi(g) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & x \\ \sin(\theta) & \cos(\theta) & y \\ 0 & 0 & 1 \end{bmatrix} . \tag{3.11}$$

Given a Lie group, or simply group, one may define a mapping that corresponds to left multiplication. To be more specific, for $h \in G$ define $L_h : G \to G$ by $L_h(g) = hg$. It is quite clear that this is a smooth mapping and that L_h has a smooth inverse map, namely $L_{h^{-1}}$. Thus, L_h is a diffeomorphism. Now, we introduce the concept of the *tangent vector*. Let $h \in G$ (we think in $\mathbb{R}^n$), then $v_h \in T_hG$, the *tangent space* at h, is an element of $\mathbb{R}^n \times \mathbb{R}^n$ of the form (v, h). v is the tangent part and h is the point of tangency. $TG \doteq \bigcup_{h \in G} T_hG$ is known as the *tangent bundle*. Of course, L_h has a tangent map $T_{g_1}L_h : T_{g_1}G \to T_{L_h(g_1)}G$. This map is given in coordinates as

$$T_{g_1}L_h(v_{g_1}) = \left(\frac{\partial L_h}{\partial g}\Big|_{g_1} v \right)_{L_h(g_1)} . \tag{3.12}$$

This verbiage is best accompanied with a continuation of examples.

Example 3.9 (Left Multiplication in $SE(1)$). *Going back to $G = SE(1)$, let $t_1 \in G$. Then for $t \in G$, $L_{t_1}(t) = t_1 + t$. L_{-t_1} is the inverse of L_{t_1}, as is clearly illustrated with the calculation $L_{-t_1}(L_{t_1}(t)) = -t_1 + (t_1 + t) = t$. The tangent map or Jacobain here is the identity* 1.

Example 3.10 (Left Multiplication in $SE(2)$). *Here, the case of $G = SE(2)$ is*

dealt with. Let $h^t = (a, b, \alpha) \in G$. *Then, for* $g^t = (x, y, \theta) \in G$ *we have*

$$L_h(g) = \begin{bmatrix} a + x\cos(\alpha) - y\sin(\alpha) \\ b + x\sin(\alpha) + y\cos(\alpha) \\ \alpha + \theta \end{bmatrix} ,$$

and the inverse of L_h *is*

$$L_{h^{-1}}(g) = \begin{bmatrix} (x-a)\cos(\alpha) + (y-b)\sin(\alpha) \\ (a-x)\sin(\alpha) - (b-y)\cos(\alpha) \\ \theta - \alpha \end{bmatrix} .$$

The tangent map or Jacobian of L_h, *which shall be denoted by* $[h]$, *is seen to be the rotation matrix*

$$[h] = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

Additionally, the tangent map or Jacobian of the inverse map $L_{h^{-1}}$, *to be denoted* $[h]^{-1}$ *(*$= [h^{-1}] = [h]^t$*), is*

$$[h]^{-1} = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix} ,$$

which is the planar rotation by $-\alpha$.

It can now be made clear that these left multiplications are coordinate changes. In the case of $G = SE(1)$, suppose that the coordinate t indicates the position of a particle relative to the origin of the real line. Then it is clear that $L_{h \doteq t_1}(t)$ is the position coordinate relative to a coordinate system whose origin is located at $h^{-1} = -t_1$. Furthermore, the tangent mapping provides the velocity of the particle t'

in the new coordinates to be $[t_1]t' = 1t' = t'$. This says that the velocity is unchanged by the new coordinate selection. Note however, that velocities are elements of the tangent bundle TG, and as such have two components, the tangent component, and the position component to which the velocity is attached. As seen in equation (3.12), the new location of the velocity is $L_{t_1}(t) = t_1 + t$, which is exactly as expected. At this juncture, the most important concept behind the reduction process can be explained. Suppose that a particle on the real line is at position t. Then the coordinate change $L_{-t}(t)$ should tell us the coordinates of this particle if the origin of the new system were to be t. Well, that position is 0, and that is precisely what $L_{-t}(t)$ is. Also, note that in this case, the velocity of the particle $[h^{-1}]t'$ is always attached to the identity of the group $e = 0$ and as such, is contained in T_eG. It has been made quite clear that L_h is smooth with respect to h. So, we can continuously and smoothly change coordinates so that the origin of the coordinate system is always on the particle. That is, by continuously altering the coordinate system, the position of the particle is always 0. It is the concept of the body frame, or in this case, the particle frame. Thus, $SE(1)$ is precisely the group needed to describe a particle frame for the mass-spring example.

As is true for the case of $G = SE(1)$, there is an interpretation of L_h for $h \in G = SE(2)$ as simply a coordinate change. Without providing a plethora of explanation, the interpretation is this. Suppose that we place a body coordinate system on a planar rigid body. Then the origin of the body coordinates lie at the centroid of the object, say (x, y), and the frame has an orientation, say θ. Together, these coordinates make up $g \in G = SE(2)$. Then $L_h(g)$ provides the origin and orientation of this body frame if the inertial frame is reselected to be h^{-1}. So, suppose we once again wish to select the frame so as to coincide with g. Then $L_{g^{-1}}(g)$ does precisely this. Also, it is mentioned that the tangent operator $[h]$ maps the velocity in the existing frame to the components of the velocity for the frame with position and orientation given by h^{-1}. Unlike the one dimensional example, where the components of the velocity were

unchanged, the components here are changed. In fact they are rotated. Translation of the frame should not affect the velocity, but rotation does. This is exactly why the tangent map for $SE(2)$ is simply a planar rotation matrix. Also note that, just as in the 1-dimensional case, the velocity in the new coordinate system $[g^{-1}]\dot{g}$ is attached to the group identity e, which is 0. That is, this velocity is contained in T_eG for this particular coordinate selection. The point is, Lie groups, at least for the purpose here, are nothing more than collections of coordinate change matrices. Also, note that $G = SE(2)$ is precisely the group needed to describe coordinate changes for the serial-link structure.

Continuing, recall now that for both of the systems being considered, the configuration manifold Q, comprised of independent generalized coordinates, has been partitioned to be $Q = G \times S$, where a point $q = (g^t, s^t)^t \in Q$ is partitioned to reflect g, the position and orientation of the mechanism at hand, and s, the internal shape of the complete system. It should now be clear that G has a Lie group structure and it is for this reason that it was labelled in this manner. G is sometimes called the group fibres. S represents what is called the *base manifold* of the *principle trivial fiber bundle*. These terms will not be discussed here except to say that researchers often speak of motion across the group fibers and motion along the base manifold. Motion along the group fibers alludes to changes in the position and orientation of the system with no change in internal configuration. Motion along the base manifold is change to internal configuration without any change to the system position and orientation. It is not a stretch of the imagination to see that there cannot be pure motion along the group fibers or the base manifold for the systems that we are interested in. That is, these are coupled. Hence, the goal is the following: to elicit change in the group variables g via change in the shape variables s. This will soon be made more clear.

First, let the definition of the left action L_g be extended to the entire configuration

manifold Q. This is accomplished with the map $\Phi_g : Q \to Q = G \times S$ given by

$$\Phi_g(q) = \Phi_g(h, s) = (L_g(h), s)$$

This is simply a coordinate change. The shape s is internal and is unaffected by a coordinate change and the definition reflects this. Now, the following definition:

Definition 3.5 (Lifted Action). *The* lifted action *is the map* $T\Phi_g : TQ \to TQ$ *defined by* $T\Phi_g(q, v) = (\Phi_g(q), T_q\Phi_g(v))$.

This is nothing more than the coordinate change once again. We now, for the purpose of later convenience, write down some explicit expressions for the lift actions.

Example 3.11 (Lift Action for $G = SE(1)$). *Firstly, let it be noted that we shall denote the lifting of both the coordinate vector* q *and its associated tangent vector* v_q *with the up arrow. e.g.,* $q \uparrow \Phi_g(q)$ *and* $v_q \uparrow T_q\Phi_g(v_q)$. *Let* $q = (g, s)^t \in Q$ *and suppose we have the tangent vector* v_q. *Then, for* $g_1 \in G$,

$$q \uparrow \Phi_{g_1}(q) = \begin{bmatrix} L_{g_1}(g) \\ s \end{bmatrix} = \begin{bmatrix} g_1 + g \\ s \end{bmatrix} ,$$
$$v_q \uparrow T_q\Phi_{g_1}(v_q) = \begin{bmatrix} [g_1] & 0 \\ 0 & 1 \end{bmatrix} v = v ,$$

Recalling from Example 3.9 that $[g_1] = 1$.

Example 3.12 (Lift Action for Serial-Link Structure). *Let* $q = (g^t, s^t)^t \in Q$ *and suppose* $h \in G$, *where* $g^t = (x, y, \theta)$ *and* $h^t = (a, b, \alpha)$. *From Example 3.10 we have*

$$q \uparrow \Phi_h(q) = \begin{bmatrix} L_h(g) \\ s \end{bmatrix} = \begin{bmatrix} a + x\cos(\alpha) - y\sin(\alpha) \\ b + x\sin(\alpha) + y\cos(\alpha) \\ \theta + \alpha \\ s \end{bmatrix} .$$

From that same example we see that

$$v_q \uparrow T_q\Phi_h(v_q) = \begin{bmatrix} [h] & 0 \\ 0 & I^{(n-1)} \end{bmatrix} v = \begin{bmatrix} [h]v_g \\ v_s \end{bmatrix} ,$$

where n is the total number of links, 0 is understood to be a matrix with zero entries of appropriate dimensions, and as stated before,

$$[h] = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

3.3.2 System Invariance and the Reduced Lagrangian

Now the definition that is of primary interest.

Definition 3.6 (Group Invariance of the Lagrangian). *A Lagrangian $L : TQ \to \mathbb{R}$ is said to be G invariant if for each h in the group G and each (q, v_q) in TQ one has*

$$L(\Phi_h(q), T_q\Phi_h(v_q)) = L(q, v_q) .$$

That is, the value of the Lagrangian is unaffected by the coordinate change.

The Lagrangian for both of the systems that this study includes are invariant with regards to their respective groups, as illustrated in the following example and theorem.

Example 3.13 (Invariance of the Mass-Spring System Lagrangian). *Continuing with the results of Example 3.11 and the Lagrangian of Example 3.6 we have, with component arguments,*

$$\begin{aligned} L(g_1 + g, s, v_g, v_s) &= \frac{1}{2}(m_1 v_g^2 + m_2(v_g + v_s)^2) - \left(\frac{1}{2}ks^2 + \frac{1}{4}ka^2s^4\right) \\ &= L(g, s, v_g, v_s) . \end{aligned}$$

This says that the mass-spring system is $G = SE(1)$ invariant.

Theorem 3.1 ($SE(2)$ invariance of the serial-link Lagrangian). *The Lagrangian of the serial-link structure is $G = SE(2)$ invariant for all n.*

Proof. We establish the invariance of the body kinetic energy $T_i = \frac{1}{2}m_{3i-2}\dot{x}_i^2 + \frac{1}{2}m_{3i-1}\dot{y}_i^2 + \frac{1}{2}m_{3i}\dot{\theta}_i^2$. Let us begin with the kinematic expression

$$\begin{bmatrix}\dot{x}_i\\ \dot{y}_i\end{bmatrix} = \begin{bmatrix}\dot{x}_1\\ \dot{y}_1\end{bmatrix} + l_1\begin{bmatrix}\sin(\theta_1)\\ -\cos(\theta_1)\end{bmatrix}\dot{\theta}_1 + \sum_{j=2}^{i-1} 2l_j\begin{bmatrix}\sin(\theta_j)\\ -\cos(\theta_j)\end{bmatrix}\dot{\theta}_j + l_i\begin{bmatrix}\sin(\theta_i)\\ -\cos(\theta_i)\end{bmatrix}\dot{\theta}_i$$

for $\dot{x}_i$ and $\dot{y}_i$, where $\theta_j = \theta_1 - \sum_{i=1}^{j-1}\phi_i$. Using this expression we focus on the sum of squares $\dot{x}_i^2 + \dot{y}_i^2$, as the term of T_i involving $\dot{\theta}$ is not affected by the lift action of $SE(2)$, which is described in Example 3.12. Recall that our group variable here is comprised of the components x_1, y_1, θ_1 and note that these variables, along with the tangent vector components corresponding to these variables, denote them v_{x_1}, v_{y_1}, and v_{θ_1}, are the only variables affected by the lift action. In fact, since θ_1, v_{x_1}, and v_{y_1} are the only components amongst these to appear in the expression for $\dot{x}_i$ and $\dot{y}_i$, we need only concern ourselves with how these variables are affected by the lift action. It is clear that $\theta_1 \uparrow \theta_1 + \alpha$ and that

$$\begin{bmatrix}\dot{x}_1\\ \dot{y}_1\end{bmatrix} \uparrow A(\alpha)\begin{bmatrix}\dot{x}_1\\ \dot{y}_1\end{bmatrix} \ , \tag{3.13}$$

where A is the 2×2 rotation matrix

$$A(\alpha) = \begin{bmatrix}\cos(\alpha) & -\sin(\alpha)\\ \sin(\alpha) & \cos(\alpha)\end{bmatrix} \ .$$

Continuing, we note that, for $1 \leq k \leq i$,

$$\sin(\theta_k) \uparrow \sin(\theta_k + \alpha) = \sin(\theta_k)\cos(\alpha) + \cos(\theta_k)\sin(\alpha) \ ;$$

$$\cos(\theta_k) \uparrow \cos(\theta_k + \alpha) = \cos(\theta_k)\cos(\alpha) - \sin(\theta_k)\sin(\alpha) \ .$$

These expressions lead to the following observation.

$$\begin{bmatrix} \sin(\theta_k) \\ -\cos(\theta_k) \end{bmatrix} \uparrow A(\alpha) \begin{bmatrix} \sin(\theta_k) \\ -\cos(\theta_k) \end{bmatrix} \ . \tag{3.14}$$

Putting together equations (3.13) and (3.14) we have

$$\begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} \uparrow A(\alpha) \begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} \ .$$

Thus, under the lift action,

$$\dot{x}_i^2 + \dot{y}_i^2 = \begin{bmatrix} \dot{x}_i & \dot{y}_i \end{bmatrix} \begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} \uparrow \begin{bmatrix} \dot{x}_i & \dot{y}_i \end{bmatrix} A(\alpha)^t A(\alpha) \begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} = \begin{bmatrix} \dot{x}_i & \dot{y}_i \end{bmatrix} \begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} = \dot{x}_i^2 + \dot{y}_i^2 \ .$$

This says that the body kinetic energy T_i is $SE(2)$ invariant and thus the Lagrangian for the entire serial-link system is invariant, as $L = T = \sum_{i=1}^{n} T_i$. □

There is a lot of machinery here and some more to come, but we are getting closer to actually discussing reduction. A hint at the reduction idea is the following. L is an energy functional and as seen with the development of the Euler-Lagrange equations, the dynamics of a system depend on this energy functional and the non-conservative generalized forces. If a Lagrangian is G invariant and the virtual work of the non-conservative forces are in some sense, to be described later, G invariant, then it stands to reason that the position and orientation coordinates of the system, or the group variables, have no real bearing on the validity of Hamilton's principle or the dynamics that result therefrom, as they can be arbitrarily changed, and in fact, set to zero without affecting the governing quantities . In these circumstances the energy associated with the system is due to its shape and its inertia, not its location. The

process of reduction is simply the realization of this observation through the resultant dynamical system described by a system of ODEs. With the next theorem we take a step toward this formal realization.

Theorem 3.2 (The Reduced Lagrangian). *Let G be a Lie group. Further suppose existence of the Lagrangian $L(q,\dot{q}) = \frac{1}{2}\dot{q}^t M(q)\dot{q} - P(q)$, where $q \in Q = G \times S$. We may rewrite L in terms of the decomposition of q. This provides us*

$$L(g,\dot{g},s,\dot{s}) = \frac{1}{2}\begin{bmatrix}\dot{g}^t & \dot{s}^t\end{bmatrix} M(g,s) \begin{bmatrix}\dot{g}\\ \dot{s}\end{bmatrix} - P(g,s)\ .$$

Then, if L is G invariant, the Lagrangian may be written

$$\begin{aligned} L(g,\dot{g},s,\dot{s}) &\doteq l(\xi,s,\dot{s}) && (3.15)\\ &= \frac{1}{2}\begin{bmatrix}\xi^t & \dot{s}^t\end{bmatrix}\begin{bmatrix} I(s) & IA(s)\\ AI(s) & m(s)\end{bmatrix}\begin{bmatrix}\xi\\ \dot{s}\end{bmatrix} - \tilde{P}(s)\ . && (3.16)\end{aligned}$$

Proof. The proof of this conjecture is not that involved. It is in fact a simple matter of using the invariance of the Lagrangian and labelling. To begin, write L in block format as follows

$$L(g,\dot{g},s,\dot{s}) = \frac{1}{2}\begin{bmatrix}\dot{g}^t & \dot{s}^t\end{bmatrix}\begin{bmatrix} M_{11}(g,s) & M_{12}(g,s)\\ M_{21}(g,s) & M_{22}(g,s)\end{bmatrix}\begin{bmatrix}\dot{g}\\ \dot{s}\end{bmatrix} - P(g,s)\ .$$

Then, since L is G invariant we have

$$\begin{aligned} L(g,\dot{g},s,\dot{s}) &= L(\Phi_{g^{-1}}(q), T_q\Phi_{g^{-1}}(v_q))\\ &= \frac{1}{2}\begin{bmatrix}([g]^{-1}\dot{g})^t & \dot{s}^t\end{bmatrix}\begin{bmatrix} M_{11}(g^{-1}g,s) & M_{12}(g^{-1}g,s)\\ M_{21}(g^{-1}g,s) & M_{22}(g^{-1}g,s)\end{bmatrix}\begin{bmatrix}[g]^{-1}\dot{g}\\ \dot{s}\end{bmatrix} - P(g^{-1}g,s)\\ &= \frac{1}{2}\begin{bmatrix}\xi^t & \dot{s}^t\end{bmatrix}\begin{bmatrix} M_{11}(e,s) & M_{12}(e,s)\\ M_{21}(e,s) & M_{22}(e,s)\end{bmatrix}\begin{bmatrix}\xi\\ \dot{s}\end{bmatrix} - P(e,s)\ . \end{aligned}$$

Making the definitions

$$I(s) \doteq M_{11}(e,s) \ , \qquad IA(s) \doteq M_{12}(e,s) \ ,$$
$$AI(s) \doteq M_{21}(e,s) \ , \qquad m(s) \doteq M_{22}(e,s) \ ,$$
$$\tilde{P}(s) \doteq P(e,s) \ , \qquad \xi \doteq [g]^{-1}\dot{g} \ ,$$

$l(\xi, s, \dot{s})$ is obtained and the proof is complete. □

Calculation of the reduced Lagrangian for a system is quite simple as the proof indicates. All that one has to do is evaluate the mass matrix M and the potential functional P at the identity element of the Lie group and swap $\dot{g}$ for ξ. The simplicity of this is now shown by carrying out the computations for the mass-spring and serial-link structures.

Example 3.14 (Mass-Spring System Reduced Lagrangian). *Note that in this example, M is constant and the potential does not depend on the group variable g. Therefore, all that is needed to obtain the reduced Lagrangian is replacement of $\dot{g}$ by ξ. So,*

$$l(\xi, s, \dot{s}) = \frac{1}{2}(m_1\xi^2 + m_2(\xi + \dot{s})^2) - \left(\frac{1}{2}ks^2 + \frac{1}{4}ka^2s^4\right) \ .$$

Example 3.15 (The Serial-Link Structure Reduced Lagrangian ($n = 2$)). *To acquire the reduced Lagrangian for $\mathcal{S}$ we shall identify the appropriate components of reduced mass-matrix by simply evaluating M of Example 3.4 at $\theta = 0$. Doing so provides*

$$I = \begin{bmatrix} 2\,m & 0 & -ml\sin(\phi_1) \\ 0 & 2\,m & -ml\,(1+\cos(\phi_1)) \\ -ml\sin(\phi_1) & -ml\,(1+\cos(\phi_1)) & 2/3\,ml^2\,(4+3\,\cos(\phi_1)) \end{bmatrix} \ ,$$

$$IA = \begin{bmatrix} ml\sin(\phi_1) \\ ml\cos(\phi_1) \\ -1/3\, ml^2\,(4 + 3\,\cos(\phi_1)) \end{bmatrix},$$

and $m = \frac{4}{3}ml^2$. *Of course,* $AI = IA^t$. *This is all the needed information, as there is no potential functional for this system.*

3.3.3 Introduction to Lie Algebras and the Adjoint Action: A Precursor to Reduction Calculations

One last tool is required before attention is turned to the reduction calculations. This tool is the idea of the Lie algebra $\eth$ associated with the Lie group G. First, the Lie algebra is defined.

Definition 3.7 (Lie Algebra). *Let* V *denote a vector space over the field* $\mathbb{F}$. *If there exists a vector valued product or bracket on this space say,* $[\cdot,\cdot] : V \times V \to V$, *that satisfies the following properties for all* $v_1\ , v_2\ , v_3 \in V$ *and for all* $c_1\ , c_2 \in \mathbb{F}$*:*

- *linearity:*

$$[c_1v_1 + c_2v_2, v_3] = c_1[v_1, v_3] + c_2[v_2, v_3]\ ;$$

- *skew-symmetry:*

$$[v_1, v_2] = -[v_2, v_1]\ ;$$

- *Jacobi identity or product rule:*

$$[v_1, [v_2, v_3]] = [[v_1, v_2], v_3] + [v_2, [v_1, v_3]]\ ,$$

then V *is said to be a Lie algebra under this product.*

Now the Lie algebra associated with the matrix Lie group $\tilde{G}$ is described. Consider a path $c : \mathbb{R} \to \tilde{G}$ such that $c(0) = I$, where I is the identity matrix and, of course,

the group identity. Then the value $\xi = c'(0)$ is defined to be an element of the Lie algebra associated with G. Here, the derivative of c is taken component-wise. Considering all such paths defines the entire vector space $\tilde{G}$. Using the commuter bracket $[A, B] = AB - BA$ for matrices one acquires the Lie algebra $\tilde{\eth}$. It should be clear that $\xi \in T_eG$ for those groups that have non-matrix counterparts such as the special Euclidean groups, orthogonal groups, etc. It is in fact the case that $\tilde{\eth} \simeq \eth \doteq T_eG$ as Lie-algebras, with the bracket on T_eG being

$$\begin{aligned}[\xi, \eta] &= T_g L_{g^{-1}}[T_e L_g \xi, T_e L_g \eta] \\ &= [g]^{-1}[[g]\xi, [g]\eta] \ ,\end{aligned}$$

for ξ and η in T_eG. The brackets $[\cdot, \cdot]$ on the right-hand side of this equation are the well-know Jacobi-Lie derivative for vector fields given by

$$[v, w]_i = \frac{\partial w_i}{\partial g} v - \frac{\partial v_i}{\partial g} w \ . \tag{3.17}$$

In coordinates, using the Einstein convention for summation over indices,

$$[\xi, \eta]_b = [g]^{-1}_{ba}[[g]\xi, [g]\eta]_a \ ,$$

where

$$[g]\xi = [g]_{(\cdot)d}\xi_d \qquad\qquad [g]\eta = [g]_{(\cdot)f}\eta_f \ .$$

We find that

$$\begin{aligned}[[g]\xi, [g]\eta]_a &= \frac{\partial([g]_{af}\eta_f)}{\partial g}([g]_{(\cdot)d}\xi_d) - \frac{\partial([g]_{ad}\xi_d)}{\partial g}([g]_{(\cdot)f}\eta_f) \\ &= \frac{\partial([g]_{af}\eta_f)}{\partial g_c}([g]_{cd}\xi_d) - \frac{\partial([g]_{ad}\xi_d)}{\partial g_c}([g]_{cf}\eta_f) \\ &= \left(\frac{\partial [g]_{af}}{\partial g_c}\eta_f\right)([g]_{cd}\xi_d) - \left(\frac{\partial [g]_{ad}}{\partial g_c}\xi_d\right)([g]_{cf}\eta_f) \ .\end{aligned}$$

Changing up the order of summation and making use of the commutative and distributive properties will provide the following expression for the Lie brackets of tangent vectors.

$$\begin{aligned}[\xi,\eta]_b &= [g]^{-1}_{ba}[[g]\xi,[g]\eta]_a \\ &= \eta_f\left([g]^{-1}_{ba}\left(\frac{\partial[g]_{af}}{\partial g_c}([g]_{cd}\xi_d)\right)\right) - \eta_f\left([g]^{-1}_{ba}\left(\left(\frac{\partial[g]_{ad}}{\partial g_c}[g]_{cf}\right)\xi_d\right)\right) \ .\end{aligned}$$

So, consider the adjoint action $ad_\xi(\cdot) \doteq [\xi,\cdot]$ on the Lie algebra. The above expression indicates how one may write the matrix associated with this operator:

$$[ad_\xi]_{bf} = [g]^{-1}_{ba}\left(\left(\frac{\partial[g]_{af}}{\partial g_c}[g]_{cd} - \frac{\partial[g]_{ad}}{\partial g_c}[g]_{cf}\right)\xi_d\right) \ . \tag{3.18}$$

All of this is made clear through examples related to the systems we have created.

Example 3.16 (The Lie Algebra se(1)). *Consider a path c through the identity of the matrix Lie group $SE(1)$, denoted $\tilde{G}$. Such a path looks like*

$$c(t) = \begin{bmatrix} 1 & g(t) \\ 0 & 1 \end{bmatrix} \ ,$$

where $g(0) = 0$. Then

$$\xi = c'(0) = \begin{bmatrix} 0 & g'(0) \\ 0 & 0 \end{bmatrix} \ .$$

Thus $\tilde{\eth}$ is comprised of elements of the form

$$\xi = \begin{bmatrix} 0 & \xi_1 \\ 0 & 0 \end{bmatrix} \ .$$

To show that every matrix of this form can be generated from some path in $\tilde{G}$, $g(t) = e^{\xi_1 t} - 1$ is selected. It is clear that $g(0) = 0$ and $g'(0) = \xi_1$. For ξ and η in $\tilde{\eth}$, it is easy to verify that the commuter bracket is the zero bracket, i.e. $[\xi,\eta] = 0$. Thus, the

Lie algebra is abelian.

$\tilde{\mathfrak{d}} \simeq T_0G = \mathfrak{d}$ *and the isomorphism is clearly* $\xi \to (\xi_1)_0$. *Also, in this case* $[g] = [g]^{-1} = 1$, *and thus equation* (3.18) *provides* $[ad_\xi] = 0$, *which is in congruence with fact that the algebra is abelian. This Lie algebra has a special denotation. It is labelled* $se(1)$.

Example 3.17 (The Lie Algebra se(2)). *Here attention is turned to the matrix version of the Lie group* $SE(2)$, *denoted* $\tilde{G}$. *As explained in Example 3.8 an element of* $\tilde{G}$ *has the form*

$$g = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & x \\ \sin(\theta) & \cos(\theta) & y \\ 0 & 0 & 1 \end{bmatrix} .$$

Thus, a parameterized path through the identity matrix may be provided by any smooth functions $x(t)$, $y(t)$, *and* $\theta(t)$, *such that* $x(0) = y(0) = \theta(0) = 0$. *Call this path* $c(t)$. *It follows that*

$$c'(t) = \begin{bmatrix} -\sin(\theta)\dot{\theta} & -\cos(\theta)\dot{\theta} & \dot{x} \\ \cos(\theta)\dot{\theta} & -\sin(\theta)\dot{\theta} & \dot{y} \\ 0 & 0 & 0 \end{bmatrix} .$$

Creating the labels $\xi_\theta = \dot{\theta}(0)$, $\xi_x = \dot{x}(0)$, *and* $\xi_y = \dot{y}(0)$, *it is seen that*

$$c'(0) = \begin{bmatrix} 0 & -\xi_\theta & \xi_x \\ \xi_\theta & 0 & \xi_y \\ 0 & 0 & 0 \end{bmatrix} .$$

These matrices make up the Lie algebra $\tilde{\mathfrak{d}}$ *associated with* $\tilde{G}$. *It is quite clear that* $(\xi_x, \xi_y, \xi_\theta)^t \in T_eG$, *where* G *is the Lie group* $SE(2)$ *and this is the manner through which the Lie-algebras for the differing forms of* $SE(2)$ *are identified with one another.*

It remains to show that given any element of T_eG*, say* $v_e = (\xi_x, \xi_y, \xi_\theta)^t$*, the matrix*

$$\begin{bmatrix} 0 & -\xi_\theta & \xi_x \\ \xi_\theta & 0 & \xi_y \\ 0 & 0 & 0 \end{bmatrix}$$

can be attained by differentiation of some path $c(t)$ *passing through the identity matrix evaluated at* 0*. It doesn't take much contriving to arrive at such a curve. The* 2×2 *skew-symmetric portion of this matrix may be obtained through the matrix exponential thereof, i.e.*

$$t \mapsto \exp\left(\begin{bmatrix} 0 & -\xi_\theta \\ \xi_\theta & 0 \end{bmatrix} t \right) = \begin{bmatrix} \cos(\xi_\theta t) & -\sin(\xi_\theta t) \\ \sin(\xi_\theta t) & \cos(\xi_\theta t) \end{bmatrix} \ .$$

The elements ξ_x *and* ξ_y *are respectively captured by the curves* $t \mapsto e^{\xi_x t} - 1$ *and* $t \mapsto e^{\xi_y t} - 1$*.*

Continuing then; let E_{ij} *be the unit matrix* $[\delta_{ik}\delta_{jl}]$*;* $k, l = 1, \ldots, 3$*. It is clear that a basis for* $\tilde{\tilde{\mathfrak{d}}}$ *is given by the skew-symmetric matrix* $f_3 = E_{21} - E_{12}$ *and the matrices* $f_1 = E_{13}$*,* $f_2 = E_{23}$*. By bracketing these matrices we acquire the structure of the entire Lie algebra due to the requisite linearity of the bracket. Doing this yields for us*

$$[f_1, f_2] = 0^{3\times 3} \ , \tag{3.19a}$$

$$[f_1, f_3] = -f_2 \ , \tag{3.19b}$$

$$[f_2, f_3] = f_1 \ . \tag{3.19c}$$

Now, let $\xi, \eta \in \tilde{\tilde{\mathfrak{d}}}$*. Then these vectors are expressed as* $\xi = \xi_x f_1 + \xi_y f_2 + \xi_\theta f_3$ *and* $\eta = \eta_x f_1 + \eta_y f_2 + \eta_\theta f_3$*. Making use of equations* (3.19) *repeatedly one arrives at the result* $[\xi, \eta] = (\xi_y \eta_\theta - \eta_y \xi_\theta) f_1 + (\xi_\theta \eta_x - \eta_\theta \xi_x) f_2$*. Therefore, in coordinates with respect*

to the basis, which is the identification with $\eth$*, one has*

$$[\xi, \eta] = \begin{bmatrix} \xi_y \eta_\theta - \eta_y \xi_\theta \\ \xi_\theta \eta_x - \eta_\theta \xi_x \\ 0 \end{bmatrix} .$$

It directly follows that

$$[ad_\xi] = \begin{bmatrix} 0 & -\xi_\theta & \xi_y \\ \xi_\theta & 0 & -\xi_x \\ 0 & 0 & 0 \end{bmatrix} . \tag{3.20}$$

Of course, this could be verified by appealing directly to equation (3.18)*. The approach here is shown, because quite often, mixing calculations in* G *and* $\tilde{G}$ *as well as* $\eth$ *and* $\tilde{\eth}$ *occurs at convenience. The Lie algebra discussed in this example has a special denotation. It is* $se(2)$*.*

3.3.4 Reduced Dynamics

Finally, we derive the reduced dynamics for the G invariant system. We begin with derivation of the so-called reduced Euler-Lagrange equations for the group variables. This simply means that we are going to rewrite the Euler-Lagrange equations in terms of the reduced Lagrangian. Recall or note the following vector relations:

$$\dot{g} = [g]\xi \ , \qquad\qquad \xi = [g]^{-1}\dot{g} \ ,$$

or coordinate relations

$$\dot{g}_b = [g]_{bc}\xi_c \ , \qquad\qquad \xi_b = [g]^{-1}_{bc}\dot{g}_c \ .$$

Using these relations and repeated use of the chain rule we may write down the pieces of the E-L equations.

$$\frac{\partial L}{\partial \dot{g}_a} = \frac{\partial l}{\partial \dot{g}_a} = \frac{\partial l}{\partial \xi_b}\frac{\partial \xi_b}{\partial \dot{g}_a} \ .$$

So that,

$$\begin{aligned}
\frac{d}{dt}\frac{\partial L}{\partial \dot{g}_a} &= \frac{d}{dt}\left(\frac{\partial l}{\partial \xi_b}\frac{\partial \xi_b}{\partial \dot{g}_a}\right) \\
&= \frac{\partial l}{\partial \xi_b}\frac{d}{dt}\left(\frac{\partial \xi_b}{\partial \dot{g}_a}\right) + \frac{d}{dt}\left(\frac{\partial l}{\partial \xi_b}\right)\frac{\partial \xi_b}{\partial \dot{g}_a} \\
&= \frac{\partial l}{\partial \xi_b}\frac{d}{dt}([g]^{-1}_{ba}) + \frac{d}{dt}\left(\frac{\partial l}{\partial \xi_b}\right)[g]^{-1}_{ba} \\
&= \frac{\partial l}{\partial \xi_b}\frac{\partial [g]^{-1}_{ba}}{\partial g_c}\dot{g}_c + \frac{d}{dt}\left(\frac{\partial l}{\partial \xi_b}\right)[g]^{-1}_{ba} \\
&= [g]^{-1}_{ba}\frac{d}{dt}\left(\frac{\partial l}{\partial \xi_b}\right) + \frac{\partial l}{\partial \xi_b}\frac{\partial [g]^{-1}_{ba}}{\partial g_c} g_{cd}\xi_d \ .
\end{aligned}$$

Also,

$$\begin{aligned}
\frac{\partial L}{\partial g_a} = \frac{\partial l}{\partial g_a} &= \frac{\partial l}{\partial \xi_b}\frac{\partial \xi_b}{\partial g_a} \\
&= \frac{\partial l}{\partial \xi_b}\frac{\partial ([g]^{-1}_{bc}\dot{g}_c)}{\partial g_a} \\
&= \frac{\partial l}{\partial \xi_b}\frac{\partial ([g]^{-1}_{bc})}{\partial g_a}\dot{g}_c \\
&= \frac{\partial l}{\partial \xi_b}\frac{\partial ([g]^{-1}_{bc})}{\partial g_a}[g]_{cd}\xi_d \ .
\end{aligned}$$

Subtracting the two terms now gives us

$$\begin{aligned}
\frac{d}{dt}\frac{\partial L}{\partial \dot{g}_a} - \frac{\partial L}{\partial g_a} &= \frac{d}{dt}\frac{\partial l}{\partial \dot{g}_a} - \frac{\partial l}{\partial g_a} \\
&= [g]^{-1}_{ba}\frac{d}{dt}\left(\frac{\partial l}{\partial \xi_b}\right) + \frac{\partial l}{\partial \xi_b}\frac{\partial [g]^{-1}_{ba}}{\partial g_c} g_{cd}\xi_d - \frac{\partial l}{\partial \xi_b}\frac{\partial ([g]^{-1}_{bc})}{\partial g_a}[g]_{cd}\xi_d \ .
\end{aligned}$$

Let us denote this quantity $(EL)_a$, as this is the Euler-Lagrange equation corresponding to the group variable g_a.

We know that $[g]^{-1}[g]$ is the identity matrix and as such, its entries have partials of value 0 with respect to any unknown. This fact is used to obtain the identity,

$$([g]^{-1}[g])_{eb} = [g]^{-1}_{ea}[g]_{ab} = \delta_{eb}.$$

So, taking the derivative of this expression provides for us

$$\begin{aligned} 0 &= \frac{\partial([g]^{-1}[g])_{eb}}{\partial g^c} \\ &= \frac{\partial([g]^{-1}_{ea}[g]_{ab})}{\partial g^c} \\ &= \frac{\partial([g]^{-1}_{ea})}{\partial g^c}[g]_{ab} + \frac{\partial([g]_{ab})}{\partial g^c}[g]^{-1}_{ea} \ . \end{aligned}$$

This implies that for any b, e, and c,

$$\frac{\partial([g]^{-1}_{ea})}{\partial g^c}[g]_{ab} = -\frac{\partial([g]_{ab})}{\partial g^c}[g]^{-1}_{ea} \ . \tag{3.21}$$

We make immediate use of this identity. Multiplying $(EL)_a$ by $[g]_{ae}$ and summing over the index a we obtain an expression for $((EL)[g])_e$. This expression is

$$\begin{aligned} \frac{d}{dt}\left(\frac{\partial l}{\partial \xi_b}\right)([g]^{-1}_{ba}[g]_{ae}) + \left(\frac{\partial l}{\partial \xi_b}\left(\frac{\partial [g]^{-1}_{ba}}{\partial g_c}(g_{cd}\xi_d)\right)\right)[g]_{ae} \\ - \left(\frac{\partial l}{\partial \xi_b}\left(\frac{\partial([g]^{-1}_{bc})}{\partial g_a}([g]_{cd}\xi_d)\right)\right)[g]_{ae} \ . \end{aligned}$$

Proceeding with the sum over b in the first term and using the above identity then yields for us,

$$\frac{d}{dt}\left(\frac{\partial l}{\partial \xi_e}\right) - \frac{\partial l}{\partial \xi_b}\left(\left(\frac{\partial [g]_{ae}}{\partial g_c}(g_{cd}\xi_d)\right)[g]^{-1}_{ba}\right) + \left(\frac{\partial l}{\partial \xi_b}\left(\left(\frac{\partial [g]_{cd}}{\partial g_a}\xi_d\right)[g]^{-1}_{bc}\right)\right)[g]_{ae} \ .$$

We further manipulate this expression by making use of commutative and distributive laws, changing summation orders, and switching the dummy indices a and c in the

third term. The result is

$$\begin{aligned}((EL)[g])_e &= \frac{d}{dt}\left(\frac{\partial l}{\partial \xi_e}\right) - \frac{\partial l}{\partial \xi_b}\left([g]^{-1}_{ba}\left(\left(\frac{\partial [g]_{ae}}{\partial g_c}[g]_{cd}\right)\xi_d\right)\right) \\ &\qquad + \frac{\partial l}{\partial \xi_b}\left([g]^{-1}_{ba}\left(\left(\frac{\partial [g]_{ad}}{\partial g_c}[g]_{ce}\right)\xi_d\right)\right) \\ &= \frac{d}{dt}\left(\frac{\partial l}{\partial \xi_e}\right) - \frac{\partial l}{\partial \xi_b}\left([g]^{-1}_{ba}\left(\left(\frac{\partial [g]_{ae}}{\partial g_c}[g]_{cd} - \frac{\partial [g]_{ad}}{\partial g_c}[g]_{ce}\right)\xi_d\right)\right) \\ &= \frac{d}{dt}\left(\frac{\partial l}{\partial \xi_e}\right) - \frac{\partial l}{\partial \xi_b}[ad_\xi]_{be} \\ &= \left(\frac{d}{dt}\left(\frac{\partial l}{\partial \xi}\right)\right)_e - \left(\frac{\partial l}{\partial \xi}[ad_\xi]\right)_e \ ,\end{aligned}$$

where we have used equation (3.18). From this statement we conclude that the Euler-Lagrange equations for the group variables expressed in terms of the reduced Lagrangian are

$$(EL)[g] = \frac{d}{dt}\left(\frac{\partial l}{\partial \xi}\right) - \frac{\partial l}{\partial \xi}[ad_\xi] \ .$$

However, $EL = \tau_g$, where τ_g is the row vector of non-conservative generalized forces acting in the group directions. Thus,

$$\boxed{\frac{d}{dt}\left(\frac{\partial l}{\partial \xi}\right) - \frac{\partial l}{\partial \xi}[ad_\xi] = \tau_g[g] \ .} \tag{3.22}$$

This equation along with the Euler-Lagrange conditions for optimality allows one to write the dynamics for the shape variable s_k in the following form:

$$\frac{d}{dt}\left(\frac{\partial l}{\partial \dot{s}_k}\right) - \frac{\partial l}{\partial s_k} = \left[\frac{d}{dt}(\frac{\partial l}{\partial \xi}) - \frac{\partial l}{\partial \xi}[ad_\xi] - \tau_g[g]\right] A_k + \tau_k \ , \tag{3.23}$$

where A_k is the kth column of the submatrix $A(s)$ of the reduced Lagrangian and τ_k is the kth component of the vector of non-conservative generalized forces acting in the direction of s_k. Going back to Theorem 3.2, it is readily discovered that

$\frac{\partial l}{\partial \xi}^t = I\xi + IA\dot{s}$. Let $p^t \doteq \frac{\partial l}{\partial \xi}$. Then

$$\boxed{\xi = I^{-1}p - A\dot{s}\ .} \tag{3.24}$$

This implies that $l(\xi, s, \dot{s}) = l_c(p, s, \dot{s}) \doteq l(I^{-1}p - A\dot{s}, s, \dot{s})$. This is known as the *constrained Lagrangian*. We now compute the left-hand side of equation (3.23) for the shape variables in terms of l_c.

$$\begin{aligned}
\frac{\partial l_c}{\partial \dot{s}_k} &= \frac{\partial l}{\partial \dot{s}_k} + \frac{\partial l}{\partial \xi_b}\frac{\partial \xi_b}{\partial \dot{s}_k} \\
&= \frac{\partial l}{\partial \dot{s}_k} - \frac{\partial l}{\partial \xi_b}A_{bk} \\
&= \frac{\partial l}{\partial \dot{s}_k} - p^t A_k\ .
\end{aligned}$$

Taking the time derivative of this expression provides us

$$\begin{aligned}
\frac{d}{dt}\frac{\partial l_c}{\partial \dot{s}_k} &= \frac{d}{dt}\frac{\partial l}{\partial \dot{s}_k} - \frac{d}{dt}(p^t A_k) \\
&= \frac{d}{dt}\frac{\partial l}{\partial \dot{s}_k} - \left[\dot{p}^t A_k + p^t \frac{d}{dt}(A_k)\right]\ .
\end{aligned}$$

Further,

$$\begin{aligned}
\frac{\partial l_c}{\partial s_k} &= \frac{\partial l}{\partial s_k} + \frac{\partial l}{\partial \xi_b}\frac{\partial \xi_b}{\partial s_k} \\
&= \frac{\partial l}{\partial \dot{s}_k} + p^t\left(-\frac{\partial A_{b(\cdot)}}{\partial s_k}\dot{s} + \frac{\partial I^{-1}_{b(\cdot)}}{\partial s_k}p\right) \\
&= \frac{\partial l}{\partial \dot{s}_k} - p^t\frac{\partial A}{\partial s_k}\dot{s} + p^t\frac{\partial I^{-1}}{\partial s_k}p\ ,
\end{aligned}$$

where the partial derivatives of vectors and matrices here are taken component-wise.

Putting it all together we have

$$\begin{aligned}
\frac{d}{dt}\frac{\partial l}{\partial \dot{s}_k} - \frac{\partial l}{\partial s_k} &= \frac{d}{dt}\frac{\partial l_c}{\partial \dot{s}_k} - \frac{\partial l_c}{\partial s_k} \\
&= \left[\frac{d}{dt}\frac{\partial l}{\partial \dot{s}_k} - \frac{\partial l}{\partial s_k}\right] - \dot{p}^t A_k - p^t \frac{d}{dt}(A_k) + p^t \frac{\partial A}{\partial s_k}\dot{s} \\
&\qquad - p^t \frac{\partial I^{-1}}{\partial s_k} p + \tau_k \\
&= \left[\dot{p}^t - p^t[ad_\xi] - \tau_g[g]\right] A_k - \dot{p}^t A_k - p^t \frac{d}{dt}(A_k) + p^t \frac{\partial A}{\partial s_k}\dot{s} \\
&\qquad - p^t \frac{\partial I^{-1}}{\partial s_k} p + \tau_k \\
&= -p^t[ad_\xi]A_k - p^t\left(\frac{\partial A_i}{\partial s}\dot{s} - \frac{\partial A}{\partial s_k}\dot{s} + \frac{\partial I^{-1}}{\partial s_k} p\right) \\
&\qquad + \tau_k - \tau_g[g]A_k \ .
\end{aligned} \tag{3.25}$$

We shall now calculate the $\frac{d}{dt}(\partial l_c/\partial \dot{s}_k) - (\partial l_c/\partial s_k)$ directly. Carrying out the block matrix multiplication implied in equation (3.15) we obtain the equation

$$l(\xi, s, \dot{s}) = \frac{1}{2}[\xi^t I \xi + \xi^t I A \dot{s} + \dot{s}^t A^t I \xi + \dot{s}^t m \dot{s}] - \tilde{P}(s) \ .$$

So,

$$l_c(p, s, \dot{s}) = l\big|_{\xi = I^{-1}p - A\dot{s}} = \frac{1}{2}\dot{s}(m - A^t I A)\dot{s} + \frac{1}{2}p^t I^{-1} p - \tilde{P}(s) \ .$$

This expression is rewritten as

$$l_c = \frac{1}{2}\dot{s}\tilde{M}\dot{s} + \frac{1}{2}p^t I^{-1} p - \tilde{P}(s) \ , \tag{3.26}$$

where $\tilde{M} = m - A^t I A$. We calculate term by term through equation (3.26).

$$\begin{aligned}\frac{\partial}{\partial \dot{s}_k}\left(\frac{1}{2}\dot{s}^t\tilde{M}\dot{s}\right) &= \frac{1}{2}\frac{\partial}{\partial \dot{s}_k}(\dot{s}_{k_1}\tilde{M}_{k_1k_2}\dot{s}_{k_2}) \\ &= \frac{1}{2}\tilde{M}_{k_1k_2}(\dot{s}_{k_1}\delta_{k_2k} + \dot{s}_{k_2}\delta_{k_1k}) \\ &= \frac{1}{2}\tilde{M}_{k_1k}\dot{s}_{k_1} + \frac{1}{2}\tilde{M}_{kk_2}\dot{s}_{k_2} \\ &= \tilde{M}_{kk_2}\dot{s}_{k_2}\ ,\end{aligned}$$

using the fact that $\tilde{M}$ is symmetric. Thus,

$$\begin{aligned}\frac{d}{dt}\frac{\partial}{\partial \dot{s}_k}\left(\frac{1}{2}\dot{s}^t\tilde{M}\dot{s}\right) &= \tilde{M}_{kk_2}\ddot{s}_{k_2} + \frac{d}{dt}(\tilde{M}_{kk_2})\dot{s}_{k_2} \\ &= \tilde{M}_{kk_2}\ddot{s}_{k_2} + \left(\frac{\partial \tilde{M}_{kk_2}}{\partial s_{k_1}}\dot{s}_{k_1}\right)\dot{s}_{k_2} \\ &= \tilde{M}_{kk_2}\ddot{s}_{k_2} + \frac{1}{2}\left(\left(\frac{\partial \tilde{M}_{kk_2}}{\partial s_{k_1}}\dot{s}_{k_1}\right)\dot{s}_{k_2} + \left(\frac{\partial \tilde{M}_{kk_1}}{\partial s_{k_2}}\dot{s}_{k_2}\right)\dot{s}_{k_1}\right)\ .\end{aligned}$$

We now calculate the other partial that is needed.

$$\begin{aligned}\frac{\partial}{\partial s_k}\left(\frac{1}{2}\dot{s}^t\tilde{M}\dot{s}\right) &= \frac{1}{2}\frac{\partial}{\partial s_k}(\dot{s}_{k_1}\tilde{M}_{k_1k_2}\dot{s}_{k_2}) \\ &= \frac{1}{2}\frac{\partial \tilde{M}_{k_1k_2}}{\partial s_k}\dot{s}_{k_1}\dot{s}_{k_2}\ .\end{aligned}$$

It follows from this equation that

$$\frac{d}{dt}\frac{\partial}{\partial \dot{s}_k}\left(\frac{1}{2}\dot{s}^t\tilde{M}\dot{s}\right) - \frac{\partial}{\partial s_k}\left(\frac{1}{2}\dot{s}^t\tilde{M}\dot{s}\right) = \tilde{M}_{kk_2}\ddot{s}_{k_2} + \frac{1}{2}C_{kk_2k_1}\dot{s}_{k_2}\dot{s}_{k_1}\ , \tag{3.27}$$

where

$$C_{kk_2k_1} = \left(\frac{\partial \tilde{M}_{kk_2}}{\partial s_{k_1}} + \frac{\partial \tilde{M}_{kk_1}}{\partial s_{k_2}} - \frac{\partial \tilde{M}_{k_1k_2}}{\partial s_k}\right)\ .$$

Moving on to the second term of equation (3.26) we have

$$\frac{\partial}{\partial \dot{s}_k}(\frac{1}{2}p^t I^{-1}p) - \frac{\partial}{\partial s_k}(\frac{1}{2}p^t I^{-1}p) = -\frac{1}{2}p^t(\frac{\partial I^{-1}}{\partial s_k})p \ . \tag{3.28}$$

Finally, for the third term of equation (3.26), we obtain

$$\frac{\partial}{\partial \dot{s}_k}(-\tilde{P}) - \frac{\partial}{\partial s_k}(-\tilde{P}) = \frac{\partial}{\partial s_k}(\tilde{P}) \ . \tag{3.29}$$

Therefore, by equations (3.27), (3.28), and (3.29),

$$\frac{d}{dt}\left(\frac{\partial l_c}{\partial \dot{s}_k}\right) - \frac{\partial l_c}{\partial s_k} = \tilde{M}_{kk_2}\ddot{s}_{k_2} + \frac{1}{2}C_{kk_2k_1}\dot{s}_{k_2}\dot{s}_{k_1} - \frac{1}{2}p^t(\frac{\partial I^{-1}}{\partial s_k})p + \frac{\partial}{\partial s_k}(V) \ , \tag{3.30}$$

and we have the equation

$$\begin{aligned} -p^t[ad_\xi]A_k - p^t(\frac{\partial A_k}{\partial s}\dot{s} - \frac{\partial A}{\partial s_k}\dot{s} + \frac{\partial I^{-1}}{\partial s_k}p) + \tau_k - \tau_g[g]A_k \\ = \tilde{M}_{kk_2}\ddot{s}_{k_2} + \frac{1}{2}C_{kk_2k_1}\dot{s}_{k_2}\dot{s}_{k_1} - \frac{1}{2}p^t(\frac{\partial I^{-1}}{\partial s_k})p + \frac{\partial}{\partial s_k}(V) \ , \end{aligned} \tag{3.31}$$

resulting from equations (3.25) and (3.30). Let $N = p^t[ad_\xi]A_k + p^t(\frac{\partial A_k}{\partial s}\dot{s} - \frac{\partial A}{\partial s_k}\dot{s})$. Then this equation becomes

$$\boxed{\tilde{M}_{kk_2}\ddot{s}_{k_2} + \frac{1}{2}C_{kk_2k_1}\dot{s}_{k_2}\dot{s}_{k_1} + N + \frac{1}{2}p^t(\frac{\partial I^{-1}}{\partial s_k})p + \frac{\partial V}{\partial s_k} = \tau_k - \tau_g[g]A_k \ .} \tag{3.32}$$

At this point another important concept must be introduced to finalize the discussion of reduced dynamics, that of a force that is invariant with respect to a Lie group.

Definition 3.8 (G-Invariant Force). *Let G be a Lie group and suppose that the configuration manifold of a given system is $Q = G \times S$, the product of group and shape spaces. Let $\tau(q, \dot{q})$ be a vector of applied forces. If $\tau(\Phi_h(q), T_q\Phi_h(v_q)) = \tau(q, v_q)[h]^{-1}$ for all $(q, v_q) \in TQ$ and $h \in G$, then τ is said to be G* invariant.

