CLASSICAL MECHANICS WITH DISSIPATIVE CONSTRAINTS

by

Shaun Russell Harker

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Mathematics

MONTANA STATE UNIVERSITY
Bozeman, Montana

June 2009

APPROVAL

of a dissertation submitted by

Shaun Russell Harker

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Dr. Tomáš Gedeon

Approved for the Department of Mathematical Sciences

Dr. Kenneth L. Bowers

Approved for the Division of Graduate Education

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U. S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Shaun Russell Harker

June 2009

ACKNOWLEDGEMENTS

There are many people whose kindness and example has inspired me, and without which I would be the worse for. I would like to thank: my father, who I could not respect more; my advisor Tomáš Gedeon for his enthusiasm, encouragement, and support over the years; Curt Vogel, for seeing past my questionable undergraduate quirks, offering me my first research job, and helping me land my first publication; Jarek Kwapisz for arranging my first talk (even though it had nothing to do with dynamics) and providing feedback on more than a few stacks of unassigned exercises from functional analysis; Marcy Barge, for proving how charming a completely inscrutable demeanor can really be; John Miller, for arranging a major research experience for me from which I have profited greatly; and Lukas Geyer, for digging both Schramm-Loewner Evolution and also *Supertramp*. And with a sense of proportion that could never hope to fit on this page, I would also like to thank the rest of my family and friends, who I love very much.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

TABLE OF CONTENTS CONTINUED

LIST OF FIGURES

ABSTRACT

The aim of this thesis is to consider the mathematical treatment of mechanical systems in the presence of constraints which are *energetically dissipative.* Constraints may be energetically dissipative due to impacts and friction. In the frictionless setting, we generalize *Hamilton's principle of stationary action*, central to the Lagrangian formulation of classical mechanics, to reflect optimality conditions in constrained spaces. We show that this generalization leads to the standard measure-theoretic equations for shocks in the presence of unilateral constraints. Previously, these equations were simply postulated; we derive them from a fundamental variational principle. We also present results in the frictional setting. We survey the extensive literature on the subject, which focusses on existence results and numerical schemes known as *time-stepping algorithms.* We consider a novel model of friction (which is more dissipative than standard Coulomb friction) for which we can give better well-posedness results than what is currently available for the Coulomb theory. To this end, we study multi-valued maps, differential inclusions, and optimization theory. We construct a differential inclusion we call the *feedback problem*, for which the multi-valued map is the solution set of a convex program. We give existence and uniqueness results regarding this feedback problem. We cast the persistent contact evolution problem of our novel model of friction into the form of a feedback problem to derive an existence result.

CHAPTER 1

INTRODUCTION

Introduction

Mechanical systems have been long studied and, for the most part, are well understood. One describes the physics with differential equations, which are then solved analytically or numerically. Typically, one of three paradigmatic choices is made for the model: either *Lagrangian*, *Hamiltonian*, or *Newtonian* mechanics. In each of these cases, one finds he has a system of ordinary differential equations to solve.

In this thesis, we consider the complications added by introducing unilateral or bilateral constraints to the system, with or without friction. A unilateral constraint on a mechanical system is one of the form

$$f(t, q) \geq 0,$$

where f is some scalar function of the position. A bilateral constraint is of the form

$$g(t, q) = 0.$$

Unilaterally constrained mechanical systems are useful in a number of applications. For example, in mathematical models of legged locomotion one needs to model the surface upon which the locomotion is performed. Another important example is

the area of virtual reality and computer graphics, where non-interpenetration constraints (which are unilateral) must be modeled. In both of these examples friction also plays an important role.

The theory of frictionless bilateral constraints poses no serious difficulty. The formulation of Lagrange was particularly well suited for it; such constraints are known as *holonomic constraints* and are standard fare in an undergraduate physics curriculum. Unilateral constraints, however, pose an entirely different problem. When one watches the evolution of mechanical system in the presence of a unilateral constraint, a moment may occur when the velocity of the system must *jump discontinuously* in order to prevent the constraint(s) from being violated. These situations are known as *impacts.*

Friction is also a difficult topic in models of unilateral constraints. In a frictional model, one proposes that non-normal forces at a unilateral constraint affect the system so that energy is dissipated (presumably in the form of heat at the contact). Rigorous mathematical results describing mechanical systems in frictional contact have had a difficult history.

Indeed, the entire mathematical theory of friction appeared doomed when the French aviator, politician, and mathematician Paul Painlevé published an example in 1895 [29] of a simple bar in frictional contact with a flat surface and the analysis revealed that Coulomb friction simply did not allow any solution that prevented the bar from accelerating into the surface.

A resolution to Painlevé's paradox was eventually furnished by David Stewart [37]. He was able to show that by allowing for impacts in situations when normal velocities of contacts at surfaces were non-negative (that is, in states of non-collision), the "bad" state that resulted in non-existence of solutions could be avoided. These became known as *tangential impacts* to indicate that the contacts were moving tangentially to the surface when impact occurs.

Using this strategy, Stewart devised an existence proof for what we call "modern Coulomb friction" (modern, because it allows for the modern resolution of Painlevé's paradox: tangential impacts). This existence proof holds for single contact situations. For multiple contacts it appears to be unknown whether the numerical methods (see Chapter 4) construct approximations which converge to solutions of the modern Coulomb model of friction.

In Chapter 6, we supply a model for friction with an existence result even for the case of multiple contacts. However, we use a friction model which is not the modern Coulomb model, and we require persistent contact.

Thesis Outline

This thesis can be divided into three main parts. The first part, comprised of Chapter 2 and Chapter 3, analyzes frictionless constraints, both unilateral and bilateral in a Hamiltonian framework. In Chapter 2, we show how a generalized approach to variational calculus yields a mathematical framework for considering frictionless

unilateral constraints. In Chapter 3, we switch to a Hamiltonian framework. We show how linear complementarity problems may be used for computations relevant to the unilaterally constrained dynamics with an inelastic constitutive law. We also show that despite all these computations (of force or impact from constraints at a give time) being well-posed, non-uniqueness problems still arise.

The second part, comprised of Chapter 4, consists of a survey of the literature for formulating and computing solutions for frictional contact problems.

The third part, comprised of Chapters 5 and 6, details a multi-contact frictional model with an existence theorem for the case of persistent contact. Our model departs from the usual Coulomb model in important ways.

Organization

Now we give more details on the contents of each chapter.

In the second chapter, we generalize the Principle of Stationary Action, central to the variational approach to physics, to reflect optimality conditions when constraints are present. We find that we obtain a measure-theoretic version of the Euler-Lagrange equations which may allow for shocks due to impacts with unilateral constraints. We compare this to the work of Moreau [24].

In the third chapter, we develop a treatment of Hamiltonian systems with frictionless unilateral and bilateral constraints. A key computational tool, the *linear complementarity problem*, is introduced. Impacts are handled using the inelastic constitutive law. We formulate so-called *force and impact problems*, and consider the

means of their solution via linear complementarity problems. We find that the friction and impact problems are well-posed, yet the full evolution problem for inelastic dynamics does not have uniqueness. We discuss how this is possible because of *right accumulations of impacts*. To illustrate this, we give an example due to Ballard [6]. We discuss conditions under which the full evolution problem is well-posed.

The fouth chapter is a survey of the theory of Coulomb friction. The formulation of this has historically had difficulties,. These difficulties date back to a non-existence example due to Painlevé in 1895, [29]. We present this example. To circumvent this non-existence result, a generalized concept of solution was introduced [37] which allowed so-called *tangential impulses.* We discuss the framework of this friction model in connection with the measure differential inclusions due to Moreau [24] and also the linear complementarity problem. We also discuss numerical aspects. Stewart [37] has shown that this new model (which we call modern Coulomb friction) admits existence of solutions for the case of a single contact.

In the fifth chapter, we formulate and analyze a differential inclusion involving a convex program, which we call a *Feedback Problem*. To this end, we must introduce set-valued functions, differential inclusions, and convex programming. We derive an existence result for feedback problems by using existence results pioneered by Filippov [12] for differential inclusions. We point out how this work generalizes to submanifolds. We give special cases for which we may also obtain uniqueness and well-posedness.

In the sixth chapter, we develop a theory of frictional contact with an existence result even for multiple contacts. We restrict ourselves to the case of persistent contact. We accomplish this by formulating friction in terms of the "feedback problem" studied in Chapter 5. Our result is the first multi-contact existence result which does not assume small friction. However, this is accomplished at the expense of departing from usual friction laws. We discuss concerns about physicality relating to this variance from modern Coulomb friction.

The seventh chapter summarizes the work done, and indicates open questions and directions for further research.

Original Results

This thesis contains original results and ideas. The most important such results and ideas follow:

- In Chapter 2, we obtain the measure-theoretic Euler-Lagrange equations for describing mechanical systems with frictionless unilateral constraints. We do this *from first principles* by inventing a generalization of the Principle of Stationary Action for Lagrangian Mechanics.
- In Chapter 5, we formulate the so-called Feedback Problem. We give an existence result and also give special cases where uniqueness and well-posedness are also obtained.
- In Chapter 6, we formulate a model for friction (in the special case of persistent contact) that admits an existence proof for the multi-contact case.

CHAPTER 2

VARIATIONAL CALCULUS APPLIED TO UNILATERAL CONSTRAINTS

Introduction

Classical physical theories are traditionally presented via the specification of a scalar function $L(t, q, \dot{q})$, known as the *Lagrangian*, whose domain is the phase space of the system. By demanding that the time-integral of the Lagrangian be stationary with respect to admissible variations, one derives the dynamical equations of motion. This is commonly called *Hamilton's Principle*, or *The Principle of Stationary Action*:

$$\delta S = \delta \int L \, dt = 0.$$

The variational approach pioneered by Lagrange is particularly well suited for problems involving bilateral frictionless constraints. Typically, these are referred to as *holonomic constraints.* We refer the reader to [13] for an introduction to the variational calculus. More recently, functional analytic methods have been used to invoke a form of Hamilton's principle to enforce elastic rebound conditions from a unilateral constraint [11].

In this paper, we consider whether variational methods can be directly applied to consider impacts which are not conservative. Since Hamilton's principle is conservative by its nature, we are forced to modify it. However, our modification is little

more than the modification one makes when finding extremal values of a function in an interval instead of the entire real line: one looks for places where the derivative vanishes *and also the endpoints of the interval.* The term "critical" is often used to describe these candidate points. Thus, we call our generalization the *principle of critical action.*

We present our generalization of Hamilton's principle which is not strictly conservative. It allows many different types of impact behaviors, including both elastic and inelastic behaviors.