The need for this definition is shown immediately. Suppose that the forces acting in the group directions involved in equations (3.24), (3.22), (3.32), namely τ_g, are G invariant. Then $\tau_g(g, s, \dot{g}, \dot{s})[g] = \tau_g(g^{-1}g, s, \xi, \dot{s}) \doteq \tilde{\tau}^t(\xi, s, \dot{s})$. Examining the effect of this invariance on the equations we arrive at the system

$$\xi = I^{-1}p - A\dot{s} \ , \tag{3.33}$$

$$\dot{p} - [ad_{I^{-1}p - A\dot{s}}]^t p = \tilde{\tau} \ , \tag{3.34}$$

$$\tilde{M}_{kk_2}\ddot{s}_{k_2} + \frac{1}{2}C_{kk_2k_1}\dot{s}_{k_2}\dot{s}_{k_1} + N + \frac{1}{2}p^t(\frac{\partial I^{-1}}{\partial s_k})p + \frac{\partial V}{\partial s_k} = \tau_k - \tilde{\tau}^t A_k \ . \tag{3.35}$$

This system of ODEs is the so-called reduced dynamics of a G invariant system. Note that the form of these equations indicate that G invaraince is a rather desirable property for a force acting in the group direction to have. For, in this case, a nice partial decoupling has been established and under some additional circumstances, namely, the G invaraince of the τ_k, the group variables are no longer present.

For both of the systems we are examining, friction is the only force that acts in the group directions. Both the control and friction forces act in the shape directions. So, upon a proper introduction of friction, we shall seek to establish that it indeed has this property under the correct assumptions and thus the resultant system dynamics will have the above form. However, for the time being, we shall assume that this is the case and proceed to illustrate the results of reduction in the presence of G invariance through the examples.

Example 3.18 (The Mass-Spring Dynamics). *By taking care of calculations as we have went along, it takes very little calculation effort to now write down the system dynamics. Firstly, note that* $I = m_1 + m_2$, $m = m_2$, *and that* $IA = m_2$. *Therefore, it is easily seen that* $I^{-1} = \frac{1}{m_1+m_2}$, $A = I^{-1}IA = \frac{m_2}{m_1+m_2}$, *and* $\tilde{M} = \frac{m_1 m_2}{m_1+m_2}$. *From these calculations alone we may proceed.*

Recall Equation (3.24), $\xi = I^{-1}p - A\dot{s}$. *Thus the equation*

$$\xi = \frac{1}{m_1 + m_2}p - \frac{m_2}{m_1 + m_2}\dot{s} \ . \tag{3.36}$$

Further, recall from Example 3.16 that $[ad_\xi] = 0$. *Thus equation* (3.22) *leads us to the equation*

$$\dot{p} = \tau_g \ .$$

Here, as will be discussed in Chapter 4, τ_g *is a dry friction expression of the form* $F(\xi, \dot{s})$. *Upon substitution of equation* (3.36) *we have*

$$\dot{p} = \tilde{F}(p, \dot{s}) \ . \tag{3.37}$$

All that remains is the equation for the dynamics of the shape variable s. *Referring to Equation* (3.32) *and referring to the calculations leading into the present example, we have that* $N = 0$, $\frac{\partial I^{-1}}{\partial s} = 0$, *and* $C_{ijk} = 0$. *Also, from an earlier example* $\frac{\partial \tilde{P}}{\partial s} = ks(1 + a^2 s^2)$. *Thus the equation*

$$\tilde{M}\ddot{s} + ks(1 + a^2 s^2) + \frac{m_2}{m_1 + m_2}F(\xi, \dot{s}) = u + F_v(\dot{s}) \ , \tag{3.38}$$

where $\tau_k \doteq u + F_l(\dot{s})$ *is the controller plus the linear viscous friction term due to the dashpot effect on the system. Equations* (3.36), (3.37), *and* (3.38), *comprise the reduced mass-spring system dynamics. Note that these equations are partially decoupled, for* ξ *drives* s, s *in turn drives* p, *and* p *drives* ξ. *The system is reduced in the sense that* g *is not needed at all. That is, in fact, the entire point. Due to the fact that the system Lagrangian is* G *invariant, or that the energy does not depend on* g, *neither do the resulting dynamics.*

Example 3.19 (The Reduced Dynamics for the Serial-Link Structure). *As above, we shall begin with equation* (3.24). *We simply need to identify the matrices* I^{-1} *and* A. *We know* I *and* IA *from Example 3.15 and so calculation is taken up*

from there. According to Maple, $I^{-1} = \frac{1}{\det(I)} W$, *where* W *is the symmetric matrix with entries*

$$
\begin{aligned}
w_{11} &= -1/3\, m^2 l^2 \left(-13 - 6\, \cos(\phi_1) + 3\, (\cos(\phi_1))^2\right) \ , \\
w_{12} &= m^2 l^2 \left(1 + \cos(\phi_1)\right) \sin(\phi_1) \ , \\
w_{13} &= 2\, m^2 l \sin(\phi_1) \ , \\
w_{22} &= 1/3\, m^2 l^2 \left(13 + 12\, \cos(\phi_1) + 3\, (\cos(\phi_1))^2\right) \ , \\
w_{23} &= 2\, m^2 l \left(1 + \cos(\phi_1)\right) \ , \\
w_{33} &= 4\, m^2 \ ,
\end{aligned}
$$

and

$$
det(I) = \frac{26}{3}\, m^3 l^2 + 4\, m^3 l^2 \cos(\phi_1) - 2\, m^3 l^2 \left(\cos(\phi_1)\right)^2 - 2\, m^3 l^2 \left(\sin(\phi_1)\right)^2 \ .
$$

Also by the magic of Maple *we obtain*

$$
A = \frac{1}{\det(I)} \begin{bmatrix} 1/3\, \left(5 + 3\, \cos(\phi_1)\right) \sin(\phi_1) l^3 m^3 \\ 1/3\, \left(-5 + 3\, (\cos(\phi_1))^2 + 2\, \cos(\phi_1)\right) l^3 m^3 \\ -2/3\, \left(5 + 3\, \cos(\phi_1)\right) m^3 l^2 \end{bmatrix} \ .
$$

Thus the first of the reduced equations is

$$
\xi = I^{-1} p - A \dot{r} \ . \tag{3.39}
$$

Moving on, recall from equation (3.20) *that*

$$
[ad_\xi] = \begin{bmatrix} 0 & -\xi_\theta & \xi_y \\ \xi_\theta & 0 & -\xi_x \\ 0 & 0 & 0 \end{bmatrix} \ .
$$

and from equation (3.39) *we have an expression for* ξ*. Thus we may write down the second reduced equation as*

$$\dot{p} = [ad_\xi]^t p + F(p, \dot{s}) \ ,$$

where it has been presupposed that $[g]^{-1}\tau_g^t = F(p, \dot{s})$ *for some appropriately chosen* F*. The fact that this is the case will be shown in Chapter 4.*

Finally, the third equation of the reduced dynamics should be stated. Clearly this is a difficult calculation to carry out. It could be done in Maple*, however there is another way to avoid dealing with this equation that will be discussed in Chapter 4.*

This example is provided to illustrate the fact that the reduction calculations are not simple for the serial-link structure, even for $n = 2$*, when performed algebraically. In fact, for* $n = 3$ *this is quite intractable. However, from a numerical computation point of view, there is not a lot to do.* Matlab*, for instance, can readily perform all the requisite numerical calculations with ease. Thus, when dealing with systems of much complexity, numerics are the way to go for performing the reduction and subsequent analysis and simulation.*

Chapter 4

Non-conservative Forces

4.1 Introduction

In the previous chapter we developed reduced mathematical models for the dynamics of the mass-spring and serial-link systems. However, these models are incomplete. Although we appropriately related the kinematics of these systems to the generalized non-conservative forces that act on them, we did not provide specific mathematical expressions or models for those forces. In this chapter we complete the system models by describing friction and control forces, and how they are modelled, in detail.

As in the case of the previous chapter, this chapter is also divided into two primary components. The first of these shall deal with the topic of friction and the second with control. The evolution of the friction component is as follows. We open the topic of friction with a discussion of classical dry-friction laws. This deliberation includes some details of the origins of these laws, the appropriateness of their use, and how they must be modified in order to use them in the context of Lagrangian dynamics. These alterations are then demonstrated by using them to partially complete the modeling of the mass-spring system. Attention is then turned to the frictional properties of the snake skin as currently understood. A general method for calculation of the cumulative frictional force acting on a link of the serial-link structure is developed.

Upon doing so, the procedure is used to generate friction models for the bodies comprising the serial-link structure that are consistent with the frictional nature of actual snake skins. Finally, we generalize the resultant frictional forces and show that they are appropriately G invariant. Thus, they may be used along with the control forces to complete the model for the serial-link structure.

The exposition on the control forces is much more straight forward and proceeds as is now indicated. To begin we develop a model that relates the tension of muscle-like actuators running parallel to the axial structure of the serial-link system to the torque that these tensions produce about the joints of adjacent links. We then explain our choice to assume that these torques may take on any values we care to specify and upon doing so, take these torques directly as the control forces for the system. Next, the technique of feedback linearization is used to demonstrate that for our systems, the shape s is completely controllable. We move on to explain how the complexity of the systems' dynamics may be reduced significantly by making use of this controllability and what advantage this holds in gaining insight into the locomotion of systems. This advantage is illustrated through use of the mass-spring system and at this juncture the study of this system is halted, narrowing the focus to the system of true interest, namely $\mathcal{S}$.

4.2 Part I: Friction Forces

4.2.1 Classical Dry Friction Laws

We begin the discussion of friction with a disclaimer. The subject of tribology, and in particular, surface friction, is rather intractable in the sense that, unlike other physical phenomena, there are no universal friction laws for media that are not continuous. Thus, for scenarios such as the friction experienced between the surfaces of two bodies in tangential contact, frictional phenomena must be handled case by case. There are, however, a few empirical observations that serve as guidelines for dry/rubbing surface

friction, like that experienced between two rigid bodies, that purportedly hold up sufficiently well when applied to a variety of engineering applications. By well, it is meant that they get the job done. These observations are as follows.

1. The frictional force F experienced between two surfaces in tangential contact is proportional to their normal load N. Labelling the constant of proportionality μ we have
$$F = \mu N \ . \tag{4.1}$$

2. The frictional force F experienced between two surfaces in tangential contact is independent of the apparent area of contact. This says that if one considers two blocks of the same material and total mass, then the cumulative friction they experience will be the same despite differing physical dimensions.

These two observations were first stated by Leonardo da Vinci, who was apparently the first to think about friction in a scientific way, designing experiments that are still in use today for measuring the phenomena. The French scientist Guillaume Amontons rediscovered these principles and they often carry his namesake. Yet another name associated with these laws is that of the French engineer, Charles Coulomb. Coulomb also reconfirmed da Vinci's observations and indeed added a third stating that friction experienced by an object in motion is independent of the sliding velocity. This would seem to be a redundant statement given the first principle, but yet it is often included in formal discussions on the topic. Most likely, Coulomb's name has come to be associated with the friction principles due to the definite contribution that he made in including something of an explanation for the observations. Coulomb hypothesized that the friction phenomena was due to asperities in the surfaces that had settled into one another like the pieces of a jigsaw puzzle. Taking this point of view the following derivation of the equation $F = \mu N$ was made.

Consider two horizontal surfaces S_1 and S_2 associated respectively with the bodies B_1 and B_2. Suppose that these surfaces are in contact with one another under the

pressure $p_1 = P/a$ due to the weight of body B_1 . Here, P represents the weight of body B_1 and a the area of contact. Suppose that surface S_1 slides over surface S_2, which is at rest. Further, suppose that the asperities of the surfaces are similarly shaped and are distributed evenly with mean distance of λ between them. So, these asperities interlock as if they were fitted. It must take a horizontal pressure to promote one asperity of S_1 over an asperity of S_2. Let us label that pressure p_2. Since the period of asperity is λ we take λ^2 to be the area taken up by one asperity. Thus the horizontal force required for promotion of S_1 is $p_2\lambda^2$. It is now clear that the average work done by this force is $p_2\lambda^3$. The potential obtained due to the incline of an asperity is $\Delta U = p_1\lambda^2 h$, where h is the average height of the asperity. As the asperity of S_1 slides down the other side of the asperity of S_2, this potential energy is *assumed* lost in heat. Since the total work W is known to be $-\Delta U$, we have that $p_2 = -(\frac{h}{\lambda})p_1$. Thus the magnitude of the frictional force per unit area is $f = (\frac{h}{\lambda})p_1$. This provides us with the series of equations

$$\frac{f}{p_1} = \frac{f\lambda^2}{p_1\lambda^2} = \frac{F}{N} = \frac{h}{\lambda} = \mu \ ,$$

where F and N are the frictional and normal force magnitudes, respectively. This says that the ratio of the magnitude of the frictional force to the magnitude of the normal force is constant, which was da Vinci's first observation.

Euler would later add the observation that under this model with the assumption of triangular asperities, $\mu = \tan(\gamma)$, where γ is the slope of a triangle edge. In fact, Euler is responsible for the use of the symbol μ for the so-called *coefficient of friction*.

To see that this "law" is fragile we merely need to pick at any of the assumptions used to arrive at it. It is clear that none of them would hold outside of an extremely controlled set of circumstances. Although many researchers have developed models that attempt to remedy one of more of the faulty presumptions, all fall short of universal appeal. Some try to satisfy inconsistencies with the addition of elastic or visco-elastic deformation of the surface asperities. Others have disregarded this

Coulomb's mode of thinking completely and have suggested the idea of adhesion due to chemical or electro-magnetic forces as a basis for developing a satisfactory theory of dry friction. Indeed, strong adhesion has been observed for materials with very clean surfaces and it is true that one may also arrive at "Coulomb's" law based on the adhesion idea. However, like the surface roughness concept, there are counterexamples to its universality. In recent times, tribologists have come to accept that all of these ideas are correct and that it is a very complex set of surface interactions that collectively constitute the phenomena of dry friction. Models have been proposed that include the effects of adhesion, deformation, surface geometry, and surface asperity statistical distribution. However, as stated before, the observations made by da Vinci are simple and purportedly hold at some level over a variety of working situations. Therefore, we adopt this observation as our kinetic dry friction "law".

Now consider the case where two solid bodies are in contact with one another but not in relative motion. A frictional reaction force will still be experienced by the bodies involved, so long as one of the bodies is subject to a force acting tangent its surface. This force is known as static friction. Obviously, since the bodies are static, this friction force must be equal and opposite to the applied tangential force. This implies that this force, unlike the kinetic friction force, is not constant.

It then follows that dry friction cannot be written as a function of the velocities of the bodies used to create it. Being more specific, consider a body lying on a fixed substratum, either resting or translating with velocity $\dot{x}$. The friction experienced by this body is multi-valued at $\dot{x} = 0$ and piecewise constant otherwise. It is often assumed is that the frictional force vanishes at $\dot{x} = 0$. The harm in doing so is probably situationally dependent and we will discuss this further as the discussion develops. The good that comes out of this is that we then have a function to work with, albeit discontinuous. So, we will adopt this convention. Call the resultant function F_d. This function is illustrated in Figure 4.1 and is described mathematically

by the expression

$$F_d(\dot{x}) = \begin{cases} -\mu_f N & \text{if } \dot{x} > 0; \\ 0 & \text{if } \dot{x} = 0; \\ -\mu_b N & \text{if } \dot{x} < 0, \end{cases}$$

where N is the normal load. For us, $N = mg$, where m is the mass of the sliding

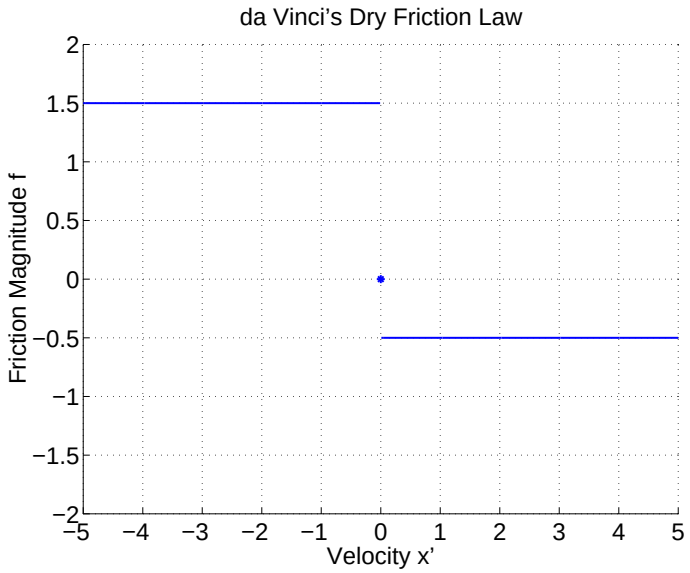


Figure 4.1: da Vinci's dry friction model F_d.

body and g is the acceleration due to gravity. Note that we have made this expression directionally dependent. That is, we apply da Vinci's law to each sliding direction independently. We have done this so that we may have an expression for the dry friction of a body whose surface has an orientation or a ratcheting effect.

Also of interest to us is the use of a simple linear expression to represent dry friction. This expression is $F_l = -\lambda\dot{x}$. Clearly this friction is not like that observed for two dry solid bodies engaged in tangential interactions. However, it provides an expression that is continuous, even in its directionally dependent form. Thus, it is very simple to work with mathematically.

We now turn attention to how to incorporate these friction expressions into the models for the mass-spring and serial-link structure systems.

Example 4.1 (Frictional Forces Acting on the Mass-Spring Structure). *As stated in the system description of Chapter 2 we assume that the large mass B_1 experiences dry friction between itself and the line on which it positioned. Due to its discontinuity, the use of the piece-wise constant dry friction function $F_d(\dot{x})$ does not allow for extrapolation of the Euler-Lagrange equations and doesn't facilitate use of a numerical integrator in the case of this system. This is not acceptable, as these tools will be used, respectively, to produce and solve the system dynamics. Hence, we seek to avoid the discontinuous nature of da Vinci's law. We approximate F_d with*

$$\tilde{F}_d(\dot{x}) = \begin{cases} -\mu_f N(\frac{2}{\pi}) \arctan(c\dot{x}) & \text{if } \dot{x} \geq 0; \\ -\mu_b N(\frac{2}{\pi}) \arctan(c\dot{x}) & \text{if } \dot{x} < 0. \end{cases}$$

This function is illustrated in Figure 4.2 along with F_d. Note that $\tilde{F}_d$ has asymptotes $f_d = -\mu_f N$ and $f_d = -\mu_b N$. Meaning that $\tilde{F}_d \to F_d$ for $\|\dot{x}\|$ large. The parameter c determines the rate of this convergence. Also, $\tilde{F}_d(0) = F_d(0) = 0$. Finally, $\tilde{F}_d$ is continuous, which solves the problems presented by F_d. Is this a legitimate approximation? We do not claim that it is. However, for lack of something better, we will use it anyway. Thus, the friction experienced by body B_1 is given by $\tau_g \doteq \tilde{F}_d(\dot{x}_1) = \tilde{F}_d(\dot{g})$ with appropriate parameters.

Also included in the description of the mass-spring system was a dash-pot effect on B_2. Dash-pot effects are often described by a linear friction expression. We adopt this description, claiming that the friction experienced by body B_2 is given by $F_l(\dot{s}) = -\lambda\dot{s}$.

These friction descriptions are in terms of the generalized coordinate vector $q^t = (g, s)$ selected for this system, and thus there is no need to generalize these forces.

Make note that τ_g is trivially a G invariant force. This is because τ_g is a function of $\dot{q}$ alone and the lift action acts as the identity on the tangent space of $SE(1)$. Thus we

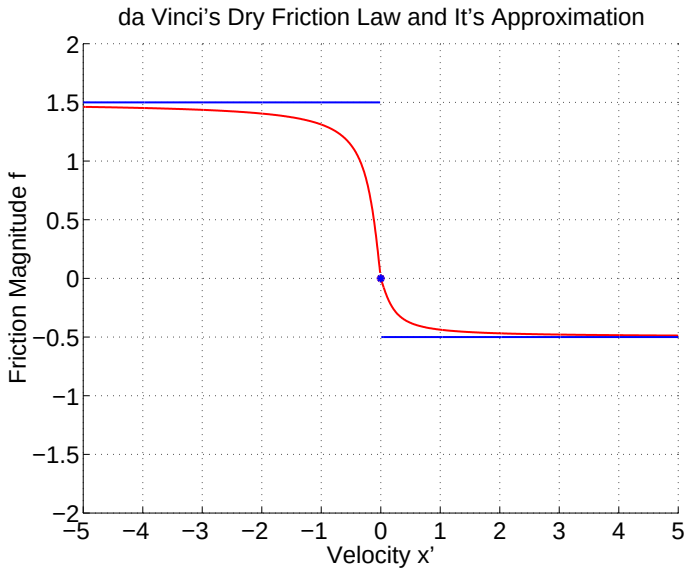


Figure 4.2: da Vinci's dry friction model F_d and its approximate $\tilde{F}_d$. Here $\mu_f = 0.5$, $\mu_b = 1.5$, and $c = 5$.

may simply replace g *by* ξ *in the expression for* τ_g. *Referring back to equation* (3.36) *we find that* $\tau_g = \tilde{F}_d(\xi) = \tilde{F}_d\left(\frac{1}{m_1+m_2}p - \frac{m_2}{m_1+m_2}\dot{s}\right)$. *This makes equation* (3.37) *of Example 3.18 complete as*

$$\dot{p} = \tilde{F}_d\left(\frac{1}{m_1+m_2}p - \frac{m_2}{m_1+m_2}\dot{s}\right) \ .$$

Also, we may nearly complete equation (3.38) *of Example 3.18 by including the expression* F_l *as follows:*

$$\tilde{M}\ddot{s} + ks(1+a^2s^2) + \frac{m_2}{m_1+m_2}\tilde{F}_d(\xi) + \lambda\dot{s} = u \ .$$

Now we consider the task of obtaining expressions for the dry friction experienced by $\mathcal{S}$. Firstly, however, one must realize that any analysis of the formulated dynamics will depend heavily upon the expression assumed to represent the tribological prop-

erties of the snake or a snake-like machine. It seems relevant then, to now review some results that indicate the tribological nature of the snake skins.

4.2.2 The Tribological Properties of Snake Skins

We begin with the results presented in a publication by Hazel et al [32]. This paper summarized the findings of experimental investigation into the morphology and micro-mechanical properties of snake skin. The experimentation was carried out via multi-mode scanning probe microscopy, a recently emerging technology that has allowed further investigation into structure and morphology at the nano-scale level. The results of this experimentation are insightful to say the least. At the nano-level the ventral portion of snake skins were found to have highly organized arrays of fibrils that elicit significant friction anisotropy. These fibrils are oriented toward the caudal portion of the animal, geometrically tapering off and raising slightly in the caudal direction. The parallel or triangular arrays of these fibrils observed then function as a ratcheting mechanism allowing the snake skin to pass easily over rough surfaces in the forward tangential direction and preventing slip in the backward tangential direction. In fact, the study indicates that the friction force experienced in the backward direction is 4 to 6 times the friction experienced in the forward direction. Further, it was found that these fibers were extremely pliable, indicating that they may displace easily when subjected to a force in the normal direction. This suggests that there may not be such a large friction force differential between the normal and forward tangential directions along the snake body. If it is the case that this micro-structure affects the friction observed at the macro level then one must cede to the idea that any point-mass dry friction law applied to the understanding of snake locomotion must be made directionally dependent in accordance with these micro-scale observations.

There is in fact some very early evidence that the conclusions of [32] are valid. Consider the work of Gray and Lissmann, two of the pioneering researchers in biomechanics and in particular the problem of snake locomotion. In [29] they considered

the friction forces involved in serpentine or undulatory movements of snakes. They grouped these forces into two categories. The first is the friction acting between the ventral portion of the snake and the ground. The second is the lateral friction experienced between the sides of the body of the snake and external resistances in the lateral direction. At the present, the ventral frictional forces will be the objects of interest. In [29] the longitudinal ventral friction forces along the path of motion are estimated. Each segment of the snake was assumed to follow that of the segment anterior to itself so that every point along the body follows the path of the head. Thus, these researchers felt that it was only necessary to make this unidirectional friction estimate. To make their calculations, a dead snake of a given weight W was dragged rectilinearly and the force required for this displacement was recorded. Assuming that the kinetic friction force experienced was proportional to the weight, in accordance with da Vinci's dry friction law, $f = \mu W$, the friction coefficient μ was calculated to be f/W. Table 4.1 indicates their findings as presented in the study. The same table may be found in their paper, with the exception of the column indicating the coefficient of friction for the backward tangential direction. Indeed, there seems to be a macro friction anisotropy.

In neither of these two studies, which are the only two that we have seen which deal with the friction experienced between a surface and snake skins, where the snakes or their skins dragged in a direction normal to the head-to-tail elongation. This needs to be done. For many make the claim that due to the scales of the snake skin, there is a large normal friction response. In fact, there are articles, such as [47], that point to [33] to validate this claim. However, Hirose *did not* measure the normal ventral friction experienced by snakes. He did discuss the matter, but upon concluding that friction in the normal direction is difficult to measure, he used the shape that he had predetermined for elicitation of snake-like motion and some fixed tangential friction coefficient to determine an expression for the normal frictional force. In other words, he selected a friction description that elicited motion from the shape he had pre-fixed,

Substratum	Length of snake (cm.)	Weight (W) of snake (g.)	Towing Force in g. Head first (T_A)	Tail first (T_P)	T_A/W	T_P/W
Glass(dry)	70	70	28		0.40	
Glass(dry)	64	80	34		0.42	
Glass(oil film)	70	70	90		1.30	
Metal (dry)	70	70	24		0.34	
Metal (dry)	78	180	60	60	0.33	0.33
Metal (oil film)	70	70	60		0.86	
Wood (dry; smooth)	64	80	27		0.34	
Cardboard (smooth)	67	83	20		0.24	
Sandpaper (fine)	64	80	52	60	0.65	0.75
Sandpaper (medium)	64	80	52	70	0.65	0.88
Sandpaper (medium rough)	64	80	40	85	0.50	1.06
Sandpaper (rough)	64	80	35	105	0.44	1.31
Sandpaper (rough)	70	70	20		0.30	
Sandpaper (rough)	66	57	25	120	0.44	2.11
Fibre mat	66	57	35	75	0.61	1.32

Table 4.1: Data collected from a ventral friction study performed by Gray and Lissmann (see [29]). An extra column has been added to the table as it is presented here. This column includes calculation of the friction coefficient for motion in the posterior direction.

not the shape that elicited motion from the friction.

The difficulty in measurement of the normal component of friction is clear to see. Firstly, due to the vertebrae structure of the snake, the body of the animal does not tend to "bend" when dragged from head-to-toe or visa-versa. On the other hand, if a force is applied to the body in the normal direction, the vertebrae will rotate with respect to one another and thus the direction of the applied force will no longer be normal. Additionally, if a snake's skin truly has anisotropic friction properties, the entire snake may begin to rotate as a normal force is applied, due to the orientational nature of the frictional force experienced. Once again, the direction of the applied force would fail to remain normal. However, it seems to us that experimentalists could probably resolve these issues in some fashion. For the complete story, the friction experienced by the snake skin needs to be measured in enough directions to determine a profile of its distribution.

4.2.3 Friction for the Planar Rigid Body

Despite this lack of information concerning the true nature of the friction experienced by the snake skin, we are in need of an expression for the residual frictional force F_i^f that acts on the body $B_i \in \mathcal{S}$. We will take the information that exists and upon extrapolation will suppose that we know and understand the nature of the friction anisotropy created via the snake skin. However, even with this assumption there remains something of a mystery concerning how one applies a dry friction law to a rigid body. In most treatments of friction, such as those seen in introductory physics texts, it is assumed that the body being studied is in motion along a line. In this case, the friction is presumed to act on the velocity of the center of mass of the body. This is an issue when we move to the planar picture. For in this setting, rotation of the body is possible and the assumption of friction acting on a single point, such as the center of mass, is in error. It does not take much to see this. Simply consider imparting a spin upon a symmetric body about its center of mass. Since the fixed-point experiences no motion and no other applied force, application of the friction law to the center of mass will indicate that the body will not experience friction and will spin forever. We know that this is not the case. As the other point masses that comprise the body rotate about the center of mass, they experience friction due to their motion relative to the surface. These frictional forces cancel one another indeed. However, they produce moments with respect to the center of mass that are not opposite in sense and thus double instead of cancelling. Hence, the body's rotation is opposed and eventually its energy dissipates and the body no longer spins. Therefore, the friction rules adopted must be applied to the entire body, not simply a single position such as the center of mass. We now provide a method for doing this for the bodies of $\mathcal{S}$ that was conceived by the author upon the inception of this research.

Consider an arbitrary element of the body $B_i \in \mathcal{S}$ with configuration coordinates (x_i, y_i, θ_i). Letting r_i^σ denote the position of a point along the body whose body

coordinates are $\mathfrak{r} = (\sigma, 0)^t$, $-l \leq \sigma \leq l$, an element may be described as the segment $r_i^{\sigma} r_i^{\sigma+\Delta\sigma}$ of length $\Delta\sigma$. Consider such an element. The velocity of any particle of this element is reasonably approximated by $\dot{r}_i^{\sigma^\star}$, where $\sigma^\star \in [\sigma, \sigma + \Delta\sigma]$. Suppose that $\Delta\sigma$ is small enough to consider the element under consideration to be essentially a particle. Further, assume the existence of a particle "law" that determines a dry friction distribution f for the body and substratum surface materials that is a function of the particle velocity. Then the friction acting on the element while in motion is approximately $f(\dot{r}_i(\sigma^\star))\Delta\sigma$. If the entire body B_i is partitioned into N such elements we obtain an estimate of the residual kinetic frictional force acting on the body by summing N estimates of this type. Using F_i^f to denote the total friction force acting on body B_i we have $F_i^f \approx \sum_{j=1}^{N} f(\dot{r}_i(\sigma_j^\star))\ \Delta\sigma_j$. So long as f is Riemann integrable we get $F_i^f = \int_{-l_i}^{l_i} f(\dot{r}_i^\sigma)\ d\sigma$ upon allowing $\|\Delta\sigma\| \to 0$. Applying this force at the body center of mass R_i, there is a corresponding moment about the centroid of B_i, denote it M_i^f, that is defined in a similar manner, where moment approximations are summed as opposed to force approximations. This moment is determined by the equation

$$M_i^f = \int_{-l_i}^{l_i} A_i \mathfrak{r}_i \times f(\dot{r}_i^\sigma)\ d\sigma\ .$$

Now, since all of the forces discussed lie in the x-y plane, the moment directions are necessarily $e^{(3)}$ and thus the sense and magnitude of the moments are sufficient to describe them. Hence, hereinafter, moments will be treated as scalars even though no notational change will ensue.

4.2.4 Friction Models for the Serial-Link Structure

Some have showed hesitation when presented this development. However, as of the moment of this writing, a good reason not to take this approach has not been provided. It is in fact the case that this same approach has independently been taken in [62] as a result of being faced with the same issue. Using this development, several directionally

dependent friction models are now introduced. A couple of these friction descriptions are taken from [62] or are adaptations thereof and the others are being proposed as possible descriptions of the friction that may in fact be induced by actual snake skin morphologies. We are the first to model such morphologies and include their effects into the dynamics of snake-like devices.

Taking the time derivative of r_i^σ we obtain

$$\dot{r}_i^\sigma = \begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} + \sigma\dot{\theta}_i \begin{bmatrix} -\sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix} . \tag{4.2}$$

The two terms of this expression are used represent the translational and rotational velocity of the element $r_i^\sigma r_i^{\sigma+\Delta\sigma}$, respectively. These velocity components are projected along the directions of the body frame axes to obtain the tangential and normal components of the velocity with respect to the body. The result is the expression

$$\begin{bmatrix} v_t^\sigma \\ v_n^\sigma \end{bmatrix} \doteq A_i^t \dot{r}_i^\sigma = A_i^t \begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma\dot{\theta}_i \end{bmatrix} . \tag{4.3}$$

We will generically label this body velocity v_b^σ with the understanding that its components depend on the body index i. With the velocity of $\Delta\sigma$ decomposed along the body coordinate frame axes, one may define differing friction characteristics along these respective directions. These tangent and normal components of friction are denoted f_t^σ and f_n^σ, respectively, and this body friction will, as in the case of the body velocity, be generically denoted f_b^σ. Upon formulation, this friction vector will be projected back into the inertial coordinate system. The result of this projection will be labelled f^σ. By describing the friction properties of the body $B_i \in \mathcal{S}$ in this manner, the friction properties in every direction relative to the body are specified under the appropriate assumptions. Examples of this are now given.

Example 4.2 (Linear Viscous Friction - After [62]). *Suppose that we specify a*

simple linear viscous friction law, say

$$f_t^\sigma = -g\Delta m c_t v_t^\sigma \ ;$$
$$f_n^\sigma = -g\Delta m c_n v_n^\sigma \ ,$$

or

$$\begin{bmatrix} f_t^\sigma \\ f_n^\sigma \end{bmatrix} = -g \begin{bmatrix} c_t & 0 \\ 0 & c_n \end{bmatrix} \begin{bmatrix} v_t^\sigma \\ v_n^\sigma \end{bmatrix} \Delta m$$

in matrix vector notation. Using the uniformity assumption on the mass of the rod provides $\Delta m = \frac{m_i \Delta\sigma}{2l_i}$. *So,*

$$\begin{bmatrix} f_t^\sigma \\ f_n^\sigma \end{bmatrix} = -g\frac{m_i}{2l_i} \begin{bmatrix} c_t & 0 \\ 0 & c_n \end{bmatrix} \begin{bmatrix} v_t^\sigma \\ v_n^\sigma \end{bmatrix} \Delta\sigma \ .$$

From this equation the approximation for the friction experienced by $\Delta\sigma$ *can be expressed in the global frame of reference as*

$$\begin{bmatrix} f_x^\sigma \\ f_y^\sigma \end{bmatrix} = A_i \begin{bmatrix} f_t^\sigma \\ f_n^\sigma \end{bmatrix} \ .$$

Thus the residual translational friction experienced by the body is

$$F_i^f = \int_{-l_i}^{l_i} \begin{bmatrix} f_x^\sigma \\ f_y^\sigma \end{bmatrix} \ d\sigma \ .$$

Performing the integration one finds that the term of the integrand corresponding to angular velocity integrates to zero and we are left with

$$F_i^f = -gm_i A_i \begin{bmatrix} c_t & 0 \\ 0 & c_n \end{bmatrix} A_i^t \begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} \ .$$

Further, due to the decomposition of the velocity into normal and tangent components, the residual rotational friction is easily expressed as

$$M_i^f = \int_{-l_i}^{l_i} \sigma f_n^\sigma \; d\sigma = -g\frac{m_i l_i^2}{3}\mu_n \dot{\theta}_i \; .$$

It is noted that this example includes the assumption of negligible tangential friction by taking $\mu_t = 0$. This is the assumption to be found in [47, 48, 46].

Example 4.3 (da Vinci's Viscous Friction - After [62]). *As another example consider applying da Vinci's friction "law" along the normal and tangent directions of the body B_i. In this case one obtains*

$$f_t^\sigma = -g\Delta m c_t sgn(v_t^\sigma) \; ;$$
$$f_n^\sigma = -g\Delta m c_n sgn(v_n^\sigma) \; ,$$

or

$$\begin{bmatrix} f_t^\sigma \\ f_n^\sigma \end{bmatrix} = -g \begin{bmatrix} c_t & 0 \\ 0 & c_n \end{bmatrix} \begin{bmatrix} sgn(v_t^\sigma) \\ sgn(v_n^\sigma) \end{bmatrix} \Delta m$$

in matrix vector notation. Thus, this expression allows one to write an expression for the residual frictional force felt by the link in terms of the Riemann integral. This expression is

$$F_i^f = \int_{-l_i}^{l_i} \begin{bmatrix} f_x^\sigma \\ f_y^\sigma \end{bmatrix} \; d\sigma \; .$$

Performing the integration one finds that

$$F_i^f = -m_i g A_i \begin{bmatrix} c_t & 0 \\ 0 & c_n \end{bmatrix} \begin{bmatrix} sat(v_t^\sigma, 0) \\ sat(v_n^\sigma, l_i\dot{\theta}_i) \end{bmatrix} \; .$$

Further, due to the decomposition of the velocity into normal and tangent components,

the residual rotational friction is easily expressed as

$$M_i^f = \int_{-l_i}^{l_i} \sigma f_n^\sigma = -\frac{1}{2}\mu_n m_i g l_i \; dzs(v_n^\sigma, l_i\dot{\theta}_i) \; .$$

The intent of these examples is quite clear. However, the end result is deceptively in violation of basic principle and practice as it regards dry friction. Consider, in the body frame, the velocity vector r_i^σ with magnitude d and orientation with respect to the body frame of γ. That is, $v_t^\sigma = d\cos(\gamma)$ and $v_n^\sigma = d\sin(\gamma)$. The slope of the line of action of this vector is given by

$$\tan(\gamma) = \frac{v_n^\sigma}{v_t^\sigma}$$

in the body frame. However, the frictional force acting on the body due to the particle at r_i^σ is, for the simple viscous Example 4.2 derived above,

$$\begin{bmatrix} f_t^\sigma \\ f_n^\sigma \end{bmatrix} = - \begin{bmatrix} c_t & 0 \\ 0 & c_n \end{bmatrix} \begin{bmatrix} v_t^\sigma \\ v_n^\sigma \end{bmatrix} \Delta m \; .$$

Let β indicate the direction of the line of action of this force (not the sense). Then the slope of the line of action in the body frame is given by

$$\tan(\beta) = \frac{c_n v_n^\sigma}{c_t v_t^\sigma} \; .$$

Therefore, for $c_n \neq c_t$, the frictional force does not act along the line of motion of the particle. i.e., $\beta \neq \gamma$. This is contrary to the standard thinking that dry friction acts along the direction of motion but in the opposite sense. These comments are illustrated by Figure 4.3.

One can develop an expression for a directionally dependent friction that, for a given particle velocity with polar coordinates (d, γ), has the same magnitude as the friction vector given by the laws presented above, however, with the correct direction.

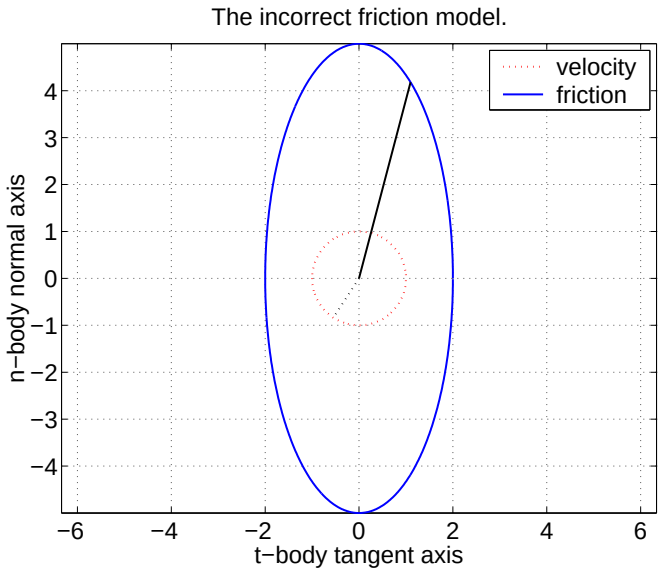


Figure 4.3: The incorrect directionally dependent linear dry friction model.

To do so, we examine the magnitude of these friction expressions. For instance, take the linear viscous friction model. Doing so we have $f_n^\sigma = -g\Delta m c_n d \sin(\gamma)$ and $f_t^\sigma = -g\Delta m c_t d \cos(\gamma)$. Thus,

$$\|f_b^\sigma\|^2 = (g\Delta m c_n d \sin(\gamma))^2 + (g\Delta m c_t d \cos(\gamma))^2 \ .$$

So, we define $\mu(\gamma) = (c_n^2 \sin^2(\gamma) + c_t^2 \cos^2(\gamma))^{\frac{1}{2}}$ and describe the friction experienced by the particle located at r_i^σ by

$$f_b^\sigma = -g\|v_b^\sigma\|\Delta m \mu(\gamma) \begin{bmatrix} \cos(\gamma) \\ \sin(\gamma) \end{bmatrix} \ .$$

In this case, the direction of the friction is the same as v_b^σ but opposite in sense and the magnitude of f_b^σ is precisely the same as the friction described in Example 4.2.

From Figure 4.4 it is clear that the intent of an elliptical directionally dependent

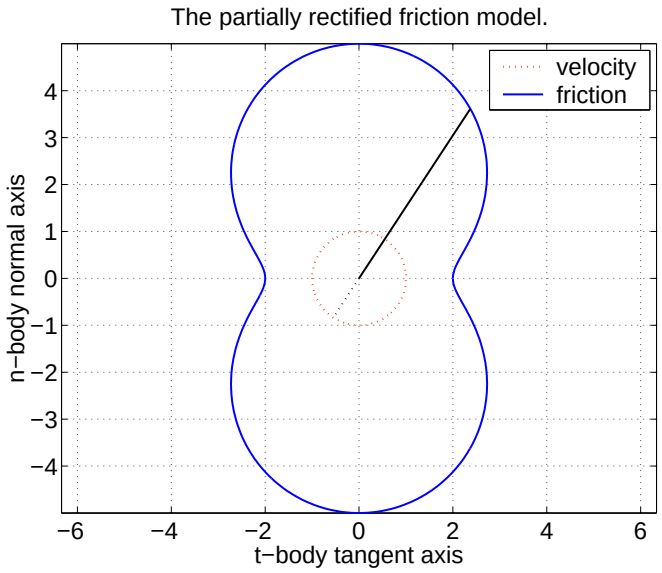


Figure 4.4: The partially corrected directionally dependent linear dry friction model.

frictional characteristic is lost by the direction correction. However, one may mathematically capture such a feature and many other sensible distributions with correct direction by simply defining a directionally dependent frictional force magnitude and not a directionally dependent force direction (i.e. the direction is set by that of the given velocity). We first describe the velocity in the body frame with polar coordinates $v_b^\sigma = (d, \gamma)$. Then, the desired frictional force magnitude is described as a function of the velocity direction, say $\mu(\gamma)$. This magnitude is placed in the direction of the velocity but with the opposite sense yielding for us the expression

$$f_b^\sigma = -g\Delta m\mu(\gamma) \begin{bmatrix} \cos(\gamma) \\ \sin(\gamma) \end{bmatrix} . \tag{4.4}$$

Note that with this description we may easily switch between a simple linear dry friction model and da Vinci's dry friction model by the inclusion or exclusion of the factor $\|v_b^\sigma\|$ in the definition of μ. Let us now completely recover the intent of

Example 4.2 but with the correct direction. Knowing that $\frac{x^2}{\mu_t^2} + \frac{y^2}{\mu_n^2} = 1$ describes an ellipse in an x-y coordinate frame, we may simply substitute the polar coordinate relations $x = d\cos(\gamma)$ and $y = d\sin(\gamma)$ into this expression and solve for d in order to arrive at the expression

$$\mu(\gamma) = \|v_b^\sigma\|\mu_t\mu_n(\mu_n^2\cos^2(\gamma) + \mu_t^2\sin^2(\gamma))^{-\frac{1}{2}} \ . \tag{4.5}$$

The resulting force distribution acting on the unit velocity ball in the body frame is given in Figure 4.5.

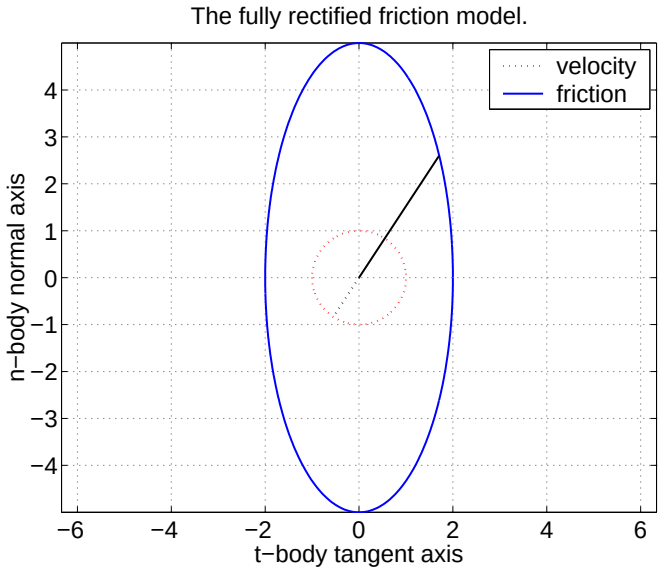


Figure 4.5: The fully corrected directionally dependent linear dry friction model.

In addition to preserving the idea that dry friction directly opposes motion, the concept of defining the directional dependence via the friction coefficient allows greater freedom in describing anisotropy. The reason is that, if one defines the directional dependance via tangent and normal decomposition, as described in Examples 4.2 and 4.3, then every other direction is defined. So, for instance, we cannot describe a forward to backward tangential friction differential as described in [29] and

[32] using this approach. Defining the dependence through equation (4.4), we can accomplish such a description and in fact can assign nearly any friction profile that we can imagine. As an example, consider altering the elliptic type friction distribution of Figure 4.5 so that there is a large friction differential between forward tangential and backward tangential directions relative to the body B_i. In this way, we try to capture the tribological properties of the snake skin. To do so, we add a scaled gaussian distribution on the interval $-\frac{\pi}{2} \leq \gamma \leq \frac{\pi}{2}$ to the basic elliptic definition of μ given by equation (4.5). The gaussian distribution with mean $\bar{\gamma}$ and standard deviation σ_γ is

$$G(\bar{\gamma}, \sigma_\gamma) = \frac{1}{\sigma_\gamma \sqrt{2\pi}} \exp\left(-\frac{\gamma - \bar{\gamma}}{2\sigma_\gamma}\right) . \tag{4.6}$$

This expression is scaled by the factor $\frac{c}{\max(G)}$ to provide a maximum distribution value of c. The distribution that we will add to the elliptical description has parameters $\bar{\gamma} = \pi$, $\sigma_\gamma = 0.5$, and $c = 5$. This distribution is illustrated in Figures 4.6 and 4.7

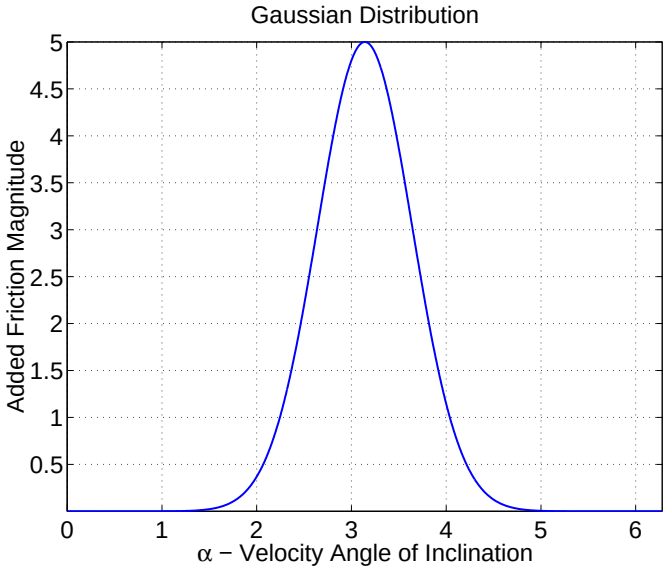


Figure 4.6: Gaussian distribution with parameters $\bar{\gamma} = \pi$, $\sigma_\gamma = 0.5$, and $c = 5$ over $0 \leq \gamma \leq 2\pi$.

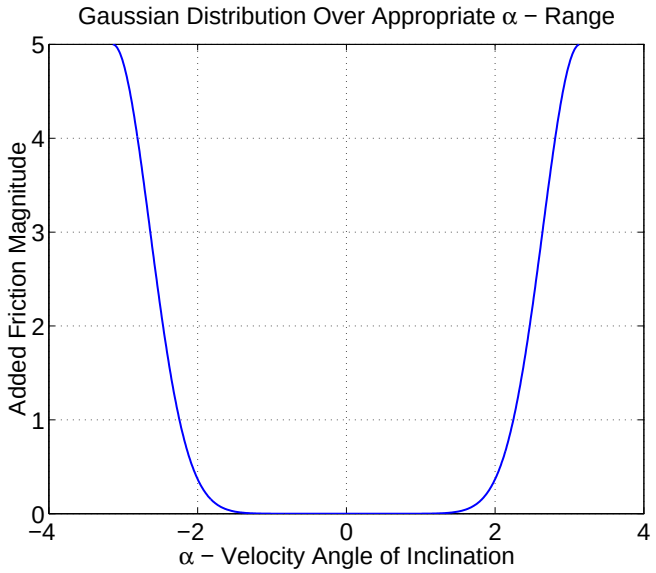


Figure 4.7: Gaussian distribution with parameters $\bar{\gamma} = \pi$, $\sigma_\gamma = 0.5$, and $c = 5$ over $-\frac{\pi}{2} \leq \gamma \leq \frac{\pi}{2}$.

Upon adding this to the elliptic definition of μ with parameters $\mu_n = 2$ and $\mu_t = 1$ one obtains the friction distribution of Figure 4.8

The drawback of this anisotropy description is that one probably cannot obtain simple expressions for F_i^f and M_i^f such as those in Examples 4.2 and 4.3. However, these quantities can be acquired numerically and an outline of the procedure is now given.

Step 1. Obtain the direction γ of the velocity r_i^σ in the body frame. To do so, calculate v_b^σ according to equation (4.3). Then, make careful use of the relation $\tan(\gamma) = \frac{v_n^\sigma}{v_t^\sigma}$ to obtain the appropriate value of the

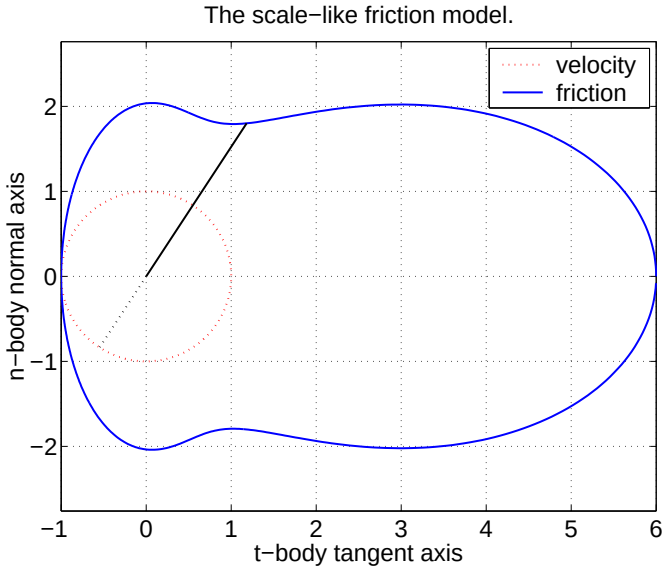


Figure 4.8: Snake scale friction distribution.

following function:

$$\gamma(v_b^\sigma) = \begin{cases} \arctan(\frac{v_n^\sigma}{v_t^\sigma}) & \text{if } v_t^\sigma > 0; \\ \frac{\pi}{2} & \text{if } v_n^\sigma > 0 \text{ and } v_t^\sigma = 0; \\ 0 & \text{if } v_n^\sigma = 0 \text{ and } v_t^\sigma = 0; \\ \frac{-\pi}{2} & \text{if } v_n^\sigma < 0 \text{ and } v_t^\sigma = 0; \\ \arctan(\frac{v_n^\sigma}{v_t^\sigma}) - \pi & \text{if } v_n^\sigma < 0 \text{ and } v_t^\sigma < 0; \\ \pi + \arctan(\frac{v_n^\sigma}{v_t^\sigma}) & \text{if } v_n^\sigma > 0 \text{ and } v_t^\sigma < 0. \end{cases} \tag{4.7}$$

Step 2. Use γ to calculate the directionally dependent friction distribution f_b^σ using equation (4.4).

Step 3. Use a quadrature rule to approximate $\int_{l_i}^{l_i} f_b^\sigma \, d\sigma$.

Step 4. Project the result of this calculation by left multiplication of A_i to

obtain F_i^f.

Step 5. Use f_n^σ of Step 2 and quadrature to approximate $M_i^f = \int_{l_i}^{l_i} \sigma f_n^\sigma \, d\sigma$.

This involves work and invariably computation time. However, sometimes work and computation time is necessary to get what is desired. We shall use this approach to obtain friction expressions for the snake-like serial-link structure.

There is but one issue left to address with generating forces in this manner. Recall that in the case of the mass-spring system, it was necessary make a continuous approximation of da Vinci's dry friction law. Here, such an approximation may or may not be necessary. To obtain a cumulative friction force for B_i the friction distribution over that body is integrated. This integration is not altered by what is occurring with a single particle of the body. Thus, if one particle is experiencing the discontinuity of da Vinci's law, the integration over the body ignores it. Hence, a da Vinci type friction distribution will only be a problem if the entire body simultaneously moves through its discontinuity point. This is not likely to happen. All the same, a factor of the form $\frac{2}{\pi}\arctan(c\|v_b^\sigma\|)$ may be included in the definition of μ to ensure that everything is continuous.

Such an approximation might be advisable anyway, since it was suggested that quadrature be used to compute the integration of the friction distribution. In order for the accuracy of the quadrature to be ensured, continuity is a minimal requirement. Certainly, due to the fact that v_b^σ is a smooth quantity with respect to σ, so long as $\gamma(v_b^\sigma)$ is smooth and the definition of $\mu(\gamma)$ is smooth, f_b^σ will be smooth and the accuracy of a quadrature rule will be assured. However, γ can fail to be smooth and in fact fails to be continuous at $v_b^\sigma = 0$. To be specific, it would seem that the real issue would come if σ corresponds to a quadrature node and $v_b^\sigma = 0$. By including the inverse tangent factor, we may eliminate this problem.

These issues may be avoided all together by using simple linear friction in lieu of da Vinci's law at the particle level.

4.2.5 Generalized Friction Forces and Their Invariance

Once F_i^f and M_i^f are constructed, we have expressions of the friction forces in terms of the coordinates $\{z_j\}_{j=1}^{3n}$ or the coordinate vector z. We seek to generalize these forces. These z_j are functions of the generalized independent coordinate vector q. i.e., $z_j = z_j(q)$. Thus, as previously shown, the generalized force associated with q_k that is due to a force acting on B_i is given by

$$Q_k = \sum_{i=1}^{n} \begin{bmatrix} (F_i^a)^t & M_i^a \end{bmatrix} \frac{\partial \vec{z}_i}{\partial q_k} \ .$$

Allowing the definition $(F^f)^t \doteq ((F_1^f)^t, M_1^f, \ldots, (F_n^f)^t, M_n^f)$ we acquire the m vector of generalized forces due to the friction acting on the system by taking the product $(F^f)^t \frac{\partial z}{\partial q}$. An expression for the gradient $\frac{\partial z_j}{\partial q}$ is stated in Example 3.4.

In addition to generalizing the forces we must also establish their group invariance so that we may use system (3.33). This is now done for the serial-link structure.

Theorem 4.1 (SE(2) Invariance of Body-Borne Directionally Dependent Friction). *Any directionally dependent frictional force that is due to the anisotropy of the serial-link structure is an $SE(2)$ invariant force.*

Proof. We begin by examining the result of applying the lift of $SE(2)$ to equation (4.2). For convenience we repeat this equation:

$$\dot{r}_i^\sigma = \begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} + \sigma \dot{\theta}_i \begin{bmatrix} -\sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix} \ .$$

Recall from the proof of Theorem 3.1 that

$$\begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} \uparrow A(\alpha) \begin{bmatrix} \dot{x}_i \\ \dot{y}_i \end{bmatrix} \ . \tag{4.8}$$

Thus we need only concern ourselves with the second term of r_i^σ. Due to the relations

$$-\sin(\theta+\alpha) = -\sin(\theta)\cos(\alpha) - \cos(\theta)\sin(\alpha)$$
$$\cos(\theta+\alpha) = \cos(\theta)\cos(\alpha) - \sin(\theta)\sin(\alpha) \ ,$$

it is seen that

$$\begin{bmatrix} -\sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix} \uparrow A(\alpha) \begin{bmatrix} -\sin(\theta_i) \\ \cos(\theta_i) \end{bmatrix} \ .$$

Since the lift in $SE(2)$ acts as the identity on $\dot{\theta}_i$ we may conclude that

$$\dot{r}_i^\sigma \uparrow A(\alpha)\dot{r}_i^\sigma \ .$$

The projection of the velocity into the body frame is given by $v_b^\sigma = A_i^t \dot{r}_i^\sigma$ and as such, applying the lifted action to v_b^σ provides us with the result that

$$v_b^\sigma \uparrow A^t(\theta_i+\alpha)A(\alpha)\dot{r}_i^\sigma = A_i^t A^t(\alpha)A(\alpha)\dot{r}^\sigma = v_b^\sigma \ .$$

This says that any friction distribution f_b^σ in the body frame based on the velocity of body particles expressed in the body coordinates is $SE(2)$ invariant. For, to express the friction distribution in the global frame we arrive at $f^\sigma = A_i f_b^\sigma$, and the resultant frictional force and moment are integrals and thus linear operators of this distribution. This means that $F_i^f \uparrow A(\alpha)F_i^f$ under the lift, as $f^\sigma \uparrow A(\theta_i+\alpha)f_b^\sigma = A(\alpha)A_i f_b^\sigma = A(\alpha)f^\sigma$. □

Due to Theorem 4.1 and Theorem 3.1 it is seen that the serial link structure dynamics take on the reduced form of equation (3.33).

4.3 Part II: Control Forces

4.3.1 A Simple Actuator Model

At this juncture we present a simple model for the forces that arise in a snake due to the tensioning of the muscles that join rib tips along the length of the snakes vertebrae structure. We do so through the serial-link structure with muscle-like actuators as described in Chapter 2.

Suppose that the muscle-like actuators of our serial-link system $\mathcal{S}$ can apply a compressional force $F^{(\cdot)}_{i_1,i_2}$ of magnitude $f^{(\cdot)}_{i_1,i_2}$ between the "rib" tips of bodies $B_{i_1}, B_{i_2} \in \mathcal{S}$. The index i_1 will indicate which body the force is acting on and i_2 will indicate the other body used in creating this force. Recall that the muscles only connect adjacent bodies and lie on one "side" of the system. Here the superscript will be either l or r to indicate whether this force is due to the actuator lying to the left or right, respectively, of the system, where the orientation is toward a decrease in body index. Thus, with the exception of the head and the tail, body $B_i \in \mathcal{S}$ has four forces acting on it. These forces are $F^r_{i,i-1}$ and $F^r_{i,i+1}$ acting at $\mathfrak{r}^r_i = (0, -l_r)^t$, and $F^l_{i,i-1}$ and $F^l_{i,i+1}$ acting at $\mathfrak{r}^l_i = (0, l_r)^t$. $\mathfrak{r}^{(\cdot)}_i$ are the rib tip coordinates in the local body coordinate system. In order to finish describing these forces a direction and sense must be assigned to them. We assume that these forces act along the line determined by the "rib" tips and in sense that indicates compression of the actuator elements. These descriptions are now expressed mathematically. It is seen that the rib tips for body B_i are located in the global coordinate system at $r^l_i = R_i + A_i\mathfrak{r}^l_i$ and $r^r_i = R_i + A_i\mathfrak{r}^r_i$. Thus, a unit vector having the same direction and sense of the forces described above is easily obtainable and a summary is provided in Table 4.2.

Recall from Chapter 3 that since we are dealing with rigid bodies, we may create a statically or dynamically equivalent system by allowing the forces described above to act at the centroid R_i of body B_i and keeping track of moments about the centroid associated with these forces. These moments are easily determined and are included

Unit Vector	Force	Moment
$u^l_{i,i-1} = \frac{r^l_{i-1} - r^l_i}{\|r^l_{i-1} - r^l_i\|}$	$F^l_{i,i-1} = f^l_{i,i-1} u^l_{i,i-1}$	$M^l_{i,i-1} = (r^l_i - R_i) \times F^l_{i,i-1}$
$u^r_{i,i-1} = \frac{r^r_{i-1} - r^r_i}{\|r^r_{i-1} - r^r_i\|}$	$F^r_{i,i-1} = f^r_{i,i-1} u^r_{i,i-1}$	$M^r_{i,i-1} = (r^r_i - R_i) \times F^r_{i,i-1}$
$u^r_{i,i+1} = \frac{r^r_{i+1} - r^r_i}{\|r^r_{i+1} - r^r_i\|}$	$F^r_{i,i+1} = f^r_{i,i+1} u^r_{i,i+1}$	$M^r_{i,i+1} = (r^r_i - R_i) \times F^r_{i,i+1}$
$u^l_{i,i+1} = \frac{r^l_{i+1} - r^l_i}{\|r^l_{i+1} - r^l_i\|}$	$F^l_{i,i+1} = f^l_{i,i+1} u^l_{i,i+1}$	$M^l_{i,i+1} = (r^l_i - R_i) \times F^l_{i,i+1}$

Table 4.2: The muscle-like actuator force and moment expressions

in the Table 4.2. We will replace our system model with this equivalent one. Thus, since the "ribs" were assumed massless, it is seen that they were only introduced in the system description to properly describe the external control forces acting on the bodies and are no longer in need of consideration. With the specification of these forces given we do two things. Firstly, the desire exists to write the forces and moments described in a matrix multiplication format. Secondly, the generalized force vector Q associated with these forces is expressed. In this way we intend to make it clear how these forces could be included directly within a simulation format.

Moving directly to the delineation of the former task, we write the composite actuator force, denoted F^a_i, as a matrix-vector multiplication. Using Table 4.2 this takes on the form

$$F^a_i = \begin{bmatrix} u^l_{i,i-1} & u^r_{i,i-1} & u^l_{i,i+1} & u^r_{i,i+1} \end{bmatrix} \begin{bmatrix} f^l_{i,i-1} \\ f^r_{i,i-1} \\ f^l_{i,i+1} \\ f^r_{i,i+1} \end{bmatrix} = u_i f_i \ ,$$

where it is noted that $F^{(\cdot)}_{1,0} = F^{(\cdot)}_{n,n+1} \doteq 0$. In a similar manner, the composite actuator

moment, denoted M_i^a, may be written

$$M_i^a = \begin{bmatrix} m_{i,i-1}^l & m_{i,i-1}^r & m_{i,i+1}^l & m_{i,i+1}^r \end{bmatrix} \begin{bmatrix} f_{i,i-1}^l \\ f_{i,i-1}^r \\ f_{i,i+1}^l \\ f_{i,i+1}^r \end{bmatrix} = m_i f_i \ ,$$

where $m_{i_1,i_2}^{(\cdot)} \doteq \|M_{i_1,i_2}^{(\cdot)}\| / f_{i_1,i_2}^{(\cdot)}$. Let

$$K = \begin{bmatrix} \begin{pmatrix} u_1 \\ m_1 \end{pmatrix} & & \\ & \ddots & \\ & & \begin{pmatrix} u_n \\ m_n \end{pmatrix} \end{bmatrix} \ ,$$

$$\tilde{f}^t = (f_1^t, \ldots, f_n^t) \ ,$$

and

$$(F^a)^t = ((F_1^a)^t, (M_1^a)^t, \ldots, (F_n^a)^t, (M_n^a)^t) \ .$$

Then, $F^a = K\tilde{f}$. Note that although $F_{i,i+1}^{(\cdot)} \neq F_{i,i-1}^{(\cdot)}$ it is the case that $f_{i,i+1}^{(\cdot)} = f_{i,i-1}^{(\cdot)}$. So, let

$$f^t = (f_{1,2}^l, f_{1,2}^r, \ldots, f_{n-1,n}^l, f_{n-1,n}^r) \ ,$$

$$S = \begin{bmatrix} I^{2\times 2} \\ I^{2\times 2} \end{bmatrix} \ ,$$

and

$$T = \begin{bmatrix} & 0^{2\times 2(n-1)} & \\ S & & \\ & \ddots & \\ & & S \\ & 0^{2\times 2(n-1)} & \end{bmatrix} .$$

Then $\tilde{f} = Tf$. So, $F^a = KTf$. All of these quantities can be constructed in an algorithmic way using a software package such as *MATLAB*.

4.3.2 Generalized Actuator Forces and Modeling Assumptions

We now seek to accomplish the later of the two pre-ascribed tasks. Once F^a is constructed, we have expressions for the actuator forces in terms of the coordinates $\{z_j\}_{j=1}^{3n}$ or the coordinate vector z. These z_j are functions of the generalized independent coordinate vector q. i.e., $z_j = z_j(q)$. As previously shown, the generalized force associated with q_k due to a force acting on body B_i is given by

$$Q_k = \sum_{i=1}^{n} \begin{bmatrix} (F_i^a)^t & M_i^a \end{bmatrix} \frac{\partial \vec{z}_i}{\partial q_k} .$$

Thus, by taking the product $(F^a)^t \frac{\partial z}{\partial q}$, the m vector of generalized actuation forces is acquired. An expression for the gradient of the z_j is stated in Example 3.4.

For the experimentation and simulation that we pursue, these actuator force formulations are not employed. Instead, we recognize that the entire intent of these actuators is to produce a bidirectional torque capability about the revolute joints of the serial-link structure. Thus, we simply assume such an actuation is possible and do not involve ourselves in how this actually occurs. More formally speaking, from the above formulation, generalized forces corresponding to the relative angles $s = (\phi_1, \ldots, \phi_{n-1})$ between links, namely $Q_4, \ldots, Q_{n+2}$, are acquired. **We separate**

the portion of τ_k due to friction, denoted in its reduced form by $\tilde{\tau}_k^f$, away from τ_k and from henceforth use τ to indicate the forces along the shape directions due to control action. Letting $\tau_k \doteq Q_{k+3}$, with $k = 1, \ldots, n-1$, we assume that we get to pick the τ_k directly and forget about any talk of actuation mechanisms.

One may bring up the very good question of "Why bother to begin with ?" The reason is the following. As indicated in the literature review, there have been many investigations into the muscle activity of snakes, the patterns of this activity, and how this activity contributes to propulsion. There have also been studies on snake locomotion via multi-body models such as the one proposed herein. Often, these studies refer to one another. However, a study that merges these efforts has not been produced to date, so much as we are aware. It would be nice to make a direct connection between the experimental biology world and the mathematical engineering world. We believe that the modeling efforts put forth in this writing are sufficiently well described so as to facilitate this possibility. Upon settling parameter issues for existing muscle models, the experimentalist could directly feedforward muscle contraction data via f and observe the resulting model motion. Likewise, it may also be a possibility for one to use the control inputs τ_k used to obtain locomotion in the serial-link structure model to determine f. f could then be compared to data. For a complete scientific picture it would seem that one of these two studies must take place at some point.

4.3.3 Feedback Linearization and the Shape Equation

We now return to the fact that the third of the equations in (3.33) can be rather convoluted for systems of much complexity. This equation is restated here, respecting force notation changes, for ease of reference and is referred to hereafter as the shape

equation.

$$\tilde{M}_{kk_2}\ddot{s}_{k_2} + \frac{1}{2}C_{kk_2k_1}\dot{s}_{k_2}\dot{s}_{k_1} + N + \frac{1}{2}p^t(\frac{\partial I^{-1}}{\partial s_k})p + \frac{\partial V}{\partial s_k} = \tau_k + \tilde{\tau}_k^f - \tilde{\tau}^t A_k \ . \tag{4.9}$$

The intractability of the shape equation is unfortunate since it is this equation that involves the control torques τ_k and ultimately we are interested in the control problem. Hence, other possibilities for control formulation must be pursued. One such possibility is to employ the so-called feedback linearization technique. This approach uses a form mocking step to recast a nonlinear problem as a linear one. It so happens that in the case of fully actuated mechanical system, this procedure can be applied in a rather straight forward manner. Let us now delineate this process.

Consider the fully actuated dynamics:

$$\Lambda(q)\ddot{q} + \Upsilon(q,\dot{q}) + U_d = \tau \ , \tag{4.10}$$

where q is the vector of independent system coordinates; Λ is a spd mass matrix; Υ is a matrix that combines other system contributions which often include centripetal and Coriolis forces, friction, gravity, and other modelled entities; U_d is some unknown contribution or disturbance; and τ is the vector of generalized forces due to control action. Consider the control

$$\tau = \Lambda(q)u + \Upsilon(q,\dot{q}). \tag{4.11}$$

Substituting this expression into the dynamics provides us the equation

$$\Lambda(q)(\ddot{q} - u) = -U_d \ ,$$

or

$$\ddot{q} = u - \Lambda(q)^{-1}U_d.$$

Thus, in the presence of perfect modeling (i.e., $U_d = 0$), we have the linear control problem

$$\ddot{q} = u \ . \tag{4.12}$$

For obvious reasons, this equation is known as Newton's equation.

Suppose that we desire q to follow some known trajectory ς. Paraphrasing, a control u is desired such that $q \approx \varsigma$ or $e \approx 0$, where $e \doteq \varsigma - q$. From this error expression it follows that

$$\ddot{e} = \ddot{\varsigma} - \ddot{q} = \ddot{\varsigma} - u.$$

Let

$$u \doteq \ddot{\varsigma} - u_e \ . \tag{4.13}$$

Then the linear stability problem

$$\ddot{e} = u_e$$

results. If this system is controllable, then we may choose any of several classical control techniques to define u_e. Define

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} e \\ \dot{e} \end{bmatrix} .$$

Then we have

$$\begin{aligned} \dot{w} = \begin{bmatrix} \dot{w}_1 \\ \dot{w}_2 \end{bmatrix} &= \begin{bmatrix} 0 & I \\ 0 & 0 \end{bmatrix} \begin{bmatrix} e \\ \dot{e} \end{bmatrix} + \begin{bmatrix} 0 \\ I \end{bmatrix} u_e \\ &= \begin{bmatrix} 0 & I \\ 0 & 0 \end{bmatrix} w + \begin{bmatrix} 0 \\ I \end{bmatrix} u_e \\ &= A_e w + B_e u_e \ . \end{aligned}$$

Let $\mathcal{C} = \begin{bmatrix} B_e & A_e B_e & \dots & A_e^{2m-1} B_e \end{bmatrix}$ denote the system's controllability matrix. See-

ing that $A_e^2 = 0$ and thus $A_e^{\bar{m}} = 0$, $\bar{m} \geq 2$, it follows that $rank(\mathcal{C}) = rank\left(\begin{bmatrix} B_e & A_eB_e \end{bmatrix}\right)$. But

$$\begin{bmatrix} B_e & A_eB_e \end{bmatrix} = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} ,$$

which is invertible, and hence $rank(\mathcal{C}) = 2m$. This implies that the system is indeed controllable! So, u_e can be selected so as to drive e to zero. Once u_e is selected, u is determined to be $u = \ddot{\varsigma} - u_e$ by equation (4.13), which in turn defines τ to be

$$\tau = \Lambda(q)(\ddot{\varsigma} - u_e) + \Upsilon(q, \dot{q}) , \tag{4.14}$$

by equation (4.11). This method of control is known in the robotics literature as computed torque control. This is because τ typically corresponds to applied torques at link joints for robot manipulators, etc. We now discuss several linear systems controls that may be incorporated into the feedback linearization scheme to bring about the tracking control objective.

Example 4.4 (Feedback Control). *We begin with the most elementary of the linear controls, namely, pole placement via state feedback. The following theorem holds:*

Theorem 4.2. *Given $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$, $\exists K \in \mathbb{R}^{m\times n}$ s.t. the eigenvalues of $A - BK$ can be assigned to arbitrary, real or complex conjugate, locations iff (A, B) is controllable.*

Since we have shown that the exact linearization of the fully actuated mechanical system is controllable, this result says that we can find a gain matrix $K \in \mathbb{R}^{m\times 2m}$ s.t. $u_e = -Kw$ stabilizes the error system. That is, we can make $e, \dot{e} \to 0$ as $t \to \infty$. Using this proportional feedback control yields the following control scheme:

$$\begin{aligned} \tau &= \Lambda(q)u + \Upsilon(q, \dot{q}) \\ &= \Lambda(q)(\ddot{\varsigma} - u_e) + \Upsilon(q, \dot{q}) \\ &= \Lambda(q)(\ddot{\varsigma} + Kw) + \Upsilon(q, \dot{q}) . \end{aligned} \tag{4.15}$$

Due to the fact that the error plant is linear, there is a straight forward way to determine a gain K that is in fact optimal in some sense. Consider the quadratic cost functional

$$J(u_e, w_0) = \int_0^\infty (w^T Q w + u_e^t R u_e)\ dt\ ,$$

where the weight matrices $Q \geq 0$ and $R > 0$ are design parameters. Making use of the calculus of variations and the so-called sweep-method it can be shown that the control that minimizes J subject to the dynamics of w is the state feedback control law described by

$$\begin{aligned} u &= -Kw\ , \\ K &= R^{-1} B_e^t \Pi\ , \end{aligned} \tag{4.16}$$

where Π satisfies the so-called algebraic Riccati equation,

$$\Pi A_e + A_e^t \Pi - \Pi B_e R^{-1} B^t \Pi + Q = 0\ .$$

In fact, this is the method we shall employ when illustrating this procedure in the context of the serial-link structure in Chapter 5. The software package MATLAB *makes this synthesis simple, since it includes routines that provide the gain K.*

Example 4.5 (PD and PID Computed Torque). *Another well-known classical control method for linear systems is the PD control or the proportional-derivative control. This control has essentially the same effect as the simple feedback control. However, the derivative is used to obtain local information about the future of the error and enables adjustment based on that prediction. The PD controller has the form $u_e = -K_v \dot{e} - K_p e$, where K_d and K_p are positive definite matrices. The resulting error system is known to be stable. The control scheme is given by*

$$\begin{aligned} \tau &= \Lambda(q) u + \Upsilon(q, \dot{q}) \\ &= \Lambda(q)(\ddot{\varsigma} - u_e) + \Upsilon(q, \dot{q}) \\ &= \Lambda(q)(\ddot{\varsigma} + K_d \dot{e} + K_p e) + \Upsilon(q, \dot{q})\ . \end{aligned} \tag{4.17}$$

This control can be appended with a smoothing or integral term. This term enables the control to adjust to constant unknown disturbances and can go quite a ways in dealing with other benign interference. We introduce the dynamic $\dot{\varepsilon} = e$ *and add a constant multiple of* ε *to the controller* u_e*. This yields*

$$\tau = \Lambda(q)(\ddot{\varsigma} + K_d\dot{e} + K_p e + K_i\varepsilon) + \Upsilon(q,\dot{q}) \ . \tag{4.18}$$

We now return to the shape equation to determine what feedback linearization produces. Indeed, by assumption, both of the systems that we have been developing have fully actuated shape variables. That is, we have a direct force input corresponding to each component of s. Thus, letting

$$\Upsilon(q,\dot{q})_k \doteq \frac{1}{2}C_{kk_2k_1}\dot{s}_{k_2}\dot{s}_{k_1} + N + \frac{1}{2}p^t(\frac{\partial I^{-1}}{\partial s_k})p + \frac{\partial V}{\partial s_k} + \tilde{\tau}^t A_k - \tilde{\tau}_k^f \tag{4.19}$$

and

$$\Lambda(q) \doteq \tilde{M} \tag{4.20}$$

we see that the shape equation may be handled by the feedback linearization technique. This provides $\ddot{s} = u$ for the shape variable dynamics and system (3.33) becomes

$$\xi = I^{-1}p - A\dot{s} \ , \tag{4.21a}$$

$$\dot{p} - [ad_{I^{-1}p - A\dot{s}}]^t p = \tilde{\tau} \ , \tag{4.21b}$$

$$\ddot{s} = u \ . \tag{4.21c}$$

This is a surprisingly simple and elegant statement of the system dynamics. Note that

1. The equation for the shape is Newton's equation. Therefore, the shape is completely controllable and we can make the shape and the shape velocity whatever we desire.

2. The shape and its velocity drives the generalized momentum equation and the momentum, shape, and shape velocity in turn drive the Lie algebra element ξ. The Lie algebra element's flow determines the group variable g. This clear relation between system states is the beauty of the partial decoupling produced by reduction.

3. The two remarks above indicate the following. Should we so desire, we could view the shape s as a control variable. Using this control we could look to manipulate g. The first of equations (4.21) indicates the coupling between the shape and the group dynamics.

4. The shape equation caste as Newton's equation can be deceptive. It would seem that in this case, s is decoupled from g. This is not so if its true that one uses a feedback on the group variable as part of a control scheme. In that situation, the control u defined by the feedback linearization, and thus τ, would depend on the group. However, the partial decoupling remains intact.

With this being said, we state that the third of these notes provide the control philosophy that we follow. We seek to determine shapes such that locomotion of the mass-spring system and the serial-link structure is achieved.

4.4 Additional Matter

To draw an end to this chapter we illustrate our control philosophy of using the shape to elicit gait from our systems. Our illustration will make use of the mass-spring system. In the case of this system, the selection of a shape that elicits locomotion is a somewhat intuitive matter. We should simply rock the small mass back and forth. Or, in mathematical terms, we select the sinusoidal shape $s = a\sin(\omega t)$. For demonstration we take $a = \frac{1}{2}$ and $\omega = 2\pi$. Further, we make the selections of Table 4.3 for the system's physical parameters. Note that the table indicates two

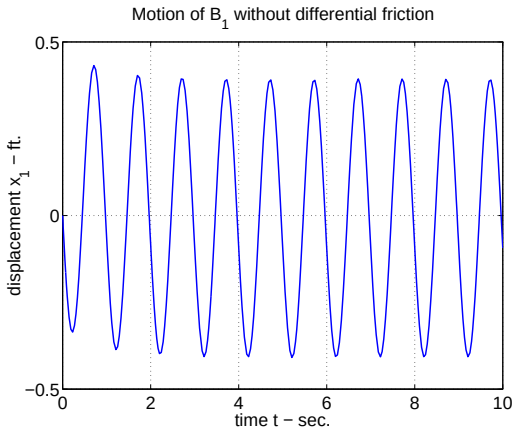
m_1 (kg)	m_2 (kg)	μ_b	μ_f	λ
1	3	10	$\frac{\mu_b}{6}/\mu_b$	0.5

Table 4.3: The mass-spring parameters

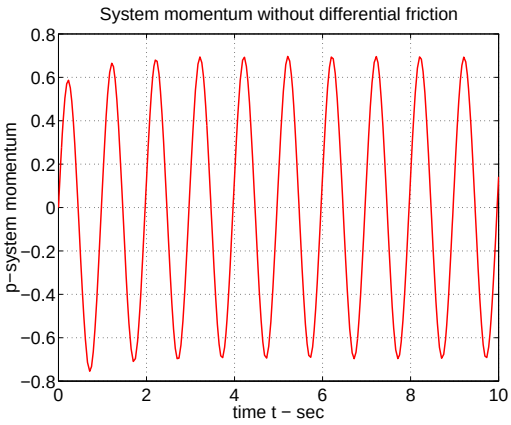
physical scenarios, one which uses a forward to backward friction differential and one that does not. The results of our feedforward shape selection are provided in Figure 4.9.

There are many things to be learned from simulations such as these. Firstly, note how essential it is for true displacement that there is a friction differential present. Obviously, if the substratum "feels" the same to B_1, rocking back and forth in a uniform manner does not result in any directional preference for displacement and the block simply oscillates about the origin of the global reference frame, as seen in Subfigures 4.9(a) and 4.9(b). Upon the introduction of a friction differential, B_1 doesn't lose nearly the amount of forward momentum to friction as backward momentum. Thus, upon displacing forward during the first half cycle of its motion, the block cannot completely return to its previous position, resulting in a net displacement thereof. This is illustrated in Subfigure 4.9(c). Also, there is a related net positive system momentum created by the friction differential, as seen in Subfigure 4.9(d). This effect comes about due to the fact that the momentum of B_1 is biased toward the positive direction, as its negative momentum is transferred to the substratum in the creation of heat and other dissipative effects. Though B_1 does attain a small negative momentum, this is easily offset by the forward momentum of B_2 as it returns to its forward thrust. Hence, the system as a whole retains a forward momentum.

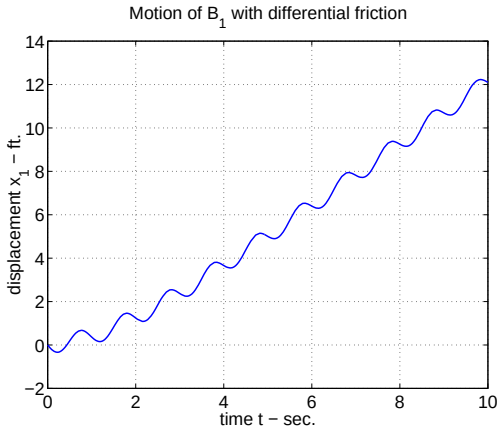
It is absolutely impossible for the B_1 to experience steady, continuous forward motion. Friction, being the only external force acting on B_1, would oppose motion, resulting in a continuous deceleration of the system. The only way to pump forward momentum back into the system is to "push backward" against the friction wall that has been created by the friction anisotropy of the surface of B_1. Similar observations



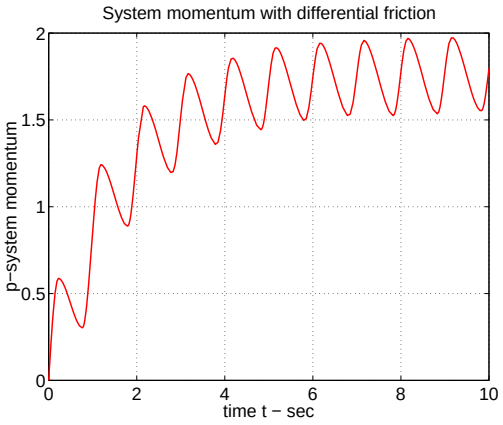
(a) Displacement of B_1 without differential friction.



(b) System momentum without differential friction.



(c) Displacement of B_1 with differential friction.



(d) System momentum with differential friction.

Figure 4.9: The resulting motion of the mass-spring system elicited through the sinusoidal feedforward shape selection.

will be made for the serial-like structure.

At this juncture we break off our study of the mass-spring system, focusing solely on the serial-link structure $\mathcal{S}$, which is the system of interest. Hopefully the mass-spring system made it an easy task for the reader to follow the model development up to this point and illustrates how this model form can be used to test shapes for their potential to produce locomotion. In the next chapter we discuss, as we did here for the mass-spring system, a feedforward gait eliciting shape for $\mathcal{S}$ and the principles underlying this shape and others that will produce locomotion.

Chapter 5

Gray's Tribute

5.1 Introduction

This chapter makes tribute to the first mechanical description of the lateral undulation mechanism of locomotion used by snakes, which was due to James Gray [28]. This tribute entails iterating the findings that he enunciated, expanding upon them, and placing them in the context of modern modelling and computation. This context will illuminate the validity of the qualitative and quantitative results of Gray. It would seem that Gray's work has gone unnoticed, misinterpreted, and in some cases unfairly criticized. This is most likely due to the fact that Gray appealed to free-body diagram type mechanical arguments and verbal descriptions of certain phenomena, rather than making use of mathematical formulae. However, a correct observation is a correct observation, no matter how it is expressed. We have reserved our judgement of Gray and by doing so have found his work to be solid, carrying with it a generality that we do not believe can be improved upon.

The chapter is divided into two logical portions. The first of these units is a statement of Gray's observations as we have come to understand them. We begin by explaining why it is the case that that any patterned motion observed in snakes must be due to periodic actuation. From there we explain in detail why the manipulation of

dry-friction cannot be the fundamental cause for steady bulk motion of the animal or $\mathcal{S}$. Although it may be redundant, for emphasis we also state this as a formal theorem and provide a proof for emphasis, as this point seems to be continually in confusion. Next, we revisit the free-body diagram analysis of Gray and the conclusions deduced thereby. These results are the necessary conditions for locomotion via undulation. Based on these criterion we recreate an example gliding form for snakes following Gray.

In the second portion of the chapter we make use of the serial-link structure and friction models to illustrate the validity and sufficiency of Gray's observations and criterion. To begin we describe how to mathematically formulate Gray's gliding form description in terms of the shape s of $\mathcal{S}$ and discuss several other shapes that have been proposed by early expositions on the matter of snake locomotion. As part of this shape construction process we point out that the highly celebrated clothoid gliding form contributed to Y. Umetani is in fact the same shape constructed by Gray. Additionally, we show that the sinusoidal shape proposed by Hirose and used everywhere today is nothing more than the smoothing of the clothoid shape by means of Fourier series expansion. Having acquired this smooth version of Gray's shape, we demonstrate that friction of the dry type cannot be used to elicit undulatory locomotion, even if the snake's skin has the highly anisotropic properties indicated by some research. Following this we introduce normal reaction forces due to lateral contacts, and the drag associated with them, by making use of viscous friction models. It is shown that these contacts allow lateral undulation to proceed as pointed out by Gray, thusly establishing sufficiency. This is expanded upon by carrying out an experiment designed by C. Gans [21, 22] to illustrate that dry friction is always debilitating to motion. Proceeding, we take a look into the roll played by the length of the snake and the parameters of the sinusoidal wave form adopted in gait generation. Included in these studies is the waveform phase shift parameter. We show that gait is lost as this shift vanishes, thus establishing the necessity of the first of Gray's

conditions. Finally, we show that tracking shapes that elicit gait produces actuation that is consistent with the remaining propulsive criterion of Gray, which concern the pattern of muscle contraction taken during locomotion.

5.2 Part I: Gray's Observations

5.2.1 The Necessity of Periodic Actuation: An Energy Argument

Consider the vertebrae-rib-actuator structure of the snake or snake-like mechanism $\mathcal{S}$. How is it possible to free the energy from muscles(actuators) for the purpose of locomotion? Like a spring, as muscle-like actuators are elongated, their potential energy is increased. This potential is released by compression. In the case of the snake it is necessary that, as one actuator associated with a joint compresses, the opposing actuator elongates. This elongation restores the potential to the system lost by the associated compression. Thus, by periodically compressing and elongating the musculature, the snake-like mechanism can continuously maintain a pool of potential energy and at the same time can continuously release energy to maintain the work required to overcome the frictional forces opposing displacement.

Being it the case that the shape or curvature of the snake or $\mathcal{S}$ is clearly determined by the compressions and elongations of the actuators, this observation alone explains the need for the shape variables to be periodic. Thus, it is seen that no theorems need to be stated or equations provided in order to make the claim that that the shape or the control inputs for $\mathcal{S}$ *must* be periodic in order for it to maintain a steady rhythmic gait. All the same, we will state this as a theorem, without proof, for emphasis.

Theorem 5.1. *The shape variables of the snake-like device must be periodic in order for the mechanism to achieve locomotion.*

These periodic shapes are referred to as or defined as *gaits* in [10, 12]. However,

we will not adopt this terminology.

5.2.2 The Inability of Snakes to Attain Steady Gait via Manipulation of Dry Friction

It is in no way possible for a snake or $\mathcal{S}$ to maintain a continuously forward bulk motion along some specified direction by manipulation of dry friction via bidirectional torque actuation about its joints. The reason for this is quite simple. If the bulk motion is steady along some direction, i.e., the velocity of this bulk motion is constant, then the net force acting on the system must be zero. That is, there can be no net frictional force. On the other hand, if all particles of the system are moving forward along some specified trajectory, then there must be a residual or net frictional force acting to oppose the motion in the direction of the bulk displacement. This is a contradiction. Hence, the conclusion is that if $\mathcal{S}$ is to attain gait, it must do so via external resistances other than friction, i.e., lateral constraints of the environment. Gray stated this conclusion in a very passive manner and contributed it to Mosauer [52]. We now state and prove this observation as a theorem in order to emphasize its definiteness and importance.

Theorem 5.2. *It is impossible for the serial-link structure $\mathcal{S}$ to attain continuously forward, steady displacement in a specified direction via the manipulation of dry-friction in the presence of free side-slip .*

Proof. To begin, let us identify the position of an arbitrary particle of the serial-link structure at time t via the arc-length parameter ϱ, where the length is measured from the head B_1 toward the tail B_n. Such a parameterization is possible being it true that the shape s at time t fixes the geometry of the structure for a given set of physical parameters. We denote this position $r(t, \varrho)$. The structure will be said to

move *continuously forward* in the direction $\vec{u} \in \mathcal{B}(0,1)$, if it is the case that

$$\langle \dot{r}(t,\varrho), \vec{u} \rangle > 0 \quad \forall t, \varrho \ .$$

We assume that the structure moves continuously forward. Under this supposition,

$$\langle F(t,\varrho), \vec{u} \rangle < 0 \quad \forall t, \varrho$$

and thus

$$\int_{\varrho} \langle F(t,\varrho), \vec{u} \rangle < 0 \quad \forall t \ ,$$

where F is the dry frictional force acting on the particle. This conclusion comes from the observation that F acts in the same direction but opposite sense of $\dot{r}$. Therefore, the net frictional force in opposition to motion in direction $\vec{u}$ cannot be balanced and consequently this motion cannot be steady, as there must be deceleration. □

Example 5.1. *As an example of Theorem 5.2, suppose that each point along the centerline of a snake follows the sinusoidal path* $r(t) = (t, \sin(t))^t$. *Then it follows that for the arc-lengths* ϱ_1 *and* ϱ_2, $r(t,\varrho_2) = r(t - t_d(\varrho_2), \varrho_1)$, *where* $t_d(\varrho)$ *is some time delay . See Figure 5.1 for an illustration. This figure may be viewed as a snapshot of the snake at a given time* t *or as a single period of the trajectory of a single point along the animal's centerline. The velocity of the motion along the path is given by* $\dot{r}(t) = (1, \cos(t))^t$ *and the average direction of bulk motion is given by the unit vector* $\vec{u} = (1,0)^t$. $\langle \dot{r}(t), \vec{u} \rangle = 1 > 0$. *Thus, the snake moves continuously forward along the direction* $\vec{u}$ *at the steady pace of* 1. *Now, further suppose that the particles of the system experience some type of dry friction, say da Vinci friction, so that*

$F = -\delta mg\mu \frac{\dot{r}}{\|\dot{r}\|}$. Then the net friction acting to oppose the motion in direction $\vec{u}$ is

$$\begin{aligned}
\int_{\varrho} \langle F\ , \vec{u} \rangle\ d\varrho &= -\rho g\mu \int_{\varrho} \frac{1}{\|\dot{r}\|} \langle \dot{r}(t, \varrho)\ , \vec{u} \rangle\ d\varrho \\
&= -\rho g\mu \int_{\varrho} \frac{1}{\|\dot{r}\|} \langle \dot{r}(t - t_d(\varrho), 0)\ , \vec{u} \rangle\ d\varrho \\
&= -\rho g\mu \int_{\varrho} \frac{1}{\|\dot{r}\|}\ d\varrho \\
&\leq -\rho g l \frac{\mu}{2}\ ,
\end{aligned}$$

where l is the total length of the snake and ρ is its mass density. So, there is a net opposing force implying deceleration in the direction of $\vec{u}$ for all time t. Hence, the velocity in the direction $\vec{u}$ cannot constantly be 1.

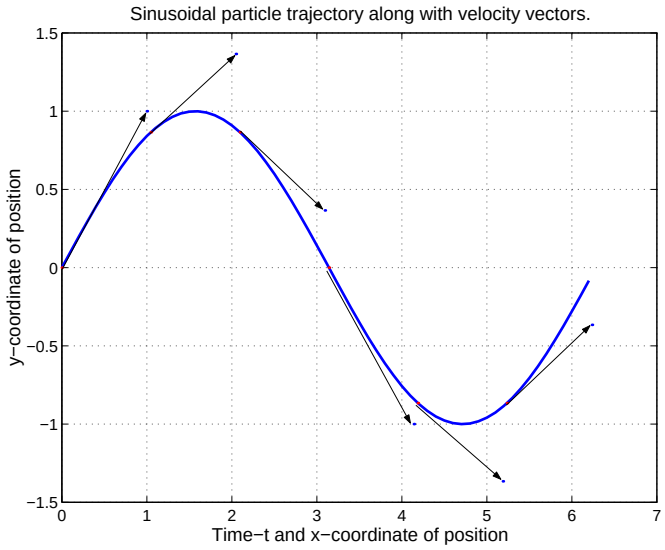


Figure 5.1: The trajectory of a particle travelling a sinusoidal path along with its velocity vectors.

5.2.3 How Normal Reaction Forces are Utilized by Snakes to Attain Steady Locomotion

Since we have established that dry friction cannot be at the root of steady propulsion, let us suppose that $\mathcal{S}$ is subject to some external resistances or constraints. How are the potential energies released by the actuation of the device transferred into net displacement via these constraints? This question was answered by Gray as follows. Consider the motion of the 3-link structure in the configuration shown in Figure 5.2. Suppose that lying normal to the center of mass of the three links are smooth rigid posts that can rotate about their z-axis. By Theorem 5.2, there exists a residual frictional force directed toward the posterior of the structure so as to oppose forward bulk motion. Call this force F^f. In order for a steady motion to take place in a particular direction, the following must be the case. The reaction forces attained through the interaction of the serial-link structure and the pegs must affect the system so as to cancel the friction force opposing motion. Label these reaction forces F_1^r, F_2^r, and F_3^r respectively in accordance with the links B_1, B_2, and B_3 that create them. Now assume that the moment τ_1 is selected so that the shape variable ϕ_1 remains constant. We seek to establish the conditions under which τ_2 may be selected so as to offset F^f.

To do so, we accept that the horizontal components of friction offset themselves over a completer period of actuation or shape due to symmetry. Thus we need only balance the reactions and F^f. Balancing the force components in the direction of motion, the lateral force components, and their moments about the joint between B_2 and B_3, provides for us the system of equations

$$-F^f - F_1^r \sin(\phi_1) + F_3^r \sin(\phi_2) = 0 \; ; \tag{5.1a}$$

$$F_2^r - F_1^r \cos(\phi_1) - F_3^r \cos(\phi_2) = 0 \; ; \tag{5.1b}$$

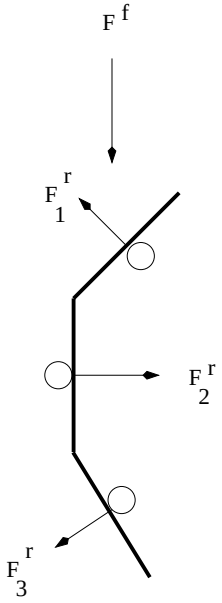


Figure 5.2: A free-body diagram for the motion of a 3-link system with rigid lateral resistance.

and

$$lF_2^r + lF_3^r - lF_1^r(1 + 2\cos(\phi_1)) = 0 \ . \tag{5.1c}$$

Solving the latter two of equations (5.1) for F_2^r and equating them provides us with the relation

$$F_1^r = F_3^r \frac{1 + \cos(\phi_2)}{1 + \cos(\phi_1)} \ .$$

Substitution of this relation into the first equation of system (5.1) and resolving this expression for F^f brings about the equation

$$\begin{aligned} F^f &= -F_3^r \left(\frac{1 + \cos(\phi_2)}{1 + \cos(\phi_1)} \sin(\phi_1) - \sin(\phi_2) \right) \\ &= -F_3^r \left(\frac{\sin(\phi_1) - \sin(\phi_2) + \sin(\phi_1)\cos(\phi_2) - \sin(\phi_2)\cos(\phi_1)}{1 + \cos(\phi_1)} \right) \\ &= -F_3^r \left(\frac{\sin(\phi_1) - \sin(\phi_2) + \sin(\phi_1 - \phi_2)}{1 + \cos(\phi_1)} \right) \ . \end{aligned}$$

It is then noted that $\tau_2 = lF_3^r$ and thus

$$F^f = -\frac{\tau_2}{l}\left(\frac{\sin(\phi_1) - \sin(\phi_2) + \sin(\phi_1 - \phi_2)}{1 + \cos(\phi_1)}\right) . \tag{5.2}$$

From this simple expression comes the most insightful observations concerning the locomotion of snakes to be made. Note the following necessary criterion for steady displacement:

Criterion 1. For the serial link structure $\mathcal{S}$ with $n = 3$, steady bulk motion *cannot* occur in a given direction if $\phi_1 = \phi_2$.

Criterion 2. If $\phi_1 < \phi_2$ and $\tau_2 > 0$, then steady forward motion along a given direction can be attained.

Criterion 3. If $\phi_1 > \phi_2$ and $\tau_2 < 0$, then steady forward motion along a given direction can be attained.

In reality, ϕ_1 need not be fixed and that τ_1 may serve to alter this angle rather than resist its change. Thus, the complete picture. For the 3-link structure, the shape variables must be periodic and must be out of phase in order for the structure to obtain steady locomotion in a specified direction by means of lateral reaction forces.

5.2.4 Using Gray's Criterion to Construct a Possible Gliding Form for Snakes

We proceed with observations that Gray used to arrive at an explanation via example of the sigmoidal form that snakes take on when exhibiting lateral undulation. Consider the sequence of configurations illustrated in Figures 5.3 and 5.4 for the serial-link structure $\mathcal{S}$ with $n = 3$ links. We label this configuration sequence C_t with $t = 1, \ldots, 12$. In configurations $C_1, \ldots, C_6$, $\phi_1 < \phi_2$. Thus, according to Criterion 2, τ_1 and τ_2 should be selected so as to increase the magnitude of the curvature ϕ_j in the

negative direction. In this way it is at least possible to sustain a steady bulk motion in the forward direction. In configurations $C_7, \ldots, C_{12}$, $\phi_1 > \phi_2$. In this case, to be in accord with Criterion 3, τ_1 and τ_2 should be selected so as to increase the magnitude of the curvature ϕ_j in the positive direction. Again, this will allow for the feasibility of a steady bulk motion. Assume then that the motion incurred by each of these configurations thrusts $\mathcal{S}$ into the successive configuration so that this sequence may be iterated. Then the structure could indefinitely retain steady bulk motion along a particular direction.

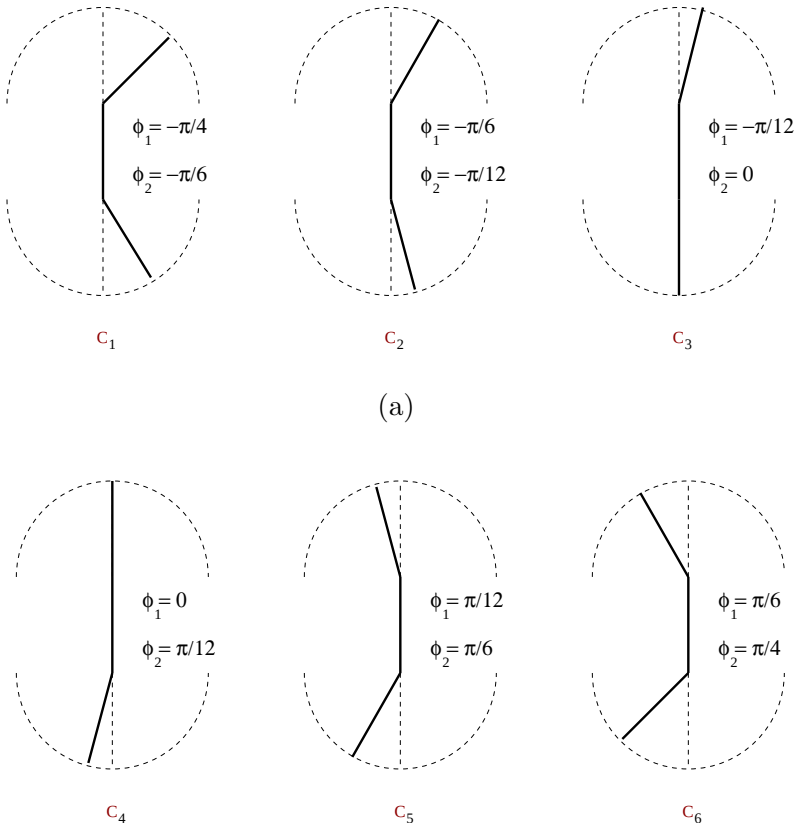
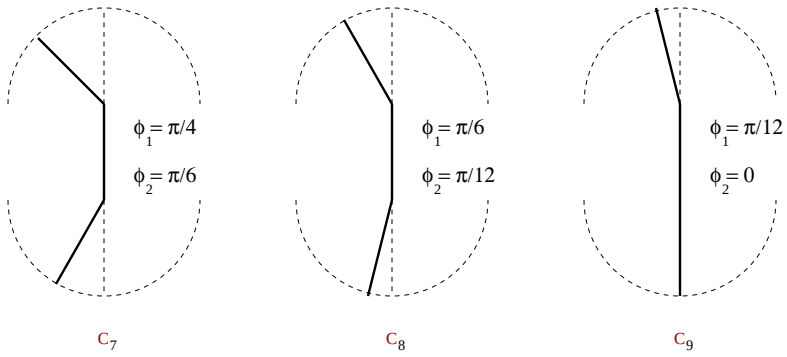
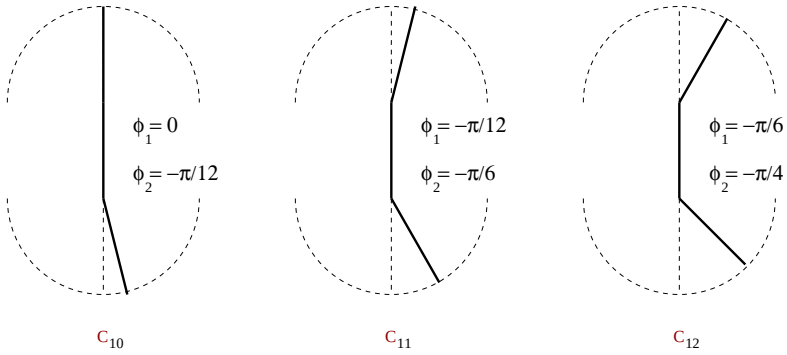


Figure 5.3: The first group of 6 out of 12 proposed configurations for locomotion. The curvature differential for these configurations is negative.



(a)



(b)

Figure 5.4: The second group of 6 out of 12 proposed configurations for locomotion. The curvature differential for these configurations is positive.

Continuing, we consider the serial-link structure $\mathcal{S}$ with $n = 14$. We select the shape of the system, $s^t = (\phi_1, \ldots, \phi_{13})$ such that successive triples of bodies; (B_1, B_2, B_3), (B_2, B_3, B_4), etc.; correspond to $C_1, \ldots, C_{12}$. By doing so, it is ensured that each portion of this structure can feasibly maintain a steady bulk motion. Furthermore, under these constraints it is seen that the result is a sigmoidal form. By assuming that each configuration provides the thrust needed to attain the following configuration, it is clear that the sigmoidal form will be repeatedly traced out by the structure. A snapshot of this form is illustrated in Figure 5.5. In this figure the

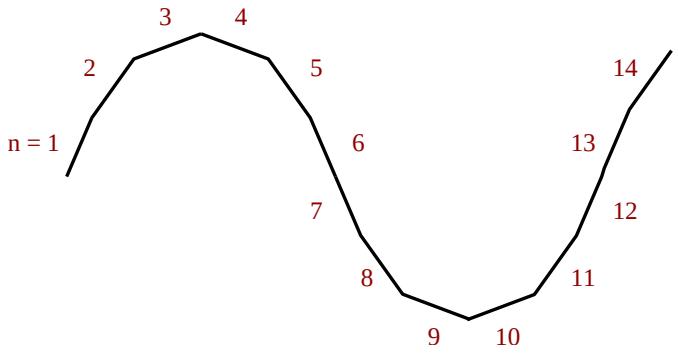


Figure 5.5: J. Gray's example gliding form.

progression of the configuration changes over time is assumed to be such that the head B_1 begins C_5. Following the head, B_2 begins C_6. This continues until B_{12}, B_{13}, and B_{14} comprise C_4. At this point it should be noted that the differential in the curvature of this curve, which is $\approx \phi_j - \phi_{j+1}$, is constant. Hence, this curve is not a sine wave. This is emphasized due to the fact that some researchers who have followed Gray's efforts, including Hirose [33], have falsely made the claim that Gray exclaimed the moving form of the snake to be a sine wave. Indeed, it is true that Gray used the term "sinusoidal" loosely and interchangeably with "sigmoidal". However, it is clear by context that the use of this term was not intended to imply the functional form $f(t) = \sin(t)$.