The thrust of this work is simple: we assume our *Principle of Critical Action*, and from it derive the so-called *measure-theoretic Euler-Lagrange* equations, which are analogues of the usual Euler-Lagrange equations, but generalized to deal with velocity discontinuities:

THEOREM 2.1. *Define*

$$\mathcal{Q} := \{q \in \mathbb{R}^d : f_\alpha(q) \geq \textit{for all } \alpha \in Z\},$$

where $\{f_\alpha\}_{\alpha \in Z}$ *is a collection of unilateral constraints satisfying the* standing hypotheses *given below. An* admissible motion $q : (t_i, t_f) \to \mathcal{Q}$ *satisfies the* Principle of Critical Action on $\mathcal{Q}$ with Lagrangian L *if and only if it satisfies the* measure-theoretic Euler-Lagrange equations*:*

$$d\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q}dt = \sum_{\alpha \in Z} (\nabla f_\alpha(t))\mu_\alpha(t), \tag{2.1}$$

where the $\{\mu_\alpha\}_{\alpha \in Z}$ *are finite non-negative scalar-valued measures on* (t_i, t_f)*, each* μ_α *supported on the set of times such that* $f_\alpha(q(t)) = 0$*.*

We must define the Principle of Critical Action and the spaces involved in order for this to be completely clear. This is done below.

Our *first principles* derivation of these measure-theoretical Euler-Lagrange equations is interesting for several reasons. Firstly, it is compelling that we have motivated the measure-theoretic Euler-Lagrange equations by making a well-motivated generalization of Hamilton's principle. Secondly, it may be possible to use this framework to answer questions regarding existence of solutions to the measure-theoretic Euler-Lagrange equations. Finally, it is an elegant result involving a healthy combination of ideas from optimality theory, functional analysis, measure theory, and the calculus of variations.

Principle of Critical Action

According to what is usually called *Hamilton's Principle*, the motion of a physical system given by a Lagrangian function L has the following extremal property: for each $t_0 < t_1$, and for each admissible variation h of the motion q such that $h(t_0) = h(t_1) = 0$,

$$\frac{d}{d\epsilon} \int_{t_0}^{t_1} L(t, q + \epsilon h, \dot{q} + \epsilon \dot{h}) \, dt = 0.$$

This principle holds even in the presence of equality constraints of the form $g(q) = 0$. In this paper, we extend the principle in such a way that we may handle constraints

of the inequality type, i.e. $f(q) \geq 0$. Our analysis also allows a considerably wider class of motions than usual. For example, we do not require the motions to be smooth, or even piecewise smooth. What we do require is that the motion $q(t)$ is absolutely continuous (AC), and that the velocity $\dot{q}(t)$ is a function of bounded variation (BV). That is, we assume the motion $q(t)$ is an AC-BV function, which we describe below.

Standing Hypotheses and Objects. For the remainder of the chapter, until otherwise specified, we make the following standing hypotheses. Let $t_i < t_f$ be *initial and final times.* We suppose $L : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is C^1, Z is a finite index set, $f_\alpha : \mathbb{R}^d \to \mathbb{R}$ is a C^3 function for each $\alpha \in Z$, and $\nabla f_\alpha \neq 0$ in a neighborhood of $f_\alpha^{-1}(\{0\})$. Finally, assume that for any $\vec{r} \in \mathbb{R}^d$, the set of constraints f_α, for which $f_\alpha(\vec{r}) = 0$, have linearly independent gradients: that is, for every $\vec{r} \in \mathbb{R}^d$, we assume

$$\{\nabla f_\alpha(\vec{r}) : f_\alpha(\vec{r}) = 0\} \text{ is a linearly independent set of vectors.}$$

Definition 2.2 (Admissible Motions.). *We say that a function $q : (t_i, t_f) \to \mathbb{R}^d$ is an* admissible motion *provided the following hold:*

1. *The function $q(t)$ is absolutely continuous.*

2. *There exists a continuous from the right function $\dot{q}(t)$ of bounded variation such that*

$$q(t) = q(t_i) \int_{t_i}^{t} \dot{q}(s)\, ds.$$

3. *The motion $q(t)$ satisfies the constraints for all times:*

$$f \circ q(t) \geq 0 \text{ for all } t \in (t_i, t_f).$$

Remark. The first and second assumptions are that $q(t)$ is what we will call an AC-BV function. We give more details concerning the AC-BV space below. For now, we mention that given any AC-BV function x we take the notation $\dot{x}$ to refer to the unique continuous from the right function of bounded variation from which we may obtain x via integration. The AC-BV space is comparable to the C^1 class of functions, for it is defined by asserting a derivative exists with a certain property. For C^1 functions we assume the derivative is continuous, and for AC-BV functions we assume the derivative has bounded variation. However, for AC-BV functions the concept of the derivative is relaxed to the almost-everywhere derivative, and we gain the technical requirement of absolute continuity in order to assure that the function may be recovered by integrating its derivative. It is straightforward to see that C^1 functions are AC functions and that C^2 functions are AC-BV functions.

Definition 2.3 (Admissible Variations.). *Let t_0, t_1 be times such that $t_i < t_0 < t_1 < t_f$. Given an admissible motion $q(t)$, we say that an AC-BV function $h(t)$ is an* admissible variation of q supported on the interval $[t_0, t_1]$ *provided that there exists a sequence of AC-BV functions (h_n) and a sequence of positive numbers (ϵ_n) such that*

1. *The function $q(t) + \epsilon_n h_n(t)$ is an admissible motion for all n.*
2. *The sequence (h_n) converges to h in the sense of AC-BV functions (to be described below).*
3. *The sequence (ϵ_n) converges to* 0.

4. *For all n, h_n is supported on the closed interval $[t_0, t_1]$. In particular, we have the endpoint conditions*

$$h_n(t_0) = h_n(t_1) = 0.$$

DEFINITION 2.4 (Principle of Critical Action.). *We now state our* Principle of Critical Action*: an admissible motion q, defined on a time interval (t_i, t_f), is said to be a* Hamiltonian trajectory *if and only if for each admissible variation h of q supported on an interval $[t_0, t_1] \subset (t_i, t_f)$, we have*

$$\frac{d}{d\epsilon}\bigg|_{\epsilon \to 0^+} \int_{t_0}^{t_1} L(t, q + \epsilon h, \dot{q} + \epsilon \dot{h})\, dt \leq 0.$$

Preliminaries

The AC-BV Function Space

DEFINITION 2.5 (Functions of Bounded Variation). *Let I be the interval $[t_0, t_1]$. We say that a function $q : I \to \mathbb{R}^d$ is a* function of bounded variation*, or that it has bounded variation, provided that it is continuous from the right and satisfies $TV[q] < \infty$. The function TV stands for* total variation *and is given by*

$$TV[q] = |q(t_0)| + \sup \sum_{i=1}^{N-1} |q(s_{i+1}) - q(s_i)|,$$

where the supremum is taken over all finite lists of times $t_0 = s_1 < s_2 < \cdots < s_N = t_1$. We denote the set of functions of bounded variation from $[t_0, t_1]$ to $\mathbb{R}^d$ by $BV([t_0, t_1], \mathbb{R}^d)$.

Remark. Usually functions with bounded variation are not assumed to be continuous from the right, so our definition is non-standard. However, we will have no occasion to deal with any functions of bounded variation which are not continuous from the right, and having to spell out this assumption repeatedly would be cumbersome.

Definition 2.6 (Borel-Stieltjes Measure). *Suppose that $x \in BV([t_0, t_1], \mathbb{R})$. We define the* Borel-Stieltjes measure corresponding to *x to be the unique finite signed measure dx such that*

$$dx([t_0, t]) = x(t) - x(t_0).$$

If we have $x \in BV([t_0, t_1], \mathbb{R}^d)$, we can produce a vector-valued Borel-Stieltjes measure, obtained in the same fashion.

Remark. We remark that to show this is well-defined is non-trivial, but we do not prove it, as it is a well-known result. In particular it is well known that BV functions are precisely those functions which may be written as the difference of monotone functions. It is also well known that continuous from the right monotone functions give rise to non-negative Borel-Stieltjes measures. One can refer to [8], for example, to put the pieces together and verify the above definition is sensible.

The following is a well-known result:

Theorem 2.7. *The space $BV([t_0, t_1], \mathbb{R}^d)$ is a Banach space when given the total variation norm.*

Proof. Signed finite measures comprise a Banach space when normed by the total-variation norm for measures (refer to [8] for more information). The mapping $x \mapsto dx$ is an isometry between the normed linear space of BV functions with the TV norm and the Banach space of signed finite measures with the total-variation norm for measures. It follows that BV is a Banach space. □

Definition 2.8 (Absolutely Continuous Functions). *We say that a function* $q : [t_0, t_1] \to \mathbb{R}^d$ *is* absolutely continuous *if there exists an integrable (in the Lebesgue sense) function* $v : [t_0, t_1] \to \mathbb{R}^d$ *such that*

$$q(t) = q(t_0) + \int_{t_0}^{t} v(t)dt.$$

The integral here is meant in the Lebesgue sense. We denote the set of all $\mathbb{R}^d$*-valued absolutely continuous functions on* $[t_0, t_1]$ *by* $AC([t_0, t_1], \mathbb{R}^d)$*. The function* $v(t)$ *is called a* Radon-Nikodym derivative *of* $q(t)$*.*

Remark. We first remark that usually we think of Radon-Nikodym derivatives as being applied to measures. In particular, a measure may have a density with respect to another measure (frequently, the Lebesgue measure). An alternative way to characterization AC functions is to define them as functions of bounded variation which generate Borel-Stieltjes measures which have densities with respect to Lebesgue measure. This is why we use the term Radon-Nikodym. We could simply say almost-everywhere derivative, but we want to emphasize that we may recover the original function through integration of its almost-everywhere derivative[1]. Note

[1]This is not the case for the so-called *devil's staircase*, for example

that Radon-Nikodym derivatives need not be unique. In particular, they come in almost-everywhere equal (with respect to Lebesgue measure) equivalence classes.

DEFINITION 2.9. *We say that a function $q : [t_0, t_1] \to \mathbb{R}^d$ is AC-BV if and only if $q(t)$ is absolutely continuous, and it has a Radon-Nikodym derivative of bounded variation. We denote the set of all such functions as AC-BV$([t_0, t_1], \mathbb{R}^d)$. By $\dot{q}$ we denote the unique continuous from the right function of bounded variation that is a Radon-Nikodym derivative of q.*

Remark. Suppose q is AC-BV; we want to see that $\dot{q}$ is well-defined. That is, we need to know that there may not be distinct continuous from the right functions of bounded variation which are both Radon-Nikodym derivatives of q. This is not particularly difficult to show, and we omit a proof.

PROPOSITION 2.10. *Let X be the space of $\mathbb{R}^d$-valued functions defined on the interval $[t_0, t_1]$ which are AC-BV:*

$$X := AC\text{-}BV([t_0, t_1], \mathbb{R}^d).$$

Then X is a Banach space when equipped with the following norm, which we call the AC-BV norm:

$$||x|| := |x(t_0)| + TV[\dot{x}]\,. \tag{2.2}$$

Moreover,

$$|x|_\infty \leq (1 + [t_1 - t_0])||x||.$$

In particular, convergence in the AC-BV norm implies uniform (and hence pointwise) convergence.