5.3 Part II: Validation and Completion of Gray's Work

5.3.1 Shape Forms

What is missing from the analysis of Gray? There are two primary, however related, components missing. Firstly, although Gray mentions that it would be possible to develop a dynamical model based on the principles that were set forth in his writing, he did not do so. Secondly, although he did provide some necessary conditions for the retention of steady motion, he did not demonstrate that they were in fact sufficient. It was assumed that by meeting the requirement in the curvature differential and by applying the torques τ_j in some appropriate way that each of the three link segments could simultaneously thrust forward and into a position such that the process could be repeated. This is was big assumption.

We are fortunate enough to have the agents to complete this study. Firstly, using the modeling tools of Chapter 3 we have already developed a dynamical model for the snake-like serial-link structure $\mathcal{S}$ that Gray described. Secondly, using our model along with the appropriate friction description and control formulation we can check the sufficiency of Gray's criterion. More specifically, being it the case that we have full controllability with respect to the shape of the structure, we may force the shape s to adhere to the configuration evolution described in Figures 5.3 and 5.4. Note however that our control ends there. We cannot ask that we have the specified shape and that appropriate torques be applied so as to cancel the net opposing frictional force. Gray did not know this. We shall have to hope that by simply keeping the shape criterion, the conditions concerning the torques is consequentially satisfied. With this being said, we will now complete the study.

To begin we shall need a functional form $s = G_s(t)$ for the shape configurations described by Gray as a function of time. This is not a difficult task. If we assume

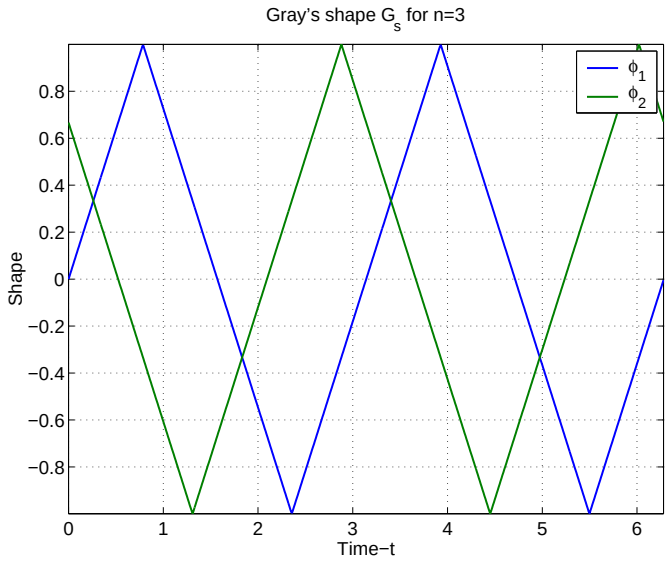


Figure 5.6: Gray's shape G_s for $n = 3$. Here $a = 1$ and $\omega = 2$.

that the C_t are time-uniform snapshots, then we may describe a period of ϕ_1 by

$$\phi_1(t) = \begin{cases} 2a(\frac{\omega}{\pi}t) \ , & \text{if } 0 \leq t < \frac{p_f}{4} \ ; \\ 2a\left(1 - \frac{\omega}{\pi}t\right) \ , & \text{if } \frac{p_f}{4} \leq t < \frac{3p_f}{4} \ ; \\ 2a\left(\frac{\omega}{\pi}t - 2\right) \ , & \text{if } \frac{3p_f}{4} \leq t < p_f \ , \end{cases} \tag{5.3}$$

where a is the amplitude of the wave, ω its frequency, and $p_f = \frac{2\pi}{\omega}$ its fundamental period. We define $\phi_1(t)$ on $\mathbb{R}$ by evaluating the restriction equation (5.3) at $t_{mod\ p_f}$. Note that the only difference between ϕ_1 and ϕ_j is a time delay. Thus, to define the remainder of the shape, we merely need to specify the appropriate time delay. To do this, we divide the fundamental period by n and expound that $\phi_j(t) \doteq \phi_1(t-(j-1)\frac{p_f}{n})$. $G_s(t)$ is now specified.

For a large number of links, $\mathcal{S}$ approaches a continuum form and its shape s tends toward the signed curvature of that form. By realizing this one can elicit many manner of shape functions that adhere to the necessary geometric conditions for motion by

starting from the continuum and moving toward the piecewise linear approximation of the form. Indeed, as shown in [62], one can actually work through a discretization over an arc-length parameterization of a form, calculating the relative angles determined by the resultant linear segments. However, this is really unnecessary. In fact, in some cases it is nearly impossible (try something as simple as sine). So long as the continuum form is periodic, all that is truly needed is the signed curvature thereof. This can be readily had in some cases. Once the curvature is determined as a function of time over a single period we may employ the same steps taken to define $G_s(t)$.

Being very specific, suppose that the path taken by a point along the form (due to the assumption that the form is periodic, all the points will follow the same path) is given by the smooth curve $r(t) = (x(t), y(t))^t$. Then the signed curvature is given by

$$K(t) = \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{\frac{3}{2}}} \ . \tag{5.4}$$

Taking $\phi_1(t) \doteq K(t)$ and $\phi_j(t) \doteq \phi_1(t-(j-1)\frac{p_f}{n})$, where again, p_f is the fundamental period of $r(t)$, a shape is defined.

Some care must be taken in applying this approach to formulating gait generating shapes. Firstly, if we select n small then $\mathcal{S}$ will not be an accurate piecewise linear approximate to the continuous form. Additionally, if the amplitude and frequency of r are too large, then the resulting shape variables ϕ_j can take on values outside of a desirable range and the links could begin to intersect one another. With this caveat as a prologue, we will develop some additional shapes using this technique that are important for comparison to G_s. We proceed directly to a relevant example.

Example 5.2 (The Sinusoidal Shape). *Suppose that we do believe that the form of a snake during steady undulatory locomotion is a sine wave. Then to determine the shape s of the snake we simply need to determine the curvature of the smooth curve*

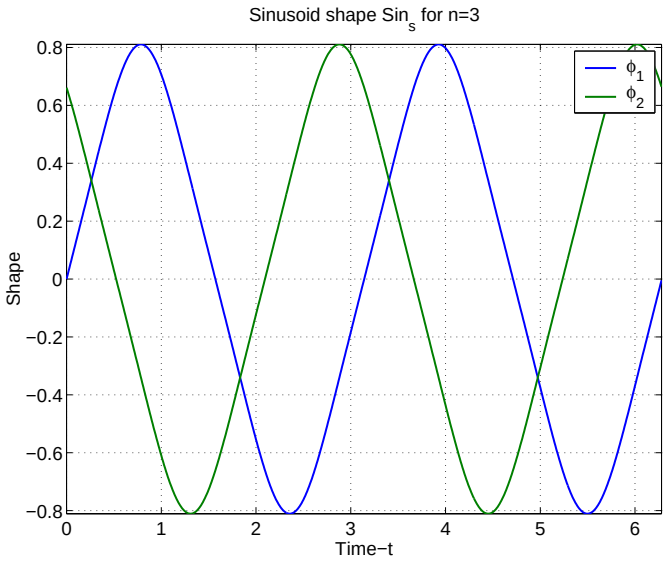


Figure 5.7: The sinusoid shape Sin_s for $n = 3$. Here $a = \frac{8}{\pi^2}$ and $\omega = 2$.

$r(t) = (t, a\sin(\omega t))^t$. *Using formula* (5.4) *we obtain*

$$K(t) = \frac{-\omega^2 a \sin(\omega t)}{(1 + (\omega a \cos(\omega t))^2)^{\frac{3}{2}}} \ .$$

To ensure that the shape angles do not become much too large due to the frequency of the sine wave, we select $a = \frac{\tilde{a}}{\omega^2}$. *In this way we obtain*

$$K(t) = \frac{-\tilde{a} \sin(\omega t)}{(1 + (\frac{\tilde{a}}{\omega} \cos(\omega t))^2)^{\frac{3}{2}}} \ ,$$

which was the expression used in related simulations. Now, let $\phi_1(t) \doteq K(t)$ *and* $\phi_j(t) \doteq \phi_1(t - (j-1)\frac{p_f}{n})$. *Then the sinusoidal shape, which we denote* $Sin_s(t)$, *is specified for any given serial-link structure size* n. *See Figure 5.7.*

Example 5.3 (The Composite Arc Shape). *As a second example of the technique, and an example that does not adhere to Criterion 1, we consider the composite arc shape. A shape that was once proposed as "the" gliding form of snakes. This shape*

is made up of two half circles pieced together. Consider the half circles

$$r_1(t) = \left(t, \sqrt{\left(\frac{p_f}{4}\right)^2 - \left(t - \frac{p_f}{4}\right)^2} \right)^t$$

and

$$r_2(t) = \left(t, -\sqrt{\left(\frac{p_f}{4}\right)^2 - \left(t - \frac{3}{4}p_f\right)^2} \right)^t \quad .$$

We make the following composite curve by adjoining these two pieces:

$$r(t) = \begin{cases} r_1(t) \ , & \text{if } 0 \leq t < \frac{p_f}{2}; \\ r_2(t) \ , & \text{if } \frac{p_f}{2} \leq t \leq p_f. \end{cases}$$

This form is a piece-wise smooth periodic function with fundamental period p_f. The form is illustrated in Figure 5.8.

It is well known that the circle has a constant curvature equal to the reciprocal of its radius. Thus, without appealing to equation (5.4) *or any other curvature formula, we have*

$$K(t) = \begin{cases} \frac{4}{p_f} \ , & \text{if } 0 \leq t < \frac{p_f}{2}; \\ -\frac{4}{p_f} \ , & \text{if } \frac{p_f}{2} \leq t \leq p_f. \end{cases}$$

These equations are defined only for a single period. As in the case of Gray's shape, we extend the definitions to $\mathbb{R}$ by evaluating the period formulations at $t_{mod\ p_f}$. Having the formula extension for the curvature of the composite arc wave we again make the definitions $\phi_1(t) \doteq K(t)$ and $\phi_j(t) \doteq \phi_1(t-(j-1)\frac{p_f}{n})$. Then the composite arc shape, which we denote $Arc_s(t)$, is specified for any given serial-link structure size n. This shape is illustrated in Figure 5.9.

Another method for producing possible gait producing shapes is simply to follow Gray's example and directly specify a curvature function. Working in this manner we do not need to have some functional expression for the continuum form. Of

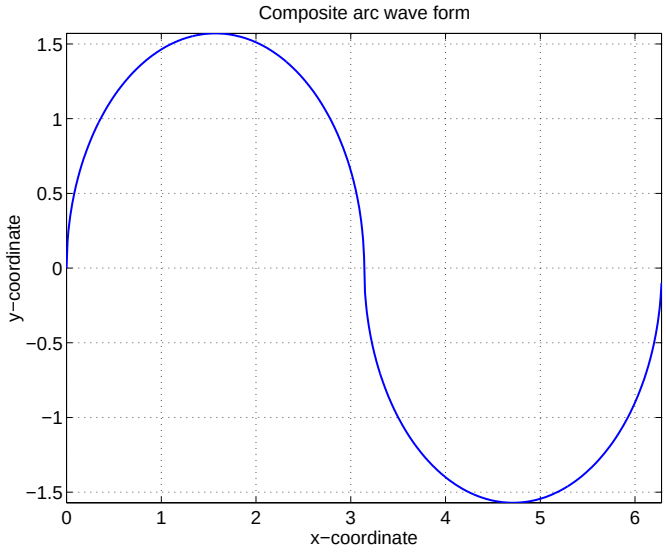


Figure 5.8: The composite arc gliding form. Here $p_f = 2\pi$.

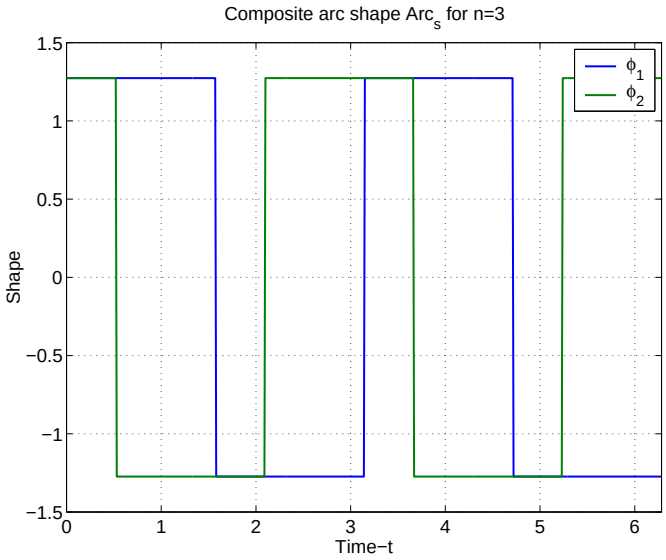


Figure 5.9: The composite arc shape Arc_s for $n = 3$. Here $p_f = 2\pi$.

course, upon specification of the curvature of a form, that curve is then specified up to position and orientation and one can use reversed engineering to determine the continuum functional form. However, for the determination of shape, the curvature is the important part and there is not a true need to find a functional expression for the continuum form. If we are truly interested in seeing it we merely need to make simulations with large n. Here are a few examples of simply specifying the curvature.

Example 5.4 (The Serpinoid Shape). *By specifying curvature directly Hirose developed his now well-known serpenoid curve and the so-called clothoid curve. Making the same observation that Gray made concerning the relationship between muscle contraction and the shape (i.e., curvature) of the structure or animal, Hirose decided that all of the forms previous to those proposed by himself were "unnatural". In particular, he pointed out that the piece-wise constant curvature, which leads to an arc or triangular shape, is discontinuous. Thus by correlation of this curvature to muscle contraction it would seem an unlikely action. The muscles would be performing a bang-bang control. This type of action, although often optimal for given criterion, is highly damaging to the structure that it operates on. He also pointed out that curvature related to the sine form is complex and that it would seem difficult to imagine muscular activity of this kind.*

To remedy the unnatural state of these suggested shapes, Hirose proposed two other curvature forms. The first of these he credits to the work of Y. Umetani. The continuum curve is described by Hirose as one for which the antagonistic muscles repeat contracting and relaxing motions alternatively every half cycle at a uniform speed. Or, a curve such that the curvature is continuous and the curvature differential is constant on a half cycle. The corresponding continuum functional form is called the clothoid curve and at the piece-wise level it corresponds to Curnu's spiral. However, note that Gray's suggestion G_s is continuous and increases and decreases at a steady rate. That is, G_s' is constant on each half cycle. So, in fact, this suggestion had already been made! ***Gray suggested the clothoid curve****! We emphasize this point*

because Hirose makes the claim in [33] that

> "After pointing out that no suggestions of great value concerning the creeping curve of the snake have previously been made, I proposed two types of curve, and stated that in terms of motor physiology these are natural and do not appear to be forced."

This was an extreme overstatement. One of the two curves that he introduced had already been suggested by the very person he claimed suggested the sinusoidal wave form.

Upon examining the clothoid curve, Hirose determined that this was still unnatural due to the fact that the curvature of this form has corners and is thus not differentiable at these points. To remedy this, Hirose suggested the form $\phi_1(t) \doteq K(t) = a\sin(\omega(t))$ *and* $\phi_j(t) \doteq \phi_1(t-(j-1)\frac{p_f}{n})$*. This way, the muscle activity would be smooth and non-complex. The continuum form corresponding to this selection of curvature is given by the equations*

$$x(\varrho) = \int_0^{\varrho} \cos(h(\bar{\varrho}))\ d\bar{\varrho} \qquad\qquad y(\varrho) = \int_0^{\varrho} \sin(h(\bar{\varrho}))\ d\bar{\varrho}\ ,$$

where $h(\bar{\varrho}) \doteq a\cos(b\bar{\varrho}) + c\bar{\varrho}$*,* ϱ *is the arc-length of the curve, and* a*,* b*, and* c *are parameters to be selected. Hirose called this curve the serpenoid curve.*

We make a comment on the serpenoid shape, which we shall denote S_s. This shape is truly the smoothing of G_s. This is now made clear. We begin with the statement of a well-known theorem.

Theorem 5.3. *Suppose that* f *is a piece-wise smooth periodic function with fundamental period* p_f*. The Fourier series of* f *is given by*

$$\tilde{f}(t) = a_0 + \sum_{k=1}^{\infty} \left(a_k \cos\left(\frac{2k\pi}{p_f}t\right) + b_k \sin\left(\frac{2k\pi}{p_f}t\right)\right)\ , \tag{5.5}$$

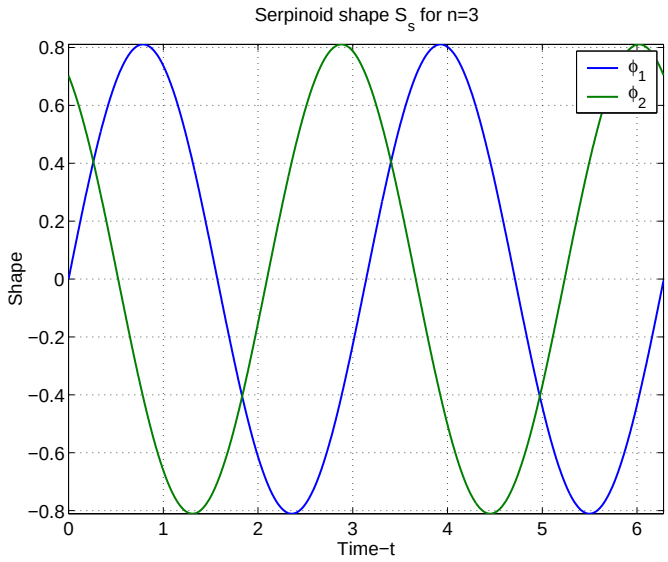


Figure 5.10: The serpinoid shape S_s for $n = 3$. Here $a = \frac{8}{\pi^2}$ and $\omega = 2$.

where

$$a_0 = \frac{1}{p_f}\int_0^{p_f} f(t)\ dt\ ,$$
$$a_k = \frac{2}{p_f}\int_0^{p_f} f(t)\cos\left(\frac{2k\pi}{p_f}t\right)\ dt\ ,$$
$$b_k = \frac{2}{p_f}\int_0^{p_f} f(t)\sin\left(\frac{2k\pi}{p_f}t\right)\ dt\ .$$

The Fourier series of f, $\tilde{f}(t)$, converges to $f(t)$ if f is continuous at t and to $\frac{f(t^-)+f(t^+)}{2}$ otherwise.

This theorem naturally implies that we may approximate the piece-wise smooth function f by the truncated series

$$\tilde{f}_{\bar{k}}(t) = a_0 + \sum_{k=1}^{\bar{k}}\left(a_k\cos\left(\frac{2k\pi}{p_f}t\right) + b_k\sin\left(\frac{2k\pi}{p_f}t\right)\right)\ .$$

Now, the component $\phi_1(t)$ of G_s is certainly continuous and piece-wise smooth peri-

odic with period $p_f = \frac{2\pi}{\omega}$. Therefore, it has a convergent Fourier series. Using *Maple* it is an easy matter to determine that the series coefficients are

$$a_0 = 0 \ , \tag{5.6}$$

$$a_k = A\frac{4}{\pi^2 k^2}\left(\cos\left(\frac{\pi k}{2}\right) - 1 - \cos\left(\frac{3\pi k}{2}\right) + (\cos(\pi\, k))^2\right) \ , \tag{5.7}$$

$$b_k = A\frac{4}{\pi^2 k^2}\left(\sin\left(\frac{\pi k}{2}\right) - \sin\left(\frac{3\pi k}{2}\right) + \cos(\pi\, k)\sin(\pi\, k)\right) \ . \tag{5.8}$$

So let us consider the approximation $\phi_1(t) \approx \tilde{f}_2(t)$. Using equation (5.6) we obtain

$$\phi_1(t) \approx \frac{8}{\pi^2} A \sin(\omega t) \ .$$

This is the form of the $\phi_1(t)$ for the serpinoid shape S_s! As a consequence of the definition of the remaining $\phi_j(t)$, the same observation holds for each component of Gray's shape G_s and Hirose's suggestion S_s. We could of course take more terms of the Fourier series for an approximation, but this added frequency content would result in what Hirose described as overly complex muscle activity.

This is interesting indeed, as it would seem that in fact no suggestions of great value concerning the shape of a snake during locomotion has been made since Gray! See Figure 5.11 for a comparison of the defining curves for G_s and S_s. It would also seem that for particular parameters the differences between the sinusoidal shape Sin_s and the serpinoid shape S_s are minute. This is illustrated in Figure 5.12. There is some difference between G_s, S_s, and Sin_s and it would seem that there is some data, such as that presented by Hirose in [33], that support the conjecture that the serpenoid shape comes closer to that of snakes than the other proposed shapes. However, it would not appear as though there is a significant difference, statistically or dynamically in these shapes. The intent of this comment is that Hirose's studies were not statistically designed experiments and although there were some differences in the data, there was no theoretical basis for interpreting the statistical significance of the

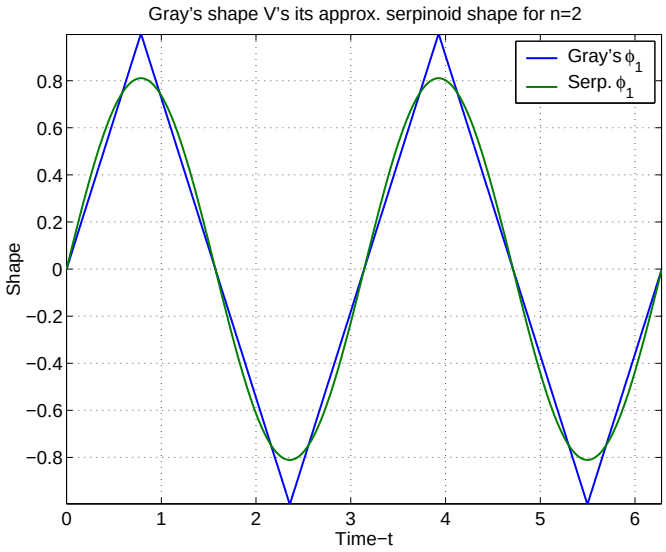


Figure 5.11: A comparison of the defining function for Gray's shape and that of its approximating serpinoid shape. Here $a = 1$ and $\omega = 2$ for Gray's shape and the amplitude is scaled appropriately by $\frac{8}{\pi^2}$ for the serpinoid shape.

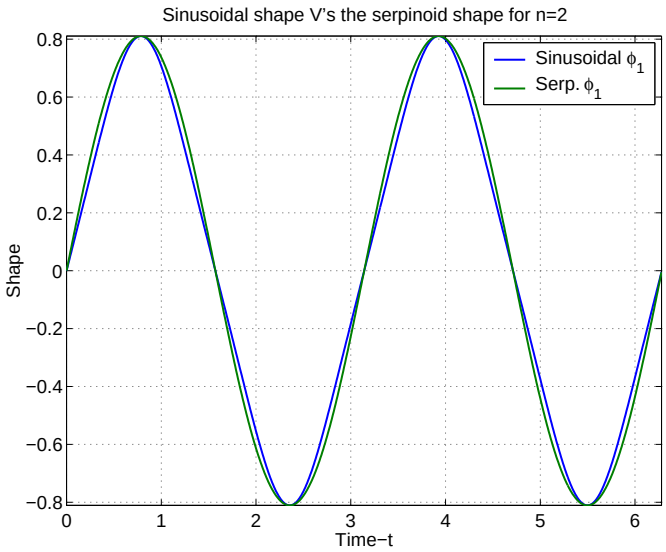


Figure 5.12: A comparison of the defining function for the sinusoid shape and that of the serpinoid shape. Here $a = \frac{8}{\pi^2}$ and $\omega = 1$.

results. Visually, the differences are so small it is doubtful that they are significant. Additionally, as we have confirmed through simulations, the locomotion achieved by Gray's, the sinusoid, and the serpinoid shapes do not differ significantly. Thus the mechanical remark. Nonetheless, the serpenoid shape is a helpful shape to utilize because it is a smooth version of Gray's shape and it has a simple functional form. Therefore, it is easier to deal with mathematically.

5.3.2 A Demonstration of the Inability of Snakes to Move via Dry-Friction

Now that we have S_s and understand its intimate relation to G_s, we may proceed to demonstrate some of Gray's points. We begin with the fact that $\mathcal{S}$ cannot attain steady forward motion along a particular direction in the presence of dry friction alone. Thus, Gray's conditions for motion are not sufficient upon removal of the lateral resistance assumptions used to derive them. This is so no matter how directionally dependent the skin or surface of the structure is. To show this we perform simulations using the equations of model (4.21), which are repeated here for convenience,

$$\begin{aligned} \xi &= I^{-1}p - A\dot{s} \ , \\ \dot{p} - [ad_{I^{-1}p - A\dot{s}}]^t p &= \tilde{\tau} \ , \\ \ddot{s} &= u \ . \end{aligned}$$

The following list summarizes the simulation set up:

1. Selecting $n = 7$, the shape S_s is tracked by $\mathcal{S}$ in the presence of uniform dry friction, elliptical dry friction with major axis in the normal directions, and the scale-like dry friction.

2. S_s has amplitude and frequency parameters. We will use $(a, \omega) = \left(\frac{8}{\pi^2}, 2\right)$. In this way, the shape corresponds to G_s with $(a, \omega) = (1, 2)$.

3. The physical parameters of $\mathcal{S}$ used for the presentation are $\rho = 6\ kg/m$, $l = 1/12\ m$, and $g = 9.8\ m/s^2$. For the sake of understanding what this means in terms of the links it is noted that each link has length $2l$. Thus, we are selecting the length to be $1/6\ m$. Using the conversion factor $1m \approx 3.28\ ft$ it is seen that each link $\approx 0.55ft$. Using this length, each link has mass $2l\rho = 1\ kg$. $1\ kg$ at gravity acceleration g provides a force of $\approx 2.20\ lb$.

4. We also require parameter values for the dry friction coefficient models. In all cases we use the simple linear elliptic model determined by equation (4.5), repeated here for convenience:

$$\mu(\gamma) = \|v_b^\sigma\|\mu_t\mu_n(\mu_n^2\cos^2(\gamma) + \mu_t^2\sin^2(\gamma))^{-\frac{1}{2}}\ .$$

In the uniform case we take $\mu_t = \mu_n = 0.5$. This value is along the order of those presented in Table 4.1. In this case equation (4.5) reduces to $\mu(\gamma) = \mu_t\|v_b^\sigma\|$ and the meaning of the word uniform is quite clear, as there is no orientational bias to this force. In the case of the elliptical frictional force we shall use equation (4.5) along with the parameter values of $\mu_t = 0.5$ and $\mu_n = 3$. Finally, in order to create the scale-like dry friction we add the gaussian term described in equation (4.6) to the elliptical model. The resultant model is

$$\mu(\gamma) = \|v_b^\sigma\|\left(\mu_t\mu_n(\mu_n^2\cos^2(\gamma) + \mu_t^2\sin^2(\gamma))^{-\frac{1}{2}} + \frac{\mu_b}{\max G}G(\bar{\gamma}, \sigma_\gamma)\right)\ .$$

In this model we use the parameters $\mu_t = 0.5$, $\mu_n = 1$, $\mu_b = 2.5$, $\bar{\gamma} = \pi$, and $\sigma_\gamma = 0.5$. This is consistent with the observations of [32, 29].

5. As explained in Chapter 4, numerical integration is required to calculate the dry frictional forces. For simulation we use the *MATLAB* routine *quad8*. This integration algorithm is a high order method that recursively employs an adaptive Newton-Cotes quadrature scheme to meet a specified relative and absolute

tolerance, *RelTol* and *AbsTol*. For our calculations, $tol \doteq [RelTol\ AbsTol] = [10^{-6}\ 10^{-5}]$.

6. We select zero initial conditions. With this selection the structure will initially experience 0 friction and will experience some bulk displacement. Additionally, it takes a few seconds for the shape tracking to converge. Thus, there is some transient behavior observed before the structure illustrates the steady-state behavior of the shape in the dry friction scenario. For the presentation we separate the transient and steady-state performance in the following manner. We run an initial simulation for 5 s in order to reach steady-state. We then take the terminal conditions q_t of this initial simulation, set $g_t = 0$, and perform another simulation spanning the time interval of $[5, 17]$. The repositioning of the group positions back to the origin provides us with a simple reference point for the resultant steady-state motion.

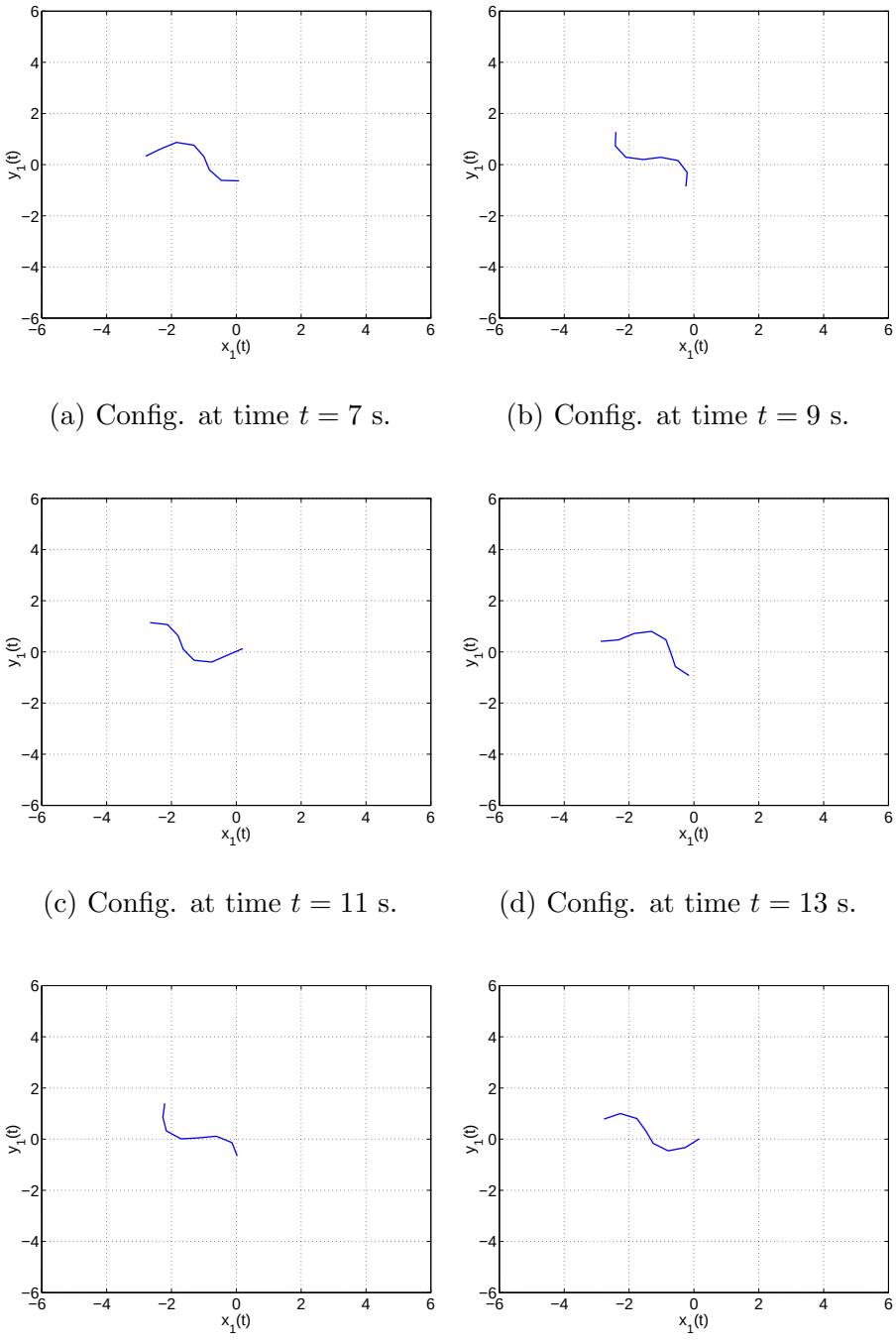
The collected evidence is now presented and interpreted. With respect to related figures, the trajectories of x_1 and p_1 are shown as solid "–" curves, the trajectories of y_1 and p_2 are dashed "- -" curves, and the trajectories of θ_1 and p_3 are displayed as dotted ":" curves. We begin with the uniform dry friction case. Table 5.1 provides some numbers while Figures 5.13 and 5.14 provide illustrations of the motion and associated quantities. Let us begin with the table. At none of the times corresponding to the 6 snap-shots presented is the position of the structure's head outside of the 1 ft ball about its steady state relocation to the origin. We see the futility of the undulations more clearly when viewing the snapshots of the entire structure corresponding to the table values. These are provided in Figure 5.13. The structure is not going anywhere once it settles in to its steady-state. If the snapshots provided are not sufficient to convince one of this, we refer the reader to Subfigures 5.14(a) and 5.14(b). In the first of these we see the initial activity of the group variables. Even during the transient behavior it is seen that the positions x_1 and y_1 return to the zero state after one fundamental period of π. As indicated in the second of these

subfigures, the head of the structure is simply swinging as the shape moves through its periodic evolution. Note that every π units the head returns to the origin with the same orientation. Perhaps more illustrative is the generalized momentum trajectories illustrated in Subfigures 5.14(c) and 5.14(d). There is no need to observe patterns here, we simply note that the magnitudes of these values are on the orders of 10^{-16} and 10^{-10} respectively, which is 0 in computational terms. Finally, Subfigures 5.14(e) and 5.14(f) illustrate the motion and velocity, respectively, of the head of $\mathcal{S}$ along the the average direction of the structure. The structure does not have any bulk motion along this direction whatsoever.

time t (s)	x_1 (ft)	y_1 (ft)	θ_1	$\|(x_1, y_1)^t\|$ (ft)
t =7	-0.19	-0.62	-0.04	0.65
t =9	-0.23	-0.58	-1.64	0.62
t =11	-0.04	-0.00	0.50	0.04
t =13	-0.36	-0.75	-0.69	0.83
t =15	-0.05	-0.40	-1.27	0.40
t =17	-0.05	-0.16	0.67	0.16

Table 5.1: Group variable values corresponding to the motion of a serial-link structure in a uniform simple dry friction environment. Related Figures: 5.13 and 5.14.

We now move on to the simulation results obtained for $\mathcal{S}$ when experiencing elliptic directionally dependent dry friction. Table 5.3 provides relevant data for the 6 snapshots of motion illustrated in Figure 5.15. As the table and figure indicates, it is again the case that at no time does the head of the structure venture beyond the bound of the 1 ft ball about the steady-state relocation to the origin. It would seem that even in the case of the directionally dependent dry friction, the structure cannot attain meaningful displacement. However, when we look a bit closer there is something different about what occurred in the presence of the directional dependence. We refer the reader to Figure 5.16. In Subfigure 5.16(b) it is clear that the head does not return to the origin of coordinates and in fact x_1 makes a steady climb away from the origin. This indication of motion is further supported by Subfigure 5.16(d).



(a) Config. at time $t = 7$ s.

(b) Config. at time $t = 9$ s.

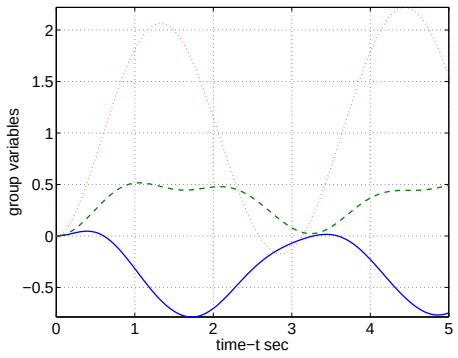
(c) Config. at time $t = 11$ s.

(d) Config. at time $t = 13$ s.

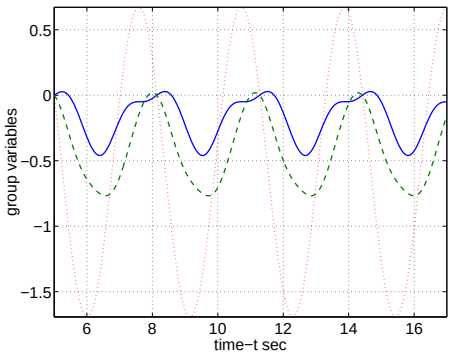
(e) Config. at time $t = 15$ s.

(f) Config. at time $t = 17$ s.

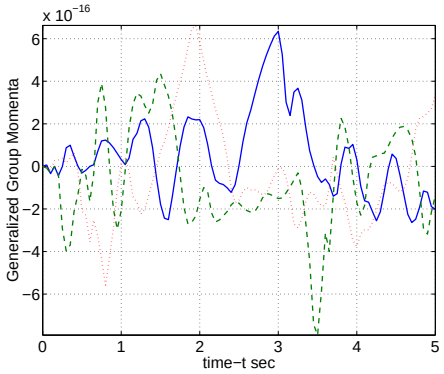
Figure 5.13: The motion of the serial-link structure tracking the Serpenoid shape S_s in a uniform simple dry friction environment. The friction parameters are $\mu_t = \mu_n = 0.5$ and the scale of the snapshots is 1 ft.



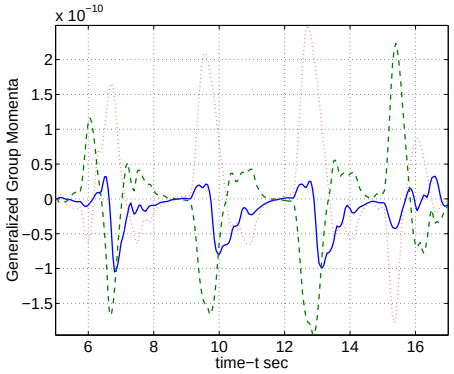
(a) Transient group behavior.



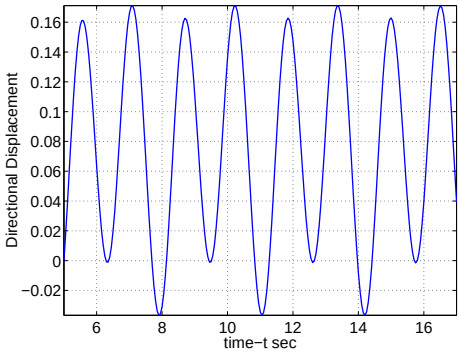
(b) Steady state behavior



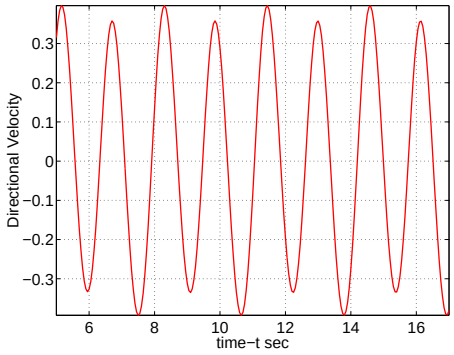
(c) Transient momenta.



(d) Steady state momenta.



(e) Directional displacement.



(f) Directional velocity.

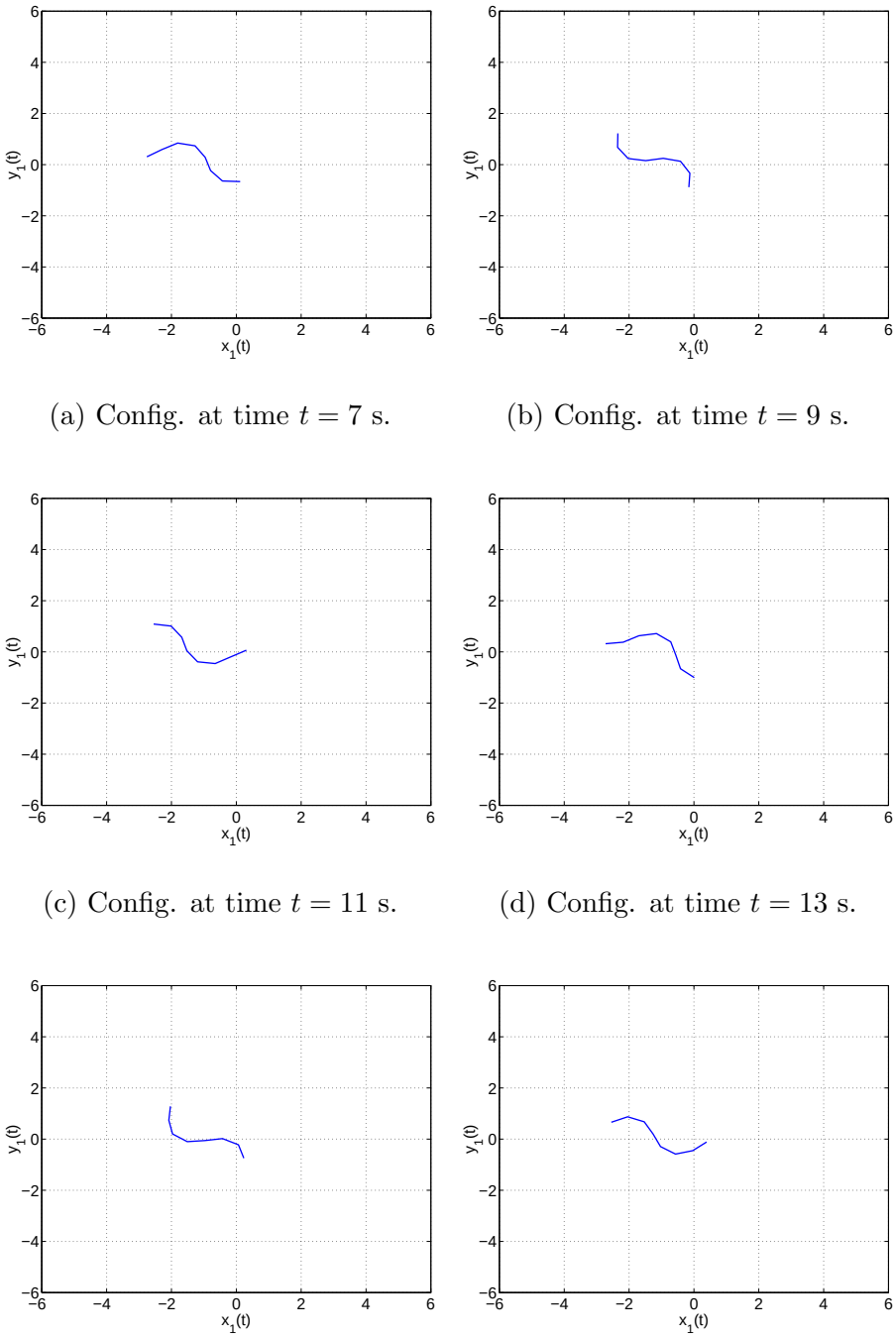
Figure 5.14: The motion of the serial-link structure tracking the Serpenoid shape S_s in a uniform simple dry friction environment continued. Individual features of the motion.

Note the positive generalized momentum component p_1 of the head. This says that the momentum of the system projected into the body frame of the head is positive along the direction tangent to the head. So, the system tends to move forward along the direction of head. However, continuous forward motion is not obtained and it is painfully clear that the directional dependence is not the impetus behind snake locomotion. We have selected friction in the normal direction to be 6 times that of the tangent direction and the payoff in momentum is almost negligible.

time t (s)	x_1 (ft)	y_1 (ft)	θ_1	$\|(x_1, y_1)^t\|$ (ft)
t =7	-0.15	-0.65	-0.04	0.67
t =9	-0.14	-0.61	-1.62	0.62
t =11	0.07	-0.07	0.49	0.10
t =13	-0.20	-0.83	-0.69	0.86
t =15	0.15	-0.48	-1.26	0.51
t =17	0.18	-0.28	0.67	0.33

Table 5.2: Group variable values corresponding to the motion of a serial-link structure that is subject to elliptic friction anisotropy. Related Figures: 5.15 and 5.16.

Interpretation of simulations for the scale-like dry friction is nearly identical. See Table 5.3 and Figures 5.17 and 5.18. The only slightly interesting difference is in the relatively calmer behavior of the generalized momenta as compared to the case of the elliptic dry friction. It would seem that the addition of the backward friction anisotropy does have a stabilizing property. This might indicate an auxiliary benefit of the morphology of snake skin, where, as mentioned in earlier discussion, a significant difference is seen between friction experienced by a snake skin along the forward and backward tangential directions. However, this friction differential is not a primary component of achieving some type of steady bulk displacement of the animal. It seems more likely that this anisotropy has been developed to aid in the effectiveness of those forms of snake locomotion that make use of static friction. Assuming that the shapes used in these simulations are reasonable models of shapes taken on by snakes, and they are, it would seem that Gray and his predecessors were absolutely correct in



(a) Config. at time $t = 7$ s.

(b) Config. at time $t = 9$ s.

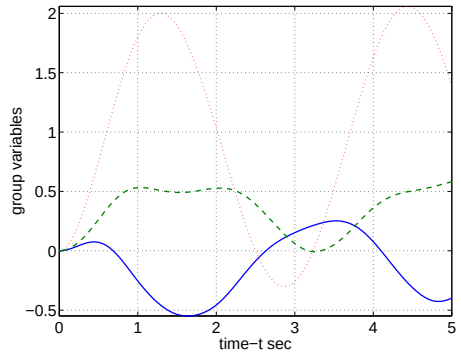
(c) Config. at time $t = 11$ s.

(d) Config. at time $t = 13$ s.

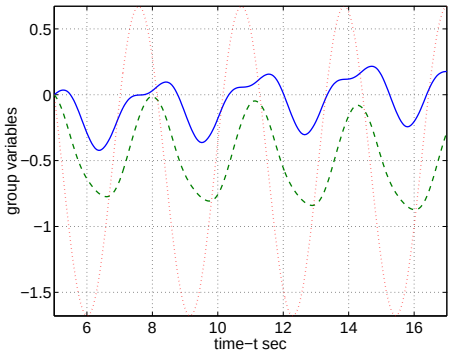
(e) Config. at time $t = 15$ s.

(f) Config. at time $t = 17$ s.

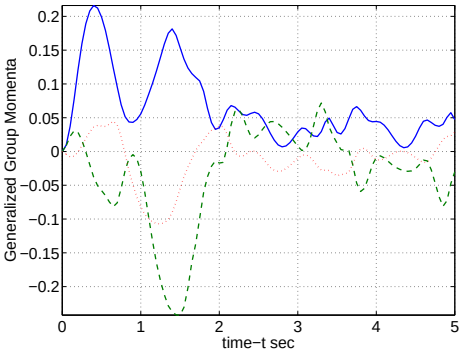
Figure 5.15: The motion of a serial-link structure that is subject to elliptic friction anisotropy. The scale of the snapshots is 1 ft.



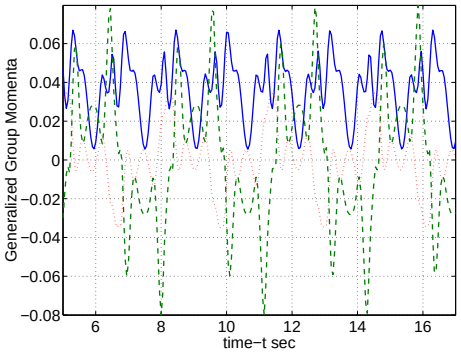
(a) Transient group behavior.



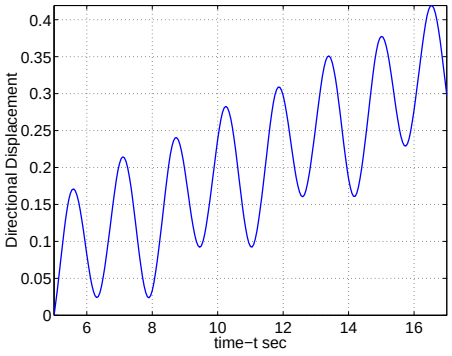
(b) Steady state behavior.



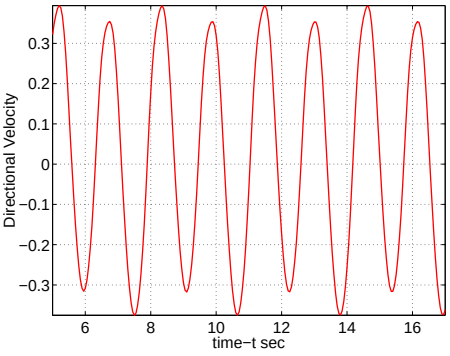
(c) Transient momenta.



(d) Steady state momenta.

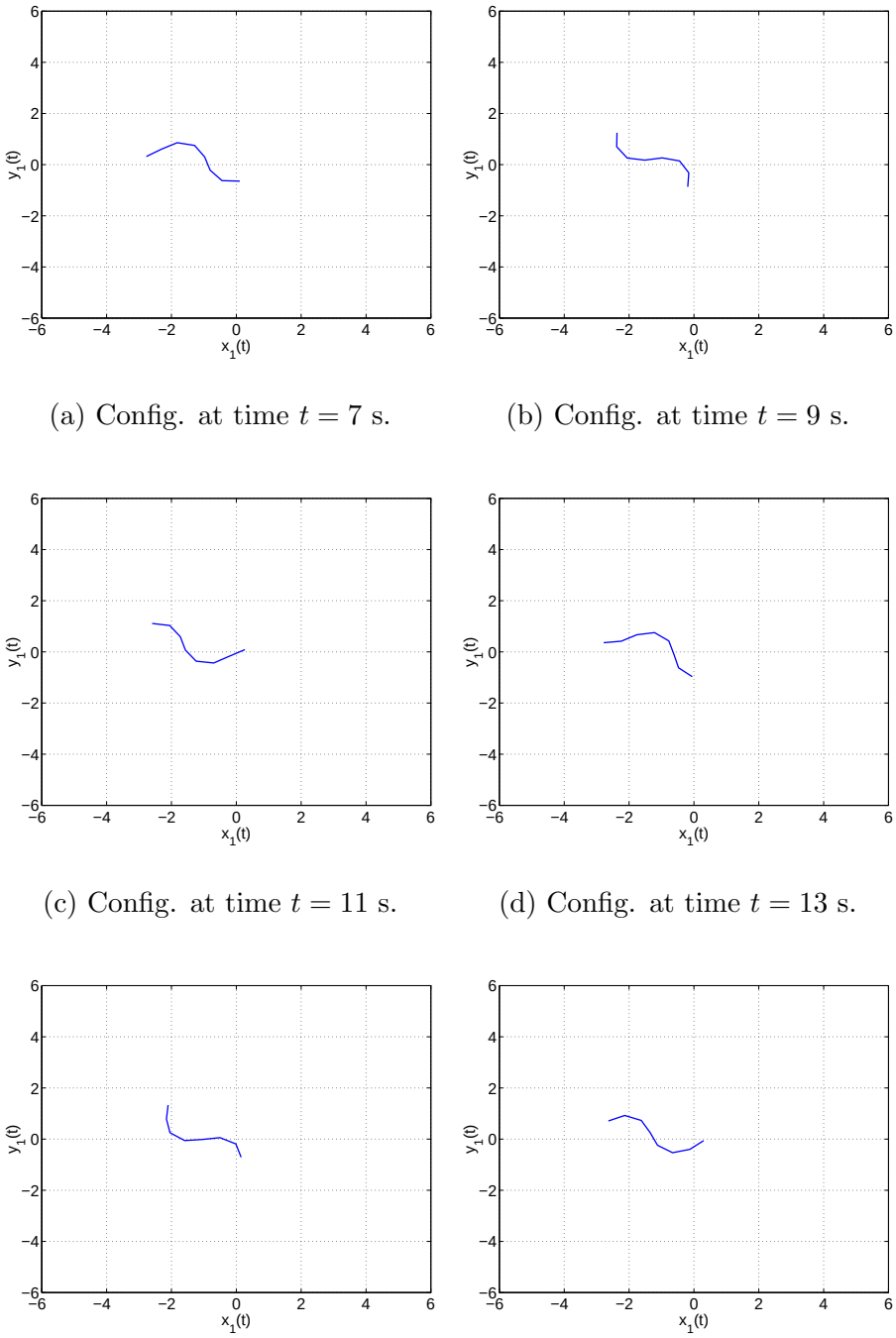


(e) Directional displacement.



(f) Directional velocity.

Figure 5.16: The motion of a serial-link structure that is subject to elliptic friction anisotropy, continued. Individual features of the motion.



(a) Config. at time $t = 7$ s.

(b) Config. at time $t = 9$ s.

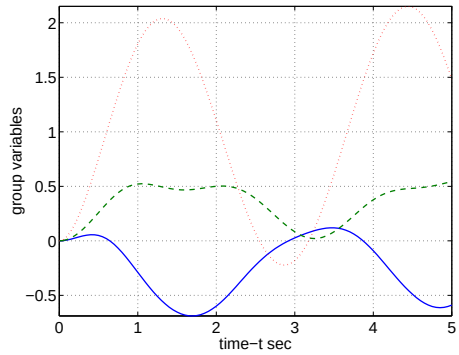
(c) Config. at time $t = 11$ s.

(d) Config. at time $t = 13$ s.

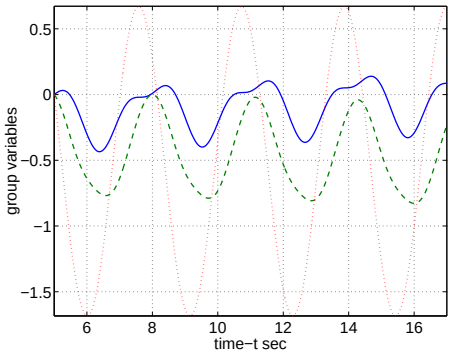
(e) Config. at time $t = 15$ s.

(f) Config. at time $t = 17$ s.

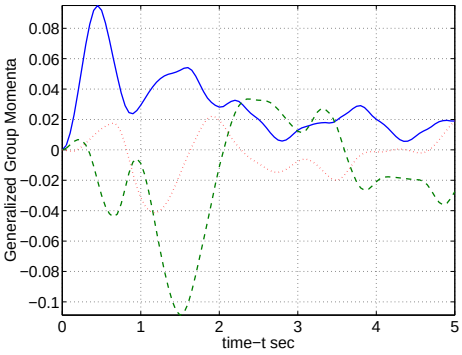
Figure 5.17: The motion of a serial-link structure that is subject to scale-like friction anisotropy. The scale of the snapshots is 1 ft.



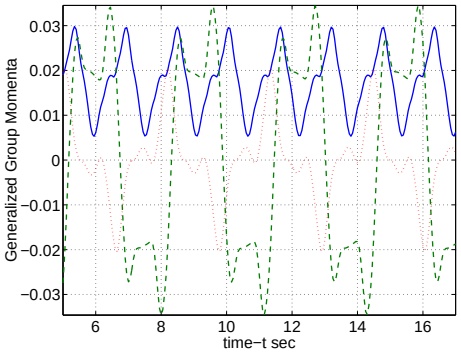
(a) Transient group behavior.



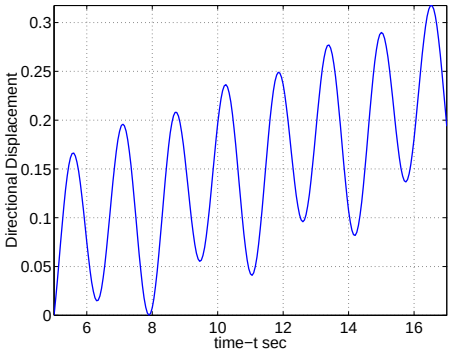
(b) Steady state behavior.



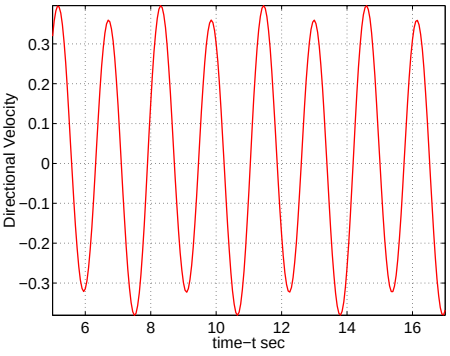
(c) Transient momenta.



(d) Steady state momenta.



(e) Directional displacement.



(f) Directional velocity.

Figure 5.18: The motion of a serial-link structure that is subject to scale-like friction anisotropy, continued. Individual features of the motion.

time t (s)	x_1 (ft)	y_1 (ft)	θ_1	$\|(x_1, y_1)^t\|$ (ft)
t =7	-0.17	-0.64	-0.04	0.66
t =9	-0.17	-0.59	-1.63	0.62
t =11	0.02	-0.04	0.49	0.05
t =13	-0.26	-0.80	-0.69	0.84
t =15	0.07	-0.45	-1.27	0.45
t =17	0.09	-0.23	0.67	0.25

Table 5.3: Group variable values corresponding to the motion of a serial-link structure that is subject to scale-like friction anisotropy. Related Figures: 5.17 and 5.18.

their statement that the motion of the snake is not primarily due to dry-friction or manipulation thereof via snake skin morphologies.

5.3.3 The Introduction of Normal Reaction Forces: Gait Obtained

As already discussed, it has been recognized that in order for $\mathcal{S}$ to move, it must be capable of eliciting substantial reaction forces normal to itself via interaction with significant protrusions from the substratum over which it moves. Examples of such protrusions being grass thickets, shrubs, and sand into which the body can partially embed itself. Gray likened these protrusions to smooth rigid cylindrical pegs.

The use of lateral resistance in propulsion generation can be observed when considering the locomotion of eels and fish. This motion occurs in a fluid, where viscous-like frictional forces are experienced. By viscous, it is meant that the pressure or reaction forces experienced by the object can be decoupled into those components that act normally and those that act tangentially. The pressure of a fluid in the normal direction acts like a wall that may be used to push against. The friction experienced by the object as the displaced fluid moves tangentially over its surface accounts for the tangent component. These viscous forces are indeed like the reaction forces that a snake could produce by interaction with a smooth peg or other fairly rigid lateral

protrusions. The normal force experienced by the snake will be approximately equal to the force it levies against the peg. Hence, the peg would act as a wall. Further, when the snake moves past the outgrowth, the dry friction resulting from their tangential contact will be minimal due to the smoothness of the peg and the anisotropy of the lateral scutes found on the animal. Thus, we may model the presence of these peg-like protrusions envisioned by Gray as viscous forces.

Models for viscous friction forces have already been produced and discussed in Examples 4.2 and 4.3 of Chapter 4. One of these is a da Vinci type friction law and the other is linear. As indicated in [62], from whence these expressions originated, these models behave similarly. Thus, for simplicity, we shall make use of the linear model of Example 4.2. Relating this model directly to the matter at hand, c_t will indicate the smoothness of the protrusions and the lateral portion of the snake or serial-link structure. c_n will be used to indicate the strength of the normal reaction. If this parameter is large, then the protrusion is like a wall. If smaller, then this "wall" either has some give or elasticity to it or may actually slide slightly when subjected to large contact forces.

Now we return to the simulation process, replacing the dry friction model with the viscous friction model. For comparative purposes, we use the same values for the parameters c_t and c_n as used for μ_t and μ_n, respectively, during the simulations with the elliptic dry friction model. In this way, the only difference between these models will be the direction in which the friction force acts.

Presented in Table 5.4 is the data for the coordinates of B_1 corresponding to the 6 motion snapshots illustrated in Figure 5.19. Indeed, the data would appear to indicate that a steady bulk motion is achieved in the presence of lateral reaction forces. This can, of course, be seen with ones own eyes by viewing the snap-shots. The most revealing evidence, however, is illustrated in the subfigures of Figure 5.20. As indicated in Subfigures 5.20(d) and 5.20(e), respectively, the structure has a steady momentum preservation with mean of ≈ 0.8 and the head travels $\approx 6\ ft$ along the

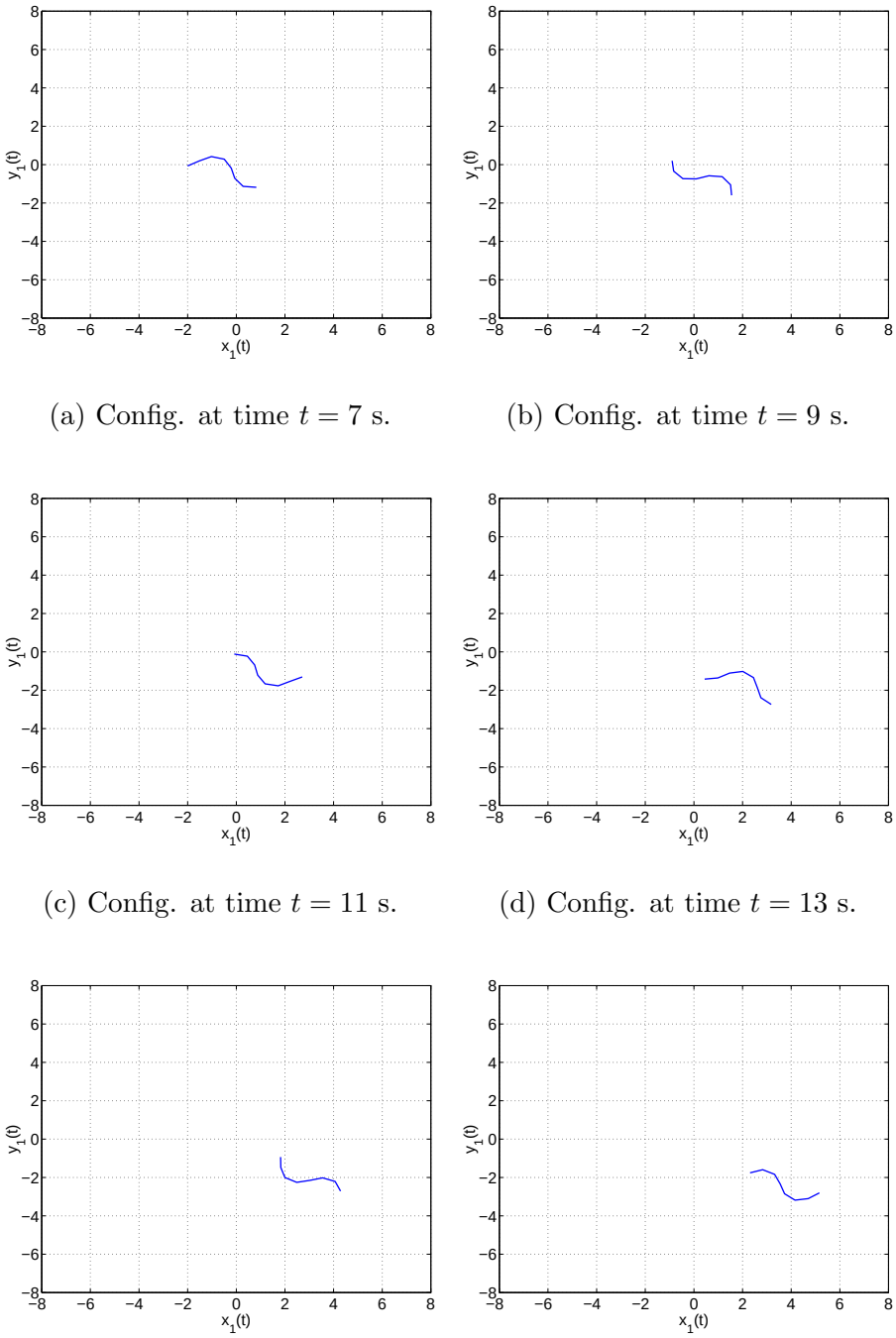
time t (s)	x_1 (ft)	y_1 (ft)	θ_1	$\|(x_1, y_1)^t\|$ (ft)
t =7	0.55	-1.16	-0.09	1.28
t =9	1.53	-1.33	-1.49	2.02
t =11	2.46	-1.42	0.43	2.84
t =13	2.96	-2.56	-0.68	3.92
t =15	4.17	-2.46	-1.15	4.84
t =17	4.93	-2.95	0.58	5.75

Table 5.4: Group variable values corresponding to the motion of a serial-link structure that is subject to lateral forces. Related Figures: 5.19 and 5.20.

steady-state direction of motion.

Furthering the discussions, not only is it the case that dry/sliding friction is not at the root of snake-like locomotion, it is in fact the case that the presence of dry friction, both ventral and lateral, will only serve to retard the potentials for motion. To demonstrate that this is so we shall now perform, through simulation, an experiment described by Carl Gans in [22]. He wrote

> "*...Consequently, lateral undulation might be expected to proceed most effectively in a frictionless system. I tested this concept by constructing a field of cylindrical, high friction, rough surfaced pegs, each of which rotated on adjustable pin bearings. A small indicator pin on each peg showed those in contact with the snake. The horizontal surface could be lubricated with wax and the bearings could be tightened to make the pens more difficult to turn. The effective friction of each peg opposing forward progress increased drastically when the propulsive contact between snake and pen shifted from rolling to sliding; snakes moved freely across the field when the pegs spun loosely, but were significantly slowed as the bearings were tightened or the floor roughened. Hence friction at the ventral surface and the lateral points definitely decreased the speed and increased the muscular effort needed for progression.*"



(a) Config. at time $t = 7$ s.

(b) Config. at time $t = 9$ s.

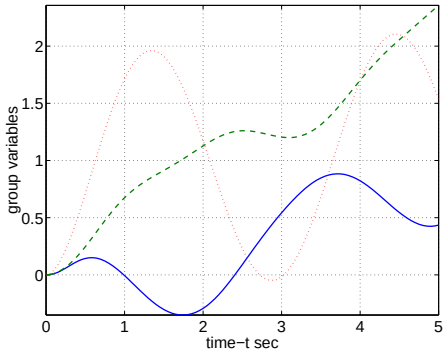
(c) Config. at time $t = 11$ s.

(d) Config. at time $t = 13$ s.

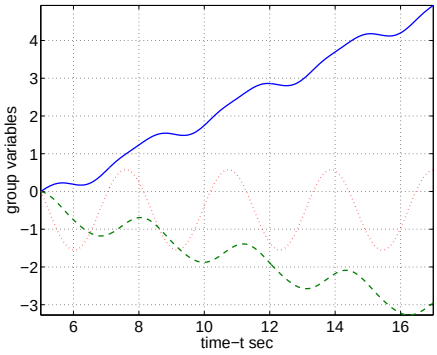
(e) Config. at time $t = 15$ s.

(f) Config. at time $t = 17$ s.

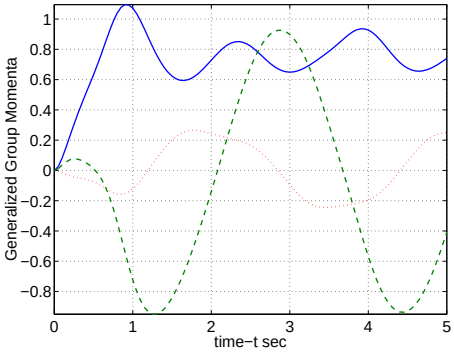
Figure 5.19: The motion of a serial-link structure that is subject to lateral forces. The scale of the snapshots is 1 ft.



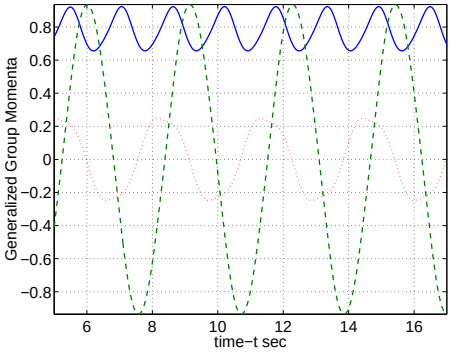
(a) Transient group behavior.



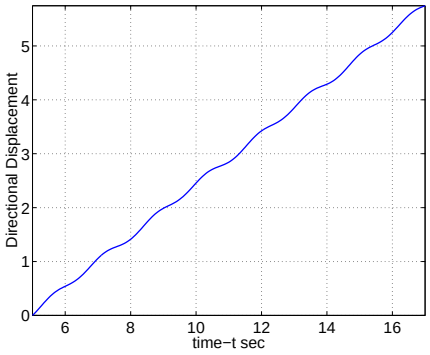
(b) Steady state behavior.



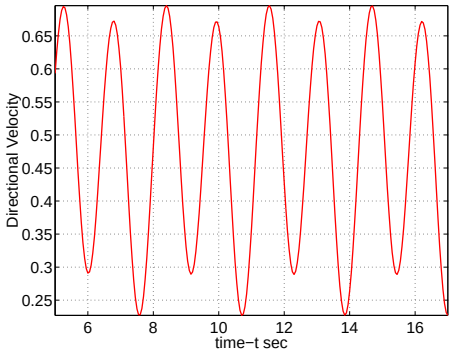
(c) Transient momenta.



(d) Steady state momenta.



(e) Directional displacement.



(f) Directional velocity.

Figure 5.20: The motion of a serial-link structure that is subject to lateral forces, continued. Individual features of the motion.

Gans cites his own work, in particular [21], as a reference relating to this experiment. Interestingly enough, however, when going to this reference Gans cites personal observations as the basis for these claims and does not provide any details (i.e. data). Undoubtedly, Gans did observe these phenomena. But simply for scientific edification we shall reinvent this experiment by using the models we have developed.

We shall use the linear viscous friction model of Example 4.2 to model both the normal reaction forces of the pegs in Gans' experiment and the friction due to tangential contact between the snake and the peg. The elliptical and scale-like linear dry-friction models are used to describe the ventral friction experienced by the snake. Thus, we combine these two models to determine an appropriate expression for the reduced frictional force $\tilde{\tau}$. Instead of tightening the bearings of pegs, we simply increase the value of c_t in the viscous friction model. Also, we have the added flexibility of being able to change the resistance imposed by the post via c_n. We will not use this flexibility. Opting instead to make the resistance quite large so as to imitate the fixed peg. Further, instead of waxing or roughening the substratum on which the snake travels, the values of μ_t, μ_n, and μ_b may simply be reduced or increased in a fitting manner. Note also that, in our case, there is always a virtual peg normal to each segment of the snakes vertebrae, like swimming in a fluid. Thus, there is no need to worry with issues such as variability due to peg spacing, etc.

Being more specific, during the first half of the simulated experiment it is assumed that the substratum on which the structure/animal lies is slick. By slick we mean that the structure experiences negligible ventral friction. Under these conditions the posts are fixed by selecting $c_n = 10$ and then we observe what occurs by successively tightening the pegs, incrementing c_t uniformly from 0 to 10 with step size $\Delta c_t = 1$. The average of the first component of the generalized momentum p_1 is used as the measure of how well the snake is capable of moving. Since the parameters of the shape used do not change from trial to trial, this is a legitimate measure of progression.

The findings of this experiment are shown in Figure 5.21. The results are con-

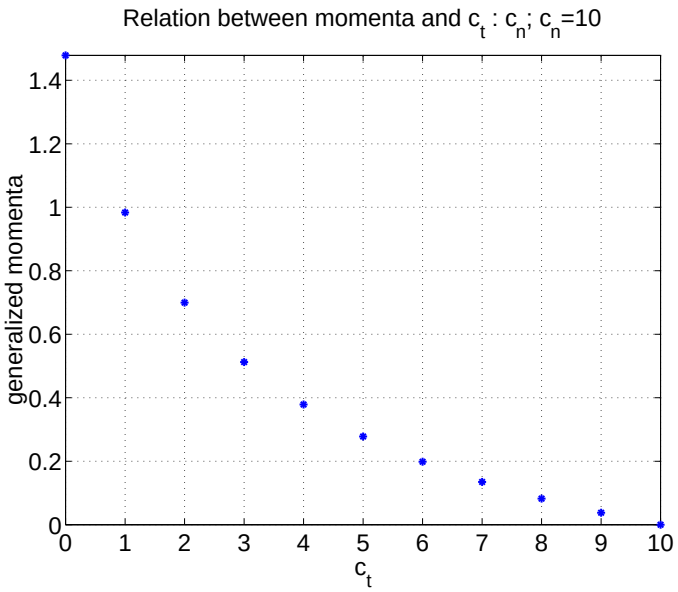


Figure 5.21: The debilitating effects of lateral friction.

sistent with Gans' observations of actual snakes. As the pegs are tightened, the locomotive capability of $\mathcal{S}$ is lost. Our results obviously fall out of the realm of the peg analogy after c_t is increased beyond a particular point. For if the actual pegs were completely tightened, c_t would not take on the value of 10 and thus the snake would still be capable of movement, albeit impaired. It is interesting to note just how sharply the momentum is decreased upon the first few increases of c_t. As pointed out in the literature, this should probably explain the morphology of the scutes present on the lateral portion of most snakes. It would seem that these scales are designed to effectively minimize c_t.

For the second portion of the experiment the focus is shifted to the debilitating effects of ventral friction. Thus we completely loosen the pegs, maintaining the settings $c_t = 0$ and $c_n = 10$ and increment $\mu = \mu_t = \mu_n$ from 0 to 10 with step size $\Delta\mu = 1$. Hence we are applying a uniform dry friction. Being it so that no true locomotive advantage was gained by use of the elliptical and scale-like frictions in our study

heretofore, it seems appropriate to do so. Again, the average of the first component of the generalized momentum vector is used as a measure of motion capability.

The findings of this experiment are shown in Figure 5.22. Just as in the case of the

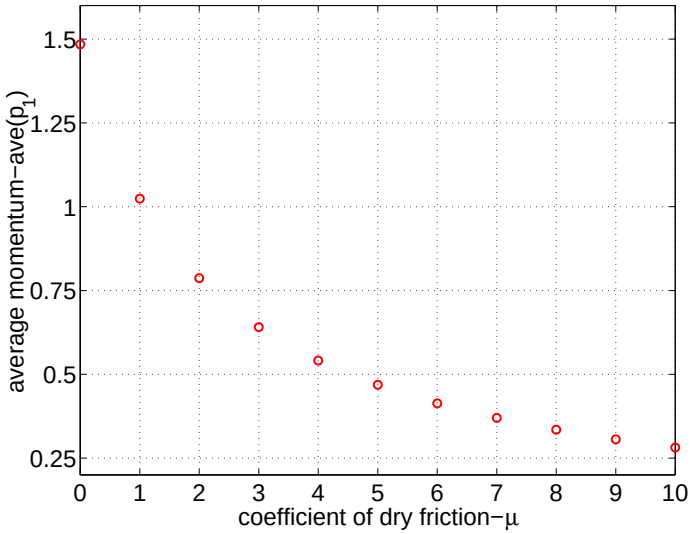


Figure 5.22: The debilitating effects of ventral friction.

presence of lateral friction, ventral friction is seen to have drastic effects on locomotive capability. Unlike the case with lateral friction however, when the coefficient of ventral friction becomes 10, thus enabling the friction to have an influence equal to that of the normal resistance, locomotion is not completely lost. It is the case though, that the average momentum becomes so low that the capability is effectively lost.

Based on these results the following conclusion is clear. The lateral undulation mode of snake locomotion is most effective in a frictionless environment.

5.3.4 The Need for a Phase Shift in Shape and Small Investigations Into Other Parameter Effects

In the previous section we established the plausibility of the assertion that a snake could only acquire gait in the presence of normal reaction forces and furthered this observation by showing that the motion proceeds more effectively in a low friction environment. With that being established we proceed to demonstrate the necessity of the phase shift applied along the shape variables for the acquisition of gait. That is, we illustrate the correctness of Criterion 1.

The serial-link structure dynamics for $n = 3$ are utilized. In this way we avoid the nonphysical situation of having the links interact with one another as the phase shift between the adjacent links is diminished. Beginning with a uniform phase shift along the structures gliding curve, this phase shift will be decremented exponentially. Being more specific we use the alteration of S_s shape defined by $\phi_1(t) \doteq K(t) = A\sin(\omega(t))$ and $\phi_j(t) \doteq \phi_1(t-(j-1)\frac{p_f}{(n-1)^k-1})$, with $k = 1, \ldots, 10$. This will allow examination of the resultant effect on the system's capability to elicit locomotion. This capability will again be measured by the average value of the steady state momentum component p_1. In order for this structure to be given as much of an advantage for successful gait as possible, we allow it to move in a frictionless environment with stiff lateral resistance. i.e., $\mu = 0$, $c_t = 0$ and $c_n = 10$. Such will be the case for the remainder of the investigations presented in this section unless otherwise stated.

The data acquired from this small study are provided in Figure 5.23 along with the spline interpolation thereof. As predicted by Criterion 1, if there is no phase shift between the differential angles of the structure, the resistance to the structure's movement cannot be equalized and the capability to maintain momentum is thus lost. One of the more interesting features of the data is the initial increase in average momentum with the decrease of phase shift. This suggests that for a given set of prefixed system and shape parameters, there is an optimal phase shift. Although this data does not illustrate it, our personal observations have shown that the optimizing

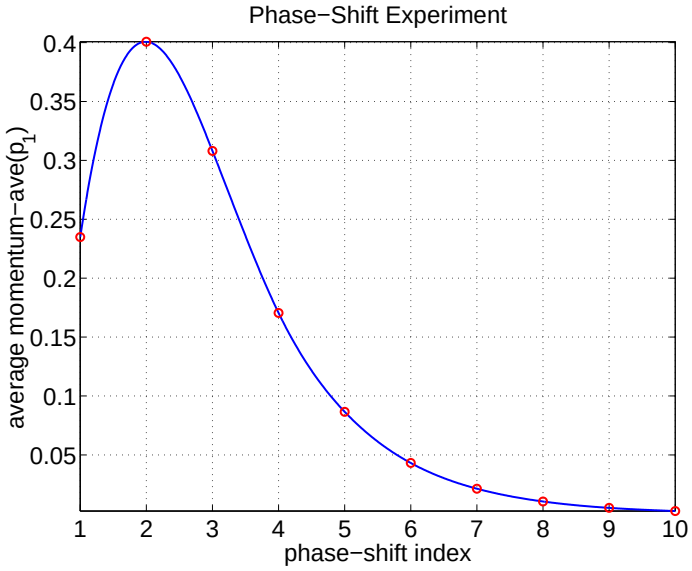


Figure 5.23: The loss of motion experienced as the shape tends toward a composite arc. i.e., a shape with zero phase-shift.

phase shift value is certainly different than the uniform selection that has been made from the onset of the shape definitions. Of course, this optimizing value is dependent on the other system and shape parameter values. One fact that remains constant, however, is Criterion 1.

Attention is now turned to another of the shape parameters, the frequency parameter ω. The relationship between locomotion and frequency of the shape variables should be quite obvious. If the structure moves 3 ft at a given frequency over a specified time period, then a distance of 6 ft can be travelled during the same span of time by doubling that frequency. That is, of course, given that the external resistances are in fact rigid and capable of returning the additional force exerted on them. All the same, for the sake of completion, we shall demonstrate that this is indeed the case. We vary ω over the range 1 to 2.5π by increments of $\Delta\omega = 0.25\pi$. At each of these values the average of p_1 is calculated. The resulting data is presented in Figure 5.24. There is nothing unexpected or distinguishing about this result. However, we would

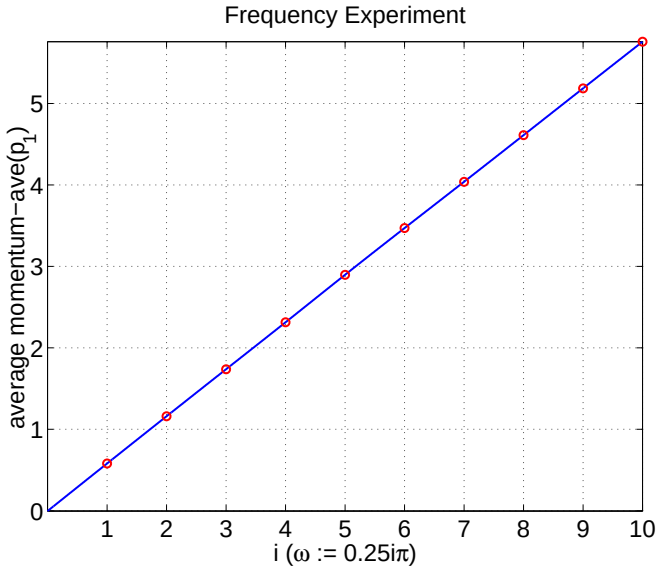
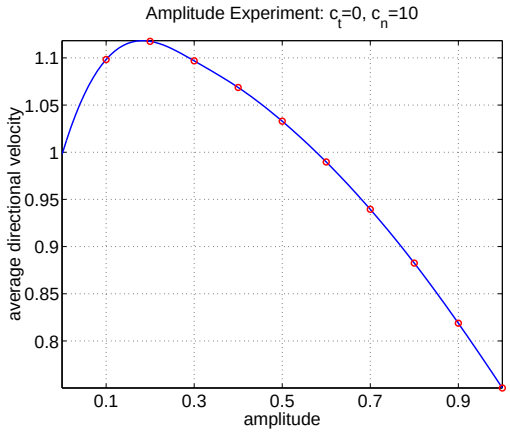


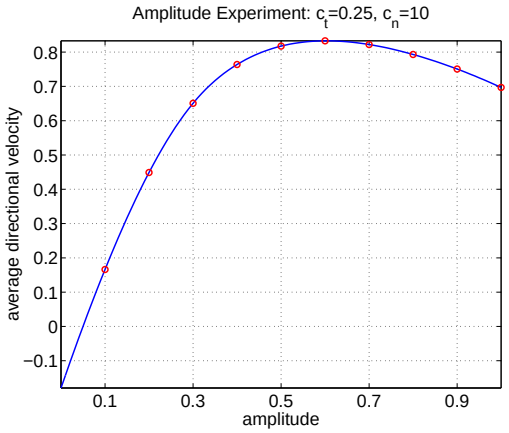
Figure 5.24: The effect of wave form frequency on momentum preservation.

like to point out that although this figure would make it appear as if there is no limit to the locomotive capacity of the structure, the reality of the matter is that the frequency would have an upper-bound determined by the capacity of the actuators driving the differential angles and the rigidity of the lateral resistances.

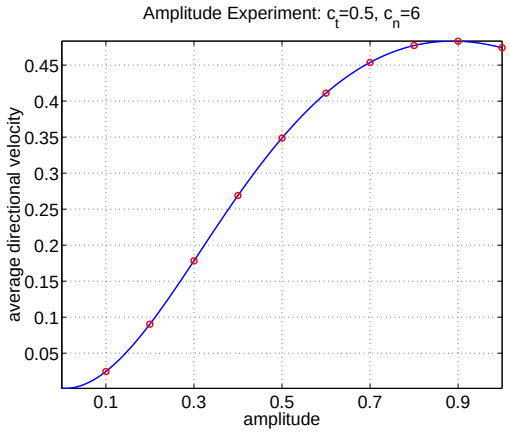
Continuing, we examine the effects of the amplitude parameter of the serpenoid shape. Upon initial experimentation with this parameter, we were a little surprised to find that the average momentum along the head direction decayed with very modest increments in amplitude. Feeling that the increased angle taken by the head of the structure due to the amplitude change may have obscured the true trend of the system momentum change, we altered our measure of progress to the average directional velocity of the head. This measure was determined for amplitude values ranging from 0.1-1 in increments of $\Delta a = 0.1$ and the results are shown in Figure 5.25(a). Just as in the case of our previous observation, the locomotive capacity of the structure quickly declines after the initial increase in the value of a. After some thought we found



(a) Ideal lateral resistance.



(b) The presence of lateral friction.



(c) Natural conditions.

Figure 5.25: The relationship between shape amplitude and directional velocity and the dependence of this relation on the lateral forces.

that the reason for this was the frictionless environment $c_t = 0$. If the structure is absolutely straight then it will not move, thus some amplitude is necessary. However, it is clear that once in motion, the least resistance is found by the 0 amplitude configuration. Thus, the increase of amplitude only served to create more resistance to motion.

However, the story is quite different if one allows a tangential resistance. In Figure 5.25(b) results are presented for the parameter value of $c_t = 0.25$ and in Figure 5.25(c) results corresponding to the assignments of $c_t = 0.5$ and $c_n = 3$, our original simulation setup values, are provided. In these cases it is seen that in the presence of only a slight tangential resistance, there is a much added utility to larger values of a. All of these figures indicate that there is an optimal amplitude, given that all the other system parameters are prefixed.

As a final glance at the large effects that the system and shape parameters can have on the structure's capability to preserve a certain amount of momentum, the effect of the size n of the structure is considered. Again, we look at average momentum along the head as a measure of the effects. Structures with 3 to 13 links are considered. The results are shown in Figure 5.26, and are quite interesting. For it would seem that for a given set of shape and physical parameters, there is an optimal size for the snake. This is sort of surprising and the reason for why this may be the case does not seem obvious. At first glance one might conclude that our results are erroneous. However, as pointed out in [22], one should reserve that judgement. Gans wrote,

> "*The literature contains a single report that suggests that considerations of the energy expended are indeed appropriate in snake locomotion. Heckrote (1967) chased various-sized specimans of the garter snake, Thamnophis s. sirtalis, through an array of pegs on one inch centers. He found that the snake's speed in terms of the number of body lengths per second increased as he used larger specimans, reached a maximum, and then decreased for the largest; this suggested that there was an optimal size for a snake.*"

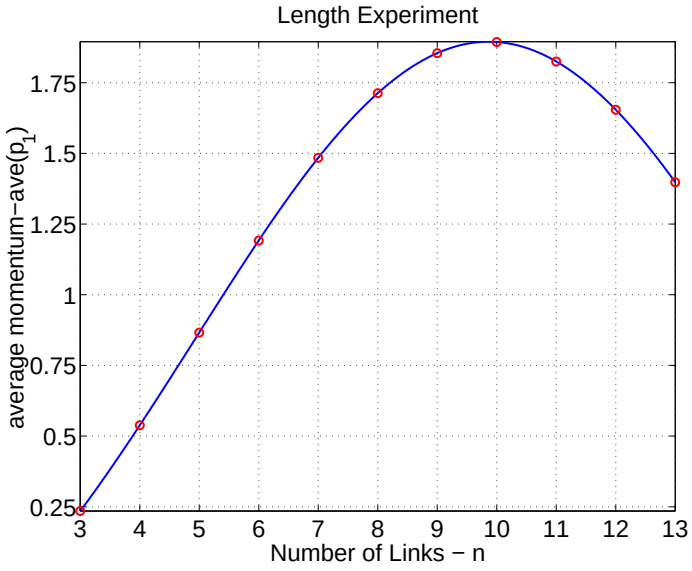


Figure 5.26: The relationship between the length of the the serial-link structure and its momentum preservation.

Quite amazing indeed. We are unaware of any results that take up the topic further, but our simulations support this claim. Also, one cannot help but notice the consistency of the simulation results with nature. Given that the links were chosen to have a length of approximately $\frac{1}{2}$ ft, the simulations indicate that the optimal length for the snake is about $5ft$. This is eerily close to the stretched length of many of these animals.

5.3.5 Actuation V's Geometry: How the Phase-shift is Utilized

To conclude our review and demonstration of the correctness of Gray's work on undulatory locomotion we investigate Criterion 2 and 3. This requires returning to the shape tracking control scheme, which is now restated for convenience. Recall that for

the desired shape ς we select

$$\tau = \Lambda(q)(\ddot{\varsigma} + Kw) + \Upsilon(q, \dot{q}) \ ,$$

where $w = (e^t, \dot{e}^t)^t$ is the vector of error and error velocities and

$$\Lambda(q) \doteq \tilde{M} \ ,$$

$$\Upsilon(q, \dot{q})_k \doteq \frac{1}{2} C_{kk_2k_1} \dot{s}_{k_2} \dot{s}_{k_1} + N + \frac{1}{2} p^t (\frac{\partial I^{-1}}{\partial s_k}) p + \frac{\partial V}{\partial s_k} + \tilde{\tau}^t A_k - \tilde{\tau}_k^f \ ,$$

with

$$\tilde{M} = m - A^t I A \ ,$$
$$C_{kk_2k_1} = (\frac{\partial \tilde{M}_{kk_2}}{\partial s_{k_1}} + \frac{\partial \tilde{M}_{kk_1}}{\partial s_{k_2}} - \frac{\partial \tilde{M}_{k_1k_2}}{\partial s_k}) \ ,$$
$$N = p^t [ad_\xi] A_k + p^t (\frac{\partial A_k}{\partial s} \dot{s} - \frac{\partial A}{\partial s_k} \dot{s}) \ .$$

It is clear that the analytical calculation of τ is rather out of the question. We have already seen that the calculation of I and A can be difficult to carry out and the resulting expression intractable. Our solution to this problem was to carry out the computations numerically with software such as *MATLAB*. One should continue in this vein and that is what we do. All of the partials and gradients indicated by the above expressions can be computed numerically with a centered difference scheme. This scheme is second order and is consequently quite accurate for sufficiently small variation in shape. In our calculations we take $\Delta s = 10^{-6}$.

Now that we have a way to calculate τ we look to establish its relationship with the curvature differential $\delta s \doteq (s_1, \ldots, s_{n-2})^t - (s_2, \ldots, s_{n-1})^t$. It was Gray's conclusion that during lateral undulation locomotion the animal should contract its musculature so as to increase signed curvature toward the right side (viewed from tail to head) of the animal along those portions of the body where the curvature differential was

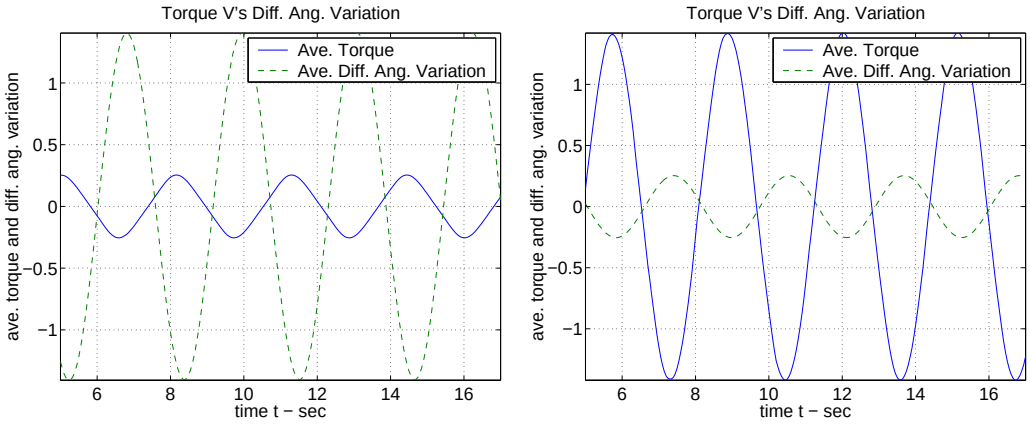
positive, thus making the torque negative. The opposite being the case for those portions where the curvature differential is negative. By doing so, he reasoned that the result would be preservation of forward momentum along the path of the animal.

It was found that the shape of $\mathcal{S}$ can be controlled and that by maintaining the shape deduced by Gray, satisfactory locomotion is achieved. So, we look to show that while this was happening, the torque to curvature differential relationship was in fact following the pattern suggested by Gray. It is somewhat difficult to devise a way to establish this relationship via simulation. However, we determined that in order to cope with the phase-shift built into the shape selection and the fact that there are differing numbers of actuation points and differential curvature components, that the average torque along the structure should be compared with the average differential curvature along the structure over time. Thus, when the snake tends to have a positive differential curvature along the length of the body, the average torque should be negative and visa versa.

To determine if this is generally the case lateral undulation was elicited via S_s with structures of lengths $n = 3, 7, 10$ and we simply plotted the average torque against the average curvature differential over time. The results of this experiment are provided in Figure 5.27. These simulation results would seem to indicate that, again, Gray was correct. He should be. The criterion he developed came through correct application of Newton's laws. Because these criterion are so solid, had our simulation experiments not agreed with them, we would have had to conclude that our simulation setup was in error.

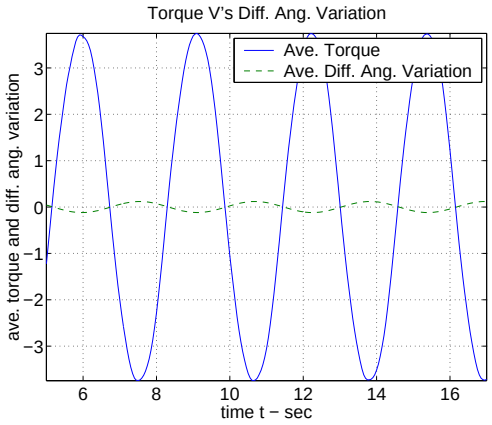
5.4 Additional Matter

The purpose of this chapter was to indicate that the basic principle of undulatory locomotion was clearly and full explained by J. Gray in [28]. His insight was founded in many years of animal locomotion research and in particular, the motion of eels and



(a) The 3 link example.

(b) The 7 link example.



(c) The 10 link example.

Figure 5.27: The relation between actuation and curvature.

fish [30, 27]. Having come to understand the method of propulsion in these animals, it was something of an easy transition to the principles of lateral undulation for the snake, for they are the same within idealized conditions. Namely, the condition of being emersed in a fluid. Concerning this idealization and others, it is felt that the most useful observations were made by J. Gray in the conclusion of his defining study. Therefore, for the sake of completion and progress, it seems more than fitting to now recall these observations. We will directly quote them, for they are stated in the most clear and elegant manner and we would risk deterioration of their value if we were to attempt to paraphrase. These observations are:

Generalization 1. "*In no case is the amplitude of the muscular contraction, or the difference in phase between successive segments, exactly the same for all segments, nor is the duration of the contraction phase of the muscular cycle necessarily the same as that of the relaxation phase. None of these facts, however, affects the main principle. The actual form of the body of a gliding snake is an expression of the precise form of the muscular cycle carried out by each group of segmental muscles and of the phase difference between the muscles of successive segments; in every case, however, gliding motion depends on the same fundamental relationship between the position of the contracting axial musculature in respect to an increase in body curvature towards the same side of the body.*"

Generalization 2. "*It has been assumed that each segment of the body is subjected to external restraint which prevents its movement in a direction normal to its own longitudinal axis. In nature, an approximation to such conditions exists when the animal is moving over a surface such as gravel of small stones, or when the animal is moving through close herbage or grass. Again, when moving over soft dry sand (into which the body tends to sink, or which tends to be heaped up at the sides by*

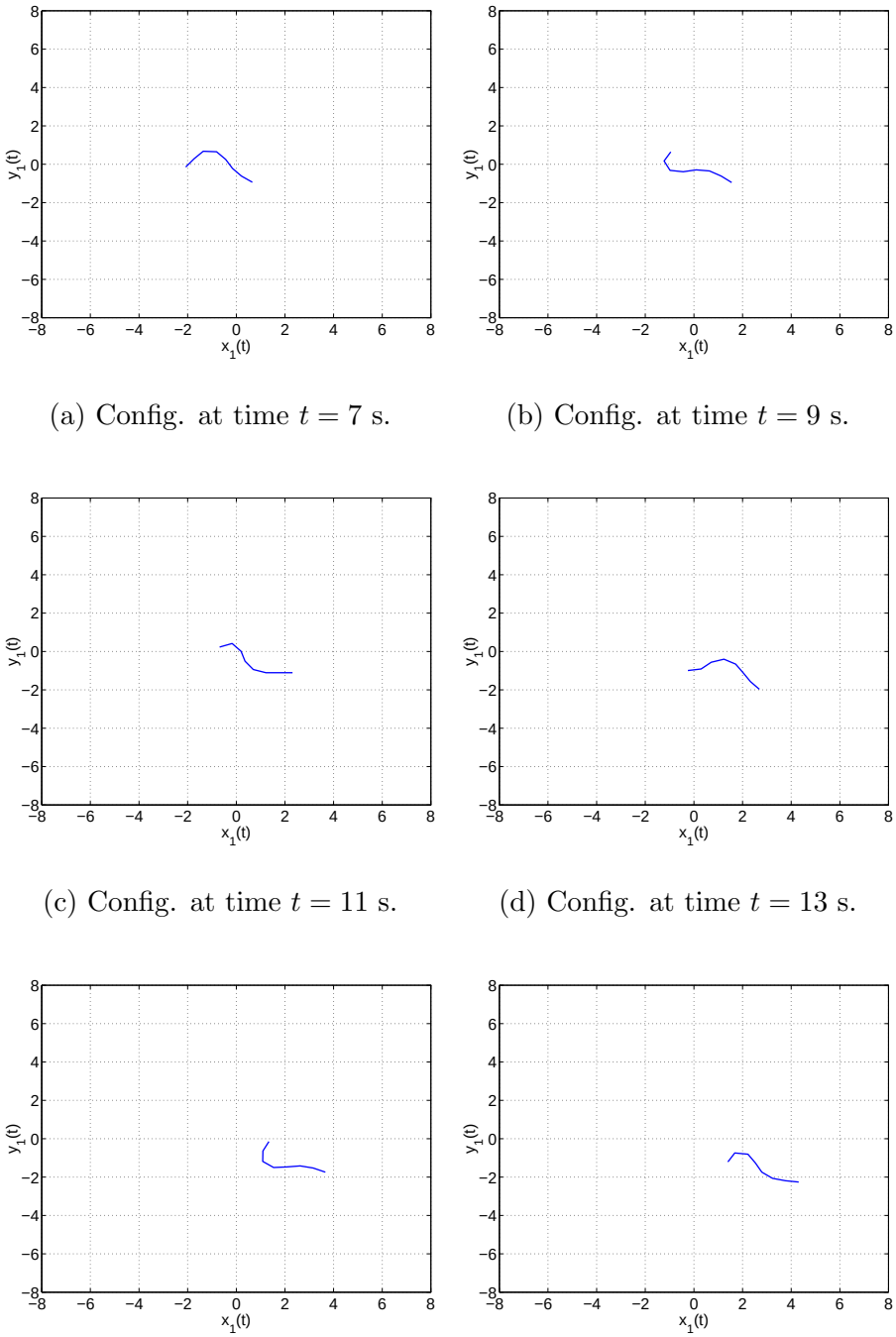
the animal's own movements) each segment is subjected to a resistance acting normally to its surface. On the other hand, the distribution of external resistances can be restricted to a limited number of points along the body,...Precisely similar arguments apply to such conditions..."

Generalization 3. "*In the idealized system ... it is possible to assume that the energy liberated, when the muscles operating across each joint pass from one phase of the contraction cycle to the next, is just sufficient to move the segment concerned from from its original position to that previously occupied by the segment immediately anterior to itself. In a living snake this condition is by no means fulfilled; more frequently, some regions provide some or all of the energy required to move other parts of the body. The fundamental principles are, however, unaffected, ...* "

It is the belief of the author that those who desire to build snake-like robot technologies in order to mimic the depth of adaptability observed amongst real snakes should heed these generalizations and make them the goals for which they strive. It is clear that the functionality that motivates the research in this area is ultimately wrapped up in these remarks. Although the use of the idealized gliding form suggested by Gray to elicit directed gait from snake-like robots is an amazing indication of the feasibility of using the lateral undulation mechanism for propulsion and should be commended, one must be careful not to become fixed to this form. For by doing so, the potentials of the mechanism will be lost.

We leave this discussion with an example. Namely, the eel. As pointed out in [27], the eel will illustrate two of the three generalizations discussed, namely Generalizations 1 and 3, when swimming. The amplitude of the differential angles increase along the vertebrae structure of the eel while in motion. It is unclear whether or not the reason for this is head stabilization or pressure manipulation. However, supposing that it is for the purpose of head stabilization, we see that the eel uses amplitude and

phase modulation along the length of this body, and by doing so, transfers the majority of the burden of locomotion effort to the posterior portion of the body. Along the same vein, for certain fish, nearly all of the propulsive effort is exerted from the posterior of the animal. To partly capture such an effect we construct the following shape: $\phi_1(t) \doteq a\sin(\omega(t))$ and $\phi_j(t) \doteq \frac{3}{2}\left(\frac{j}{n-1}\right)\phi_1(t-(j-1)\frac{p_f}{n})$. Using the same parameters as those used to generate Figures 5.19 and 5.20, we track this shape and observe the locomotion. The results are provided in Figures 5.28 and 5.29. As can clearly be seen in the motion snapshots, breaking the serpenoid form did not destroy locomotion and we were able to stabilize the anterior portion of the structure (Note that the anterior always remains straight).



(a) Config. at time $t = 7$ s.

(b) Config. at time $t = 9$ s.

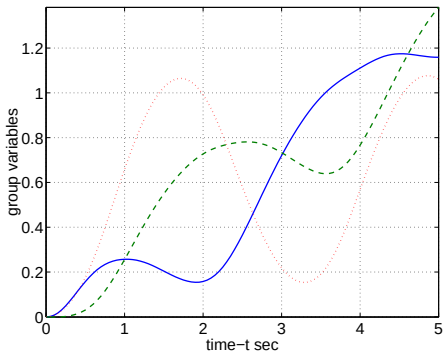
(c) Config. at time $t = 11$ s.

(d) Config. at time $t = 13$ s.

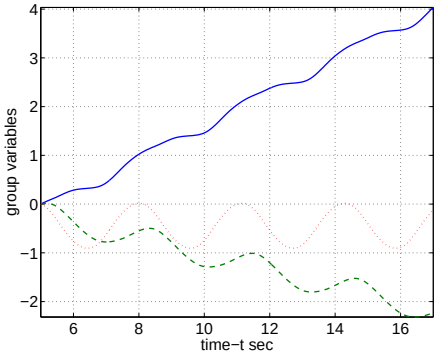
(e) Config. at time $t = 15$ s.

(f) Config. at time $t = 17$ s.

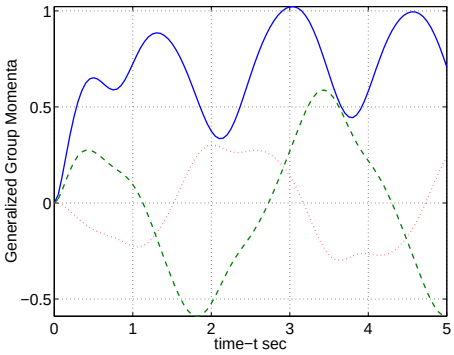
Figure 5.28: The motion of a serial link structure mimicking an eel-like shape.



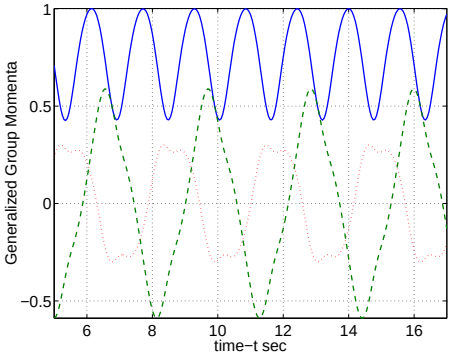
(a) Transient group behavior.



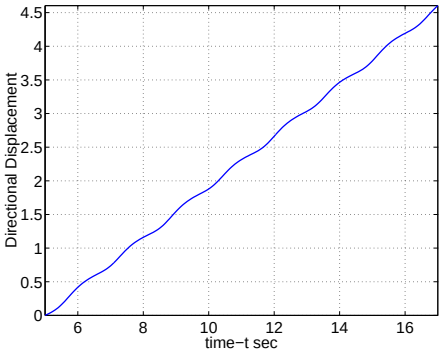
(b) Steady state behavior.



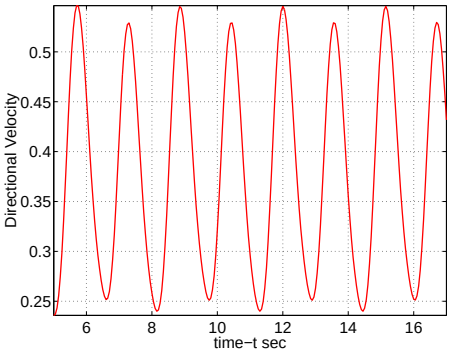
(c) Transient momenta.



(d) Steady state momenta.



(e) Directional displacement.



(f) Directional velocity.

Figure 5.29: The motion of a serial link structure mimicking an eel-like shape, continued. Individual features of the motion

Chapter 6

Numerical Optimal Control

6.1 Introduction

In this last chapter we considered the fundamental mechanism of the lateral undulation form of snake locomotion, carefully noting the generality of it. Using this mechanism we constructed an example gliding form for $\mathcal{S}$ after the construction of Gray and illustrated its locomotive capability in simulation. Hirose had claimed that this form was quite in line with that of actual snakes when introduced to an approximately uniform substratum. However, as emphasized by Gray, the shape or differential angle relations that he developed were simply a way to realize a sigmoidal form and by doing so, a way to indicate why snakes take such a form during locomotion. There are a number of similar periodic relationships that elicit the same effect, as indicated by our use of the sinusoidal wave form to derive differential angle relations which were qualitatively identical to the sinusoidal relations though not of that form. Additionally, even within the adoption of the sinusoidal shape variables there is much room for adaptation that has never been utilized. For instance, as indicated at the closing of Chapter 5, it is quite possible that the sinusoidal shape variables could exploit differing amplitudes and phase-shifts. We illustrated this with the eel example. Thus the question, what is optimal? What form should the snake's

shape variables take on for locomotion in a relatively uniform environment? Are the sinusoidal shape variable forms with uniform amplitude and phase-shift indeed what is natural? These questions have not, at the time of this writing, been answered from an analytic perspective.