Proof. First we establish that $||\cdot||$ is a norm on X. Clearly $||0|| = 0$. For $\kappa > 0$, $|\kappa x| = \kappa\,||x||$. Now we show that for $x, y \in$ AC-BV, $||x+y|| \leq ||x|| + ||y||$:

$$\begin{aligned} ||x+y|| &= |x(t_0)+y(t_0)| + \mathrm{TV}\,[\dot{x}+\dot{y}] \\ &\leq |x(t_0)| + \mathrm{TV}\,[\dot{x}] + |y(t_0)| + \mathrm{TV}\,[\dot{y}] = ||x|| + ||y||. \end{aligned}$$

We've used the subadditivity of the total variation functional here. Now we show that X is a Banach space when equipped with this norm. To this end, suppose that (x_n) is a sequence of points in X which is Cauchy with respect to the AC-BV norm $||\cdot||$. We show that (x_n) converges to a point in X in the sense of the AC-BV norm.

Since (x_n) is Cauchy in the AC-BV norm, it follows that $(\dot{x}_n)$ is Cauchy in the TV norm. By Theorem 2.7, BV is a Banach space with the TV norm. It follows that $(\dot{x}_n)$ converges in the TV norm to some function we call $\dot{x}$. Since (x_n) is Cauchy in the AC-BV norm, the sequence $(x_n(t_0))$ converges. Call the value it converges to $x(t_0)$. Define

$$x(t) := x(t_0) + \int_{t_0}^{t} \dot{x}(s)ds.$$

We claim (x_n) converges to x in the AC-BV norm. Notice that for each n,

$$x_n(t) = x_n(t_0) + \int_{t_0}^{t} \dot{x}_n(s)ds.$$

Hence

$$||x - x_n|| = |x(t_0) - x_n(t_0)| + \mathrm{TV}[\dot{x} - \dot{x}_n] \to 0 \text{ as } n \to \infty.$$

Hence X is a Banach space with the given norm. All that remains is to show that the AC-BV norm dominates the uniform norm. To see this, we bound $|\dot{x}|_\infty$ for $x \in X$. In particular, $|\dot{x}|_\infty \leq \mathrm{TV}[\dot{x}] \leq ||x||$. Now see that $|x|_\infty \leq |x(t_0)| + (t_1 - t_0)|\dot{x}|_\infty$. Since $|x(t_0)| \leq ||x||$, we may combine all these expressions to get the bound

$$|x|_\infty \leq (1 + [t_1 - t_0])||x||.$$

It follows that AC-BV convergence implies uniform convergence. □

Remark. We have now introduced the major norms of this work. Our notation for them will be consistent. We denote the usual finite-dimensional Euclidean norm as $|\cdot|$, the uniform (or supremum) norm as $|\cdot|_\infty$, the total variation norm as $\mathrm{TV}[\cdot]$, and the AC-BV norm as $||\cdot||$.

PROPOSITION 2.11. *Suppose $x \in AC([t_0, t_1], \mathbb{R}^d)$ and $y \in BV([t_0, t_1], \mathbb{R}^d$. Then*

$$\int_a^b x \cdot dy = \lim \sum x(t_i) \cdot (y(t_{i+1}) - y(t_i)) = \lim \sum x(t_{i+1}) \cdot (y(t_{i+1}) - y(t_i)),$$

where the limit is over finer partitions $\{t_i\}$ of $[a, b]$. Moreover, the same result holds if we assume instead that x is BV and y is AC.

Proof. It is known [8] that if the Riemann integral exists, then the Lebesgue integral is equal to it. We show the Riemann integral does exist, from which the formulas above follow straightforwardly. The key to this is that if we produce a very fine partition $\{t_i\}$ of $[a, b]$, then $x(t)$ will be nearly constant on the intervals $[t_i, t_{i+1}]$. This forces the upper and lower Riemann sums associated with the partition to be close (and getting

closer as we select finer partitions). It is straightforward to verify (by monotonicity of integrals) that the upper and lower Riemann sums are upper and lower bounds for the value of the Lebesgue integral $\int_a^b x \cdot y$. It follows that the upper and lower Riemann sums must both converge (as we make the partition finer) to the value of the Lebesgue integral. Because the upper and lower Riemann sums converge to the value of the Lebesgue integral, the Riemann sums of the proposition must converge as well, by the squeeze theorem of limits.

Now we show the moreover part. We show the sums

$$\sum x(t_i^*) \cdot (y(t_{i+1}) - y(t_i))$$

converge. Here, t_i^* are arbitrary choices $t_i^* \in [t_i, t_{i+1}]$.

Given any partition, we define $\{m_i\}$ and $\{M_i\}$ to be choices for t_i^* which minimize or maximize the Riemann sum on that partition, respectively. These choices describe the lower and upper Riemann sums, which we show converge for reasons similar to before.

We determine the difference between upper and lower Riemann sums:

$$\begin{aligned}\left|\sum x(m_i) \cdot (y(t_{i+1}) - y(t_i)) - \sum x(M_i) \cdot (y(t_{i+1}) - y(t_i))\right| \\ \leq \left|\sum (x(m_i) - x(M_i)) \cdot (y(t_{i+1}) - y(t_i))\right|.\end{aligned}$$

We want to show the right-hand side tends to zero for finer partitions. We may apply the Cauchy-Schwarz inequality to the right-hand side in order to obtain

$$\left|\sum (x(m_i) - x(M_i)) \cdot (y(t_{i+1}) - y(t_i))\right| \leq \sum |x(m_i) - x(M_i)| \ \sup \ |y(t_{i+1}) - y(t_i)|.$$

Because x is BV, there exists a constant $K > 0$ which is independent of the partitions, such that for every partition we have

$$\sum |x(m_i) - x(M_i)| < K.$$

Hence,

$$\sum |x(m_i) - x(M_i)| \sup |y(t_{i+1}) - y(t_i)| \leq K \sup |y(t_{i+1}) - y(t_i)|,$$

which tends to 0 as we make finer partitions since y is AC (and uniformly continuous on $[a, b]$). Hence the upper and lower Riemann sums converge, and the formulas must hold. □

Given a measurable vector-valued function x and a vector-valued measure dy, we will define the measure $x \cdot dy$ to be the unique measure $\tilde{\mu}$ such that $\tilde{\mu}(E) = \int_E x \cdot dy$. The preceding proposition gives conditions for when we may use the Riemann approach to integration.

THEOREM 2.12 (Product Rule.). *Let I be the interval $[t_0, t_1]$. Suppose $x \in BV([t_0, t_1], \mathbb{R}^d)$, and $y \in AC([t_0, t_1], \mathbb{R}^d)$. The Borel-Stieltjes measure $d(x \cdot y)$ is related to the Borel-Stieltjes measures dx and dy in the following manner:*

$$d(x \cdot y) = x \cdot dy + y \cdot dx.$$

Proof. We consider prove only the $d = 1$ case, as the generalization to higher d is straightforward.

Let $[a,b] \subset [t_0,t_1]$. We show that

$$\int_a^b d(xy) = \int_a^b x\,dy + \int_a^b y\,dx.$$

The left hand side may be evaluated as a simple expression: products of BV functions are again BV and we have $\int_a^b d(xy) = [d(xy)]([a,b]) = x(b)y(b) - x(a)y(a)$.

The right hand side may be evaluated as follows, using Proposition 2.11:

$$\begin{aligned}\int_a^b xdy + \int_a^b ydx &= \lim \sum \left(x(t_i)\left[y(t_{i+1}) - y(t_i)\right] + y(t_{i+1})\left[x(t_{i+1}) - x(t_i)\right]\right) \\ &= \lim \sum \left(x(t_{i+1})y(t_{i+1}) - x(t_i)y(t_i)\right) \text{ which telescopes} \\ &= \lim x(b)y(b) - x(a)y(a) = x(b)y(b) - x(a)y(a)\end{aligned}$$

Hence the left and right sides are indeed equal, and thus $d(xy)$ and $xdy+ydx$ must be the same measure, as they agree on a subset of measurable sets which generate the Borel sigma algebra. □

THEOREM 2.13 (Integration by Parts). *Suppose* $x \in AC\text{-}BV([t_0,t_1],\mathbb{R}^d)$ *and* $y \in BV([t_0,t_1],\mathbb{R}^d)$. *Then*

$$\int_{t_0}^{t_1} \dot{x}(t)\cdot y(t)dt = x(t)\cdot y(t)|_{t_0}^{t_1} - \int_{t_0}^{t_1} x(t)\cdot dy(t) \tag{2.3}$$

Proof. In particular, x is AC. Integrating the conclusion of Theorem 2.12 over $[t_0,t_1]$ yields

$$\int_{t_0}^{t_1} y(t)\cdot dx(t) = x(t)\cdot y(t)|_{t_0}^{t_1} - \int_{t_0}^{t_1} x(t)\cdot dy(t),$$

which in turn yields Equation (2.3) after substituting $dx = \dot{x}dt$, which is valid since dx is absolutely continuous with respect to the Lebesgue measure dt. □

Tangent Cones, Fréchet Derivatives, and Criticality

DEFINITION 2.14. *A* closed cone *in a Banach space X is a non-empty subset C that is closed, convex, and invariant under multiplication by positive scalars.*

DEFINITION 2.15. *Let M be a closed subset of a Banach space X with norm $|\cdot|_X$. Let $x \in M$. We say that $v \in X$ is a* geometric tangent vector at x in M *provided that there exists a sequence of points $(x_n) \subset M$ tending to x and a constant $\kappa \geq 0$ such that*

$$v = \kappa \lim_{n \to 0} \frac{x_n - x}{|x_n - x|_X},$$

where the limit is understood as convergence in X. For $\kappa = 1$, we say it is a unit geometric tangent vector. *We define the* tangent cone at x of M, *to be denoted $T_x M$, to be the closure of the convex hull of the geometric tangent vectors at x of M.*

PROPOSITION 2.16. *Let X be a Banach space. Suppose M is closed subset of X. Let $x \in M$. The tangent cone $T_x M$ is a closed cone in the Banach space X.*

Proof. This follows straightforwardly from the definition. In particular, $T_x M$ is defined to be the closure of the convex hull of the geometric tangent vectors. The convex hull of the geometric tangent vectors is a cone, since the set of geometric tangent vectors is closed under multiplication by non-negative scalars. One can readily verify that the closure of a cone is again a cone, indeed a closed cone. □

PROPOSITION 2.17. *Suppose that (q_n) tends to q in a closed subset M of a Banach space X. Suppose that (ϵ_n) is a sequence of positive numbers tending to zero. Assume*

that for some $v \in X$, we have

$$v = \lim_{n\to\infty} \frac{q_n - q}{\epsilon_n},$$

where the limit implies convergence in X. Then $v \in T_qM$.