In this chapter, an optimal control methodology is developed for determining the optimal gait eliciting shapes for under-actuated systems with symmetries. As pointed out in the work of Ostrowski, the class of systems that fall into this framework is rather large and diverse. Biological examples are abundant, including worms, amoeba, fish, eels, and of course the topic we are currently investigating, snakes. Satellite reorientation and other applications also fall under this guise. It is noted that Ostrowski has also made efforts in this way, presenting a discussion of an optimal control methodology in [56].

To initiate the development of the methodology, Gaussian quadrature is introduced. This will be followed by the development of implicit Runge-Kutta methods for solving ODEs that are based on collocation. Using the Gaussian quadrature nodes as collocation points we arrive at the high order Gaussian integrators. Next, the traditional optimal control problem is introduced. Using the union of Gaussian quadrature and Gaussian integration, a high order discretization of the optimization problem is obtained. We then develop optimality conditions from this discretization.

After obtaining the optimality conditions in a very general setting, it is shown how they can be adapted with minimal effort to apply to the dynamics of locomotion systems with symmetries. There are two primary tools used in the adaptation. One of these, as indicated before, is to view the shape variables as control variables. The other is to use the idea of the periodic derivative to perform the minimization of the cost over the space of periodic functions. After providing a good deal of detail concerning the resulting optimality conditions and their solution we then use the method to determine the optimal gait pattern for the $\mathcal{S}$ and explore the changes in the optimal solution which occur due to changes in parameter values.

6.2 Gaussian Quadrature

Consider the function $f \in \mathcal{C}[t_0, t_f]$. We wish to determine s nodes $c_1, \ldots, c_s$ and s weights $b_1, \ldots, b_s$ such that the approximation

$$\int_{t_0}^{t_f} f(t)\ dt \approx \sum_{i=1}^{s} b_i f(c_i)$$

is as accurate as possible. Being it the case that the space of polynomials $\mathcal{P}$ is dense in $\mathcal{C}[t_0, t_f]$, we will measure accuracy by how well this approximation performs for $f \in \mathcal{P}$. It is noted that the approximation scheme has $2s$ unknowns that we are free to select. Therefore, it makes sense to choose these unknowns so that the approximation is exact for the subspace $\mathcal{P}_{2s-1}$ of polynomials comprised of polynomials with degree $\leq 2s-1$. For, due to the fact that integration is a linear operator, this task may be accomplished by asking that the approximation is exact for the basis $\{1, t, \ldots, t^{2s-1}\}$ of this subspace. This provides $2s$ conditions with which to determine the $2s$ unknowns for the integration approximation. Consider the following examples.

Example 6.1 (A One Term Approximation). *Suppose that $s = 1$ and $[t_0, t_f] = [-1, 1]$. Then the desire is to find c_1 and b_1 such that the approximation*

$$\int_{-1}^{1} f(t)\ dt \approx b_1 f(c_1)$$

is exact for all $f \in \mathcal{P}_1$. This requires exactness of the approximation for $f = 1, t$. Hence, we solve the two equations

$$b_1 = \int_{-1}^{1} 1\ dt = 2\ ;$$
$$b_1 c_1 = \int_{-1}^{1} t\ dt = 0\ .$$

Doing so reveals that the most accurate one term approximation is provided by the parameters $b_1 = 2$ and $c_1 = 0$, and that this approximation is $2f(0)$.

Example 6.2 (A Two Term Approximation). *Suppose that $s = 2$ and retain the interval of integration $[t_0, t_f] = [-1, 1]$. Then the desire is to find c_1, c_2 and b_1, b_2 such that the approximation*

$$\int_{-1}^{1} f(t)\ dt \approx b_1 f(c_1) + b_2 f(c_2)$$

is exact for all $f \in \mathcal{P}_3$. This requires exactness of the approximation for $f = 1, t, t^2, t^3$. Hence, we solve the system of equations

$$\begin{aligned} b_1 + b_2 &= \int_{-1}^{1} 1\ dt = 2\ ; \\ b_1 c_1 + b_2 c_2 &= \int_{-1}^{1} t\ dt = 0\ ; \\ b_1 c_1^2 + b_2 c_2^2 &= \int_{-1}^{1} t^2\ dt = \frac{2}{3}\ ; \\ b_1 c_1^3 + b_2 c_2^3 &= \int_{-1}^{1} t^3\ dt = 0\ . \end{aligned}$$

For this system the solution cannot simply be read from the equations. However, there is a solution and it is $b_1 = b_2 = 1$, $c_1 = -\frac{\sqrt{3}}{3}$, and $c_2 = \frac{\sqrt{3}}{3}$. Thus, the most accurate two term approximation to the integral is given by $f(-\frac{\sqrt{3}}{3}) + f(\frac{\sqrt{3}}{3})$.

The latter of these two examples makes it clear that for large s, solution of the resulting system of equations for the approximation parameters is not feasible. However, there is something special about the values of c_i found in these two examples. These values are roots of the Legendre polynomials of degrees 1 and 2, respectively. The Legendre polynomials comprise an orthonormal basis for $\mathcal{P}$ with respect to the L^2 norm on $\mathcal{P}[-1, 1]$. It turns out that for any given s, c_i should be selected as the roots of the Legendre polynomial of order s, and b_i should be assigned the value $b_i \doteq \int_{-1}^{1} L_i\ dt$, where L_i is the ith Lagrange interpolating polynomial corresponding to the c_i. Thus, the following theorem.

Theorem 6.1 (Gaussian Quadrature). *Let $c_1, \ldots, c_s$ be the roots of the Legendre polynomial of degree s and define*

$$b_i = \int_{-1}^{1} \prod_{\substack{i=1 \\ i \neq j}}^{s} \frac{(t - c_i)}{(c_j - c_i)} \, dt$$

for $i = 1, \ldots, s$. Then

$$\int_{-1}^{1} f(t) \, dt \approx \sum_{i=1}^{s} b_i f(c_i)$$

is exact for $f \in \mathcal{P}_{2s-1}$.

The proof of this theorem is omitted. However, the crux of the argument is that the Legendre polynomials are orthonormal. See [68] for more detail.

The Gaussian quadrature formula may be applied for integration over any connected compact subset of $\mathbb{R}$, $[t_0, t_f]$. This is accomplished by making use of the coordinate transformation

$$\sigma(t) = \frac{2t - t_0 - t_f}{t_f - t_0} \ .$$

By Fubini's theorem,

$$\int_{t_0}^{t_f} f(t) \, dt = \frac{t_f - t_0}{2} \int_{-1}^{1} f\left(\frac{(t_f - t_0)\sigma + t_0 + t_f}{2} \right) \, d\sigma \ , \tag{6.1}$$

which may be approximated by Gaussian quadrature. Let us consider the following examples.

Example 6.3 (A One Term Approximation: Reprise). *Let $[0, 1]$ be the interval of interest and suppose that we desire a 1 term approximation to the integral*

$$\int_{0}^{1} f(t) \, dt \ .$$

Applying equation (6.1) *we see that this integral is equivalent to*

$$\frac{1}{2}\int_{-1}^{1} f\left(\frac{\sigma+1}{2}\right)\, d\sigma\ .$$

Then, according to Example 6.1, we have $f(\frac{1}{2})$*, the midpoint value, as the best* 1 *term approximation to this integral. We make note that the weight has been altered to* 1 *and the nodal value to* $\frac{1}{2}$*.*

Example 6.4 (A Two Term Approximation: Reprise). *Again, let* $[0,1]$ *be the interval of interest. However, now we shall require a* 2 *term approximation to the integral*

$$\int_{0}^{1} f(t)\, dt\ .$$

As before, this integral is equivalent to

$$\frac{1}{2}\int_{-1}^{1} f\left(\frac{\sigma+1}{2}\right)\, d\sigma\ .$$

According to Example 6.2,

$$\frac{1}{2}f\left(\frac{-\sqrt{3}+3}{6}\right)+\frac{1}{2}f\left(\frac{\sqrt{3}+3}{6}\right)\ ,$$

is the best 2 *term approximation to this integral. It is noted that the weights have been altered to* $\frac{1}{2}$ *and the nodal values to* $\frac{\sqrt{3}\pm 3}{6}$*.*

As indicated by this discussion. If we like, we may redefine the gaussian nodes and weights by the relations

$$\begin{aligned} b_i &\leftarrow \frac{b_i}{2}\ , \\ c_i &\leftarrow \frac{c_i+1}{2}\ , \end{aligned}$$

and translate integration on arbitrary compact connected intervals to integration on

$[0, 1]$ instead of reverting back to integration over the interval $[-1, 1]$. This choice is quite common and since it holds advantage for our development, we make these assignments.

One may ask how accurate Gaussian quadrature is in general. Although it is the case that this technique provides exact results for $\mathcal{P}_{2s-1}$, this is not a guarantee of satisfactory performance when it is applied to functions that are not contained in this space. The following theorem addresses this issue in part.

Theorem 6.2 (Gaussian Quadrature Error Estimate). *Let $f \in \mathcal{C}^{2s}[-1,\ 1]$ and let c_i, b_i; $i = 1, \ldots, s$; be the Gaussian quadrature nodes and weights. Then*

$$\int_{-1}^{1} f(t)\ dt - \sum_{i=1}^{s} b_i f(c_i) = \frac{f^{2s}(\zeta)}{(2s)!} \|\mathcal{L}_s\|_{L^2}\ ,$$

where $\zeta \in (-1,\ 1)$ and $\mathcal{L}_s$ is scalar multiple of the Legendre polynomial of degree s.

For a proof of this theorem, see [68]. The result basically says that if the function f is sufficiently smooth or does not change "much" over the interval of integration, then the approximation is quite good.

This theorem hints to us that the quadrature method should not be used for entire intervals of integration. Instead, the interval of integration should be partitioned into subintervals with Gaussian quadrature applied to each of these. Being more specific, consider the partition of the interval $[t_0, t_f]$ determined by the nodal points $t_k = t_0 + k\Delta t$, $k = 0, \ldots, N$, with $\Delta t = \frac{t_f - t_0}{N}$. Using this partition we may write

$$\int_{t_0}^{t_f} f(t)\ dt = \sum_{k=1}^{N} \int_{t_{k-1}}^{t_k} f(t)\ dt\ ,$$

and Gaussian quadrature may be applied to each of the integrals comprising the sum to arrive at a very accurate approximation to the integral over the larger time span. The result is

$$\int_{t_0}^{t_f} f(t)\ dt \approx \Delta t \sum_{k=1}^{N} \sum_{i=1}^{s} b_s f(\tau_{ki})\ ,$$

where $\tau_{ki} \doteq t_{k-1} + c_i \Delta t$. By localizing the integration, we take the fullest advantage of Theorem 6.2.

6.3 Collocation Methods and the Implicit Runge-Kutta Scheme

Here, implicit Runge-Kutta (RK) integration methods based on collocation are introduced in a unified manner. Consider the ODE $x' = f(t, x)$, subject to the initial condition $x(0) = x_0$, with an interval of integration $[t_0, t_f]$. This interval of integration is partitioned into N subintervals using the uniformly spaced nodes $\{t_k\}_{k=0}^{N}$ with the typical condition $t_{k-1} < t_k \ , k = 0, \ldots, N$. On the subinterval $[t_{k-1}, t_k]$, s collocation points, $\tau_{ki} = t_{k-1} + c_i \Delta t$, $0 \leq c_1 < c_2 < \ldots < c_s \leq 1$, are selected. A polynomial $\phi(t)$ with the following properties, called the collocation conditions, is sought:

$$\phi(t_{k-1}) = x_{k-1} \ ,$$

$$\phi'(\tau_i) = f(\tau_i, \phi(\tau_i)) \ , \forall i \ .$$

Let $\tau = t_{k-1} + \sigma \Delta t \ , \ 0 \leq \sigma \leq 1$ and consider the Lagrange interpolating polynomial

$$L_j(\tau) = \prod_{\substack{i=1 \\ i \neq j}}^{s} \frac{\tau - \tau_i}{\tau_j - \tau_i}$$

or

$$\begin{aligned} L_j(\tau) = L_j(t_{k-1} + \sigma \Delta t) &= \prod_{\substack{i=1 \\ i \neq j}}^{s} \frac{t_{k-1} + \sigma \Delta t - t_{k-1} - c_i \Delta t}{t_{k-1} + c_j \Delta t - t_{k-1} + c_i \Delta t} \\ &= \prod_{\substack{i=1 \\ i \neq j}}^{s} \frac{\sigma - c_i}{c_j - c_i} \\ &= L_j(\sigma) \ . \end{aligned}$$

Then the $s-1$ degree polynomial uniquely determined by $\phi'(\tau) \doteq \sum_{j=1}^{s} L_j(\sigma) f^{kj}$, $f^{kj} \doteq f(\tau_i, \phi(\tau_i))$, and the condition $\phi(t_{k-1}) = x_{k-1}$, satisfies the collocation conditions.

Now that an expression for ϕ' is available, some simple integrations will lead to an implicit RK method. Integrating ϕ' on $[t_{k-1}, \tau_i]$ provides us the equation

$$\begin{aligned}\phi(\tau_i) - \phi(t_{k-1}) &= \sum_{j=1}^{s} f^{kj} \int_{t_{k-1}}^{t_{k-1}+c_i\Delta t} L_j(\tau)\ d\tau \\ &= \Delta t \sum_{j=1}^{s} f^{kj} \int_{0}^{c_i} L_j(\sigma)\ d\sigma\ .\end{aligned}$$

In a similar manner, integration of ϕ' on $[t_{k-1}, t_k]$ yields for us the equation

$$\begin{aligned}\phi(t_k) - \phi(t_{k-1}) &= \sum_{j=1}^{s} f^{kj} \int_{t_{k-1}}^{t_k} L_j(\tau)\ d\tau \\ &= \Delta t \sum_{j=1}^{s} f^{kj} \int_{0}^{1} L_j(\sigma)\ d\sigma\ .\end{aligned}$$

Let $a_{ij} \doteq \int_0^{c_i} L_j(\sigma)\ d\sigma$ and $b_j \doteq \int_0^1 L_j(\sigma)\ d\sigma$. Then the following RK scheme ensues:

$$x^{ki} \doteq \phi(\tau_i) = x_{k-1} + \Delta t \sum_{j=1}^{s} a_{ij} f^{kj}\ ,\ i = 1, \ldots, s\ .$$

$$x_k \doteq \phi(t_k) = x_{k-1} + \Delta t \sum_{j=1}^{s} b_j f^{kj}\ .$$

The RK scheme can be conveniently represented by the Butcher diagram:

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s} \\ c_2 & a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline & b_1 & b_2 & \cdots & b_s \end{array}\ .$$

The most commonly used implicit RK methods based on collocation are those that

correspond to certain quadrature methods. In particular, Gaussian quadrature provides us with a collocation RK method. The method is determined by using the Gaussian quadrature points c_i over the interval $[0, 1]$ to define the collocation nodes τ_i, for each of the subintervals determined by discretization. This is illustrated by the following example.

Example 6.5 (Gaussian RK Schemes). *Taking $s = 1$ and referring back to Example 6.3 we obtain the implicit midpoint scheme*

$$\begin{array}{c|c} \frac{1}{2} & \frac{1}{2} \\ \hline & 1 \end{array}$$

and for $s = 2$ we have

$$\begin{array}{c|cc} \frac{3-\sqrt{3}}{6} & \frac{1}{4} & \frac{3-2\sqrt{3}}{12} \\ \frac{3+\sqrt{3}}{6} & \frac{3-2\sqrt{3}}{12} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

after referencing Example 6.4.

There are a couple of properties of Gaussian RK methods that should be noted. We now state these properties as theorems, but do not provide proofs, as these statements are common knowledge and a proof may be found in many texts that deal with the subject.

Theorem 6.3 (Order). *The order of an s-stage Gaussian Runge-Kutta method is $2s$.*

The proof of this theorem follows from the precision of Gaussian quadrature, which is $2s$. This says that the implicit midpoint rule is a second order ODE solver.

Theorem 6.4 (Stability). *Gaussian Runge-Kutta methods are A-stable.*

This theorem follows from the implicit nature of these methods and tells us that our discretization step size need not be taken exceedingly small in order to guarantee

that the scheme does not diverge. This allows the method to be used on mildly stiff problems, though it is ill-suited for very stiff problems.

We directly show how the techniques of quadrature and their corresponding collocation RK methods can be used to solve optimal control problems.

6.4 The Optimal Control Problem

Consider the standard optimal control problem. Given an autonomous system with dynamics $\dot{x} = f(x, u)$ and initial state x_0, we wish to select a control input u so as to drive the state trajectory to a specified final state x_{t_f} at some specified final time t_f, in an optimal manner. By optimal we shall mean that the desire is to select $u \in \mathcal{A}$, $\mathcal{A}$ being some admissible set of controls, such that a cost functional of the form

$$J_{t_f}(x_0, u(t)) = g(x(t_f)) + \int_0^{t_f} (l(x(t)) + v(u(t)))\ dt\ , \tag{6.3}$$

is minimized subject to the conditions

$$\frac{d}{dt}x(t) = f(x(t), u(t))\ ; \tag{6.4}$$

$$x(0) = x_0\ . \tag{6.5}$$

It is typical to make use of the calculus of variations in conjunction with the theorem of Lagrange multipliers to derive necceassary optimality conditions for the solution to this minimzation problem. However, this results in a two-point boundary-value (TPBV) problem that is rather intractable. In most cases, other than the linear dynamics case, an analytical solution of the TPBV problem is impossible to obtain. In these situations the typical approach to procuring a solution is to discretize the optimality conditions using a uniform mesh and then solve the resulting non-linear system via Newton's method. This is a very clever idea. However, this approach does not always include assurances of accuracy and stability. If the system is stiff,

the mesh norm may have to be selected to be small so as to compensate for this system property. This selection in turn leads to the issue of error accumulation and propagation.

A new approach that works toward the resolution of these complications has been proposed [35]. The philosophy is to discretize the system dynamics and the cost functional before deriving optimality conditions. In this way, Gaussian quadrature and the corresponding implicit RK scheme may be utilized to ensure accuracy and stability in the approximation. We will make use of this approach and now discuss the details thereof.

Applying composite Gaussian quadrature to the objective function (6.3) and applying the corresponding implicit RK method to the constraining ODE (6.4) we arrive at the following discrete optimization problem: Defining $u_N \doteq \{u^{ki}\}_{k,i=1}^{N,s} \in \mathcal{A}$, where $u^{ki} \doteq u(\tau_{ki})$ and $\mathcal{A}$ is the space of admissible control sequences, we desire

$$\min_{u_N \in \mathcal{A}} J^N(u_N) \ ,$$

where

$$J^N(u_N) = g(x^N) + \Delta t \sum_{k=1}^{N} \sum_{i=1}^{s} b_i (l(x^{ki}) + v(u^{ki})) \ , \tag{6.6}$$

subject to the constraints

$$\begin{aligned} f^{ki} &= f(x^{ki}, u^{ki}) \ , \\ x^{ki} \doteq \phi(\tau_{ki}) &= x_{k-1} + \Delta t \sum_{j=1}^{s} a_{ij} f^{kj} \ , \\ x_k \doteq \phi(t_k) &= x_{k-1} + \Delta t \sum_{j=1}^{s} b_j f^{kj} \ , \end{aligned}$$

with $k = 1, \ldots, N$ and $i = 1, \ldots, s$. To solve this constrained optimization problem

the multipliers p^k and p^{ki} are introduced and used to define Lagrangian L,

$$\begin{aligned} L(x^{\cdot}, u^{\cdot,\cdot}, f^{\cdot,\cdot}; p^{\cdot}, p^{\cdot,\cdot}) = g(x^N) &+ \sum_{k=1}^{N}\sum_{i=1}^{s} b_i(l(x^{ki}) + v(u^{ki}))\Delta t \\ &+ \Delta t \sum_{k=1}^{N}\sum_{i=1}^{s} b_i\, p^{ki} \cdot (f(x^{ki}, u^{ki}) - f^{ki}) \\ &- \Delta t \sum_{k=1}^{N} p^k \cdot \left(\frac{x_k - x_{k-1}}{\Delta t} - \sum_{i=1}^{s} b_i f^{ki} \right) \ . \end{aligned}$$

The Lagrange multiplier theorem may now utilized to determine necessary conditions for the optimization of L. We begin by calculating the needed derivatives of L. Let us predicate these calculations by noting that

$$\frac{\partial x^{ki}}{\partial f^{mq}} = \delta_{mk} a_{iq} \Delta t \ , \tag{6.7}$$

and

$$\frac{\partial x^{ki}}{\partial x^{m}} = \delta_{(m+1)k} \ , \tag{6.8}$$

where δ is the Kronecker delta. Proceeding with the variable f^{mq} it is seen that

$$\begin{aligned} \frac{\partial L}{\partial f^{mq}} = &\sum_{k=1}^{N}\sum_{i=1}^{s} \left(b_i l_x(x^{ki}) \frac{\partial x^{ki}}{\partial f^{mq}} \right) \Delta t - b_q (p^{mq})^t \Delta t \\ &+ \sum_{k=1}^{N}\sum_{i=1}^{s} \left(b_i\, p^{ki} \cdot f_x(x^{ki}, u^{ki}) \frac{\partial x^{ki}}{\partial f^{mq}} \right) \Delta t + b_q (p^m)^t \Delta t \ , \end{aligned}$$

where equation (6.8) has been employed. Summing over the index k, observing equation (6.7), and combining a couple of terms leads to us to the equation

$$\frac{\partial L}{\partial f^{mq}} = \sum_{i=1}^{s} (b_i\, p^{mi} \cdot (f_x(x^{mi}, u^{mi}) + l_x(x^{mi})) a_{iq} \Delta t^2) - b_q (p^{mq})^t \Delta t + b_q (p^m)^t \Delta t \ .$$

Making the assignment, $g^{mi} \doteq f_x^t(x^{mi}, u^{mi}) p^{mi} + l_x^t(x^{mi})$, we finally arrive at the

expression

$$\frac{\partial L}{\partial f^{mq}} = \sum_{i=1}^{s} (b_i (g^{mi})^t a_{iq} \Delta t^2) - b_q (p^{mq})^t \Delta t + b_q (p^m)^t \Delta t \ . \tag{6.9}$$

Now consider the argument p^{mq}. It is easy to see that

$$\frac{\partial L}{\partial p^{mq}} = \sum_{k=1}^{N} \sum_{i=1}^{s} b_i (f(x^{ki}, u^{ki}) - f^{ki})^t \frac{\partial p^{ki}}{\partial p^{mq}} \Delta t \ .$$

Summing over k and i yields for us the equation

$$\frac{\partial L}{\partial p^{mq}} = b_q (f(x^{mq}, u^{mq}) - f^{mq})^t \Delta t \ , \tag{6.10}$$

as $\frac{\partial p^{ki}}{\partial p^{mq}} = \delta_{mk} \delta_{qi} I$. Continuing, the argument p^m is considered next.

$$\frac{\partial L}{\partial p^m} = \Delta t \sum_{k=1}^{N} \left(\frac{x_k - x_{k-1}}{\Delta t} - \sum_{i=1}^{s} b_i f^{ki} \right)^t \frac{\partial p^k}{\partial p^m} \ .$$

This equation, upon observing that $\frac{\partial p^k}{\partial p^m} = \delta_{mk} I$ and summing over k, resolves to the expression

$$\frac{\partial L}{\partial p^m} = \Delta t \left(\frac{x_m - x_{m-1}}{\Delta t} - \sum_{i=1}^{s} b_i f^{mi} \right)^t \ . \tag{6.11}$$

Moving on, x^m is considered. The partial differentiation of L with respect to this particular variable is carried out and presented here on the term level, as there are cases, and they are more easily summarized in this manner. For $m = 1, \ldots, N-1$;

$$\begin{aligned} \frac{\partial}{\partial x^m} \sum_{k=1}^{N} \sum_{i=1}^{s} b_i l(x^{ki}) \Delta t &= \sum_{k=1}^{N} \sum_{i=1}^{s} b_i l_x(x^{ki}) \frac{\partial x^{ki}}{\partial x^m} \Delta t \\ &= \sum_{i=1}^{s} b_i l_x(x^{(m+1)i}) \Delta t \ ; \end{aligned} \tag{6.12}$$

where summation over k has been performed and relation (6.8) observed. For $m =$

$0, \ldots, N-1$;

$$\begin{aligned}\frac{\partial}{\partial x^m}\sum_{k=1}^{N}\sum_{i=1}^{s} b_i (p^{ki})^t f(x^{ki}, u^{ki})\Delta t &= \sum_{k=1}^{N}\sum_{i=1}^{s} b_i (p^{ki})^t f_x(x^{ki}, u^{ki})\frac{\partial x^{ki}}{\partial x^m}\Delta t \\ &= \sum_{i=1}^{s} b_i (p^{(m+1)i)})^t f_x(x^{(m+1)i}, u^{(m+1)i})\Delta t \ ;\end{aligned} \tag{6.13}$$

where again, summation over k has been performed and relation (6.8) observed. Also, noting that $\frac{\partial x^k}{\partial x^m} = \delta_{mk} I$, one finds that

$$-\frac{\partial}{\partial x^m}\sum_{k=1}^{N} p^k \cdot x^k = -(p^m)^t \ , \ m = 1, \ldots, N \ ; \tag{6.14}$$

$$\frac{\partial}{\partial x^m}\sum_{k=1}^{N} p^k \cdot x^{k-1} = (p^{m+1})^t \ , \ m = 0, \ldots, N-1 \ . \tag{6.15}$$

From equations (6.12), (6.13), and (6.14) and the definition of g^{ik} it follows that

$$\frac{\partial L}{\partial x^m} = \begin{cases} \sum_{i=1}^{s} b_i (g^{1i})^t \Delta t + (p^{m+1})^t \ , & \text{if } m = 0 \ ; \\ \sum_{i=1}^{s} b_i (g^{(m+1)i})^t \Delta t - (p^m)^t + (p^{m+1})^t \ , & \text{if } m = 1, \ldots, N-1 \ ; \\ -(p^N)^t + g_x(x^N) \ , & \text{if } m = N \ . \end{cases} \tag{6.16}$$

Finally, differentiation of L with respect to u^{mq} provides us the equation

$$\begin{aligned}\frac{\partial L}{\partial u^{mq}} &= \sum_{k=1}^{N}\sum_{i=1}^{s} b_i v_u(u^{ki})\frac{\partial u^{ki}}{\partial u^{mq}}\Delta t + \sum_{k=1}^{N}\sum_{i=1}^{s} b_i (p^{ki})^t f_u(x^{ki}, u^{ki})\frac{\partial u^{ki}}{\partial u^{mq}}\Delta t \\ &= b_q(v_u(u^{mq}) + (p^{mq})^t f_u(x^{mq}, u^{mq}))\Delta t \ ,\end{aligned} \tag{6.17}$$

where the sums have been taken and the fact that $\frac{\partial u^{ki}}{\partial u^{mq}} = \delta_{mk}\delta_{qi} I$ has been used. With all of the requisite partials of L we may proceed to use the theorem of Lagrange

multipliers. The directional derivative or variation of L in direction h is

$$\delta L[h] = \sum_k \frac{\partial L}{\partial x^k} h_{x^k} + \sum_{ki} \frac{\partial L}{\partial u^{ki}} h_{u^{ki}} + \sum_{ki} \frac{\partial L}{\partial f^{ki}} h_{f^{ki}} + \sum_k \frac{\partial L}{\partial p^k} h_{p^k} + \sum_{ki} \frac{\partial L}{\partial p^{ki}} h_{p^{ki}} \ .$$

The Lagrange multiplier theorem allows us to treat the directions of the arguments of L as independent. It is known then that in order to obtain an optimal value of the functional L, $\delta L[h] = 0$ must hold for all directions h. This then implies, along with the pseudo independence of the argument directions, that $\frac{\partial L}{\partial (\cdot)} = 0$. These are the necessary optimality conditions.

The necessary conditions extrapolated from the theorem of Lagrange multipliers are now given form. Let us begin with equation (6.9). Setting this expression equal to zero and transposing provides us the equation

$$\sum_{i=1}^{s} (b_i g^{mi} a_{iq} \Delta t^2) - b_q p^{mq} \Delta t + b_q p^m \Delta t = 0 \tag{6.18}$$

$$\leftrightarrow \Delta t \sum_{i=1}^{s} (\frac{b_i}{b_q} g^{mi} a_{iq}) - p^{mq} + p^m = 0 \ \ \text{(divide by } \Delta t \text{ and } b_q) \tag{6.19}$$

$$\leftrightarrow p^{mq} = p^m + \Delta t \sum_{i=1}^{s} (a_{(s+1)-q,(s+1)-i} g^{mi}) \quad \left(\frac{b_i}{b_q} a_{iq} = a_{(s+1)-q,(s+1)-i} \right) \ . \tag{6.20}$$

Equating expression (6.11) with zero and transposing yields the equation

$$\frac{x_m - x_{m-1}}{\Delta t} - \sum_{i=1}^{s} b_i f^{mi} = 0 \ . \tag{6.21}$$

Continuing, the equation

$$f^{mq} = f(x^{mq}, u^{mq}) \tag{6.22}$$

is obtained by equating expression (6.10) with zero and transposing. Next, by setting

expression (6.16) equal to zero and transposing it is found that

$$\begin{cases} \begin{cases} -p^{m+1} = \Delta t \sum_{i=1}^{s} b_i g^{1i} \ , & \text{if } x^0 \text{ is free;} \\ x^0 = x_0 \ , & \text{if } x^0 \text{ is fixed;} \end{cases} , & \text{if } m = 0; \\ -\frac{p^{m+1}-p^m}{\Delta t} = \sum_{i=1}^{s} b_i g^{(m+1)i} \ \text{(divide by } \Delta t) \ , & \text{if } m = 1, \ldots, N-1; \\ \begin{cases} p^N = g_x^t(x^N) \ , & \text{if } x^N \text{ is free;} \\ x^N = x_{t_f} \ , & \text{if } x^N \text{ is fixed;} \end{cases} , & \text{if } m = N; \end{cases} \cdot \tag{6.23}$$

Finally, setting expression (6.17) equal to zero provides us with the condition

$$v_u^t(u^{mq}) + f_u^t(x^{mq}, u^{mq}) p^{mq} = 0$$

or

$$u^{mq} = \psi(x^{mq}, p^{mq}) \ , \tag{6.24}$$

for appropriately defined ψ. These necessary optimality conditions are now summarized, for appropriate index ranges, as follows:

$$\begin{cases} \frac{x^m - x^{m-1}}{\Delta t} = \sum_{i=1}^{s} b_i f^{mi} \ ; \\ f^{mq} = f(x^{mq}, u^{mq}) \ ; \\ x^{mq} = x_{m-1} + \Delta t \sum_{j=1}^{s} a_{qj} f^{mj} \ , \end{cases} \tag{6.25a}$$

$$\begin{cases} -\frac{p^{m+1}-p^m}{\Delta t} = \sum_{i=1}^{s} b_i g^{(m+1)i} \ ; \\ g^{mq} = f_x(x^{mq}, u^{mq}) p^{mq} + l_x(x^{mq}) \ ; \\ p^{mq} = p^m + \Delta t \sum_{i=1}^{s} (a_{(s+1)-q,(s+1)-i} g^{mi}) \ , \end{cases} \tag{6.25b}$$

$$u^{mq} = \psi(x^{mq}, p^{mq}) \ , \tag{6.25c}$$

$$\begin{cases} -p^{m+1} = \Delta t \sum_{i=1}^{s} b_i g^{1i} \ ; \\ x^0 = x_0 \ , \end{cases} \tag{6.25d}$$

$$\begin{cases} p^N = g_x(x^N) \ ; \\ x^N = x_{t_f} \ . \end{cases} \tag{6.25e}$$

These equations may be solved via Newton's method, or some other root finder, to determine the optimal control and state approximations.

6.5 Selection of Optimal Gaits

In this section an adaptation of the optimal control methodology introduced that allows determination of optimal locomotive shapes for underactuated systems with symmetries is presented. The development is predicated by the following preparatory discussion.

The shape s is viewed as the control variable and equations (4.21) as the dynamics for the systems we now focus on. These equations are repeated here for convenience:

$$\xi = I^{-1}p - A\dot{s} \ , \tag{6.26a}$$

$$\dot{p} - [ad_{I^{-1}p - A\dot{s}}]^t p = \tilde{\tau} \ . \tag{6.26b}$$

Defining $x \doteq (g, p)^t$ and recalling that $\dot{g} = [g]\xi$, it is seen that equations (6.26) provide us with a system plant of the form $\dot{x} = f(x, s, \dot{s})$. Unlike the form we assumed for the dynamics presupposed in the previous development of the previous section, this plant depends explicitly on $\dot{s}$, the derivative of the shape.

Furthermore, for locomotion systems, we wish to restrict our attention to those shapes that are periodic so that we stay in line with the notion that repeated and patterned contraction and relaxation of the muscle-like actuators is a requisite for a steady rhythmic motion. Thus, the condition of $s(t + p_f) = s(t)$, $\forall t \in \mathbb{R}$, where p_f is

the fundamental period selected is imposed. This narrows the admissible control space $\mathcal{A}$. In the more generic development it was tacitly assumed that the admissible control space was, to the extent which still allows use of optimization criterion, unrestricted.

With these two remarks in mind we follow the strategy of the previous section. The cost functional

$$J(s) = g(x_{t_f}) + \int_0^{t_f} (l(x) + v(s, \dot{s}, \ddot{s}))\ dt\ , \tag{6.27}$$

along with the dynamics,

$$\dot{x} = f(x, s, \dot{s})\ ,$$
$$x_0 = x(0)\ ,$$

are discretized using the implicit R-K scheme with $s = 1$. When $s = 1$, $c_1 = \frac{1}{2}$, $a_{1,1} = \frac{1}{2}$, and $b_1 = 1$. Thus, the discrete cost criterion

$$J^N = g(x^N) + \Delta t \sum_{k=1}^{N} (l(x^k) + v(s^k, \dot{s}^k, \ddot{s}^k))\ , \tag{6.28}$$

and the discrete dynamics

$$f^k = f(x^k, s^k, \dot{s}^k)\ , \tag{6.29a}$$
$$x^k = x_{k-1} + \frac{\Delta t}{2} f^k\ , \tag{6.29b}$$
$$x_k = x_{k-1} + \Delta t f^k\ . \tag{6.29c}$$

Here, the notations s^k, $\dot{s}^k$, and $\ddot{s}^k$ indicate the respective functions s, $\dot{s}$, and $\ddot{s}$, evaluated at the gaussian node $\tau_{k1} = t_k + \frac{1}{2}\Delta t$. This is the midpoint rule and is known to be a second order numerical integration scheme.

We wish to determine the unknown controls $\{s^k\}_{k=1}^N$ such that the functional J^N is minimized subject to the constraint imposed by the discrete dynamics. In addition

to the explicit appearance of $\dot{s}$ in the system plant, this functional has also been altered to depend on the unknowns $\dot{s}^k$ and $\ddot{s}^k$, $k = 1, \ldots, N$, which makes it different, albeit only slightly, from the cost of the previous section. It is noted that we may use the values s^k along with center difference schemes to determine approximations of these quantities. The estimates are

$$\dot{s}^k \approx \frac{s^{k+1} - s^{k-1}}{2\Delta t} \tag{6.30a}$$

$$\ddot{s}^k \approx \frac{s^{k+1} - 2s^k + s^{k-1}}{\Delta t^2} \ . \tag{6.30b}$$

These approximations are known to be second order and as such their use will not diminish the order of the implicit midpoint integrator. Therefore, the approximations are used and henceforth, when we write $\dot{s}^k$ or $\ddot{s}^k$, it is to their approximates that we refer and it should be understood that these quantities are linear functions of the s^k. By making this approximation, v depends only on s^k and we return to framework of the previous section.

To minimize J^N over $\vec{s} \doteq ((s^1)^t, \ldots, (s^N)^t)^t \in \mathcal{A}$, subject to the discrete dynamics of equation (6.35), we use the method of Lagrange multipliers. Let $\{p_k\}_{k=1}^N$ and $\{p^k\}_{k=1}^N$ be a vector-valued sequences of Lagrange multipliers. Using these multipliers the appended cost functional or Lagrangian is defined. This functional is,

$$\begin{aligned} L = g(x^N) &+ \Delta t \sum_{k=1}^{N} (l(x^k) + v(s^k, \dot{s}^k, \ddot{s}^k)) \\ &+ \Delta t \sum_{k=1}^{N} (p^k)^t (f(x^k, s^k, \dot{s}^k) - f^k) \\ &+ \Delta t \sum_{k=1}^{N} p_k^t \left(\frac{x_k - x_{k-1}}{\Delta t} - f^k \right) \ . \end{aligned} \tag{6.31}$$

This cost is minimized without constraint over the set of periodic sequences of $\mathcal{A}$ of fundamental period p_f. The unknowns are the state sequence x_k, the costate

sequences p_k and p^k, the control sequence s^k, and the vector field sequence f^k. The theorem of Lagrange multipliers guarantees that unconstrained minimization of the appended cost is equivalent to minimization of the original cost with constraint.

Before discussing the necessary optimality conditions for the minimization of L it must be explained how the constraint on the admissible space of controls is enforced. Some of the parameters of the problem must be chosen appropriately. Firstly, t_f should be selected so as to be a multiple of the fundamental period p_f, say $t_f = qp_f$. Secondly, Δt or N should be selected so that $\bar{k}\Delta t = p_f$ for some iterate $\bar{k}$. Then it is demanded that $s^k = s^{(k \mod \bar{k})}$. This demand is imposed through the periodic derivative. By periodic derivative our intended meaning is that calculation of $\dot{s}^k$ and $\ddot{s}^k$ make use of the relations $s^k = s^{(k \mod \bar{k})}$. Continuing with the vector notations, we make the assignments

$$\dot{\vec{s}} \doteq ((\dot{s}^1)^t, \ldots, (\dot{s}^N)^t)^t \ ;$$
$$\ddot{\vec{s}} \doteq ((\dot{s}^1)^t, \ldots, (\dot{s}^N)^t)^t \ ,$$

and

$$\vec{s}_i \doteq ((s^{(i-1)\bar{k}+1})^t, \ldots, (s^{i\bar{k}})^t)^t \ ;$$
$$\dot{\vec{s}}_i \doteq ((\dot{s}^{(i-1)\bar{k}+1})^t, \ldots, (\dot{s}^{i\bar{k}})^t)^t \ ;$$
$$\ddot{\vec{s}}_i \doteq ((\ddot{s}^{(i-1)\bar{k}+1})^t, \ldots, (\ddot{s}^{i\bar{k}})^t)^t \ ,$$

for $i = 1, \ldots, q$. The latter of these expressions represent vectors containing the q periods of the sequence $\{s^k\}_{k=1}^N$ and its first and second derivative approximations.

We also define the following $\bar{k} \times \bar{k}$ matrices:

$$D_1 = \begin{bmatrix} 0^{(n-1)} & I^{(n-1)} & & -I^{(n-1)} \\ -I^{(n-1)} & \ddots & \ddots & \\ & \ddots & & \\ & & & I^{(n-1)} \\ I^{(n-1)} & & -I^{(n-1)} & 0^{(n-1)} \end{bmatrix} \tag{6.32}$$

and

$$D_2 = \begin{bmatrix} 2I^{(n-1)} & -I^{(n-1)} & & -I^{(n-1)} \\ -I^{(n-1)} & \ddots & \ddots & \\ & \ddots & & \\ & & & I^{(n-1)} \\ -I^{(n-1)} & & -I^{(n-1)} & 2I^{(n-1)} \end{bmatrix} . \tag{6.33}$$

It should then be clear by making reference to equations (6.30a) and (6.30b) as well as the periodicity relations that the following equations hold

$$\dot{\vec{s}}_i = D_1 \vec{s}_i \ ; \tag{6.34a}$$

$$\ddot{\vec{s}}_i = -D_2 \vec{s}_i \ . \tag{6.34b}$$

Now, the necessary conditions for optimality may, for the most part, be read from (6.25). Beginning with the equations (6.25a) we have

$$f^k = f(x^k, s^k, \dot{s}^k) \ , \tag{6.35a}$$

$$x^k = x_{k-1} + \frac{\Delta t}{2} f^k \ , \tag{6.35b}$$

$$x_k = x_{k-1} + \Delta t f^k \ . \tag{6.35c}$$

Solving equation (6.35c) for f^k yields for us the equation

$$f^k = \frac{x_k - x_{k-1}}{\Delta t} \ . \tag{6.36}$$

Equation (6.36) is substituted into equation (6.35b) to determine that

$$x^k = \frac{x_{k-1} + x_k}{2} \ .$$

This equation along with equation (6.36) are then substituted into equation (6.35a) to provide us with a single equation for the discrete dynamics, namely

$$\frac{x_k - x_{k-1}}{\Delta t} = f\left(\frac{x_{k-1} + x_k}{2}, s^k, \dot{s}^k\right) \ . \tag{6.37}$$

These dynamics are accompanied by the initial condition $x_0 = x(0)$ taken from equation (6.25d).

Next, consider equation set (6.25b), which provides us with

$$-\frac{p_k - p_{k-1}}{\Delta t} = g^k \ , \tag{6.38a}$$

$$g^k = f_x^t(x^k, s^k, \dot{s}^k)p^k + l_x^t(x^k) \ , \tag{6.38b}$$

$$p^k = p_k + \frac{\Delta t}{2} g^k \ . \tag{6.38c}$$

Substitution of equation (6.38a) into equation (6.38c) will provide for one the relation

$$p^k = \frac{p_k + p_{k-1}}{2} \ . \tag{6.39}$$

Substitution of equation (6.39) into equation (6.38b) and then placing the result in equation (6.38a) yields to us the discrete costate dynamics

$$\frac{p_{k-1} - p_k}{\Delta t} = f_x^t\left(\frac{x_{k-1} + x_k}{2}, s^k, \dot{s}^k\right)\left(\frac{p_k + p_{k-1}}{2}\right) + l_x^t\left(\frac{x_{k-1} + x_k}{2}\right) \ . \tag{6.40}$$

These dynamics are accompanied by the terminal condition $p_N = g_x(x_N)$, as indicated by equation (6.25e).

The remaining optimality condition is equation (6.25c). Due to the fact that v is now both implicitly and explicitly dependent upon $\vec{s}$, the derivation of this condition must be considered anew. The condition comes through the demand that $\delta L[h_{\vec{s}}] = 0$. This derivative is fairly simple to calculate. However, notation can be awkward. The following definitions are made in hope that the discussion concerning the calculation of this variation will flow most easily. Let

$$y \doteq \sum_k v^k \ ,$$

where

$$v^k \doteq v(s^k, \dot{s}^k, \ddot{s}^k) \ .$$

Further, upon making the assignments

$$\begin{aligned} pf_{s^k} &\doteq (p^k)^t \frac{\partial f}{\partial s^k}(x^k, s^k, \dot{s}^k) \ , \\ pf_{\dot{s}^k} &\doteq (p^k)^t \frac{\partial f}{\partial \dot{s}^k}(x^k, s^k, \dot{s}^k) \ , \end{aligned}$$

the associations

$$\begin{aligned} pf_{\vec{s}} &\doteq (pf_{s^1}, \ldots, pf_{s^N}) \ , \\ pf_{\dot{\vec{s}}} &= (pf_{\dot{s}^1}, \ldots, pf_{\dot{s}^N}) \ , \end{aligned}$$

are made. The variation of interest is then

$$\begin{aligned} \delta L[h_{\vec{s}}] = \Delta t \left(\frac{\partial y}{\partial \vec{s}} h_{\vec{s}} + \frac{\partial y}{\partial \dot{\vec{s}}} \dot{h}_{\vec{s}} + \frac{\partial y}{\partial \ddot{\vec{s}}} \ddot{h}_{\vec{s}} \right) \\ + \Delta t \left(pf_{\vec{s}} h_{\vec{s}} + pf_{\dot{\vec{s}}} \dot{h}_{\vec{s}} \right) \ . \end{aligned} \tag{6.41}$$

Rewriting this variation so as to reflect the periodicity of $\vec{s}$ provides us with

$$\begin{aligned}\delta L[h_{\vec{s}}] = \Delta t \sum_{i=1}^{q} \left(\frac{\partial y}{\partial \vec{s}_i} h_{\vec{s}_i} + \frac{\partial y}{\partial \dot{\vec{s}}_i} \dot{h}_{\vec{s}_i} + \frac{\partial y}{\partial \ddot{\vec{s}}_i} \ddot{h}_{\vec{s}_i} \right) \\ + \Delta t \sum_{i=1}^{q} \left(p f_{\vec{s}_i} h_{\vec{s}_i} + p f_{\dot{\vec{s}}_i} \dot{h}_{\vec{s}_i} \right) \ ,\end{aligned}$$

or, appealing to the periodicity relations along with the difference relations of equations (6.34a) and (6.34b),

$$\begin{aligned}\delta L[h_{\vec{s}}] = \Delta t \sum_{i=1}^{q} \left(\frac{\partial y}{\partial \vec{s}_i} h_{\vec{s}_1} + \frac{\partial y}{\partial \dot{\vec{s}}_i} D_1 h_{\vec{s}_1} - \frac{\partial y}{\partial \ddot{\vec{s}}_1} D_2 h_{\vec{s}_1} \right) \\ + \Delta t \sum_{i=1}^{q} \left(p f_{\vec{s}_i} h_{\vec{s}_1} + p f_{\dot{\vec{s}}_i} D_1 h_{\vec{s}_1} \right) \ .\end{aligned}$$

From here the final optimality condition may be stated. It is necessary that $\delta L[h_{\vec{s}}] = 0$ for all appropriate directions $h_{\vec{s}} \in \mathcal{A}$. That is to say that it should be the case that at a minimizer

$$\left[\sum_{i=1}^{q} \left(\frac{\partial y}{\partial \vec{s}_i} + \frac{\partial y}{\partial \dot{\vec{s}}_i} D_1 - \frac{\partial y}{\partial \ddot{\vec{s}}_i} D_2 \right) + \sum_{i=1}^{q} \left(p f_{\vec{s}_i} + p f_{\dot{\vec{s}}_i} D_1 \right) \right] h_{\vec{s}_1} = 0 \ ,$$

for all appropriate directions $h_{\vec{s}_1}$. Hence we have the necessary condition

$$\sum_{i=1}^{q} \left(\frac{\partial y}{\partial \vec{s}_i} + \frac{\partial y}{\partial \dot{\vec{s}}_i} D_1 - \frac{\partial y}{\partial \ddot{\vec{s}}_i} D_2 \right) + \sum_{i=1}^{q} \left(p f_{\vec{s}_i} + p f_{\dot{\vec{s}}_i} D_1 \right) = 0 \ .$$

The optimality conditions for the periodic synthesis are now summarized as follows:

$$\frac{x_k - x_{k-1}}{\Delta t} = f\left(\frac{x_{k-1} + x_k}{2}, s^k, \dot{s}^k \right) \ ; \tag{6.42a}$$

$$x_0 = \text{given} \ ; \tag{6.42b}$$

$$\frac{p_{k-1}-p_k}{\Delta t} = f_x^t\left(\frac{x_{k-1}+x_k}{2}, s^k, \dot{s}^k\right)\left(\frac{p_k+p_{k-1}}{2}\right) + l_x^t\left(\frac{x_{k-1}+x_k}{2}\right) \ ; \tag{6.42c}$$

$$p_N = g_x^t(x_N) \ ; \tag{6.42d}$$

$$\sum_{i=1}^{q}\left(\frac{\partial y}{\partial \vec{s}_i} + \frac{\partial y}{\partial \dot{\vec{s}}_i}D_1 - \frac{\partial y}{\partial \ddot{\vec{s}}_i}D_2\right) + \sum_{i=1}^{q}\left(pf_{\vec{s}_i} + pf_{\dot{\vec{s}}_i}D_1\right) = 0 \ . \tag{6.42e}$$

Formulae for the optimality conditions summarized in (6.42) using specific functionals v, g, and l are now delineated. To begin, because we are concerned primarily with obtaining a final positional state, we shall select $l(x) = 0$ and

$$g(x) = \frac{1}{2}\alpha\sum_a (x^a - x_d^a)^2 \ , \tag{6.43}$$

where x^a represents the positional coordinates of interest and x_d^a is the desired final value of these states. Additionally, a very simple functional v is selected, although others could be used.

$$v(s^k, \dot{s}^k, \ddot{s}^k) \doteq \frac{1}{2}(w_1(s^k)^2 + w_2(\dot{s}^k)^2 + w_3(\ddot{s}^k)^2) \ , \tag{6.44}$$

where w_1, w_2, and w_3 are scalar weights. This definition for v implies

$$y = \frac{1}{2}\sum_{i=1}^{q}(w_1(\vec{s}_i)^2 + w_2(\dot{\vec{s}}_i)^2 + w_3(\ddot{\vec{s}}_i)^2) \tag{6.45}$$

as an expression for y. Using this expression and the difference relations (6.34a) and (6.34b) along with the periodicity relations, it is clear that

$$\begin{aligned}
\frac{\partial y}{\partial \vec{s}_i} &= w_1(\vec{s}_i)^t \ , \\
\frac{\partial y}{\partial \dot{\vec{s}}_i} &= w_2(\dot{\vec{s}}_i)^t = w_2(\vec{s}_1)^t D_1^t \ , \\
\frac{\partial y}{\partial \ddot{\vec{s}}_i} &= w_3(\ddot{\vec{s}}_i)^t = -w_3(\vec{s}_1)^t D_2 \ .
\end{aligned}$$

We now return to equation set (6.42) with these functional choices and their conse-

quences and state the equations to be during our investigations. These equations are

$$\frac{x_k - x_{k-1}}{\Delta t} = f\left(\frac{x_{k-1}+x_k}{2}, s^k, \dot{s}^k\right) \ ; \tag{6.46a}$$

$$x_0 = \text{given} \ ; \tag{6.46b}$$

$$\frac{p_{k-1} - p_k}{\Delta t} = f_x^t\left(\frac{x_{k-1}+x_k}{2}, s^k, \dot{s}^k\right)\left(\frac{p_k + p_{k-1}}{2}\right) \ ; \tag{6.46c}$$

$$p_N = g_x(x_N) \ ; \tag{6.46d}$$

$$q\left(w_1 I^{\bar{k}\times\bar{k}} - w_2 D_1^2 + w_3 D_2^2\right)\vec{s}_1 + \sum_{i=1}^{q}\left(p f_{\vec{s}_i}^t - D_1 p f_{\dot{\vec{s}}_i}^t\right) = 0 \ . \tag{6.46e}$$

In order to determine the minimizer $(\{x_k\}, \{p_k\}, \{s^k\})$ of L, equations (6.46) must be solved. One approach to this task is to use Newton's method to determine a zero of the variation δL. The appeal of using Newton's method is that, if the approximate gets within a sufficiently small ball of a root, the method is quadratically convergent. However, there are a couple of issues with using Newton's method in this instance, one of which is the following. Newton's method requires a derivative of the function whose root is sought. This requires taking a second derivative of f. Even if finite difference approximations are used, the resulting difference operators are not easily described, especially when the requirement of periodicity is introduced.

In lieu of this, we employ a fixed-point method or time marching scheme. This method and the manner in which it is implemented is now explained.

1. The method is initiated by selecting an initial shape sequence $\{s^k\}_{k=1}^{\bar{k}}$ over the fundamental period p_f. We shall denote this initial shape s_0. This sequence is then repeated over the time interval selected to arrive at the full sequence, denoted in vector form, $\vec{s}$.