Proof. To avoid triviality assume $v \neq 0$. Note that the X-norm is continuous in X. Thus,

$$\frac{|q_n - q|_X}{\epsilon_n} \to |v|_X \text{ as } n \to \infty.$$

It follows that

$$\frac{\epsilon_n |v|_X}{|q_n - q|_X} \to 1 \text{ as } n \to \infty.$$

Using this fact, we may write

$$v = \lim_{n\to\infty} \frac{q_n - q}{\epsilon_n} \frac{\epsilon_n |v|_X}{|q_n - q|_X} = |v|_X \lim_{n\to\infty} \frac{q_n - q}{|q_n - q|_X},$$

which shows that v is a geometric tangent vector with $\kappa = |v|_X$, and hence $v \in T_qM$. □

Definition 2.18. *Suppose that X is a Banach space and $J : X \to \mathbb{R}$ is a scalar function. Suppose that for some $x \in X$, there exists a bounded linear functional ρ such that for every vector v,*

$$\rho(v) = \lim_{h\to 0^+} \frac{J(x + hv) - J(x)}{h}.$$

Then we say that J is Fréchet differentiable *at x and that its Fréchet derivative is the bounded linear functional ρ. We denote the Fréchet derivative at x in the direction v by $\delta J_x(v)$. That is, $\rho(v) = \delta J_x(v)$.*

DEFINITION 2.19. *Suppose X is a Banach space. Suppose $J : X \to \mathbb{R}$ is Fréchet differentiable. Let M be a closed subset of X. We say that a point $x \in M$ is* critical *with respect to J on M if either $\delta J_x(T_xM) \geq 0$ or $\delta J_x(T_xM) \leq 0$. The former case we will refer to as* positive critical*, the latter as* negative critical*. If x is both positive critical and negative critical with respect to J, we say it is* stationary*.*

Admissible Motions and their Variations

The purpose of the following section is to examine the definitions we gave earlier of admissible motions and variations, and compare them to tangent cone concepts. This will enable us to characterize Hamiltonian trajectories using the machinery of tangent cones and negative criticality.

DEFINITION 2.20. *Suppose $a, b \in \mathbb{R}^d$. Define $\mathcal{M}((t_0, a), (t_1, b))$ to be the set of AC-BV functions $\{q(t)\}$ such that $q(t_0) = a$, $q(t_1) = b$, and for all $\alpha \in Z$, all $t \in [t_0, t_1]$, $f_\alpha(q(t)) \geq 0$.*

We will henceforth refer to $\mathcal{M}((t_0, a), (t_1, b))$ simply as M whenever it is clear from context.

DEFINITION 2.21. *Suppose $q \in M := \mathcal{M}((t_0, a), (t_1, b))$. Define the following collections of bounded linear functionals on $C([t_0, t_1], \mathbb{R}^d)$:*

$$\begin{aligned} \mathcal{C}_q := \; & \{h \mapsto \nabla f_\alpha \cdot h(t) : \alpha \in Z \text{ and } t \in [t_0, t_1] \text{ such that } f_\alpha \circ q(t) = 0\} \\ & \cup \; \{h \mapsto \pm \hat{e}_i \cdot h(t_0) : i = 1, 2, \cdots d\} \\ & \cup \; \{h \mapsto \pm \hat{e}_i \cdot h(t_1) : i = 1, 2, \cdots d\}. \end{aligned}$$

The notation $\hat{e}_i$ refers to the canonical basis vector in $\mathbb{R}^d$. We remark that $\mathcal{C}_q$ depends on the choice $[t_0, t_1] \subset (t_i, t_f)$, though it does not actually depend on the

choice of endpoint conditions a and b. We do not write this dependence, and it should always be clear from context.

DEFINITION 2.22. *Suppose that X is a Banach space and S is a collection of bounded linear functionals on X. We define* the cone of S *to be the nonempty closed convex set*

$$S_* := \{x \in X : \textit{For every } \ell \in S, \textit{ we have that } \ell(x) \geq 0\}.$$

DEFINITION 2.23. *Suppose X is a Banach space and $S \subset X$. Denote the set of bounded linear functionals on X as X^*. We define* the dual cone of S *to be the nonempty set*

$$S^* := \{\ell \in X^* : \textit{For every } x \in S, \textit{ we have that } \ell(x) \geq 0\}.$$

It is straightforward to see that S^* is convex. We also note that D^* is closed in both the strong and weak topologies on X^* (we define these notions below in case the reader is unfamiliar).*

We now state a major result whose proof is postponed:

LEMMA 2.24. *Let $M := \mathcal{M}((t_0, a), (t_1, b))$; define $\mathcal{C}_q$ as in Definition 2.21. Then the cone of $\mathcal{C}_q$ is precisely the uniform closure of the tangent cone of M at q:*

$$\overline{T_q M} = \textit{ cone of } \mathcal{C}_q.$$

Remark. This result is difficult. We prove it below as Lemma 2.62.

PROPOSITION 2.25. *The set of admissible motions on (t_i, t_f) are those AC-BV functions $q(t)$ whose restrictions to the domains of the form $[t_0, t_1] \subset (t_i, t_f)$ are in $\mathcal{M}((t_0, q(t_0)), (t_1, q(t_1)))$.*

Proof. The admissible motions are simply the AC-BV functions obeying the constraints. By assumption q is AC-BV; we only need to show it obeys the constraints. Since the restrictions to each closed interval $[t_0, t_1] \subset (t_i, t_f)$ obey the constraints, it follows that the constraints are everywhere obeyed on (t_i, t_f). □

LEMMA 2.26. *Let $q : (t_i, t_f) \to \mathbb{R}^d$ be an admissible motion. An AC-BV function $v(t)$ is an admissible variation of q supported on $[t_0, t_1]$ if and only if v is supported on an interval $[t_0, t_1] \subset (t_i, t_f)$ and restricted to this interval v is a geometric tangent vector at $q|_{[t_0,t_1]}$ in $\mathcal{M}((t_0, q(t_0)), (t_1, q(t_1)))$.*

Proof. Suppose first that $v(t)$ is an admissible variation of q. We will abuse notation and use q and v to also refer to their restrictions to $[t_0, t_1]$. We show that $v(t)$ is a geometric tangent vector at q in M. From Definition 2.3, there must exist a sequence of AC-BV functions (v_n) tending to v and a sequence of positive real numbers (ϵ_n) tending to 0 such that $q(t) + \epsilon_n v_n(t)$ is an admissible motion for all n. Also, $v_n(t_0) = v_n(t_1) = 0$, so we can even conclude $q_n(t) := q(t) + \epsilon_n v_n(t) \in M$. Note that (q_n) converges to q in the AC-BV sense. Define $\kappa := |v|$. We claim that

$$v = \kappa \lim_{n\to\infty} \frac{q_n - q}{|q_n - q|},$$

which would prove v is a geometric tangent vector in the sense of Definition 2.15. This claim is true, since

$$\kappa \lim_{n\to\infty} \frac{q_n - q}{|q_n - q|} = |v| \lim_{n\to\infty} \frac{v_n}{|v_n|} = |v| \cdot \frac{v}{|v|} = v,$$

where we have used the fact that $|v| > 0$ (which we may assume to avoid triviality) implies that the projection to the unit sphere $x \mapsto x/|x|$ is continuous in a neighborhood of v.

Now the converse. Suppose that v is a geometric tangent vector at q in M. Note that $v(t_0) = v(t_1) = 0$. By Definition 2.15, there exists a sequence $(q_n) \subset M$ converging to q and a nonnegative scalar $\kappa \geq 0$ such that

$$v = \kappa \lim_{n\to\infty} \frac{q_n - q}{|q_n - q|}.$$

Define

$$v_n = \kappa \frac{q_n - q}{|q_n - q|} \text{ and } \epsilon_n = \frac{|q_n - q|}{\kappa}.$$

Observe that (v_n) converges to v in the AC-BV sense. Note that since $q_n(t_0) = a$ and $q_n(t_1) = b$ for all n, $v_n(t_0) = v_n(t_1) = 0$ for all n. Observe that (ϵ_n) is a nonnegative sequence of scalars converging to 0 (since (q_n) converges to q in the AC-BV sense). By defining $v_n(t) = 0$ for $t \notin [t_0, t_1]$, we may define the sequence of admissible motions (q_n) such that $q_n := q + \epsilon_n v_n$. It is straightforward to verify that the (q_n) really are admissible motions. According to Definition 2.3, v is an admissible variation. $\square$

Measure-Theoretic Euler-Lagrange Equations

The Action Functional

DEFINITION 2.27. *Recall L is the Lagrangian function $L(t, q, \dot{q})$ which is C^1 in its arguments, by the standing hypotheses. We define the* action functional *J as a real-valued function on AC-BV$([t_0, t_1], \mathbb{R}^d)$ given by*

$$J(q) = \int_{t_0}^{t_1} L(t, q, \dot{q})\, dt. \tag{2.4}$$

LEMMA 2.28. *The action functional J given by Equation (2.4) is Fréchet differentiable.*

Proof. Let $x \in \text{AC-BV}([t_0, t_1], \mathbb{R}^d) =: X$. Define the functional δJ_x such that for any $v \in X$,

$$\delta J_x(v) := \lim_{h \to 0^+} \frac{J(x + hv) - J(x)}{h}.$$

By the definition of Fréchet differentiability, our goal is to show δJ_x is a bounded linear functional on X. To this end we compute:

$$\delta J_x(v) = \lim_{h \to 0^+} \int_{t_0}^{t_1} \frac{1}{h} \left[L(t, x + hv, \dot{x} + h\dot{v}) - L(t, x, \dot{x}) \right] dt.$$

Using differentiability of L, we may obtain

$$\delta J_x(v) = \lim_{h \to 0^+} \int_{t_0}^{t_1} \frac{1}{h} \left[h \frac{\partial L}{\partial x} \cdot v + h \frac{\partial L}{\partial \dot{x}} \cdot \dot{v} + E(t, h) \right] dt, \tag{2.5}$$

where $E(t, h)$ is the error of the linear approximation. We need to establish several properties of E. Since linear approximations of differentiable functions have sublinear error, we know that for each $t \in [t_0, t_1]$,

$$\lim_{h\to 0^+} \frac{E(t,h)}{h} = 0.$$

Claim. When $h > 0$ is sufficiently small, then $E(t,h)/h$ has a uniform bound for all $t \in [t_0, t_1]$.

Since L is continuously differentiable, it is locally Lipschitz continuous, and hence $\Delta L/h$ is bounded by some constant for all $t \in [t_0, t_1]$ and sufficiently small $h > 0$. Since the derivatives $\frac{\partial L}{\partial x}$ and $\frac{\partial L}{\partial \dot{x}}$ are continuous, they are also bounded. Similarly v and $\dot{v}$ are bounded, as they are of bounded variation. It follows that for some $C > 0$, $E(t,h)/h \leq C$ for all $t \in [t_0, t_1]$, sufficiently small $h > 0$. *This completes the proof of the claim.*

We may simplify (2.5) further to obtain

$$\delta J_x(v) = \lim_{h\to 0^+} \int_{t_0}^{t_1} \left[\frac{\partial L}{\partial x} \cdot v + \frac{\partial L}{\partial \dot{x}} \cdot \dot{v} + \frac{E(t,h)}{h} \right] dt,$$

which becomes

$$\delta J_x(v) = \int_{t_0}^{t_1} \left[\frac{\partial L}{\partial x} \cdot v + \frac{\partial L}{\partial \dot{x}} \cdot \dot{v} \right] dt + \lim_{h\to 0^+} \int_{t_0}^{t_1} \frac{E(t,h)}{h}\, dt.$$

The limit of the second integral is zero. To see this, we must appeal to the Lebesgue Dominated Convergence Theorem. For this to apply, we need to see that the integrand is dominated, for sufficiently small $h > 0$, by an integrable function $g(t)$. Since $E(t,h)$ is bounded by some constant for all $t \in [t_0, t_1]$ and sufficiently small $h > 0$, the theorem applies (with the dominating function $g(t)$ chosen to be that bounding constant) and we may move the limit inside.

We arrive at

$$\delta J_x(v) = \int_{t_0}^{t_1} \left[\frac{\partial L}{\partial x} \cdot v + \frac{\partial L}{\partial \dot{x}} \cdot \dot{v} \right] dt.$$

At this point we see that δJ_x is a linear functional on v. It remains to be seen that it is bounded.

Observe that $\frac{\partial L}{\partial \dot{x}}(t, x(t), \dot{x}(t))$ is a BV function of t. Denote by $d\frac{\partial L}{\partial \dot{x}}$ the associated Borel-Stieltjes measure. Applying integration by parts, Theorem 2.3,

$$\delta J_x(v) = \int_{t_0}^{t_1} \frac{\partial L}{\partial x} \cdot v dt + \left[\frac{\partial L}{\partial \dot{x}} \cdot v(t_1) - \frac{\partial L}{\partial \dot{x}} \cdot v(t_0) \right] - \int_{t_0}^{t_1} v \cdot d\frac{\partial L}{\partial \dot{x}}. \tag{2.6}$$

From this formula we obtain

$$\delta J_x(v) \leq$$
$$\left[\left| \frac{\partial L}{\partial x} \right|_\infty + \left| \frac{\partial L}{\partial \dot{x}}(t_1, x(t_1), \dot{x}(t_1)) \right| + \left| \frac{\partial L}{\partial \dot{x}}(t_0, x(t_0), \dot{x}(t_0)) \right| + \left| d\frac{\partial L}{\partial x} \right| ([t_0, t_1]) \right] |v|_\infty.$$

Note that $\frac{\partial L}{\partial x}$ is bounded. Observe that since $\frac{\partial L}{\partial \dot{x}}$ is BV, the total variation measure $\left| d\frac{\partial L}{\partial x} \right|$ is finite. Hence for each $x \in X$, there exists $\tilde{M} > 0$ such that

$$\delta J_x(v) \leq \tilde{M} |v|_\infty. \tag{2.7}$$

By Proposition 2.10, this implies $\delta J_x(v) \leq \tilde{M}(1 + [t_1 - t_0])||v||$, so we see that δJ_x is a bounded linear functional. Hence J is Fréchet differentiable at x. □

COROLLARY 2.29. *Let $\tilde{q}$ be an admissible motion. Let q be the restriction of $\tilde{q}$ to $[t_0, t_1] \subset (t_i, t_f)$. The Fréchet derivative of the action functional $\delta J_q : AC\text{-}BV([t_0, t_1], \mathbb{R}^d) \to \mathbb{R}$ admits a unique continuous extension to a bounded linear functional $\ell : C([t_0, t_1], \mathbb{R}^d) \to$*

$\mathbb{R}$. *By continuous extension, we mean that ℓ is bounded with respect to the uniform metric, and $\ell(v) = \delta J_q v$ for all $v \in AC\text{-}BV([t_0, t_1], \mathbb{R}^d) \subset C([t_0, t_1], \mathbb{R}^d)$.*

Proof. Allowing that v may be any element of $\mathrm{C}([t_0, t_1], \mathbb{R}^d)$, Equation (2.6) furnishes such an extension. Boundedness follows from Equation (2.7). Since the space $\text{AC-BV}([t_0, t_1], \mathbb{R}^d)$ is a dense subset of $\mathrm{C}([t_0, t_1], \mathbb{R}^d)$ in the uniform metric, there can be at most one continuous extension. □

PROPOSITION 2.30. *An admissible motion $q : (t_i, t_f) \to \mathbb{R}^d$ is Hamiltonian if and only if for every $[t_0, t_1] \subset (t_i, t_f)$, the restriction $q|_{[t_0,t_1]}$ is negative critical in $M := \mathcal{M}((t_0, q(t_0)), (t_1, q(t_1)))$ with respect to J:*

$$\delta J_q(T_q M) \leq 0.$$

Remark. Observe that we are abusing notation here: $T_q M$ doesn't technically make sense, since $q \notin M$. Rather, the restriction of q to $[t_0, t_1]$ is what should be considered. The same goes for δJ.

Proof. First we show that if an admissible motion q is Hamiltonian, then

$$\delta J_q(T_q M) \leq 0.$$

Here, $M = \mathcal{M}((t_0, q(t_0)), (t_1, q(t_1)))$, for some $[t_0, t_1] \subset (t_i, t_f)$. Since q is Hamiltonian, we know by Definition 2.4 that for every admissible variation h of q supported on

$[t_0, t_1]$,

$$\left.\frac{d}{d\epsilon}\right|_{\epsilon\to 0^+} \int_{t_i}^{t_f} L(t, q+\epsilon h, \dot{q}+\epsilon\dot{h})\, dt \leq 0. \tag{2.8}$$

By Lemma 2.28, the action functional J is Fréchet differentiable at q, and hence we have

$$\left.\frac{d}{d\epsilon}\right|_{\epsilon\to 0^+} \int_{t_0}^{t_1} L(t, q+\epsilon h, \dot{q}+\epsilon\dot{h})\, dt = \delta J_q(h).$$

Thus, Equation (2.8) is asserting $\delta J_q(h) \leq 0$ for all admissible variations h of q supported on $[t_0, t_1]$. By Lemma 2.26, the restrictions of these admissible variations are precisely the geometric tangent vectors of $q|_{[t_0,t_1]}$ in M. Thus we have shown that $\delta J_q(h) \leq 0$ for all geometric tangent vectors h in T_qM. Since T_qM is the closed convex hull of the geometric tangent vectors at $q|_{[t_0,t_1]}$ in M, and because δJ_q is a bounded linear functional – and hence continuous – this proves $\delta J_q(T_qM) \leq 0$, or in other words, $q|_{[t_0,t_1]}$ is negative critical in M with respect to J.

Now the converse. We suppose that q is admissible and for every $[t_0, t_1] \subset (t_i, t_f)$, we have $\delta J_q(T_qM) \leq 0$, where $M := \mathcal{M}((t_0, q(t_0)), (t_1, q(t_1)))$. We show that $q(t)$ is Hamiltonian. To this end we let $h(t)$ be an admissible variation of q supported on some set $[t_0, t_1] \subset (t_i, t_f)$. We show Equation (2.8) holds. By Lemma 2.26, $h \in T_qM$. Hence $\delta J_q(h) \leq 0$ by hypothesis. This is implies Equation (2.8), so Definition 2.4 is satisfied. □

Lemma 2.31. *Let $M := \mathcal{M}((t_0, a), (t_1, b))$. The action functional J is negative critical at $q \in M$ if and only if the extension of δJ_q to $C([t_0, t_1], \mathbb{R}^d)$ satisfies*

$$\delta J_q(\overline{T_qM}) \leq 0.$$

Proof. Supose J is negative critical at $q \in M$. According to Lemma 2.29, the action functional J admits a Fréchet derivative at q which has a continuous extension to $C([t_0, t_1], \mathbb{R}^d)$. Hence $\delta J_q(T_qM) \leq 0$ implies that $\delta J_q(\overline{T_qM}) \leq 0$ by continuity of δJ_q with respect to the uniform metric.

Now the converse. Clearly, $T_qM \subset \overline{T_qM}$. Hence $\delta J_q(\overline{T_qM}) \leq 0$ implies the inequality $\delta J_q(T_qM) \leq 0$. □

Statement of Main Result

In this section, we state and prove our main result. We begin with a preliminary result characterizing of motions obeying the Principle of Critical Action:

THEOREM 2.32. *An admissible motion $q : (t_i, t_f) \to \mathbb{R}^d$ is Hamiltonian if and only if for every $[t_0, t_1] \subset (t_i, t_f)$, there exists a finite measure μ supported on $\mathcal{C}_q$ such that for each $h \in AC\text{-}BV([t_0, t_1], \mathbb{R}^d)$,*

$$\delta J_q[h] + \int \ell(h)\, d\mu(\ell) = 0.$$

Remark. This proof uses Lemma 2.45, which is proved below in an independent section.

Proof. First assume q is a Hamiltonian trajectory. Let $[t_0, t_1] \subset (t_i, t_f)$. Let $M := \mathcal{M}((t_0, q(t_0)), (t_1, q(t_1)))$. We show there exists a finite measure μ with the desired property.

Let q also denote its restriction to $[t_0, t_1]$. By Lemma 2.30, J is negative critical at $q \in M$. By Lemma 2.31, this implies that $-\delta J_q[h] \geq 0$ for all $h \in \overline{T_qM}$. By Lemma

2.24, $\overline{T_qM}$ is the cone of $\mathcal{C}_q$: hence, we see that $-\delta J_q$ is nonnegative on the cone of $\mathcal{C}_q$. It follows that $-\delta J_q$ is in the dual cone of the cone of $\mathcal{C}_q$. By Lemma 2.45, there exists a finite measure μ supported on $\mathcal{C}_q$ satisfying the conclusion of the theorem.

Now the converse; suppose that for every $[t_0, t_1] \subset (t_i, t_f)$ there exists a finite non-negative measure μ supported on $\mathcal{C}_q$ such that $-\delta J_q(h) = \int_{\mathcal{C}_q} \ell(h)\, d\mu(\ell)$ for each $h \in T_qM$. We show that q is a Hamiltonian trajectory.

Appealing to Lemma 2.30, we show that for every $[t_0, t_1] \subset (t_i, t_f)$ we have that $q|_{[t_0,t_1]}$ (we shall abuse notation and simply call this q) is negative critical in $M := \mathcal{M}((t_0, q(t_0)), (t_1, q(t_1)))$. Let $[t_0, t_1] \subset (t_i, t_f)$. To show q is negative critical on M we appeal to the definition and show that

$$\delta J_q(T_qM) \leq 0.$$

Let $h \in T_qM$. Observe $T_qM \subset \overline{T_qM}$. Thus $h \in \overline{T_qM}$. We show $\delta J_q(h) \leq 0$. By Lemma 2.24, $\overline{T_qM}$ is the cone of $\mathcal{C}_q$. It follows that h is in the cone of $\mathcal{C}_q$, and hence $\ell(h) \geq 0$ for all $\ell \in \mathcal{C}_q$. By hypothesis, there exists a non-negative measure μ supported on the cone of $\mathcal{C}_q$ such that $-\delta J_q(h) = \int_{\mathcal{C}_q} \ell(h) d\mu(\ell)$. Since $\ell(h) \geq 0$ for all $\ell \in \mathcal{C}_q$, the integrand is non-negative. Hence $\delta J_q(h) \leq 0$. The converse is shown. □

Now we are ready to state and prove the main result of this chapter:

THEOREM 2.33. *The function $q : (t_i, t_f) \to \mathbb{R}^d$ is a Hamiltonian trajectory if and only if it satisfies the* measure-theoretic Euler-Lagrange equations*:*

$$d\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} dt = \sum_{\alpha \in Z} (\nabla f_\alpha(t)) \mu_\alpha(t), \tag{2.9}$$

where the $\{\mu_\alpha\}_{\alpha\in Z}$ *are finite non-negative scalar-valued measures on* (t_i, t_f)*, each* μ_α *supported on the set of times such that* $f_\alpha(q(t)) = 0$*.*

Proof. Suppose first that q is a Hamiltonian trajectory. We show the measure-theoretic Euler-Lagrange equations hold.