2. Next, the iteration process begins. Using s_0, equation (6.46a) is solved for the sequence $\{x_k\}_{k=1}^{N}$. This solution is denoted x^0 and the successive solutions

of this equation x^j. Note that a superscript with index j is used instead of a subscript with index k on x in order to clarify that we are referring to an iteration of the optimal control solution method rather than an iterate of the sequence $\{x_k\}_{k=1}^N$.

3. Once x^0 and s_0 are formed, equation (6.46c) can be solved for the sequence $\{p_k\}_{k=0}^N$. This solution is denoted p^0 and the notation p^j is used for successive iterations. Again, note that a superscript with index j is used for clarity.

4. Having formed x^0 and p^0, these sequences along with s_0 are applied to create the second term, $pf \doteq \sum_{i=1}^q \left(pf^t_{\vec{s}_i} - D_1 pf^t_{\dot{\vec{s}}_i} \right)$, of equation (6.46e).

5. Finally, consider the following fixed point iteration scheme in s_j:

$$\frac{s_{j+1} - s_j}{\delta t} + q\left(w_1 I^{\bar{k}\times\bar{k}} - w_2 D_1^2 + w_3 D_2^2 \right) s_{j+1} + pf(s_j) = 0 \ . \tag{6.47}$$

δt is some small quantity that has meaning only in the context of this scheme and has nothing to do with the earlier discritizations. From this scheme we obtain s_{j+1} and create $\vec{s}$. We then return to step 2 and continue to repeat these steps until acceptable convergence is attained.

Before moving on with discussion of implementation note that the scheme of step 5 is consistent with the optimal control or gait objective. For suppose that $x^j \to x^*$, $p^j \to p^*$, and $s_j \to s_*$. Then s_* is a fixed point of the equation implied by (6.47) and we arrive at the expression

$$q\left(w_1 I^{\bar{k}\times\bar{k}} - w_2 D_1^2 + w_3 D_2^2 \right) s_* + pf(s_*) = 0 \ , \tag{6.48}$$

indicating that (x^*, p^*, s_*) satisfy the necessary optimality conditions of equation set (6.46).

There is a little bit of detail concerning the implementation of this iteration scheme that should be discussed. Firstly, concerning step 2 we note that the vector field f

is non-linear and the iterates x_k^j are implicitly defined. Thus, we cannot simply determine a closed form expression that will allow time-marching of the state x. To determine the iterate x_k^j from x_{k-1}^j Newton's method is used to search out a root of the equation

$$E(x) \doteq \frac{x - x_{k-1}^j}{\Delta t} - f\left(\frac{x_{k-1}^j + x}{2}, s_j^k, \dot{s}_j^k\right) \ . \tag{6.49}$$

Of course Newton's method requires an initial guess at the root $x = x_k^j$ that is sought. Since f is a continuous vector field, if Δt is sufficiently small, x_{k-1}^j should be near x_k^j. Thus by using the initial guess $x = x_{k-1}^j$ it only takes a few iterations of Newton's method before its quadratically convergent nature takes hold and the approximate arrives at the value desired. In fact, we only use 3 Newton iterations in our implementation.

Once x^j is obtained it is a simple task to determine p^j. Since equation (6.46c) is linear in the iterates p_k^j, equation (6.46d) along with x_N^j are used to calculate p_N^j, and then p_k^j is marched backward in time.

The iteration scheme of equation (6.47) is also designed to be linear in the iterates s_j. So, after x^j and p^j have been captured, one may solve for s_{j+1} from s_j in a single iteration. Let us discuss this scheme in a bit more detail. If we should rewrite equation (6.47) as follows:

$$\begin{aligned} s_{j+1} &= s_j - \delta t \left(q\left(w_1 I^{\bar{k}\times\bar{k}} - w_2 D_1^2 + w_3 D_2^2\right) s_{j+1} + p f(s_j)\right) \\ &= s_j - \delta t \, d_j \ , \end{aligned}$$

it is seen that this scheme has the form of a typical optimization scheme with search direction d_j and step length δt. This is only in form due to the fact that the search direction,

$$d_j \doteq q\left(w_1 I^{\bar{k}\times\bar{k}} - w_2 D_1^2 + w_3 D_2^2\right) s_{j+1} + p f(s_j) \ ,$$

is dependent on both s_j and s_{j+1}. One should recognize d_j as an approximation to

the gradient of L with respect to $\vec{s}$, and from this point of view, this scheme is a rendition of the method of steepest descent. As such, it is relatively clear that d_j is a legitimate search direction. However, as is well known, one must be careful with step length selection, δt. If one moves too far, or not far enough, along the chosen direction, the iteration scheme may diverge or become stagnant. Often it is the case that some adaptive method of selecting δt from iteration to iteration is employed. It is the line search and trust region methods that we speak of [55, 5]. We do not employ such a scheme in our implementation. Instead, we simply select δt conservatively, say, on the order of 10^{-4}. We found that although the scheme does not illustrate rapid convergence behavior, it does make sufficient progress toward convergence.

Now that the optimal control method and its implementation have been described and discussed, attention is turned to its use in determining optimal gaits and answering some scientific questions of interest. The next section is devoted to this end.

6.6 Optimal Control Results

This section includes excerpts of several "experiments" in optimal locomotive shape selection for the serial-link structure $\mathcal{S}$. A large portion of the experimentation presented is oriented toward discovering the effects that particular optimization design parameters can have on performance. In general, the selection of these parameters can be very important and thus such an effort is justified. We begin with a set of parameters that appear to achieve, at some level, the basis task of directed locomotion. From that point we will graduate to some small experiments directed at either improving the performance of the resulting control or altering the shape in some specified way while preserving the performance. At all times, however, our focus will be on the trends of the resulting shape. After exploring optimization design, we will then use our methodology to explore some issues of more general scientific interest. In particular, we explore the effects that lateral forces can have on optimal locomotive

shapes and to what extent propulsion and locomotive shape depend on the length of the device.

6.6.1 The Initial Experiment and Results

To begin, the values of physical parameters of $\mathcal{S}$ and its environment that are to be used for the majority of the experimentation are specified. With exception to the size of $\mathcal{S}$, which shall be taken to be $n = 5$, we will take these values to be the same as those used for the simulation experiments of Chapter 5. It is hoped that this will allow for some comparison and intuition concerning the validity of experimental results.

With an instance of the physical system selected, the goal of the initial control design experiment may be described. The requests are levied by means of the cost functional parameters. Firstly, it is required that the shape s of $\mathcal{S}$ be selected to be periodic with fundamental period p_f. The second request is that this shape must elicit a displacement of the head B_1 with zero system initial conditions from an initial location at the origin of coordinates (i.e. $(x_1, y_1)^t = (0, 0)$) to a point that lies away from the origin of coordinates in one of the following eight directions: east, southeast, south, southwest, west, northwest, north, and northeast. The points that lie along these directions are selected with ordinate values of magnitude 0 and 3. The ordinate values represent length in the units of ft. Thus, these points lie on a box of side length 3 ft centered at the origin. The value of 3 was selected due to the fact that the structure that has been selected has a total longitudinal length of ≈ 2.5 ft. So, by this selection, we are starting out with a somewhat modest request that the structure crawl just over a body length in a specified direction. The directions described, the coordinates of the corresponding terminal points along those directions, as well as the notation that will be used hereafter to refer to these directions and their associated points, are summarized in Table 6.1.

The choice of parameter values related to the optimization method which shall be

	South	Southeast	East	Northeast	North
Notation	S	SE	E	NE	N
$(x_1, y_1)_d$	$(0, -3)$	$(3, -3)$	$(3, 0)$	$(3, 3)$	$(0, 3)$

	Northwest	West	Southwest
Notation	NW	W	SW
$(x_1, y_1)_d$	$(-3, 3)$	$(-3, 0)$	$(-3, -3)$

Table 6.1: Directions used in the control experiments, their related notations, and the specific destination coordinates.

used during this initial experiment is provided in Table 6.2. The initial shape iterate

Functional Parameters	α	w_1	w_2	w_3
Parameter Values	500	1	1	0.1

Discretization Parameters	t_f	Δt	N	p_f
Parameter Values	5	0.05	100	1

Table 6.2: Cost functional, discretization, and periodicity parameters used for optimal shape experiments.

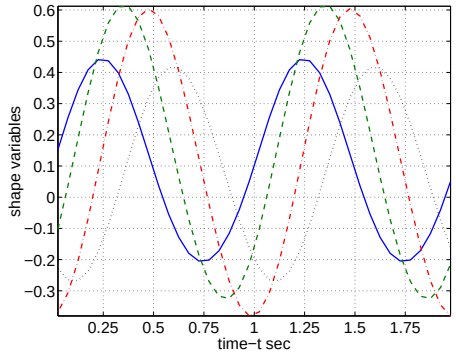
is selected to be $s_0 = \{(\phi_1^k, \phi_2^k, \phi_3^k, \phi_4^k)^t\}_{k=1}^{\bar{k}}$, where $\phi_j^k \doteq 0.1 \sin(2\pi t_k - 0.7(j-1))$. The reasoning for doing so should be transparent. The selection of the amplitude and phase-shift parameters of this shape are rather random.

Performing minimization for each of the directions of Table 6.1 we arrived at a minimizing shape sequence $\{s^k\}_{k=1}^{\bar{k}}$, which will simply be called the optimal shape s, state sequence $\{x_k\}_{k=1}^N$, and costate sequence $\{p_k\}_{k=1}^N$. Figures 6.1, 6.2, and 6.3 illustrate two periods of the optimal shape and the resulting behavior of the states x_1 and x_2 (y_1) upon driving the dynamics (6.26) with a cubic-spline interpolation of the optimal shape. The shape only covers a time span of a single period and thus was repeated as necessary to cover the time span of the feedforward integration. For these simulations we integrated the dynamics over the time span of 0 to 15 sec. Thus

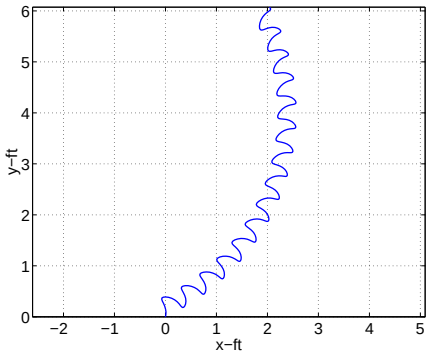
the optimal shape was repeated $3q$ times.

It would seem that, for the most part, the initial experiment was a successful one. Figure 6.1, which illustrates the eastern directions, is the most convincing of the plates. However, the others also indicate promise. In the case of the eastern directions, it is noted that the general direction of the head displacement seems correct. Especially when one keeps in mind that the directive under which the shape operated was to deliver the head to the position coordinates of Table 6.1 at the end of precisely 5 periods. If we only look at the direction of the first five periods of the head trajectory it is seen that the head was on a direct course toward the desired terminal points.

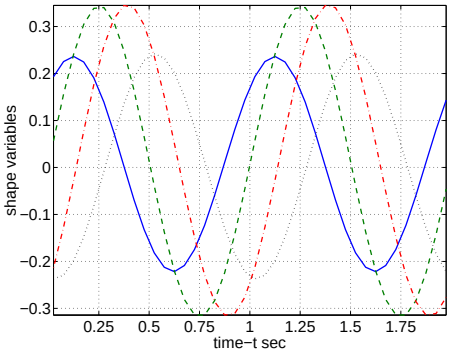
The results for the western directions of SW, W and NW, presented in Figure 6.2, were not quite as in line with the requests that have been made. However, there was more going on with these simulations than an initial glance will allow one to perceive. Recall the form of the initial shape iterate s_0 or more precisely, the form of ϕ_j^k. There is a -0.7 s time delay or phase shift along the index j. For an actual snake, this delay represents the wave of muscle contraction which propagates along the length of the animal in order to produce locomotion. The fact that the delay is negative indicates that the wave begins at the head B_1 and travels toward the tail B_n of the structure or the animal. As can be seen for the eastern directions, upon convergence to the optimal shape, there was still a presence of the phase shift or delay and the propagation in the direction of the tail also remained. However, when looking at the optimal shapes for the western directions we find something that is most interesting. Indeed, there was still a phase shift or time delay present among the shape variables. However, the ordering of the shape curves were opposite to that of the initial iterate. That is, the propagation of the wave moved from the tail B_n to the head B_1. This observation means quite a lot. Note that due to the initial condition $x_0 = \vec{0}$ and the initial shape iterate s_0, the structure $\mathcal{S}$ was positioned to move head first in the eastern direction. However, it would make sense that if the structure needs to move



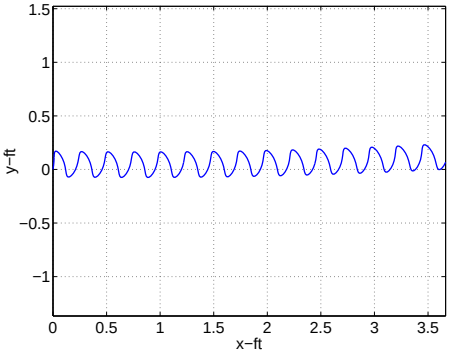
(a) Optimal NE shape.



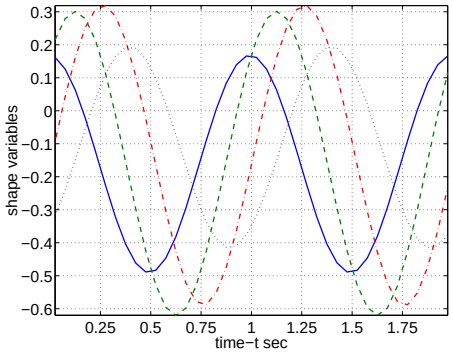
(b) Feedforward motion.



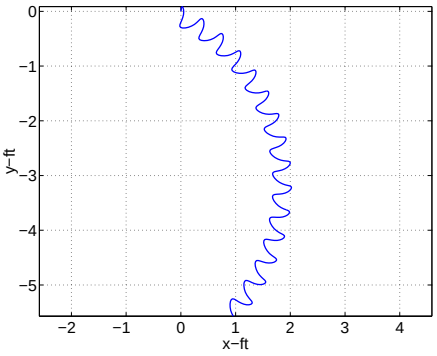
(c) Optimal E shape.



(d) Feedforward motion.

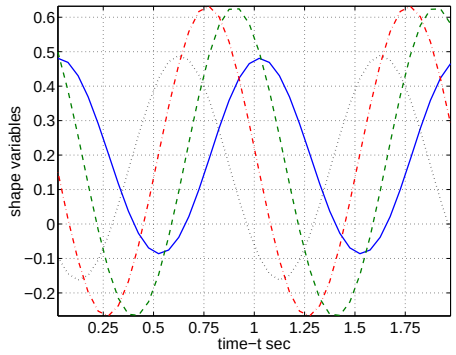


(e) Optimal SE shape.

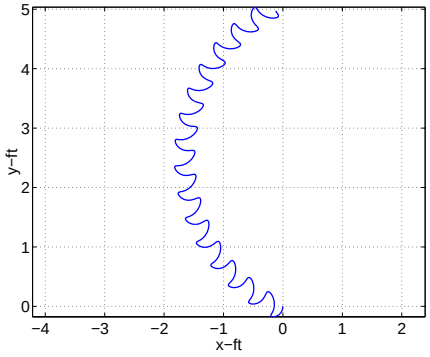


(f) Feedforward motion.

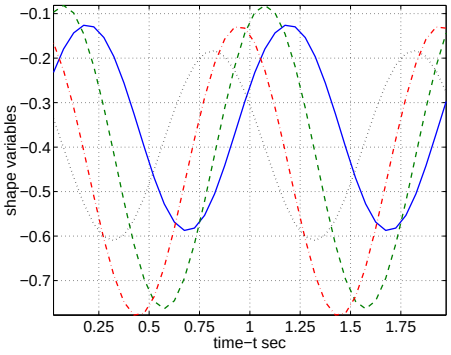
Figure 6.1: The optimal shapes for producing snake locomotion in the eastern directions along with the feedforward motion of B_1 corresponding to each of them.



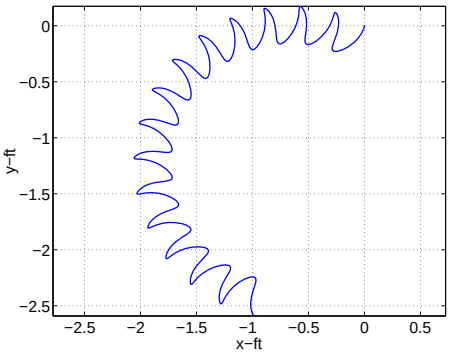
(a) Optimal NW shape.



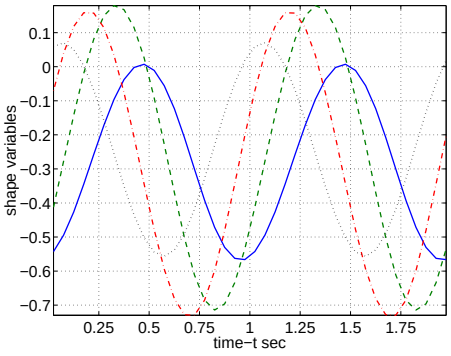
(b) Feedforward motion.



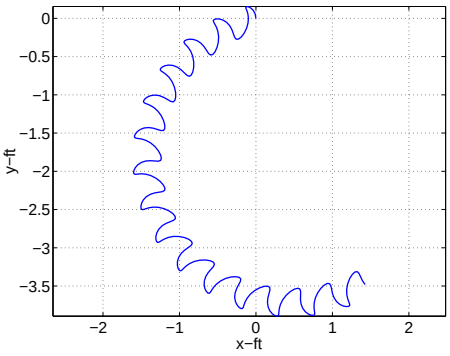
(c) Optimal W shape.



(d) Feedforward motion.

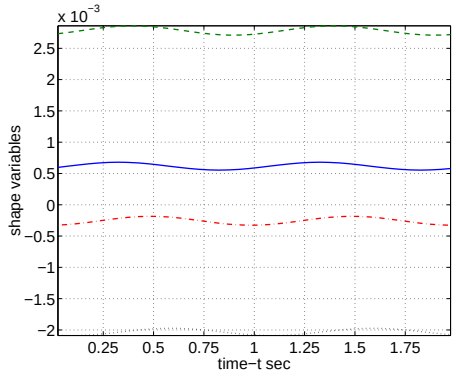


(e) Optimal SW shape.

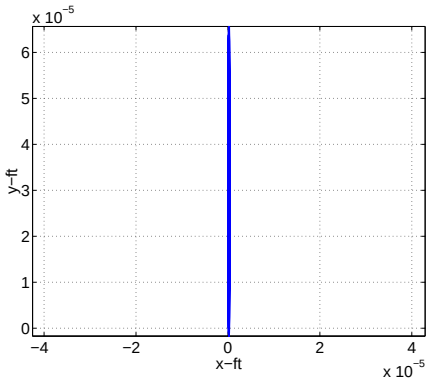


(f) Feedforward motion.

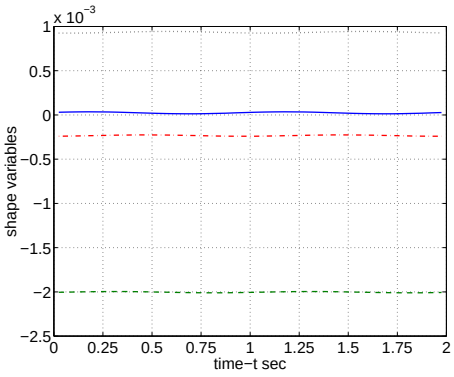
Figure 6.2: The optimal shapes for producing snake locomotion in the western directions along with the feedforward motion of B_1 corresponding to each of them.



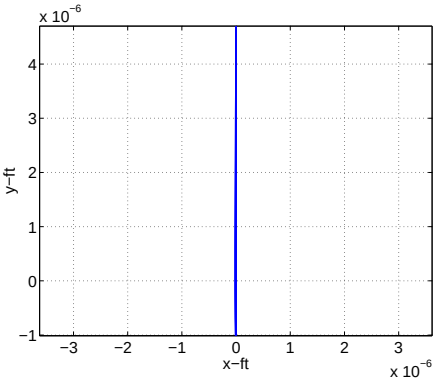
(a) Optimal N shape.



(b) Feedforward motion.



(c) Optimal S shape.



(d) Feedforward motion.

Figure 6.3: The optimal shapes for producing snake locomotion in the pole directions along with the feedforward motion of B_1 corresponding to each of them.

in a western direction it would not be optimal to move in the eastern direction, turn around, and then come back toward the west. What is optimal, given the initial conditions, is for the structure to treat its tail B_n as if it were its head B_1 and simply travel tail first in the western direction. For in our formulation, the head is qualitatively no different from the tail. This is precisely what the optimal shape method has determined. Instead of moving head first, the shape was selected so as to produce a tail first locomotion, dragging the head toward the desired direction.

As already mentioned, the resulting head trajectories seem to indicate that there is some accuracy lost in this transition. For the first five cycles of the head trajectory in the western directions did not seem nearly as in line with their the target as those for the eastern directions. But still, the fact that the method we are employing is robust enough to completely change the direction of shape wave propagation along the structure when asked to move in the direction opposite that for which it is initially oriented, is quite impressive.

Lastly, there is the pole directions N and S. Again, the results for these direction are found in Figure 6.3. In these cases $\mathcal{S}$ failed to move in the desired direction. In some sense, this is not surprising. For, based on the the initial data x_0, we were requesting that the structure move normal to itself. That would seem difficult to do. In the case of the parameter selection for this initial experiment, it was apparently too difficult. The optimal thing for $\mathcal{S}$ to do was nothing, as indicated by the optimal shape and corresponding head trajectories. Although the head did move toward the target, it did not expend any effort to do so. Since both contribute to the cost, the results seem to indicate that the effort that it would take to make the directed motion outweighs the cost of zero displacement.

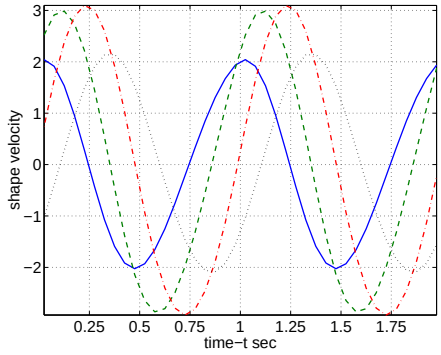
In addition to the failure of the initial control design in the pole directions, it must be admitted that the displacements observed for the other directions were also less than completely satisfactory. For although there was a correct directed displacement, the structure fell short of its target and did not accomplish a displacement of

approximately one body length. In our next experiment we shall attempt to remedy this.

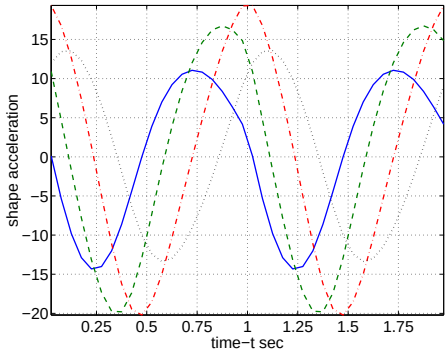
Before leaving this initial effort, however, we would like to indicate that there seems to be a recognizable pattern for the optimal shape. Firstly, the shape forms seem to be very sinusoidal. This observation is in part confirmed by plotting the shape velocity and acceleration. Figure 6.4 contains plots of the optimal shape velocity (left column) and optimal shape acceleration (right column) corresponding to the eastern directions under the initial experimental setup. Note that these velocities and accelerations are also sinusoidal. This observation would seem to provide support for Hirose's claim that sinusoidal curvature changes are the most natural [33].

Secondly, there are amplitude modulations and phase shifts. In light of (5.2), both are expected. The author personally had anticipated amplitude modulations that brought about shape trajectories similar to those of an eel or fish. In all cases considered we found that the amplitude of the structure's differential angles would be selectively increased from the head of the system to the midpoint thereof, just as in the case of the eel. However, at that point, the optimization method was clever enough to determine that instead of continuing this increase along the length of the body, it was just as effective to then reverse the trend, decreasing the amplitude along the posterior portion of the structure. This makes very good sense. For the ability of the structure to equalize its opposing friction is found in the magnitude of the curvature differential, not necessarily its sign. Hence, by decreasing the amplitude of the differential angles along the posterior of the system, the same curvature differential magnitude is obtained as would be produced by a continued amplitude increase, but the actuation effort used to obtain this differential is reduced.

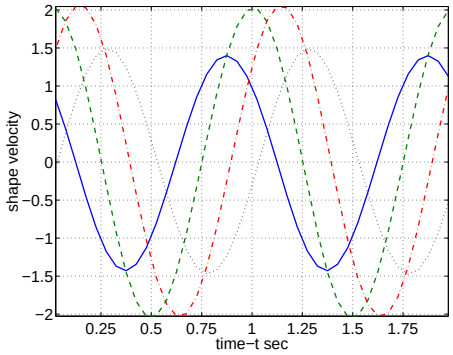
The phase shift seems to work toward grouping together, in some sense, ϕ_1 with ϕ_2 and ϕ_3 with ϕ_4. Further phase shifting and the amplitude modulation tend to ensure that ϕ_2 and ϕ_3 are out of phase with ϕ_1 and ϕ_4, respectively, and that the trajectories of ϕ_2 and ϕ_3 have an amplitude such that they approximately intersect



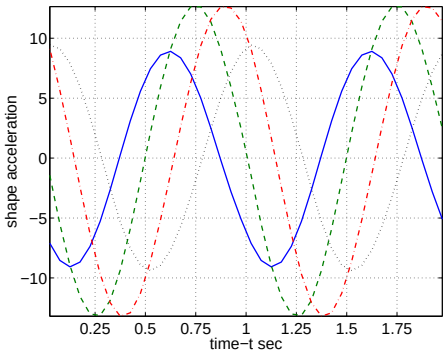
(a) Optimal NE velocity.



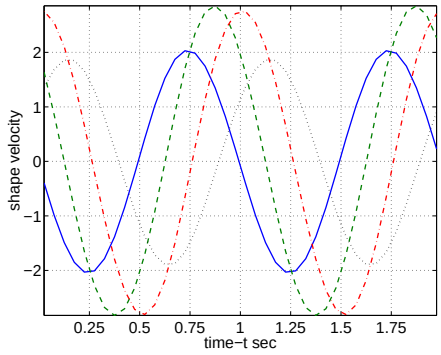
(b) Optimal NE acceleration.



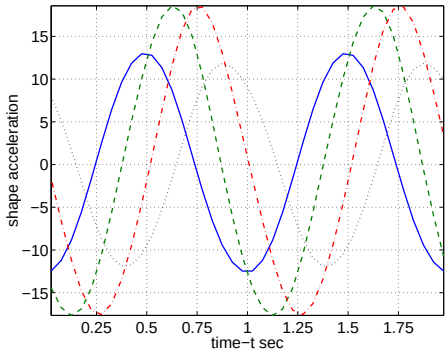
(c) Optimal E velocity.



(d) Optimal E acceleration.



(e) Optimal SE velocity.



(f) Optimal SE acceleration.

Figure 6.4: The optimal shape velocities and accelerations used for producing snake locomotion in the eastern directions.

the maximum and minimum values of ϕ_1 and ϕ_4, respectively.

Finally, the trajectory of the head appears to oscillate about circular trajectories. The radius of curvature of these trajectories and the direction of the principle unit normal vector or curvature vector are clearly related to the bias of the optimal shape trajectories away from zero. This makes perfect sense. Recall that the shape variables actually represent a curvature differential. Thus, if these variables have a constant bias in the positive direction the structure will arc left along its length. If the shape variables have a constant bias in the negative direction, then the structure will arc right along its length. The magnitude of the constant bias corresponds to the radius of curvature of the arc formed by the body and that of the trajectory ultimately followed by the head. Other than this bias or offset, there doesn't appear to be any difference in the optimal shapes used for turning or straight path motion. In fact, taking the point-of-view that lines are degenerate circles, we can claim that it appears that the centerline path of the structure always follows a circular path to reach the desired target, altering the radius of curvature appropriately to change the direction. This is in complete agreement with the heuristic steering approach taken by Hirose et. al. [18], and the feedback direction controls proposed by Ostrowski and McIsaac [46] and Saito et. al. [62].

It is recalled that the German team of Linnemann, Paap, Klaassen, and Vollmer had illustrated that the optimal trajectories for a serial-link structure driven by wheels were circular arcs [43]! Apparently, even though the nature of the serial-link structure we consider requires a non-constant curvature to elicit gait, it is still most optimal to take circular paths, varying away only as necessary to make use of the undulatory propulsive mechanism.

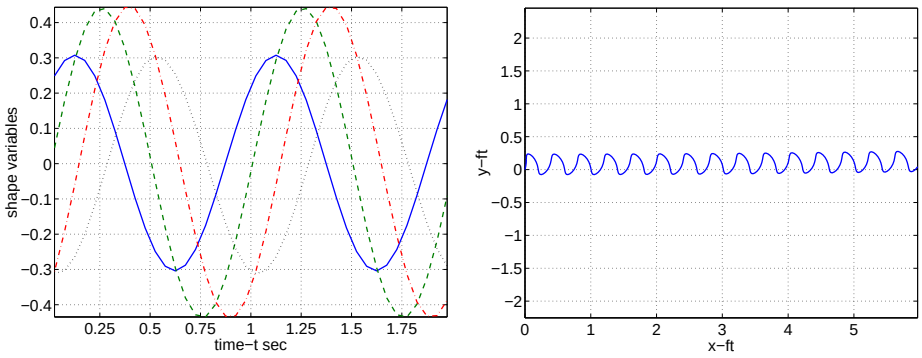
6.6.2 Experimentation with the Design Parameter α

The second control design experiment that we perform investigates the role that the parameter α can play in the performance of the resulting optimal shape/controller.

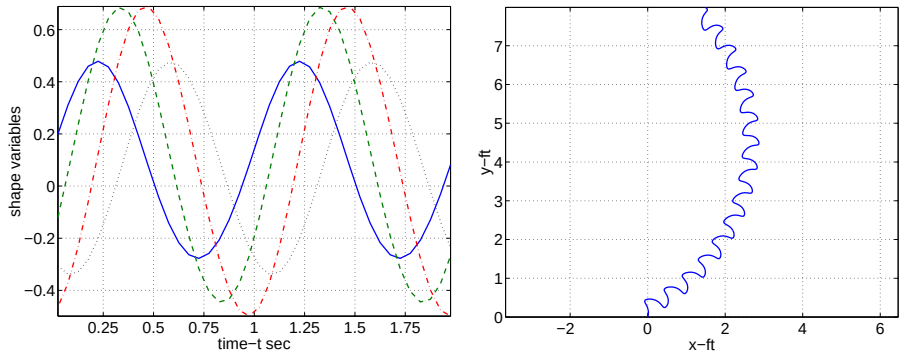
This parameter effectively penalizes the discrepancy between the coordinates of the final position of the head and the coordinates desired. It was noted that the performance of the optimal shapes obtained from the design of Table 6.2 were slightly unsatisfactory in the sense that the head did not make it to the desired coordinates specified in Table 6.1. In the case of the pole directions, the resulting optimal shapes were a complete failure with respect to motion. So, it makes sense that we should try increasing α to close this gap between the desired performance and the performance obtained.

The directions of N, NE, and E are selected to experiment with. We could, for completion, work with all of the directions. However, it does not seem as though this is or should be necessary. Firstly, although there is no difference between the tail and the head of the structure qualitatively speaking, there is a difference from the point of view that, although perhaps at a crude level, the resulting motions of the serial-link structures are likened to those of an actual snake. Therefore, head first motions are preferable. The initial experiment indicated that when the structure is asked to move in a direction more in line with the tail, it travels tail first. Thus we shall avoid asking that the structure move in such directions, which are the western directions. Secondly, it is rather clear that the directions S and SE are, with the exception of sign, no different than the respective directions N and NE directions. For instance, the optimal shape for the SE direction should be the negative of the optimal shape for the NE direction. Thus, we no longer concern ourselves with these directions either.

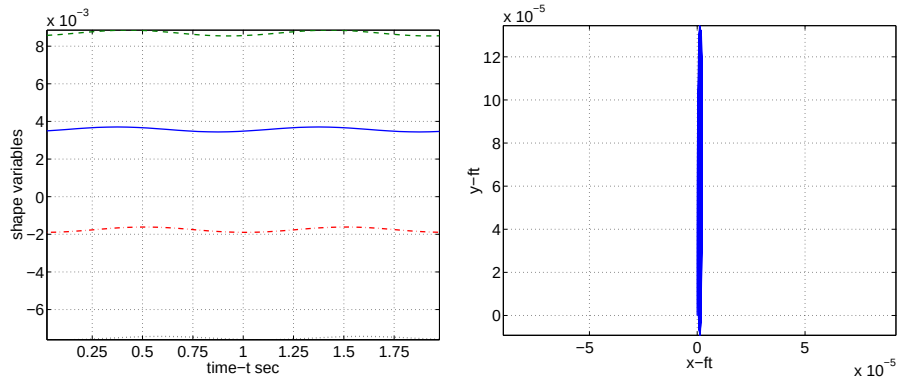
For each direction we obtain optimal shapes corresponding to modification of the value of α from 500 to 1000 and then 1500. Again, to gauge the performance of the minimizing shape, we examine plots of the optimal shape along with the corresponding feedforward head trajectory over the time span of 15 s. The results for $\alpha = 1000$ are provided in Figure 6.5 and those for $\alpha = 1500$ are provided in Figure 6.6. At first glance of the results for $\alpha = 1000$ it would seem that this change had little impact.



(a) Optimal E shape. (b) Feedforward motion.

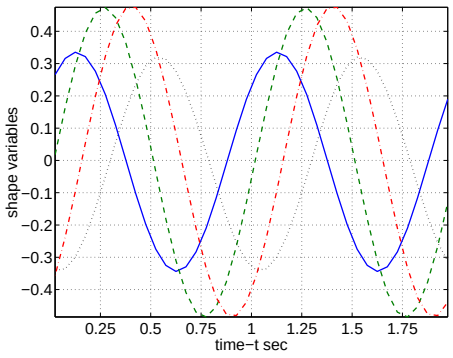


(c) Optimal NE shape. (d) Feedforward motion.

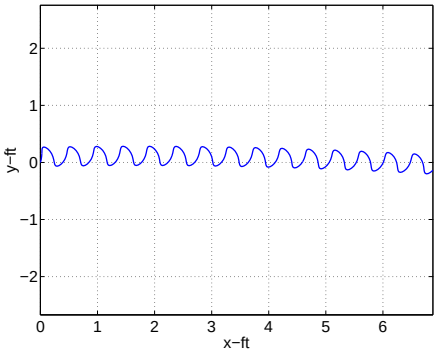


(e) Optimal N shape. (f) Feedforward motion.

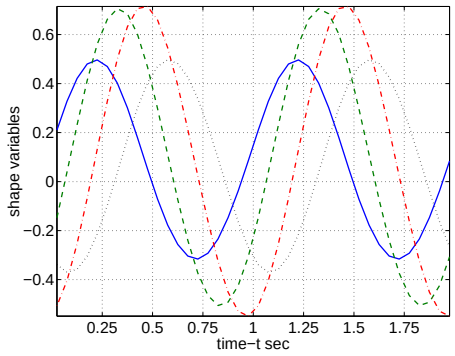
Figure 6.5: The optimal shapes for producing snake locomotion in the specified directions along with the feedforward motion of B_1 corresponding to each of them. Here α has been increased to 1000.



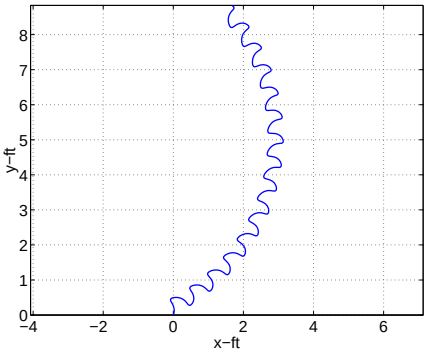
(a) Optimal E shape.



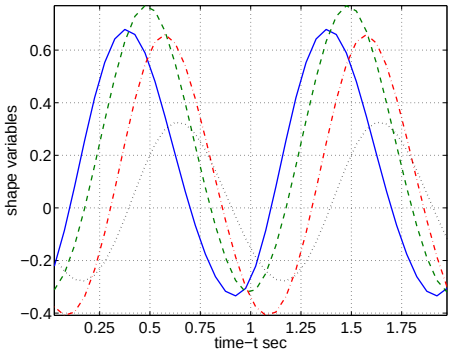
(b) Feedforward motion.



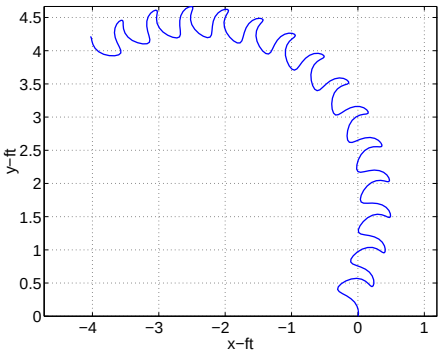
(c) Optimal NE shape.



(d) Feedforward motion.



(e) Optimal N shape.



(f) Feedforward motion.

Figure 6.6: The optimal shapes for producing snake locomotion in the specified directions along with the feedforward motion of B_1 corresponding to each of them. Here α has been increased to 1500.

The E and NE shapes and head trajectories look about the same as those initially obtained and the structure still did not move in the N direction. However, upon closer examination it is seen that there are differences. To begin, note that in the case of the directions E and NE the structure travelled further in its allotted 5 s. For $\alpha = 500$ the head did not even travel 1.5 $feet$ in the eastern direction over the time span of 5 periods. However, in the case of $\alpha = 1000$, the head moved ≈ 2 ft in a 5 period span of time. A distance differential can also be observed for the NE direction. Further, note that in the case of $\alpha = 500$, allowing the structure to continue its pattern of motion beyond the time period specification of $t_f = 5$ s did not result in the head making it to its desired termination point. However, looking at the results for direction E with $\alpha = 1000$, we see that the shape had been selected in such a way so as to align the coordinates of the head with its rendezvous point of $(3, 0)$. So, the shape in this case took B_1 further and put it in line with the specified directive, even though it still failed to ultimately obtain it. That is an improvement. In the case of direction NE, similar improvements were observed. However, it was still the case that it is more optimal for the structure not to move at all than it is for the head to travel normal to itself.

The results of the design with $\alpha = 1500$ show that further improvement can be made in the selection of the shape by increasing α. Again, in the case of direction E and direction NE, the structure traveled further in a 5 period time span and was more in line to actually intersect the rendezvous points of those directions upon continuation of the motions. Note that in the case of the E direction B_1 actually travelled a whole body length in the 5 period time span. Also, in the case of the NE direction, the head came much closer to $(3, 3)$ than before. Additionally, we finally got the head to move in the northern direction. In fact, the head moved over a body length in a 5 period time span and upon continuation the head trajectory intersected its target coordinates of $(0, 3)$. From this simulation it became clear why this behavior was not elicited from the previous designs. When travelling in the northern direction

the amplitude of the shape variable ϕ_1 increases significantly when compared to the trajectory obtained for ϕ_1 in the other directions. This, of course, increased the cost incurred by the shape and led it to outweigh the cost of not moving in the specified direction.

The conclusion of this experiment is that we must select α large to obtain the desired performance. We could increase this value further and continue to seek improvements of the types we have seen. However, we shall accept the results of the alteration of the value of α from 500 to 1500 as sufficient and move on to experiment with the effects of other parameters.

6.6.3 Experimentation with the Initial Fixed-Point Scheme Iterate

The matter of selecting the initial shape iterate s_0 is now given further consideration. It is good measure to experiment with initial iterates other than that previously selected, namely those that do not contain the bias toward locomotion, to determine what effects these selections can make on the optimal shape obtained upon convergence of the optimization method. The possibility of multiple local minima is always present. So, it could perhaps be the case that different, and perhaps, better shapes could be obtained should it be so that a different initial shape iterate is used.

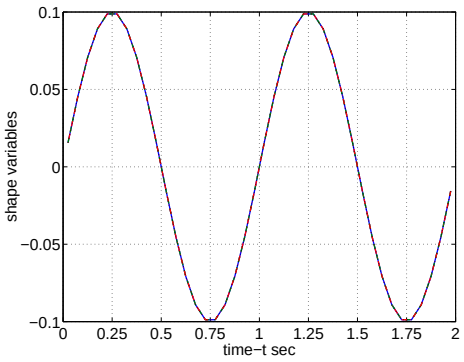
In the attempt to obtain a glimpse of what variability is present in the resulting optimal shape we try three initial shape iterates that are different from the one used up to this point and ask that the structure move in direction E. We only work with this single direction hereafter, for there is little to be learned concerning the optimal shape by altering directions at this point. The first alteration comes in destroying the phase shift in the initial iterate. Thus, we make the selection $s_0 = \{(\phi_1^k, \phi_2^k, \phi_3^k, \phi_4^k)^t\}_{k=1}^{\bar{k}}$, where $\phi_j^k \doteq 0.1 \sin(2\pi t_k)$. Two periods of this iterate are shown in Figure 6.7(a). The optimal shape that resulted from this initial guess at the shape is shown in Figure 6.7(b) and the the corresponding feedforward head trajectory in Figure 6.7(c).

It was found that there is some difference between the optimal shape trajectories determined here and those resulting from the previous initial shape iterate. Namely, it seems that the amplitude modulation is slightly different. However, the fact that the amplitude of the inner shape variables is significantly larger than that of the outer shape variables remains unchanged. Other than this alteration and a change in bias, however, there does not appear to be any other differences between the results and their performance with regards to the displacement of the head is similar. This observation confirms the essential nature of the phase shift for locomotion. The same type of phase shift modulations occur, even though no hint at the need for a phase shift was provided in the initial shape iterate.

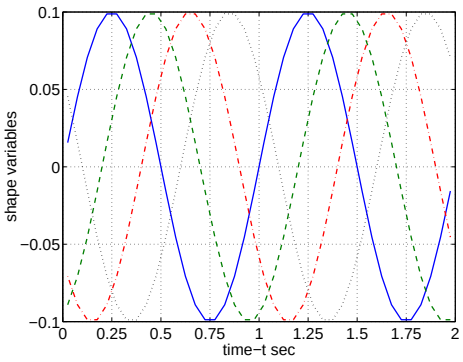
We next address the question of whether or not it may be beneficial to provide an initial iterate with a more structured phase shift than previously used. In particular, we would like to know if the uniform phase shift used in defining the shapes of Chapter 5 would provide a different optimal shape sequence. This uniformity was used to guarantee that at each instance in time the structure would conform its shape to a corresponding time instance of a complete period of a travelling wave. So, this inquiry is used to get a hint at whether there was any advantage in this selection of a uniform phase shift along a period. The initial iterate to be used is $s_0 = \{(\phi_1^k, \phi_2^k, \phi_3^k, \phi_4^k)^t\}_{k=1}^{\bar{k}}$, where $\phi_j^k \doteq 0.1 \sin(2\pi(t_k - \frac{(j-1)}{n}))$. The same illustrations as provided for the initial shape iterate with no phase-shift are provided in the right column of Figure 6.7. As in the previous experiment, the resulting optimal shape trajectories did differ from those previously obtained in both amplitude modulation and bias. However, again it was the case that with the exception of these differences, there does not appear to be any significant changes in the results obtained for an initial shape iterate with uniform spacing along a period and that obtained for initial iterate with the a randomly selected uniform phase shift of -0.7.

Due to the fact that the outcome of these experiments do not illustrate any great advantage or disadvantage in biasing the initial iterate with a phase shift, we do

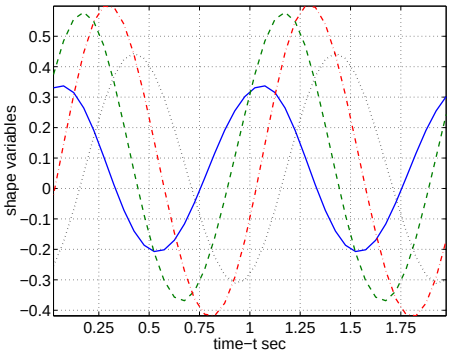
not see any purpose in continuing to do so and thus adopt an initial iterate with no phase shift for experimentation henceforth. As our final trial in experimenting with the effects of de-biasing the initial shape iterate, we remove the sinusoidal functional form from the initial iterate. By doing so, we hope to ensure that the sinusoidal functional form of all the optimal shapes determined up to this point was not simply a repercussion of beginning with this form. Thus we select the initial shape iterate $s_0 = \{(\phi_1^k, \phi_2^k, \phi_3^k, \phi_4^k)^t\}_{k=1}^{\bar{k}}$, where $\phi_j^k \doteq 0.01$. This is about the flattest periodic initial iterate that could be imagined. The zero initial iterate could be selected, but this is a singular point of the dynamics and results in the zero sequence $\{s_j\}_{j=0}^{\infty}$, $s_j = 0$. Thus, we start a very small bias in the curvature. Figure 6.8 illustrates the results of this choice. Subfigure 6.8(a) is a plot of the initial shape iterate, Subfigure 6.8(b) shows the result of 2 iterations of the fixed-point procedure, and Subfigures 6.8(c) and 6.8(d) illustrate the optimal shape and feedforward trajectory obtained for the head of the structure. The shape after two iterations of the method is provided to illustrate how quickly the shape began to evolve away from the constant initial iterate. Interestingly enough, the resultant optimal shape and its performance were almost precisely the same as that obtained for the sinusoidal initial iterate with no phase shift. In fact, we would guess that they are in fact identical with the difference being that one of the two of these shapes is a bit closer to exact convergence than the other. Thus, the conclusion is that absolutely nothing was gained with respect to the resulting optimal shape by selecting an initial shape iterate of sinusoidal form, and that the approximately sinusoidal results that have been observed up to this point were not due to such a selection. One advantage, however, that was gained by choosing the sinusoidal initial iterate was increased convergence speed. It takes a significantly longer time to reach convergence when beginning with a constant iterate. For this reason, we shall use a sinusoidal initial iterate for the remainder of the experiments.



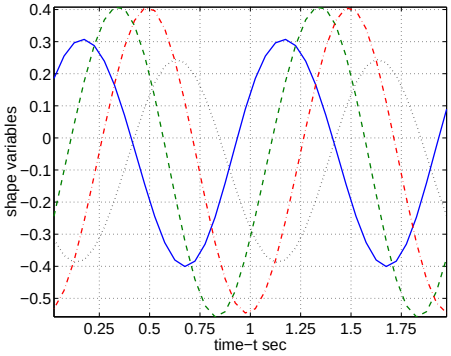
(a) s_0 with no shift.



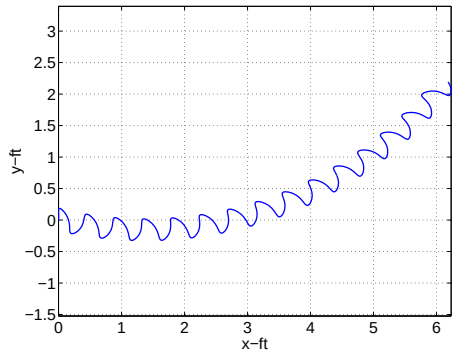
(b) s_0 with uniform shift.



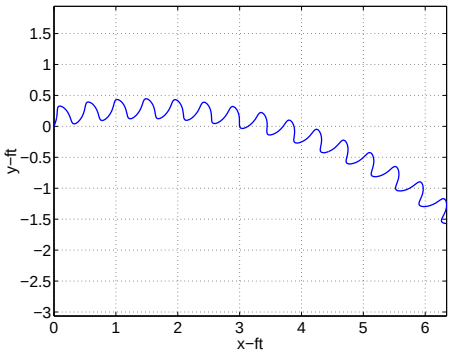
(c) Optimal s, no shift.



(d) Optimal s, uniform shift.

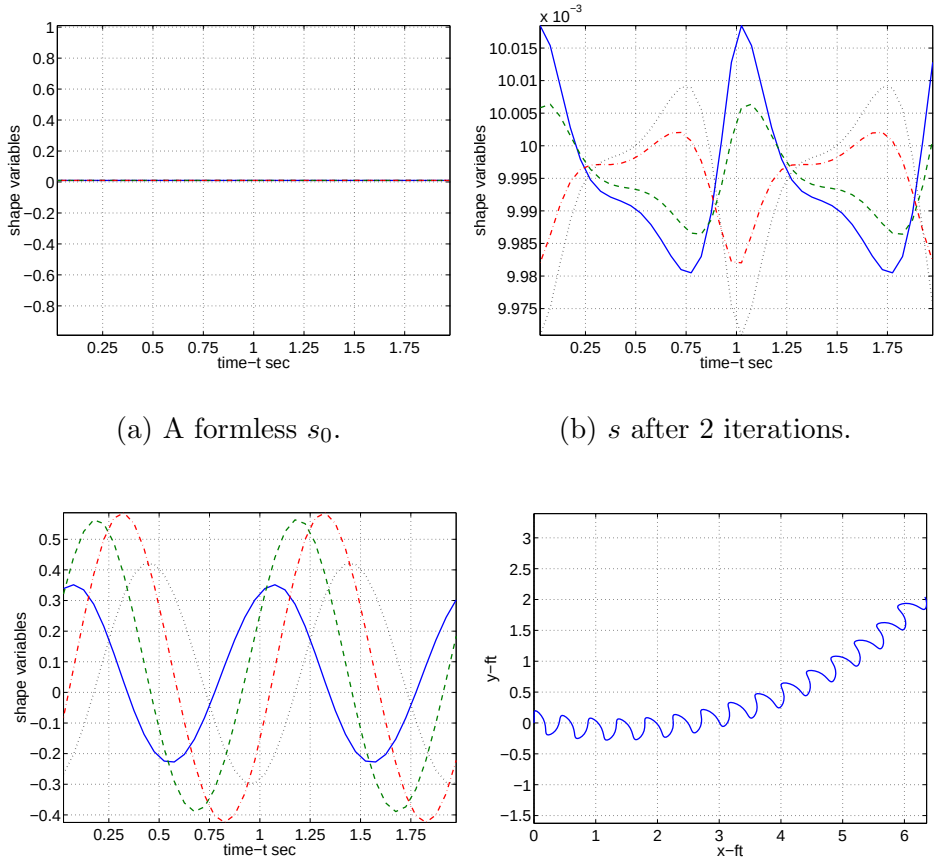


(e) B_1 motion, no shift.



(f) B_1 motion, uniform shift.

Figure 6.7: Optimality results for initial shape iterates with phase-shift changes.



(a) A formless s_0.

(b) s after 2 iterations.

(c) The optimal s.

(d) Motion of B_1.

Figure 6.8: Optimality results for an initial shape iterate with no form.

6.6.4 The Effects of the Shape/Control Weights

Since we have satisfied the question of whether or not the results we obtained in the first few experiments were overly influenced by the selection of the initial shape iterate s_0, we may return to concerning ourselves with the effects of the cost design. We have already experimented slightly with the parameter α of g. Thus it remains to determine the possible sensitivity of the resulting optimal shape sequence to the selection of the parameters w_1, w_2, and w_3 of v.

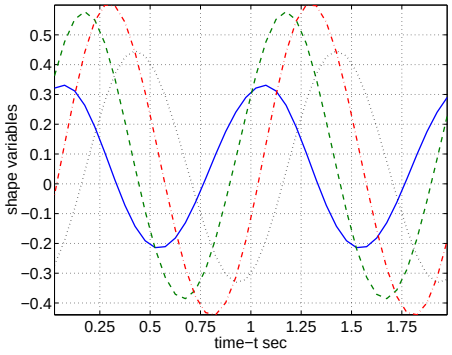
Before moving directly to blind experimentation, it is probably worthwhile to make some observations concerning the purpose of these weights. Going back to the definition of the functional v of equation (6.44), it is seen that these weights serve as penalties on the magnitudes of s, $\dot{s}$ and $\ddot{s}$. Being it the case that $\dot{s}$ and $\ddot{s}$ describe the nature of s, all of these weights penalize the shape s. More specifically, the parameters w_2 and w_3, being the weights on the derivatives of s, serve to control the smoothness of s. Now, it would make sense that s needs to be smooth, being it the case that the optimization method employs discretization. Also, following the logic of [33], this should also be the case for actual snakes due to physiological considerations. Thus, these weights should not be taken too small.