Let $\epsilon > 0$. Choose $t_0 = t_i+\epsilon$ and $t_1 = t_f-\epsilon$. Define $M := \mathcal{M}((t_0, q(t_0)), (t_1, q(t_1)))$. By Equation (2.6), for each $h \in \text{AC-BV}([t_0, t_1], \mathbb{R}^d)$, we have

$$\delta J_q(h) = \int_{t_0}^{t_1} \frac{\partial L}{\partial q} \cdot h \, dt - \int_{t_0}^{t_1} h \cdot d\frac{\partial L}{\partial \dot{q}} + \ell_1(h(t_0)) + \ell_2(h(t_1)), \tag{2.10}$$

where ℓ_1 and ℓ_2 are some linear functionals on $\mathbb{R}^d$.

By Lemma 2.29, δJ_q has a unique extension to $C([t_0, t_1], \mathbb{R}^d)$; we abuse notation and continue to refer to this extension as δJ_q. It follows from the Riesz Representation Theorem, that δJ_q may be represented (via isomorphism) by a finite signed d-dimensional vector-valued measure. This is known as a Riesz representation:

$$-\delta J_q = d\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} dt + \nu, \tag{2.11}$$

where ν is a measure supported on the times $\{t_0\}\cup\{t_1\}$, such that $\int h \, d\nu = \ell_1(h(t_0))+ \ell_2(h(t_1))$.

By Theorem 2.32, for each $[t_0, t_1] \subset (t_i, t_f)$, there exists a non-negative finite measure μ supported on $\mathcal{C}_q$ such that for each $h \in \text{AC-BV}([t_0, t_1], \mathbb{R}^d)$,

$$\delta J_q[h] + \int_{\mathcal{C}_q} \ell(h) \, d\mu(\ell) = 0. \tag{2.12}$$

We remark here that δJ_q and $\mathcal{C}_q$ are dependent on the choice of $[t_0, t_1]$, although this dependence is not written.

Since every functional in $\mathcal{C}_q$ is continuous with respect to the uniform metric, it follows that (2.12) is also a formula for the unique extension of δJ_q to $C([t_0, t_1], \mathbb{R}^d)$. We consider it in this context from now on.

We now consider the Choquet integral (see Lax [19] for a discussion on Choquet theory) $\int_{\mathcal{C}_q} \ell(h)\, d\mu(\ell)$. We may break the integral into $|Z| + 1$ parts:

$$\int_{\mathcal{C}_q} \ell(h)\, d\mu(\ell) = \sum_{\alpha \in Z} \int_{\mathcal{C}_q^\alpha} \ell(h)\, d\mu(\ell) + \tilde{\nu},$$

where $\mathcal{C}_q^\alpha$ is the set of constraints $\{h \mapsto \nabla f_\alpha|_{q(t)} \cdot h(t)\}_{t \in f_\alpha^{-1}(\{0\})}\}$, and $\tilde{\nu}$ is a vector-valued measure supported on $\{t_0\} \cup \{t_1\}$. We can decompose this even further:

$$\int_{\mathcal{C}_q} \ell(h)\, d\mu(\ell) = \sum_{\alpha \in Z} \int_{t_0}^{t_1} \nabla f_\alpha|_{q(t)} \cdot h(t)\, d\mu_\alpha(t) + \tilde{\nu},$$

where μ_α is the restriction of μ to subsets of $\mathcal{C}_q^\alpha$ for $\alpha \in Z$. We write $\mu_\alpha(t)$ rather than $\mu_\alpha(\ell)$ since once we know α, t identifies ℓ. Hence we may regard each μ_α as a measure on $[t_0, t_1]$, which must be supported on those times such that $f_\alpha(t) = 0$. Accordingly, we may understand the bounded linear functional on $C([t_0, t_1], \mathbb{R}^d)$ given by $\int_{\mathcal{C}_q} \ell(h)\, d\mu(\ell)$ as its Riesz representation, which is

$$-\delta J_q = \sum_{\alpha \in J} \nabla f_\alpha \mu_\alpha + \tilde{\nu}. \tag{2.13}$$

Recall the convention about products of measurable functions and measures being again measures.

By Corollary 2.29 , δJ_q admits a unique extension to a bounded linear functional on $C([t_0, t_1], \mathbb{R}^d)$. Looking at Equations (2.11) and (2.13), and noting that the μ_α are

supported on the appropriate sets, we see that on $[t_0, t_1]$, we have

$$d\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q}dt + \nu = \sum_{\alpha \in Z} (\nabla f_\alpha(t))\mu_\alpha(t) + \tilde{\nu}.$$

If we restrict to $(t_0, t_1) \subset [t_0, t_1]$, then ν and $\tilde{\nu}$ vanish. We have shown the measure-theoretic Euler-Lagrange equations hold on (t_0, t_1). It is straightforward to see that as we let ϵ tend to 0 that we may arrange that the μ_α extend to measures satisfying the conclusion of the theorem.

Now the converse. Suppose that $q(t)$ satisfies the measure-theoretic Euler-Lagrange equations (2.9) where the $\{\mu_\alpha\}_{\alpha \in Z}$ are finite non-negative scalar-valued measures on (t_i, t_f), each μ_α supported on the set of times such that $f_\alpha(q(t)) = 0$. We wish to show that q is a Hamiltonian trajectory. Observe that the measure-theoretic Euler-Lagrange equations are satisfied on the intervals $[t_0, t_1] \subset (t_i, t_f)$ *modulo a measure ν supported on the endpoints*. Indeed, in the first part of the proof may be taken in reverse to show that satisfaction of the measure-theoretic Euler-Lagrange equations on $[t_0, t_1]$ modulo such a measure is equivalent to the condition for each $h \in \text{AC-BV}([t_0, t_1], \mathbb{R}^d)$,

$$\delta J_q[h] + \int \ell(h)\, d\mu(\ell) = 0,$$

for some non-negative scalar measure μ supported on $\mathcal{C}_q$. Applying Theorem 2.32, we see that q is a Hamiltonian trajectory. □

An Example

Example. We consider the one dimensional dynamical system given by the Lagrangian $L = \frac{1}{2}\dot{q}^2$ and the single inequality constraint $q \geq 0$. Here, $q = 0$ represents a "surface" and the Lagrangian is that of the usual point mass of Newtonian physics.

PROPOSITION 2.34. *The Hamiltonian trajectories in this example are precisely those trajectories that are piecewise linear with kinks only occurring at the times for which* $q = 0$.

Proof. We write the measure-theoretical Euler-Lagrange equations:

$$d\dot{q} = \mu(t),$$

where $\mu(t)$ is a measure supported on those times t for which $q(t) = 0$. Hence for times when $q(t) > 0$, we must have $\dot{q}$ remaining constant for nearby times; this results in a locally linear trajectory. If the position collides with the constraint $q(t) \geq 0$, then μ can and must deliver an impulse in order to prevent violation of $q(t) \geq 0$. If this impulse causes $\dot{q}$ to become positive, then $q(t) > 0$ thereafter. Otherwise, the final velocity $\dot{q}$ may be zero, in which case we may have $q(t) = 0$ after impact for an indeterminate amount of time. At any time during this period of rest, μ may deliver another impulse to cause $q(t) > 0$, after which no further impacts are possible. In every situation, we see that the trajectory $q(t)$ is piecewise linear with kinks only occurring at times t when $q(t) = 0$. (Indeed, only at times when $q(t)$ is either impacting or releasing from the constraint.) □

Having given our main results, we turn now to the technical proofs we have postponed. Afterwards, we give a brief summary of what we have done and discuss related work.

Technical Proofs

We now begin a long section of proofs. There are two goals: Lemma 2.45 and Lemma 2.62.

Lemma 2.45 is a fairly straightforward exercise in functional analysis. Lemma 2.62, however, is far more complicated. The essential idea is the following: rather than showing that the uniform closure of T_qM (see the statement of the lemma to see what M refers to) is the cone of $\mathcal{C}_q$ directly, we show first that the uniform closure of T_qM is the tangent cone at q of the uniform closure of M, where the tangent cone is considered in the space of continuous functions. That is, we show first that

$$\overline{T_qM} = T_q\bar{M}.$$

We also show

$$T_q\bar{M} = \text{ the cone of } \mathcal{C}_q.$$

Together, these show Lemma 2.62.

Unfortunately, these are not simple facts to prove either! We will require many tools in order to prove these set equalities.

For the remainder of this work, drop the Standing Hypothesis; from now on we will spell everything out and not refer to globally defined objects.

Functional Analysis

DEFINITION 2.35. *Suppose X is a linear space. We say that a subset S of X is a* convex *if for all $a, b \in S$, and all $\alpha \in [0,1]$,*

$$\alpha a + (1-\alpha)b \in S.$$

Given any subset S of X, we define the convex hull *of S to be the set*

$$co\, S := \{x \in X : x = \sum_{k=1}^{N} \alpha_k x_k, \text{ where } x_k \in S,\ \alpha_k \geq 0 \text{ for all } k, \text{ and } \textstyle\sum_{k=1}^{N} \alpha_k = 1\}.$$

DEFINITION 2.36. *Suppose S is a subset of a linear space. We define the* positive span of S *to be the collection of all finite combinations of elements of S with non-negative coefficients.*

DEFINITION 2.37. *Suppose X is a normed linear space. We say that X is a* locally convex topological space *if X is equipped with a Hausdorff topology such that addition is continuous, multiplication by scalars is continuous, and every open set containing the origin contains a convex open set containing the origin.*

DEFINITION 2.38. *Suppose that X is a Banach space. We define* the dual of X*, which we denote X^*, to be the Banach space of bounded linear functionals on X, with the norm*

$$|\ell| := \sup_{|x|=1} |\ell(x)| \text{ for all } \ell \in X^*.$$

We define the weak* topology *on X^* to be the coarsest topology on X^* such that the bounded linear functionals*

$$\{\rho \in X^{**} : \text{For some } x \in X, \text{ for every } \ell \in X^*,\ \rho(\ell) = \ell(x)\}$$

are all continuous.

THEOREM 2.39 (Hahn-Banach [19]). *Suppose that X is a locally convex topological vector space. If C is a convex subset of X and x is a point not in the closure of C, then there exists $c \in \mathbb{R}$ and a continuous linear functional ℓ such that $\ell(x) < c \leq \ell(C)$. Moreover, if C contains the origin then c can be taken to be* 0.