At the same time, allowing the value of these parameters to be quite large, will not permit the shape to exhibit variability and will force it to tend toward the most smooth of objects, a straight line. It has already been observed that the resulting optimal shape sequence values appear to have components that are approximately sinusoidal in form. So we will appeal to this form, $\phi(t) = A\sin(\omega t)$, to illustrate this point further. Using the calculations $\dot{\phi}(t) = A\omega\cos(\omega t)$ and $\ddot{\phi}(t) = -A\omega^2\sin(\omega t)$ it is clear that $\|\phi\|_\infty = A$, $\|\dot{\phi}\|_\infty = A\omega$, and $\|\ddot{\phi}\|_\infty = A\omega^2$. So, the cost v may be thought of as a quantity enveloped by a correct multiple of the form $\bar{v} = w_1A^2 + w_2A^2\omega^2 + w_3A^2\omega^4$. In the case of the method devised herein, the frequency related quantity ω is fixed. Up to this point in the experimentation this value has been 2π. Thus, for a given set of weights, the only way to reduce $\bar{v}$ and consequently v is to decrease the amplitude

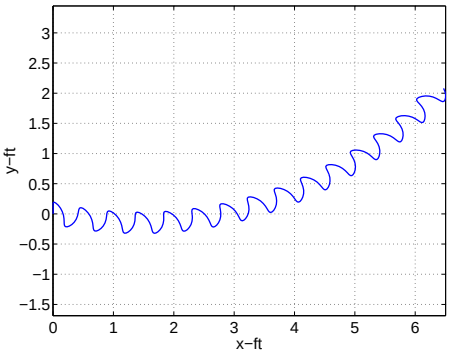
A. If the weights are too high, in particular w_3, then A must be taken very small to keep $\bar{v}$ small and the sinusoidal form of the shape components consequently gets compressed into a horizontal line. It is for this reason in fact that w_3 was selected to be smaller for the initial trials of the method than the other two weights. This tradeoff may be the point of blame for the low amplitude results we have obtained for the optimal shape up to this point and is more than likely at the root of the need to select a large value of the parameter α to obtain reasonable performance.

In the next experiments that we shall perform we seek to effectively reduce the restriction on the amplitude of the high frequency shape by reducing the values of the parameters w_2 and w_3. By looking at $\bar{v}$ it is seen that one need only question the effects of the values of these parameters, for they are the coefficients of the dominant terms of v. We begin by fixing w_1 and w_3 at their default values of 1 and 0.1. respectively, and varying w_2. The first alteration is the reduction of w_2 from 1 to 0.5 and the second is a further reduction to 0.1. The results of performing this experiment are provided in Figure 6.9. The top of this figure, Subfigures 6.9(a) and 6.9(a), correspond to the optimal shape and feedforward head trajectory for $w_2 = 0.5$. The bottom of the figure, Subfigures 6.9(c) and 6.9(d), contain the same information corresponding to $w_2 = 0.1$. It takes little more than a glance at the optimal shapes presented in these figures to determine that for each decrease made in the value of w_2 there is a corresponding increase in shape amplitude. However, the decrease is only very slight and is hardly meaningful. Thus we move on to working with the parameter w_3.

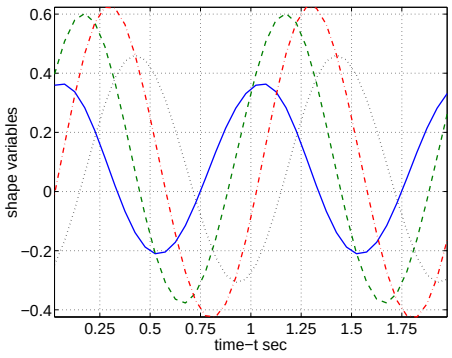
In the case of alteration of w_3, the value of w_2 is returned to 1 held fixed along with w_1. Then we decrease the value of w_3. The first reduction is from the value of 0.1 to 0.05 and the second reduction alters the value to 0.01. We provide the results of making these changes in Figure 6.10. The format of this figure is the same as that provided for the trials of w_2. Just as in the case of w_2, although there was an increase in amplitude correlating to the decrease in the parameter value, this increase is unsubstantial.



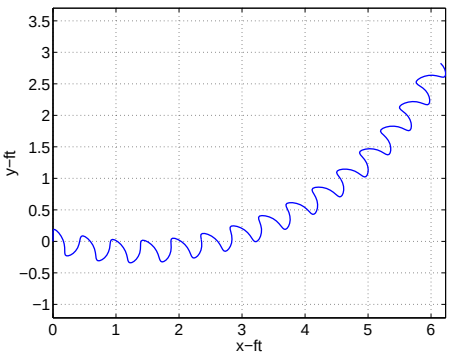
(a) Optimal s, $w_2 = 0.5$.



(b) Motion of B_1.

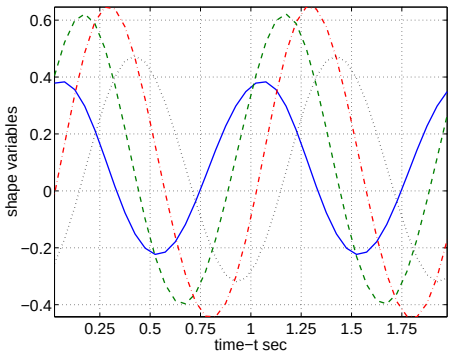


(c) Optimal s, $w_2 = 0.1$.

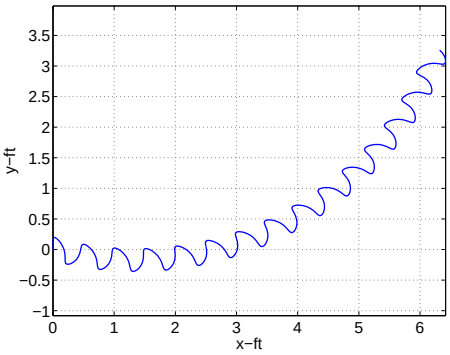


(d) Motion of B_1.

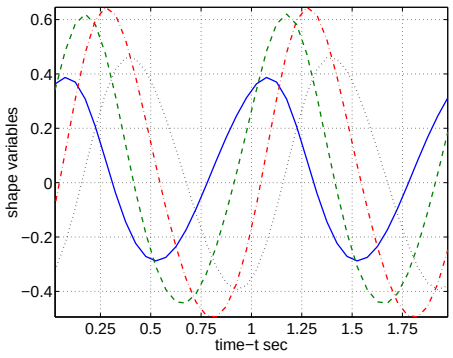
Figure 6.9: Optimality results corresponding to decrements in w_2.



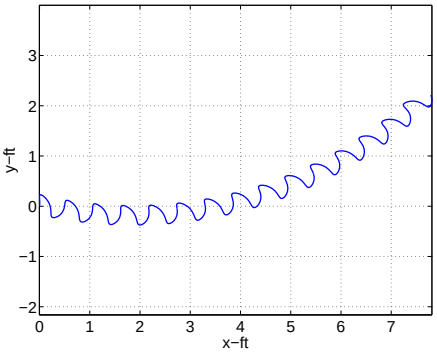
(a) Optimal s, $w_3 = 0.05$.



(b) Motion of B_1.



(c) Optimal s, $w_3 = 0.01$.



(d) Motion of B_1.

Figure 6.10: Optimality results corresponding to decrements in w_3.

These results could mean one of two things. Going back to the expression $\bar{v}$, it is clear that this quantity is much more easily reduced through the amplitude value A than by means of w_2 and w_3, for each term is quadratic in this factor. Thus, for a large frequency such as 1, the modest reduction of w_2 and w_3 may not be capable of compensating for the cumulative effect the smaller amplitude has on the cost v. Of course, one could try more radical reduction of these parameters. However, as we have already stated, this destroys the regulation of the shape and this is undesirable.

On the other hand, it may simply be the case that for a high frequency mode of operation, the structure operates more optimally in a low amplitude state. Of course, if we appeal, as we did at the outset of this experimentation, to the argument that regularization of the shape derivatives makes sense from the point-of-view of muscular physiology, then we should take this conclusion as the acceptable one. Since we can take this view, we do. Furthermore, we shall henceforth make fixed the settings $w_1 = 1$, $w_2 = 1$, and $w_3 = 0.05$.

6.6.5 Experimentation with Shape Frequency

Having concluded at the close of the previous section that reasonable regularization of the shape is appropriate it would now seem that, keeping all other factors fixed, should the minimum of the cost remain approximately the same, a decrease in frequency would result in an increase in amplitude. We come to this conclusion by once again referring to the expression $\bar{v}$. In fact, due to the presence of higher order terms in the frequency related parameter ω in this expression, it would seem that each incremental decrease would demand consecutively larger significant increases in the amplitude. We shall now carry out an experiment to determine whether or not this is the case. Of course, there is no good reason to believe that the optimal cost should remain the same, and we will do nothing to try to enforce such a condition. This was strictly a supposition.

The experiment is quite simple. We increase t_f from 5 to 6 and select p_f values

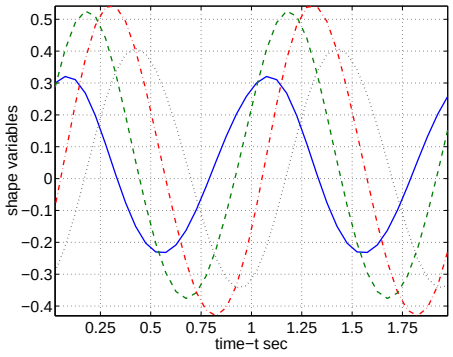
of 1, 1.5, and 2 as trial parameter values. For each of these values we determine the optimal shape. These shapes along with the corresponding feedforward head trajectories obtained are provided in Figure 6.11. The rows of these tiled illustrations correlate respectively to the trial values of p_f as listed.

The results indicate that our expectations concerning the resulting effects of the shape frequency were sound. Each time the frequency is lowered compensation is made by a visibly significant increase in amplitude. It is probably the case that the cost functional value does remain approximately the same because the objective of translation to the terminal coordinates of $(3, 0)$ remains unchanged and a very tight restriction has been placed on this restriction via α.

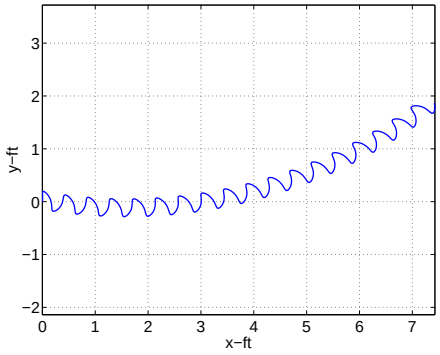
We must comment on the fact that some strange behavior occurred in the case of the optimal shape component ϕ_1 for $p_f = 2$. See Figure 6.11(d). This behavior is more than likely due to the decrease in the regularization of the shape velocity from $w_3 = 0.1$ to $w_3 = 0.05$. This illustrates the negative effects that can occur due to discretization in the presence of non-smoothness. Ultimately, the method that we employed did not fail. As seen in Figure 6.11(f), the structure did move in the eastern direction. However, a stronger value of w_3 would probably have resulted in a more convincing optimal shape.

6.6.6 The Effects of Changing Demands

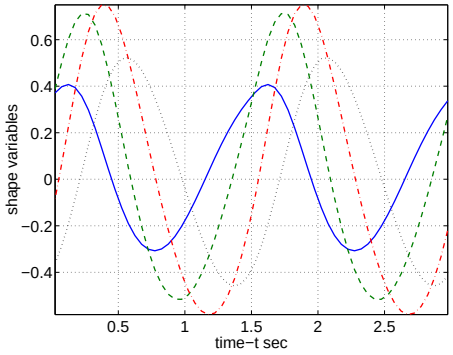
One of the earlier observations that was a bit unsettling was that $\mathcal{S}$ was never capable of completely delivering B_1 to the desired destination coordinates under the demands of the cost. Most likely it was the contribution of v to the optimization cost that created this threshold. This would explain why the value of α had to be taken large to elicit motion in the northern direction. Previously, the cost incurred by the shape must have dominated the total cost. Likewise, for the other directions. For a large value of α, the contribution of g probably dominated the cost until the head was delivered within a ball of a given radius about the destination coordinates. Upon



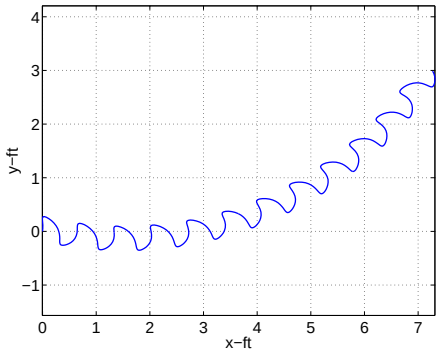
(a) Optimal s, $p_f = 1$.



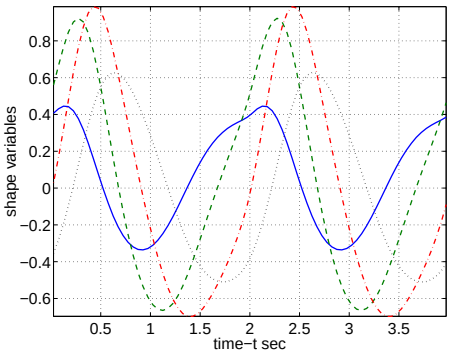
(b) Motion of B_1.



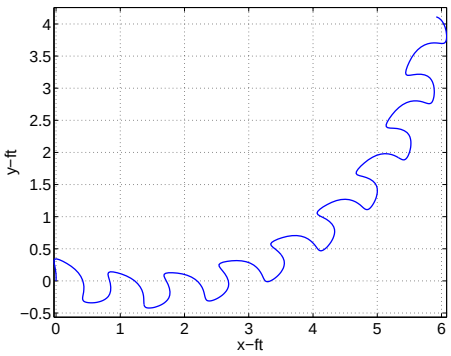
(c) Optimal s, $p_f = 1.5$.



(d) Motion of B_1.



(e) Optimal s, $p_f = 2$.



(f) Motion of B_1.

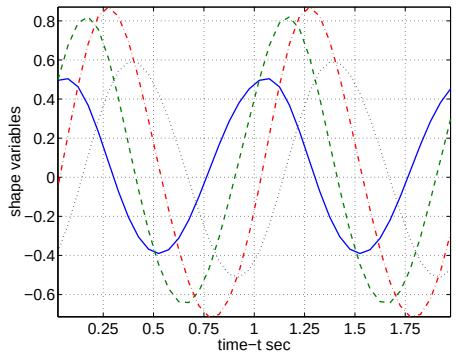
Figure 6.11: Optimality results corresponding to decreases in frequency.

breaking the threshold of the ball the cost of the shape over the specified time interval must have exceeded the cost incurred by g and thus the optimal thing to do was to produce a shape that almost delivered the head to the destination but didn't quite make it.

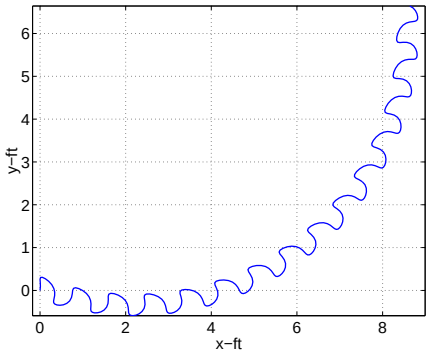
As stated before, we could attempt to circumvent this situation by increasing the value of α beyond its currently fixed value of 1500 to determine if this is in fact the case. In lieu of this we will illustrate the correctness of this notion by asking that the structure deliver its leading link to a point even further along the the eastern direction than we had previously. The optimal thing to do should be to strike a balance between the effort of the shape and the demand that the head make it to its destination. In doing so, it should be the case that the structure will exert the effort to travel further along the eastern direction than it did previously, although still falling short of the desired target.

The destination coordinates of $(5, 0)^t$ are used in place of $(3, 0)^t$ for the first trial, and the coordinates $(7, 0)^t$ are used for the second. The resulting optimal shape and feedforward head trajectories are provided in Figure 6.12. The top row of the tiled illustrations correspond to the desired displacement of length 5 ft and the bottom row corresponds to the desired displacement of 7 ft.

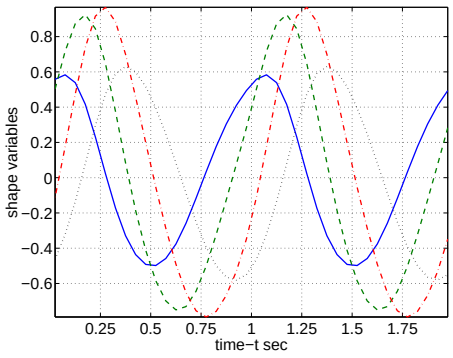
It would appear that things are precisely as we had believed. When we asked the structure to move 5 ft instead of 3 ft over the period of 5 s it responded by increasing its shape amplitude, moving approximately 4 ft and reaching the desired destination upon continuation 1 s thereafter. When asked to move still further, the structure again increased its shape amplitude just slightly, allowing it to displace its head at the rate of approximately 1 ft/sec . However, it is noted that we quickly push the compliance of the structure to our demands toward its limit by means of these changes. In the case of the $7ft$ request, the structure did not near its target in the continuation. This suggests that the structure was truly being asked to perform outside its capabilities under the given penalties. It is probable that the structure



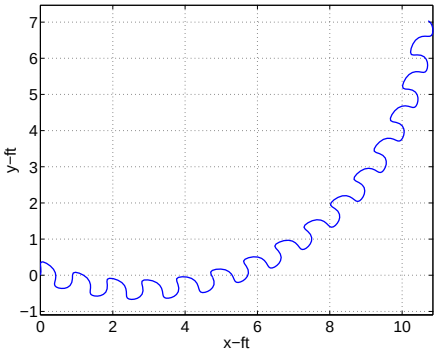
(a) Optimal s, 5 ft request.



(b) Motion of B_1.



(c) Optimal s, 7 ft request.



(d) Motion of B_1.

Figure 6.12: Optimality results corresponding to increased displacement demands.

needs more leverage in some form to perform what is asked of it. This will bring us to our final two experiments. But before moving on, note that henceforth we will ask that the structure displace its head from the origin to the destination coordinates of $(5, 0)^t$, as this request resulted in satisfactory shape performance during the present experimentation and we seek to push the shape to adapt as we progress.

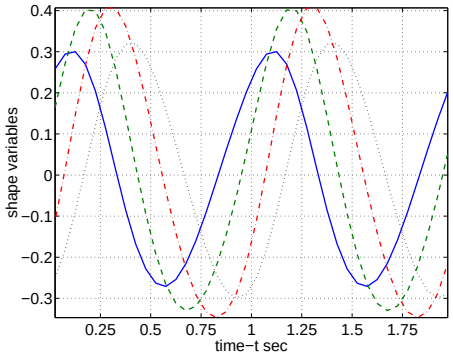
6.6.7 The Effects of the Friction Pool on Shape and Structure Capability

Since the outset of the research contained herein, it has been clear that the ability of snakes or snake-like structures to achieve locomotion lies in the coupling of the geometry of the structure and the nature of the friction pool created by the contact between the structure and its environment. Having worked thoroughly through the basic principles and ideas presented by Gray and having found them to be sound, it has become clear that, of the interactions of the snake with its environment, it is the interactions that involve the lateral portion of the animal that are responsible for the geometry or shape of the snake that we observe while they are performing the lateral undulation mode of motion.

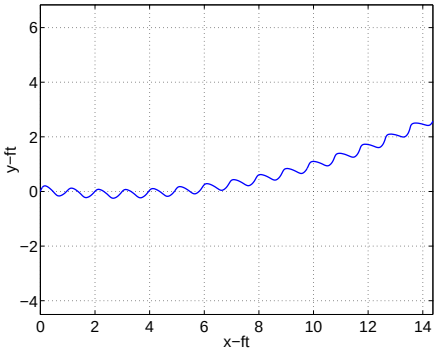
Since coming to this realization we have been eager to search out a satisfactory answer to the following secondary question. How, or should, the generic sigmoid locomotory shape of the snake be altered as the nature of these lateral reaction forces acting normally to the body change? Recall from the discussions of Chapter 4 that as the magnitudes of the coefficients of the viscous friction model change, so does the direction of the resultant friction force. More specifically, it has been demonstrated that as the ratio $\frac{c_t}{c_n}$ approaches 1 from below, the locomotive capability of a fixed instance of the serpinoid shape dissolves rapidly. It is wondered if it is possible for the shape to, in part, make up for this loss by alteration of its geometry in accordance with directional changes of the friction.

The experiment used to examine the query is quite simple. We fix the current

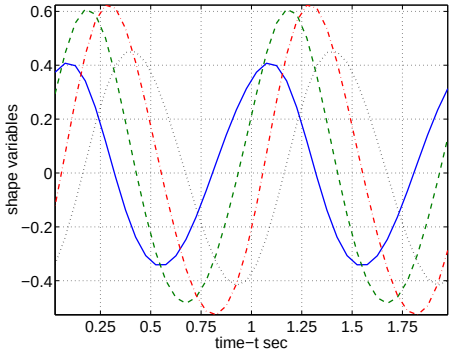
value of the viscous friction parameter c_t at 0.5 and then vary its counterpart c_n over the set of values $\{10, 5, 2.5, 1\}$, observing the outcome of the optimal shape and feedforward head trajectories. The results of these experiments are shown in the rows of Figures 6.13 and 6.14, respectively. Looking at these plots we see that the true



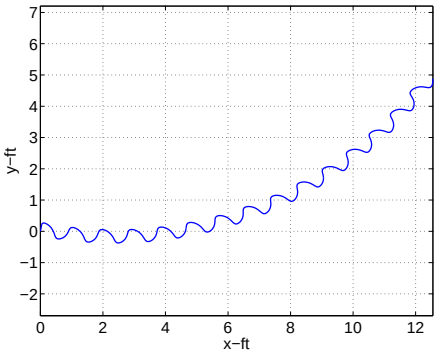
(a) Optimal s, $c_n : c_t = 10 : 0.5$.



(b) Motion of B_1.



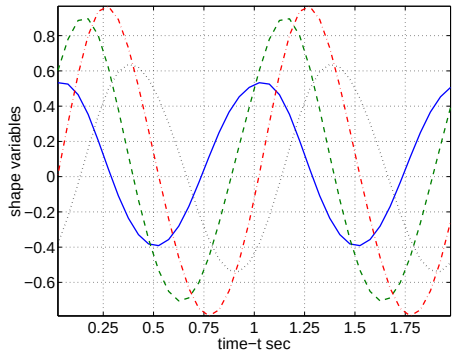
(c) Optimal s, $c_n : c_t = 5 : 0.5$.



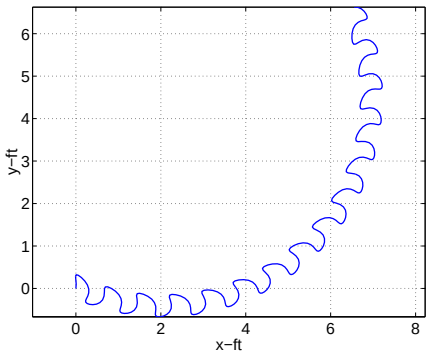
(d) Motion of B_1.

Figure 6.13: The effects of lateral forces on locomotion. Results indicating the pure necessity of a high normal reaction to lateral friction force ratio in the attainment of directed locomotion.

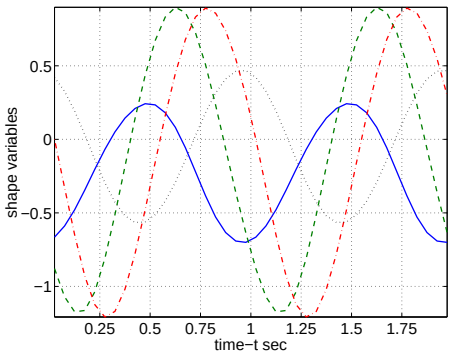
character of the optimal shape remained the same, as has been the case throughout our experiments. It is the case that, as friction ratio decreases, the structure, finding it more difficult to accomplish its designated task, makes small adaptations in the



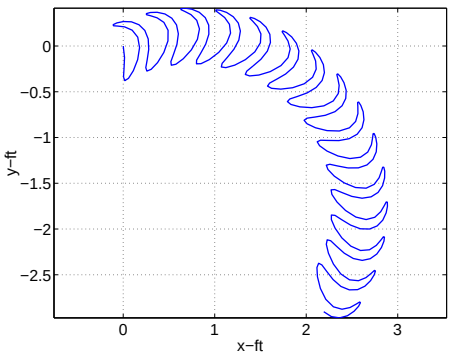
(a) Optimal s, $c_n : c_t = 2.5 : 0.5$.



(b) Motion of B_1.



(c) Optimal s, $c_n : c_t = 1 : 0.5$.



(d) Motion of B_1.

Figure 6.14: The effects of lateral forces on optimal locomotion. Results indicating the pure necessity of a high normal reaction to lateral friction force ratio in the attainment of directed locomotion.

way of scaling and bias. But generically speaking, the optimal form of the structure is quite fixed or mathematically speaking, appears global. It would be a pointless endeavor to speak of these now redundant alterations.

There are only a couple of notes about these results that we believe merit mention. Firstly, for a high friction coefficient ratio, the structure actually accomplished its prescribed task of displacing its head to the destination of $(0, 5)^t$ in the allotted time limit of 5 s for the first time, illustrating both speed and accuracy. Secondly, at the other end of the spectrum the performance of the optimal shape was a complete failure. The structure simply flailed around like a fish out of water. This is precisely what it was. A fish knows without thought that the only way to propagate itself is through lateral undulation. However, in the absence of the pressure differential created by a pool of fluid, the motion is futile. Dry friction anisotropy cannot provide aid. A snake cannot attain steady bulk motion in a specified direction without significant normal reaction forces and we have, for the first time to our knowledge, mathematically demonstrated (at least for this instance of the system) that this is the case without assuming anything about the geometry of the serial-link structure.

6.6.8 A Look at the Role of the Size of the Serial-Link Structure

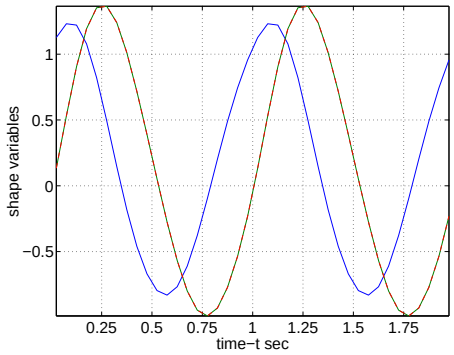
For our last experiment we investigate the role of the size n of $\mathcal{S}$ with regards to optimal locomotive shapes. Having noticed in Chapter 5 that the size can have a dramatic effect on the amount of momentum the structure can preserve, we would expect that by adding links to the structure it may more easily be able to accomplish the displacement tasks that we devise. Additionally, having only utilized the 5-link structure up to this point, we do not know whether or not the generic character of the optimal shape that has come to be redundant by now transcends this structure size. Of course, the number of configurations that the structure could possibly take on increases exponentially, base 2, with the number of links minus 1. Hence, perhaps

it is possible that the optimal tendencies could change with an increase in the number of links.

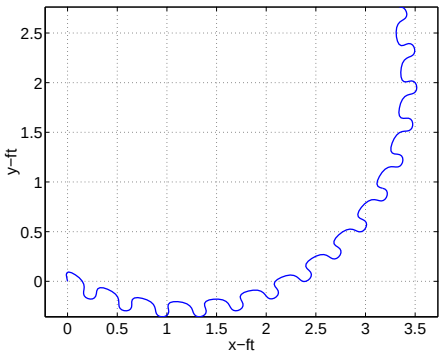
Three different structure sizes, $n = 3, 6, 10$, are considered. The interest in the 3-link structure comes out of its use in Gray's analysis. The 6-link structure provides an example near to that of the 5-link structure that we have examined heretofore but one that has an even number of links. It seems interesting to see what happens with the "extra" shape variable. The 10 link structure provides a sample of what happens on the higher end of things and a second sample of an odd number of shape variables.

Before presenting the results, it should be noted that for the 10-link example the discretization parameters are altered from those of Table 6.2. In order to speed up the numerics we select a more coarse mesh, selecting $\Delta t = 0.1$.

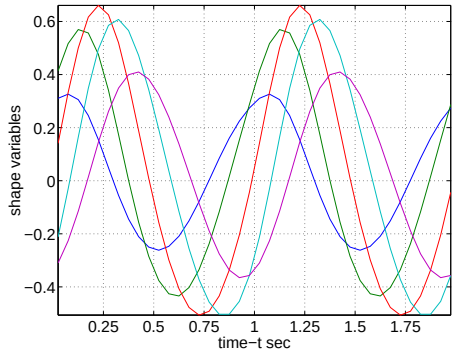
The shape and the head trajectory for each trial value of n are presented respectively in the rows of Figure 6.15. Just as in the case of the 5-link structure, it was the optimal tendency of the structure to develop a phase shift between the shape variables, and to accentuate the amplitude of the shape variables toward the medial portion of the structure. Again, the increased potential of $\mathcal{S}$ to preserve momentum with higher link counts is observed. Note that in the case of the 10-link structure the destination target of $(5, 0)^t$ was reached in the allotted 5 *sec* time period. This is only the second occurrence of complete success in the accomplishment of the objective amongst all the experimental trials. Recall that in the case of the previous success the friction ratio $\frac{c_t}{c_n}$ was near ideal being $\frac{0.5}{10}$. However, here we used the more realistic ratio of $\frac{0.5}{3}$. This indicates for the second time now that effective structures need to be long and more than likely this is the answer to adaptation to the friction environment. The vertebra count of actual snakes, as discussed in Chapter 2 would seem to support this observation.



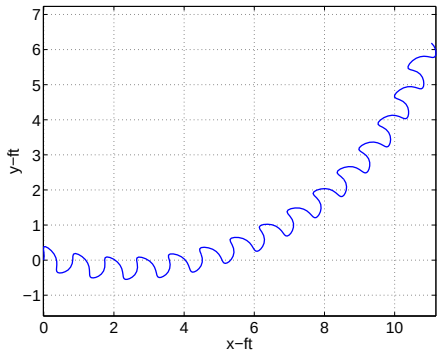
(a) Optimal s, $n = 3$.



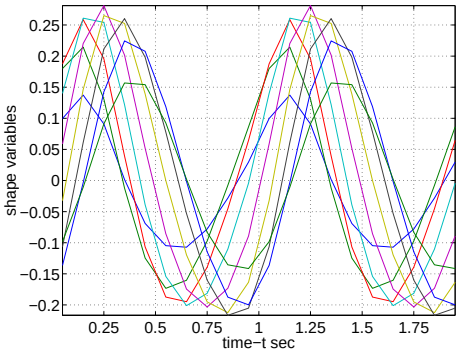
(b) Motion of B_1.



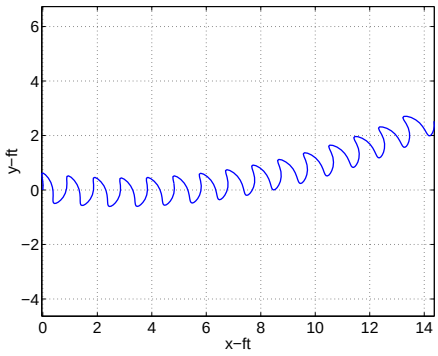
(c) Optimal s, $n = 6$.



(d) Motion of B_1.



(e) Optimal s, $n = 10$.



(f) Motion of B_1.

Figure 6.15: The effects of length on optimal locomotion. Results indicating that elongation aids in efficiency.

Chapter 7

Conclusions and Future Directions

7.1 Summary and Conclusions

Let us recapitulate this effort and put the results into perspective. In Chapter 2 we described the serial-link structure and explained its purpose in roughly modeling the morphology of snakes as locomotive devices. In Chapter 3 we examined how Lagrangian mechanics could be used to formulate the dynamics of the serial link structure and how these dynamics could be put into a reduced form using the inherent symmetries of the system and the forces acting on it. In Chapter 4 we finished our modeling discussions by providing details concerning how to model friction and actuator forces. In the discussions on friction we illustrated how one could incorporate the idea of snake skin with anisotropic friction characteristics into the dynamics and showed that such forces were indeed symmetric and thus could be used in the context of the reduction described in the previous chapter. Additionally, our control discussions indicated the fact that the shape of the serial-link structure was completely controllable and we discussed a method for tracking arbitrary shapes. By doing so we were able to explain how the shape equation portion of the reduced dynamics could be effectively ignored and how the shape itself could be regarded as the control for the system, focusing our efforts on the appropriate selection of shape for the acquisition

of gait. That is essence of the first half of this work.

In the second half of our efforts we approached the problem of determining snake shapes for the purpose of the production of undulatory locomotion from two fronts. The first of these, presented in Chapter 5, involved rediscovering the work of J. Gray. Essentially, we reinvented his study on undulatory locomotion and discussed how he was able to deduce a shape that would produce locomotion by application of a few very simple and exceedingly general criterion that embody the mechanism of the undulatory form of locomotion. We further detailed how the smoothing of this example form gives rise to the so-called serpenoid shape that is used by nearly everyone who has successfully completed a study on the subject of the autonomous control of snake-like machines. However, upon doing so, we were careful to point out, as did Gray, that this form is only an example of how snakes make use of the more general criterion underlying lateral undulation. This point was illustrated via an eel-like example.

The second approach to formulating gait evoking shapes involved setting up an optimal control problem utilizing periodic shape forms as control inputs and was developed in the first half of Chapter 6. There were many reasons for taking this approach. One such reason was to determine if the serpenoid shape was in some sense optimal. Hirose had presented data in his work [33] indicating that this shape was quite in line with the shape actually used by snakes while in motion over a uniform substratum. Based on this data, Hirose concluded that the serpenoid curve was "the" natural choice of gliding form used by snakes for locomotion. Many researchers have adopted this perspective. However, in light of Gray's criterion, it seemed doubtful to us that this form was the most natural. For only phase modulations are employed by this shape to create the necessary curvature differential for motion. It appeared more natural to us that the snake should also take advantage of amplitude modulation in its acquisition of locomotion, further accentuating curvature differential and increasing its capability to preserve a steady forward momentum. It was believed

that the need for such adjustment would be heightened as the friction pool provided by lateral resistance was weakened. By developing a methodology for determining optimal shapes for locomotion, we were able to let mathematics decide what form provides the most suitable balance between effort exertion and propulsive capacity.

What we found during our optimal shape investigations, which were presented in the latter half of Chapter 6, was a balance between our belief that amplitude modulation was natural for taking the greatest advantage of Gray's criterion and the need to decrease this amplitude for the purpose of minimizing the energy exerted by actuation. As suspected, a constant shape amplitude is not consistent with the objective of preserving as much steady-state momentum during motion as possible. Thus, the serpenoid gliding form is not the most natural in the sense of satisfying standard optimality criterion. It is more productive to take advantage of both amplitude and phase-shift modulations along the length of the structure. This falls in line with Gray's criterion and further validates our conclusion that it is more fitting to give attention to the analysis that led Gray to the introduction of the composite clothoid gliding form rather than focusing on this form itself as "the" locomotive form of snakes.

Turning focus away from the differences between optimal shapes and the serpenoid shape, it would seem that the use of a bias in conjunction with low amplitude sinusoidal shape inputs is quite consistent with optimality results. In this regard, our results agree with the conclusions of Hirose. Additionally, it appears to us that the feed-back mechanisms currently employed in [46] and [62] for the direction control of the snake-like serial-link structure are on the right track.

7.2 Future Directions

With regards to future directions it must be said that the topic of snake locomotion and in particular, the use of lateral undulations for obtaining gait, is still wide open

in many respects. The scientific investigations that relate to and have a significant bearing on this topic, which need to and should be performed, are too many to mention. In addition to the shear number of issues, the spectrum of disciplines that need to be involved in their deliberation is immense. Thus we will focus our attention here on those matters that may be taken up directly by control engineers and mathematicians and which directly relate to the conclusions of this study.

Immediately, studies into the liberation of the undulatory locomotive mechanism should be taken up using the serial-link structure model. Shape patterns of the generic form

$$\phi_j(t) = a(t,j)\sin(\omega t + d(t,j)) + b(t,j) \ .$$

should be considered as a means of localizing undulations along the centerline of the structure. Simple experimental investigations through simulation could indicate quite a lot. For instance, we illustrated at the close of Chapter 5 that $a(t,j)$ could be selected in such a way so as to stabilize the head of the serial-link structure. Through this function alone one could experiment with the idea of stabilizing the tail and medial portions of the structure while still maintaining steady directed motion with the other portions of the structure. Though the latter will likely call for use of longer structures. If one could make these possibilities functional, then the time component of the amplitude function could be applied to pass the localization of the undulations to any point along the length of the device while still maintaining directed motion. This would certainly begin to capture the true capabilities of the snake's hyper-redundancy, as indicated by the generalizations of Gray quoted at the closing of Chapter 5.

Further, $b(t,j)$ could be used to not only steer and direct the serial-link structure as a bulk object, but also to do so locally. This feature along with localization of the propulsive mechanism could allow the a serial-link structure of sufficient length to not only pass through narrow openings but also the narrow winding passages that snakes appear to navigate effortlessly.

It is unclear at this point whether or not there is some advantage in a continuum representation of the snake. Although, with the number of links that snakes actually possess and the relatively small length thereof, these animals can appear to be somewhat like inextensible cables. The author had at one point conceived the notion that the snake should be modelled in precisely that way, as an inextensible cable. Upon doing so, one could assume that through actuation about the "infinite" number of vertebrae, the curvature of the animal could be controlled, just as the relative angles of the discrete serial link structure are controlled. Hence, the shape would be a solitary signed curvature functional $K(t, s)$ parameterized by time and the arc-length of the snake's winding curve. A form for this functional could be taken. Namely, we could easily adapt the sinusoidal form used for the differential angles of serial link structure so as to apply to the continuum by simply replacing the joint parameter j with the arc-length parameter s to arrive at the functional

$$K(t, s) = a(t, s) \sin(\omega t + d(t, s)) + b(t, s) \ .$$

It turns out that this was precisely the technique applied by Burdick and Chirikjian to treat the motion control of kinematic hyper-redundant manipulators and locomotion devices [13, 12, 14, 10]. Additionally, as indicated by F. Matsuno and S. Hara [45], there could be some control formulation advantages inherent in the use of a distributed parameter system model. Whether this is the case or not, it seems like it would be good measure to consider the potentials of the continuum approach to lateral undulation and snake locomotion in general.

Each of the queries just described could and should, for the sake of simplicity, be taken up in a uniform planar environment with an assumption of a viscous type interaction law that will allow independent choice of lateral reaction and friction forces. However, after appropriately exploring these possibilities, the uniform plane assumption should be dropped along with the friction assumption. As expressed before, pragmatists care about the terrain adaptability and stability properties of

the snake. Thus, what a serial-link structure can do under ideal conditions is in some sense beside the point. As pointed our by Gans [21], the lateral undulation mechanism hardly requires a continuum of lateral reaction forces. For this reason, this same mechanism can be used by the animal to climb trees and produce directed motion in the presence of nothing more than a series of rigid pegs, as demonstrated now several times by those who perform experiments with actual snakes. Hence, some effort needs to be exerted to capture this adaptability to and use of the environment. As the matter currently stands, there is not a single study in which the snake-like mechanism uses its environment for the generation of the lateral reaction forces needed for motion. In every case, including our own, it is either assumed that the capacity to elicit significant normal reaction forces is some how built into the structure, such as no side-slip wheels and sled mechanisms, or that the structure is emersed in a fluid. Ultimately, this mold must be broken. Only then will we begin to develop the understanding needed to capture those features of the snake that have such great potential in engineering technologies.

We believe that these tasks we have discussed can and will be done given sufficient time and effort. We look forward to contributing our own efforts in this regard and to sharing in the understanding of those works that are produced by the many talented scientists who have also taken up the study of the analysis and control of snake-like locomotors.

References

[1] R.D. Arnell, P.B. Davies, J. Halling, and T.L.Whomes. *Tribology: Principles and Design Applications*. Springer-Verlag New York Inc., 1991.

[2] Michael Artin. *Algebra*. Printice Hall, Upper Saddle River, New Jersey, 1991.

[3] Uri M. Ascher and Linda R. Petzold. *Computer Methods for Ordanary Differentail Equations and Differential-Algebraic Equations*. Society for Industrial and Applied Mathematics, 1998.

[4] Nakhle H. Asmar. *Partial Differential Equations and Boundary Value Problems*. Prentice Hall, Upper Saddle River, New Jersey, 2000.

[5] M. Asghar Bhatti. *Practical Optimization Methods: With Mathematica Applications*. Springer-Verlag New York, Inc.; TELOS, 2000.

[6] Bharat Bhushan. *Principles and Applications of Tribology*. John Wiley and Sons, Inc., 1999.

[7] Frank Philip Bowden and David Tabor. *Friction: An Introduction to Tribology*. The Science Study Series. Robert E. Krieger Publishing Company, Inc., Malibar, Florida, 1982.

[8] U. Brechtken-Manderscheid. *Introduction to the Calculus of Variations*. Chapman & Hall, 1991.

[9] Richard L. Burden and J. Douglas Faires. *Numerical Analysis*. Brooks/Cole Publishing Company, ITP, sixth edition, 1997.

[10] J.W. Burdick, J.Radford, and G.S. Chirikjian. A "sidwinding" locomtion gait for hyper-redundant robots. In *Proceedings of the 1993 IEEE International Conference on Robotics and Automation*, pages 101–106, Atlanta, Georgia, May 1993.

[11] F.L. Chernous'ko. The motion of a multilink system along a horizontal plane. *Journal of Applied Mathematical Mechanics*, 64(1):5–15, 2000.

[12] Gregory S. Chirikjian and Joel W. Burdick. Hyper-redundant robot mechanisms and their applications. In *Proceedings of the 1991 IEEE/RSJ International Workshop on Intelligent Robotics and Systems*, pages 185–190, Osaka, Japan, November 1991.

[13] Gregory S. Chirikjian and Joel W. Burdick. Kinematics of hyper-redundant robot locomotion with applications to grasping. In *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*, pages 720–725, Sacramento, California, April 1991.

[14] G.S. Chirikjian and J.W. Burdick. A geometric approach to hyper-redundant manipulator obsticle avoidance. *Journal of Mechanical Design*, 114:580–585, December 1992.

[15] H. Date, Y. Hoshi, and M. Sampei. Locomotion control of a snake-like robot based on dynamic manipulability. In *Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robotics and Systems*, pages 2236–2241, 2000.

[16] H. Date, Y. Hoshi, M. Sampei, and Shigeki Nakaura. Locomotion control of a snake-like robot with constraint force attenuation. In *Proceedings of the American Control Conference*, pages 113–118, 2001.

[17] Kevin Dowling. Limbless locomotion: Learning to crawl. In *Proceedings of the 1999 IEEE International Conference on Robotics and Automation*, pages 3001–3006, Detroit, Michagan, May 1999.

[18] Gen Endo, Keigi Togawa, and Shigio Hirose. Study on self-contained and terrain adaptive active cord mechanism. In *Proceedings of the 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1399–1405, 1999.

[19] Charles Fox. *An Introduction to the Calculus of Variations*. Oxford University Press, London: Geoffrey Cumberlege, 1950.

[20] Edwardo F. Fukushima, Shigio Hirose, and Takeo Hayashi. Basic manipulation considerations for the articulated body mobile robot. In *Proceedings of the 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 386–393, Victoria, B.C., Canada, October 1998.

[21] Carl Gans. Terrestial locomotion without limbs. *American Zoologist*, 2:167–182, 1962.

[22] Carl Gans. *Biomechanics: An Approach to Vertebrate Biology*. The University of Michigan Press, Ann Arbor, 1974.

[23] Carl Gans and D. Baic. Regional specification of reptilian scale surfaces: Relation of texture and biologic role. *Science*, 195:1348–1350, 1977.

[24] J.-P. Gasc, D. Cattaert, C. Chasserat, and F. Clarac. Propulsive action of a snake pushing against a single site: Its combined analysis. *Journal of Morphology*, 201:315–329, 1989.

[25] Herbert Goldstein. *Classical Mechanics*. Addison-Wesley Series in Advanced Physics. Addison-Wesley Publishing Company, Inc., 1950.

[26] Ian A. Gravagne and Ronald L. Woodfin. Mine-sniffing robotic snakes and eels: Fantasy or reality. Preprint, 8 pages.

[27] J. Gray. Studies in animal locomotion: I. the movement of fish with special reference to the eel. *Journal of Experimental Biology*, 10(1):88–104, 1932.

[28] J. Gray. The mechanism of locomotion in snakes. *Journal of Experimental Biology*, 23(2):101–120, 1946.

[29] J. Gray and H.W. Lissmann. The kinetics of locomotion of the grass-snake. *Journal of Experimental Biology*, 45:354–367, 1949.

[30] James Gray. *Animal Locomotion*. The World Naturalist. W.W. Norton & Company, Inc., New York, 1968.

[31] Donald T. Greenwood. *Classical Dynamics*. Dover Publications, Inc., 1997.

[32] J. Hazel, M. Stone, M.S. Grace, and V.V. Tsukruk. Nanoscale design of snake skin for reptation locomotions via friction anisotrophy. *Journal of Biomechanics*, 32:477–484, 1999.

[33] Shigeo Hirose. *Biologically Inspired Robots: Snake-Like Locomotors and Manipulators*. Oxford University Press, 1993.

[34] Shigio Hirose and Akio Morishima. Design and control of a mobile robot with an articulated body. *The International Journal of Robotics Research*, 9(2):99–114, 1990.

[35] Kazufumi Ito and Istvan Lauko. Numerical methods for optimal feedback synthisis. Preprint, 17 pages.

[36] Bruce C. Jayne. Mechanical behavior of snake skin. *Journal of Zoology*, 214:125–140, 1988.

[37] Bruce C. Jayne. Muscular mechanisms of snake locomotion: An electromyographic study of lateral undulation of the Florida banded water snake (nerodia fasciata) and the yellow rat snake (elaphe obsoleta). *Journal of Morphology*, 197:159–181, 1988.

[38] C.T. Kelly. *Iterative Methods for Linear and Nonlinear Equations*, volume 16 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics, 1995.

[39] Scott D. Kelly and Richard M. Murray. Geometric phases and robotic locomotion. Technical Report 94-014, California Institute of Technology, Pasadena, California, September 1994. Submitted to J. Robotic Systems, Sep. 94.

[40] Bernhard Klaassen and Karl L. Paap. Gmd-snake2: A snake-like robot driven by wheels and a method for motion control. In *Proceedings of the 1999 IEEE International Conference on Robotics and Automation*, pages 3014–3019, Detroit, Michagan, May 1999.

[41] P.S. Krishnaprasad and D.P. Tsakiris. G-snakes: Nonholonomic kinematic chains on lie groups. Technical Report 94-27, University of Maryland, College Park, Maryland, 1994. Submitted to: The 33rd IEEE Conference on Decision and Control, Lake Buena Vista, Florida, December 14-16, 1994.

[42] Frank L. Lewis and Vassilis L. Syrmos. *Optimal Control*. John Wiley and Sons, Inc., second edition, 1995.

[43] Ralf Linnemann, Karl L. Paap, Bernhard Klaassen, and Jürgen Vollmer. Motion control of a snakelike robot. In *IEEE Proceedings of the Third European Workshop on Advanced Mobile Robots*, pages 1–8, Zurich, Switzerland, September 1999.

[44] Jerrold E. Marsden and Tudor S. Ratiu. *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, volume 17 of *Texts in Applied Mathematics*. Springer-Verlag New York Inc., 1994.

[45] Fumitoshi Matsuno and Shinji Hara. Bottom-up and top-down approaches to dynamics of hyper-redundant mechanical systems. In *Proceedings of the 1999 IEEE International Conference on Systems, Man, and Cybernetics*, pages 172–177, Tokyo, Japan, October 1999.

[46] K.A. McIsaac, A.K. Das, J.M. Esposito, and J.P. Ostrowski. A hierarchical, modal approach to hybrid systems control of autonomous robots. In *Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robotics and Systems*, pages 1020–1025 vol. 2, Takamatsu, Japan, October 2000.

[47] K.A. McIsaac and J.P. Ostrowski. A geometric approach to gait generation for eel-like locomotion. In *Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robotics and Systems*, pages 2230–2235 vol. 3, Takamatsu, Japan, October 2000.

[48] K.A. McIsaac and J.P. Ostrowski. Motion planning for dynamic eel-like robots. In *Proceedings of the 2000 IEEE International Conference on Robotics and Automation*, pages 1695–1700 vol. 2, San Francisco, California, April 2000.

[49] G. Migadis and K.J. Kyriakopoulos. Design and forward kinematic analysis of a robotic snake. In *Proceedings of the 1997 IEEE International Conference on Robotics and Automation*, pages 3493–3498, Albuquerque, New Mexico, April 1997.

[50] Brad R. Mood and Carl Gans. Kinematics, muscular activity and propulsion in gopher snakes. *The Journal of Experimental Biology*, 201:2669–2684, 1998.

[51] Desmond F. Moore. *Principles and Applications of Tribology*, volume 14 of *International Series in Materials Science and Technology*. Pergamon Press, 1975.

[52] Walter Mosauer. On the locomotion of snakes. *Science*, 76:583–585, 1932.

[53] Hideo Nakamura, Takefumi Shimada, and Hisato Kobayashi. An inspection robot for feeder cables: Snake like motion control. In *Proceedings of the 1992 IEEE International Conference on Industrial Electronics, Control, Instramentation, and Automation*, pages 849–852, San Diego, California, November 1992.

[54] Henk Nijmeijer and Arjan van der Schaft. *Nonlinear Dynamical Control Systems*. Springer-Verlag New York, Inc., 1990.

[55] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer-Verlag New York, Inc., second edition, 1999.

[56] James P. Ostrowski, Jaydev P. Desai, and Vijay Kumar. Optimal gait selection for nonholonomic locomotion. *The International Journal of Robotics Research*, 19(3):225–237, 2000.

[57] James Patrick Ostrowski. *The Mechanics and Control of Undulatory Robotic Locomotion*. PhD thesis, California Institute of Technology, 1996.

[58] Jim Ostrowski. Computing reduced equations for robotic systems with constraints and symmetries. *1999 IEEE Transactions on Robotics and Automation*, 15(1):111–123, 1999.

[59] J.P. Ostrowski and J.W. Burdick. The geometric mechanics of undulatory robotic locomotion. *The International Journal of Robotics Research*, 17(7):687–701, 1998.

[60] Karl L. Paap, Frank Kirchner, and Bernhard Klaassen. Motion control scheme for a snake-like robot. In *Proceedings of the 1999 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, pages 59–63, Monterey, California, November 1999.

[61] Hans Sagan. *Introduction to the Calculus of Variations*. Dover Publications, Inc., New York, 1992.

[62] M. Saito, M. Fukaya, and T. Iwasaki. Serpentine locomotion with robotic snakes. *IEEE Control Systems Magazine*, 22(1):64–81, 2002.

[63] K. Sarrigeorgidis and K.J. Kyriakopoulos. Motion control of the n.t.u.a. robotic snake on a planar surface. In *Proceedings of the 1998 IEEE International Conference on Robotics and Automation*, pages 2977–2982, Leuven, Belgium, May 1998.

[64] Ahmed A. Shabana. *Computational Dynamics*. John Wiley and Sons, Inc., 1994.

[65] Ahmed A. Shabana. *Dynamics of Multibody Systems*. Cambridge University Press, second edition, 1998.

[66] Yansong Shan and Yoram Koren. Design and motion planning of a mechanical snake. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(4):1091–1100, 1993.

[67] Stephanie Frank Singer. *Symmetry in Mechanics: A Gentle, Modern Introduction*. Burkhauser, 2001.

[68] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*, volume 12 of *Texts in Applied Mathematics*. Springer-Verlag New York Inc., second edition, 1993.

[69] J.A. Williams. *Engineering Tribology*. Oxford University Press Inc., New York, 1994.

[70] Rainer Worst and Ralf Linnemann. Construction and operation of a snake-like robot. In *Proceedings of the 1996 IEEE International Joint Symposia on Intelligence and Systems*, pages 164–169, Rockville, Maryland, November 1996.