LEMMA 2.40. *Suppose X is a Banach space. Let S be a nonempty collection of bounded linear functionals on X. The dual cone of the cone of S is the weak* closure of the positive span of S.*

Proof. Suppose ℓ is not in the weak* closure of the positive span of S. We show that ℓ is not in the dual cone of the cone of S. By the definition of the weak* topology, there exists a neighborhood U of ℓ of the form

$$U = \bigcap_{k=1}^{n} \{\rho \in X^* : |\rho(x_k) - \ell(x_k)| < \epsilon\}$$

which is disjoint from the positive span of S. Observing that U is a convex set with an interior point ℓ disjoint from positive span of S (which is again convex), we appeal to the hyperplane separation (Hahn-Banach) theorem to produce $x \in X$ such that $\ell(x) < 0$ and $\rho(x) \geq 0$ for every ρ in the positive span of S. We see that the x we have produced in this way is in the cone of S. It follows then that ℓ is not in the dual cone of the cone of S. This proves the first inclusion.

Now the opposite inclusion. Suppose ℓ is in the weak* closure of the positive span of S; we show it is in the dual cone of the cone of S. Let S_* denote the cone of

S. We must show that for every $x \in S_*$, $\ell(x) \geq 0$; so let $x \in S_*$. For every $\rho \in S$, $\rho(x) \geq 0$. It follows immediately that for every ρ in the positive span of S, $\rho(x) \geq 0$. We know that every weak*-neighborhood U of ℓ intersects the positive span of S. In particular, for every $\epsilon > 0$, the neighborhood

$$U_\epsilon := \{\rho \in X^* : |\rho(x) - \ell(x)| < \epsilon\}$$

intersects the positive span of S. Hence for every $\epsilon > 0$, $\ell(x)$ is ϵ-close to some $\rho(x) \geq 0$ for some ρ in the positive span of S. Thus, $\ell(x) \in (-\epsilon, \infty)$ for each $\epsilon > 0$. It follows that $\ell(x) \geq 0$, whence ℓ is in the dual of S_*. □

Definition 2.41 (Weak Integral). *Suppose μ is a measure on some measure space Y. Suppose that X is a locally convex topological space (LCT space). Given $x \in X$ and a measurable function $f : Y \to X$, we say that*

$$x = \int f \, d\mu$$

provided that for every continuous linear functional ℓ on the space X,

$$\ell(x) = \int \ell \circ f \, d\mu.$$

Theorem 2.42 (Choquet [19]). *Let C be a convex compact subset of a locally convex topological space. Then any point $x \in C$ admits a probability measure μ supported on the extreme points of C such that*

$$x = \int e \, d\mu(e),$$

where the integral is to be interpreted in the weak sense.

THEOREM 2.43 (Alaoglu [19]). *Suppose X is the dual of a Banach space. The closed unit ball in X is compact in the weak* topology.*

LEMMA 2.44. *Suppose that $\mathcal{S}$ is a set of points in an LCT space. The closure of the positive span of $\mathcal{S}$ is equivalently given by the union of all nonnegative multiples of the closure of the convex hull of $\mathcal{S}$, that is,*

$$\overline{\operatorname{span}_+ \mathcal{S}} = \bigcup_{\lambda \geq 0} \lambda \overline{\operatorname{co} \mathcal{S}}.$$

Proof. It follows readily from the definition that the positive span of $\mathcal{S}$ is nothing more than the union of all nonnegative multiples of the convex hull of $\mathcal{S}$:

$$\operatorname{span}_+ \mathcal{S} = \bigcup_{\lambda \geq 0} \lambda \operatorname{co} \mathcal{S}.$$

Since the closure of a union contains the union of closures,

$$\bigcup_{\lambda \geq 0} \lambda \overline{\operatorname{co} \mathcal{S}} \subset \overline{\bigcup_{\lambda \geq 0} \lambda \operatorname{co} \mathcal{S}} = \overline{\operatorname{span}_+ \mathcal{S}}.$$

This proves one inclusion; we now show

$$\overline{\operatorname{span}_+ \mathcal{S}} \subset \bigcup_{\lambda \geq 0} \lambda \overline{\operatorname{co} \mathcal{S}}.$$

To this end, we show that whenever $\ell \in X^*$ is in the closure of the positive span of S, then it is also contained in the union of the nonnegative multiples of the closure of the convex hull of S.

Every neighborhood of ℓ intersects the positive span of S. We show that every neighborhood of ℓ intersects $\lambda \cdot \operatorname{co} \mathrm{S}$ for some $\lambda \geq 0$. Let U be a neighborhood of ℓ. It

intersects the positive span of S; suppose the point of intersection is $\rho = \sum_{i=1}^{n} \lambda_i \rho_i$ for positive coefficients λ_i and some ρ_i in S. Define $\lambda = \sum \lambda_i$. Observe that $\rho = \lambda \cdot \sum \frac{\lambda_i}{\lambda} \rho_i$, which is a convex combination of points in S. Hence $\rho \in \lambda \operatorname{co} \mathrm{S}$. □

LEMMA 2.45. *Suppose that $\mathcal{S}$ is a uniformly bounded set of bounded linear functionals on a Banach space X. Every point ℓ in the dual cone of the cone of $\mathcal{S}$ admits a finite measure μ supported on S that represents ℓ in the sense of Choquet, that is,*

$$\ell = \int e \, d\mu(e).$$

Proof. Let ℓ be in the dual cone of the cone of $\mathcal{S}$. By Lemma 2.40, ℓ is in the weak* closure of the positive span of S. By Lemma 2.44, we see that $\ell \in \lambda \overline{\operatorname{co} \mathcal{S}}$ for some $\lambda \geq 0$. Observe that $\overline{\operatorname{co} \mathcal{S}}$ (where the overline means weak* closure) is bounded and weak* closed; hence it is a weak* closed subset of some ball, which by Alaoglu's theorem is weak* compact. It follows that $\overline{\operatorname{co} \mathcal{S}}$ is weak* compact, being a closed subset of a compact space, and hence Choquet's theorem applies: for any $\rho \in \overline{\operatorname{co} \mathcal{S}}$ there exists a probability measure μ supported on $\mathcal{S}$ such that

$$\rho = \int e d\mu(e).$$

This integral is meant in the weak sense, where we consider X^* (the dual of X) to have the weak* topology (this makes X^* an LCT space and the definition applies). Accordingly, for every $x \in X$,

$$\rho(x) = \int e(x) d\mu(e).$$

In particular, $\ell = \lambda\rho$ for some $\rho \in \overline{\text{co}\,\mathcal{S}}$. If μ is the probability corresponding to ρ in the Choquet sense, then $\lambda\mu$, which is a finite measure, corresponds to ℓ in the Choquet manner. This is the conclusion of the lemma. □

Miscellaneous Results

The rest of the chapter is devoted to the proof of Lemma 2.62. We introduce some tools we will need for later. We begin with the inverse function theorem:

THEOREM 2.46 (Inverse Function Theorem [26]). *Suppose A is an open subset of $\mathbb{R}^d$ and that $F : A \to \mathbb{R}^d$ is a C^r function. If $DF|_q$ is non-singular for some $q \in A$, then there exist neighborhoods U and V of q and $F(q)$ respectively such that F is a C^r bijection from U to V admitting a C^r inverse.*

We will have occasion to use "bump" functions and partitions of unity in the coming work. For now, we define bump functions and point out they exist:

DEFINITION 2.47. *We say that $\phi : [t_0, t_1] \to \mathbb{R}$ is a* bump function *if it is C^∞ and takes values in the unit interval $[0, 1]$. Given any two disjoint closed subspaces A and B of $[t_0, t_1]$, it is a well-known theorem (see [26], for example) that there exists a bump function ϕ defined on this interval such that $\phi|_A = 0$ and $\phi|_B = 0$.*

The following is straightforward, and we will use it without reference:

LEMMA 2.48. *The pasting lemma applies to AC-BV functions. That is, suppose if $x \in$ AC-BV$(A, \mathbb{R}^d)$ and $y \in$ AC-BV$(B, \mathbb{R}^d)$, where A and B are closed intervals with*

a non-empty intersection. Suppose that x and y agree on $A \cap B$. Then there is an AC-BV function $z \in$ AC-BV$(A \cup B, \mathbb{R}^d)$ such that $x = z|_A$ and $y = z|_B$.

In order to understand the AC-BV space, we will occasionally need to know facts about the TV norm. The following comes in useful:

LEMMA 2.49. *Let $x, y \in BV([t_0, t_1], \mathbb{R})$. The TV norm satisfies the following:*

1. *It obeys the inequality*

$$TV[xy] \leq |x|_\infty \, TV[y] + |y|_\infty \, TV[x].$$

2. *It is submultiplicative:*

$$TV[xy] \leq TV[x] \, TV[y].$$

Proof. We show **1**. Appealing directly to the definition of the TV norm,

$$\mathrm{TV}[xy] = |x(t_0)y(t_0)| + \lim \sum |x(t_{i+1})y(t_{i+1}) - x(t_i)y(t_i)| \, .$$

Notice that

$$\begin{aligned} |x(t_{i+1})y(t_{i+1}) - x(t_i)y(t_i)| &\leq |x(t_{i+1})y(t_{i+1}) - x(t_{i+1})y(t_i)| \\ &\quad + |x(t_{i+1})y(t_i) - x(t_i)y(t_i)| \\ &\leq |x|_\infty |y(t_{i+1}) - y(t_i)| + |y|_\infty |x(t_{i+1}) - x(t_i)| \, . \end{aligned}$$

Also,

$$|x(t_0)y(t_0)| \leq |x|_\infty |y(t_0)| + |y|_\infty |x(t_0)| \, .$$

Observe that **1** follows straightforwardly upon substitution of the last two inequalities into the definition of $\mathrm{TV}[xy]$.

Now we show **2**. Consider first the case where x and y are non-negative monotone non-decreasing functions. For monotone non-decreasing functions z, it is straightforward from the definition that $\mathrm{TV}[z] = |z|_\infty = |z(t_1)|$. Then we have shown **2** for this special case (indeed, equality holds for this special case).

In the general case, we may represent x and y as the difference of continuous from the right nonnegative monotone non-decreasing functions $x^+ - x^-$ and $y^+ - y^-$ respectively in such a way that $\mathrm{TV}[x] = \mathrm{TV}[x^+] + \mathrm{TV}[x^-]$, and similarly $\mathrm{TV}[y] = \mathrm{TV}[y^+] + \mathrm{TV}[y^-]$. (This is a well known characterization of functions of bounded variation; it follows from the Jordan and Hahn decompositions of signed measures and the correspondence between BV functions and signed measures. See [8]).

We calculate:

$$\begin{aligned}
\mathrm{TV}[xy] &= \mathrm{TV}[(x^+ - x^-)(y^+ - y^-)] = \mathrm{TV}[x^+y^+ - x^+y^- - x^-y^+ + x^-y^-] \\
&\leq \mathrm{TV}[x^+y^+] + \mathrm{TV}[x^+y^-] + \mathrm{TV}[x^-y^+] + \mathrm{TV}[x^-y^-] \\
&= \big(\mathrm{TV}[x^+] + \mathrm{TV}[x^-]\big)\,\big(\mathrm{TV}[y^+] + \mathrm{TV}[y^-]\big) \\
&= \mathrm{TV}[x]\,\mathrm{TV}[y].
\end{aligned}$$

□

Remark. The above lemma holds even if x and y are vector or matrix-valued, and the product xy is interpreted in the sense of matrix multiplication. We omit the proof

of this straightforward generalization. However, to avoid ambiguity, we point out that we use the Euclidean norm $|v|$ for vectors $v \in \mathbb{R}^d$, and for matrices in $\mathrm{Mat}(\mathbb{R}^d, \mathbb{R}^{d_2})$ we use the standard operator norm (which we will again refer to as Euclidean):

$$|A| := \sup_{|x|=1} |Ax|.$$

The uniform and total variation norms of time-varying vectors and matrices can then be derived using these Euclidean norms at the fixed times.

DEFINITION 2.50. *We introduce* little-oh notation*: given two sequences of positive real numbers (a_n) and (b_n), we say*

$$a_n = o(b_n)$$

if and only if

$$\lim_{n \to 0} \frac{a_n}{b_n} = 0.$$

Taylor Theorem Results

The following four lemmas are extensions of Taylor's theorem which apply to function spaces:

LEMMA 2.51. *Suppose that $f : U \to \mathbb{R}^{d_2}$ is a C^1 mapping, where U is an open subset of $\mathbb{R}^d$. Suppose $q \in$ AC-BV$([t_0, t_1], \mathbb{R}^d)$ such that $q(t) \in U$ for all $t \in [t_0, t_1]$. Suppose (q_n) is a sequence of functions in AC-BV$([t_0, t_1], \mathbb{R}^d)$ such that $q_n(t) \in U$ for all $t \in [t_0, t_1]$. Suppose further that (q_n) tends to q in the AC-BV sense. Then*

$$Df|_{q(t)}(q_n(t) - q(t)) = f(q_n(t)) - f(q(t)) + E_n(t),$$

where functions $E_n : [t_0, t_1] \to \mathbb{R}^{d_2}$ *satisfy*

$$\lim_{n\to\infty} \frac{|E_n(t)|}{||q_n - q||} = 0,$$

where the limit is meant in the L_1 *sense:*

$$\int_{t_0}^{t_1} |E_n| dt = o(||q_n - q||).$$

Proof. Assume first that $d_2 = 1$.

By the definition of differentiability,

$$\lim_{n\to 0} \frac{|E_n(t)|}{|q_n(t) - q(t)|} = 0 \text{ for all } t \in [t_0, t_1].$$

It is straightforward to verify that

$$\frac{|E_n(t)|}{|q_n(t) - q(t)|} \leq \left|\nabla f|_{q(t)}\right| + \frac{|f(q_n(t)) - f(q(t))|}{|q_n(t) - q(t)|}.$$

Since f is C^1, it is also Lipschitz continuous (at least when restricted to some neighborhood of $\{q(t) : t \in [t_0, t_1]\}$ within which a tail end of the sequence (q_n) take their values), hence both terms of the above inequality are bounded. Thus, $\frac{|E_n(t)|}{|q_n(t)-q(t)|}$ is bounded by a constant. Consequently, the Lebesgue dominated convergence theorem applies, and we have that

$$\lim_{n\to\infty} \int_{t_0}^{t_1} \frac{|E_n(t)|}{|q_n(t) - q(t)|} \, dt = 0.$$

Since $|q_n(t) - q(t)| \leq ||q_n - q||$, we have

$$\frac{|E_n(t)|}{||q_n - q||} \leq \frac{|E_n(t)|}{|q_n(t) - q(t)|} \text{ for all } t \in [t_0, t_1].$$

It follows from the monotonicity of integrals that

$$\lim_{n\to\infty} \int_{t_0}^{t_1} \frac{|E_n(t)|}{||q_n - q||}\, dt = 0,$$

which is the conclusion we seek.

Now suppose that $d_2 > 1$. The function f may be decomposed into f_k, $k = 1, 2, \cdots, d_2$, such that $f_k := \pi_k \circ f$. (Here π_k is the canonical projection onto the kth coordinate.) The above argument applies to each of the f_k, and it is straightforward to see that the result then extends to f. □

LEMMA 2.52. *Suppose that $f : U \to \mathbb{R}^{d_2}$ is a C^2 mapping, where U is an open subset of $\mathbb{R}^d$. Suppose $q \in$ AC-BV$([t_0, t_1], \mathbb{R}^d)$ such that $q(t) \in U$ for all $t \in [t_0, t_1]$. Suppose (q_n) is a sequence of functions in AC-BV$([t_0, t_1], \mathbb{R}^d)$ such that $q_n(t) \in U$ for all $t \in [t_0, t_1]$. Suppose further that (q_n) tends to q in the AC-BV sense. Then*

$$Df|_{q(t)}(q_n(t) - q(t)) = f(q_n(t)) - f(q(t)) + E_n(t),$$

where functions $E_n : [t_0, t_1] \to \mathbb{R}^d$ satisfy

$$\lim_{n\to\infty} \frac{E_n(t)}{||q_n - q||} = 0,$$

where the limit is meant in the BV sense:

$$TV[E_n] = o(||q_n - q||).$$

Proof. Assume first that $d_2 = 1$.

Note that $E_n(t_0) = 0$ and E_n is AC-BV. We may bound TV$[E_n]$ using the formula

$$\text{TV}[E_n] \leq \int_{t_0}^{t_1} |\dot{E}_n(t)| dt.$$

We can give an expression for $\dot{E}_n(t)$:

$$\dot{E}_n(t) = \dot{q}^T(t) Hf|_{q(t)}(q_n(t) - q(t)) - (Df_{q_n(t)} - Df_{q(t)})\dot{q}_n(t)$$

Using Lemma 2.51, we have

$$Df_{q_n(t)} - Df_{q(t)} = Hf|_{q(t)}(q_n(t) - q(t)) + G_n,$$

where $|G_n|_1 = o(||q_n - q||)$. Substituting this into the expression for $\dot{E}_n$ leads to:

$$\dot{E}_n(t) = (\dot{q}(t) - \dot{q}_n(t))^T Hf|_{q(t)}(q_n(t) - q(t)) + G_n\dot{q}_n.$$

We may thus bound $|\dot{E}_n|$:

$$\int_{t_0}^{t_1} |\dot{E}_n| dt \leq (t_1 - t_0)|\dot{q} - \dot{q}_n|_\infty \left|Hf|_{q(t)}\right|_\infty |q_n - q|_\infty + |\dot{q}_n|_\infty \int_{t_0}^{t_1} |G_n| dt.$$

Observe that $|\dot{q}_n|_\infty$ is bounded, $\left|Hf|_{q(t)}\right|_\infty$ is bounded, $|\dot{q} - \dot{q}_n|_\infty$ tends to zero, and $|q_n - q|_\infty \leq ||q - q_n||$. Also, notice $\int |G_n| dt = o(||q_n - q||)$. Hence, both terms are $o(||q - q_n||)$ and the conclusion of the lemma follows.

Now suppose that $d_2 > 1$. The function f may be decomposed into f_k, $k = 1, 2, \cdots, d_2$, such that $f_k := \pi_k \circ f$. The above argument applies to each of the f_k, and it is straightforward to see that the result then extends to f. □

LEMMA 2.53. *Suppose that $F : U \to \mathbb{R}^{d_2}$ is a C^3 mapping, where U is an open subset of $\mathbb{R}^d$. Suppose $q \in$ AC-BV$([t_0, t_1], \mathbb{R}^d)$ such that $q(t) \in U$ for all $t \in [t_0, t_1]$. Suppose (q_n) is a sequence of functions in AC-BV$([t_0, t_1], \mathbb{R}^d)$ such that $q_n(t) \in U$ for*

all $t \in [t_0, t_1]$. Suppose further that (q_n) tends to q in the AC-BV sense. Then

$$DF|_{q(t)}(q_n(t) - q(t)) = F(q_n(t)) - F(q(t)) + E_n(t),$$

where functions $E_n : [t_0, t_1] \to \mathbb{R}^{d_2}$ satisfy

$$\lim_{n\to\infty} \frac{E_n(t)}{||q_n - q||} = 0,$$

where the limit is meant in the AC-BV sense:

$$||E_n|| = o(||q_n - q||).$$

Proof. First we we prove the special case of $d_2 = 1$. Define

$$\phi(x, y) := DF|_x(y - x) - [F(y) - F(x)]. \tag{2.14}$$

See that ϕ is C^2 in both of its arguments.

Observe that $E_n(t) = \phi(q_n(t), q(t))$. We show that $||E_n|| = o(||q_n - q||)$. By the definition of the AC-BV norm,

$$||E_n|| = |E_n(t_0)| + \mathrm{TV}[\dot{E}_n].$$

It is straightforward to see that $|E_n(t_0)| = o\left(||q_n - q||\right)$. We show that $\mathrm{TV}[\dot{E}_n] = o\left(||q_n - q||\right)$, which then demonstrates $||E_n|| = o(||q_n - q||)$.

By the chain rule,

$$\dot{E}_n(t) = \phi_x(q_n(t), q(t)) \cdot \dot{q}_n(t) + \phi_y(q_n(t), q(t)) \cdot \dot{q}(t).$$

Taking partial derivatives of (2.14), we obtain

$$\begin{aligned} \phi_x(x,y) &= (HF)|_x(y-x), \\ \phi_y(x,y) &= DF|_x - DF|_y. \end{aligned}$$

Here HF denotes the Hessian, which is the symmetric matrix of second derivatives. Using these formulas, we write

$$\dot{E}_n(t) = \big((HF)|_{q(t)}(q_n(t)-q(t))\big)\cdot\dot{q}_n(t) + \big(DF|_{q(t)} - DF|_{q_n(t)}\big)\cdot\dot{q}(t).$$

Since $DF|_q$ is C^2 in q, we apply Lemma 2.52 to obtain

$$DF|_{q(t)} - DF|_{q_n(t)} = HF|_{q(t)}(q(t)-q_n(t)) + G_n(t),$$

where $\mathrm{TV}[G_n] = o\left(||q_n - q||\right)$.

Substituting this yields

$$\dot{E}_n(t) = \big((HF)|_{q(t)}(q_n(t)-q(t))\big)\cdot(\dot{q}_n(t)-\dot{q}(t)) + G_n(t)\cdot\dot{q}(t). \tag{2.15}$$

We show the two terms on the right-hand side of Equation (2.15) both have total variations which are $o\left(||q_n - q||\right)$. Once this is accomplished, it follows that $||E_n|| = o(||q_n - q||)$, and the result will be proven.

We show $\mathrm{TV}[\big((HF)|_{q(t)}(q_n(t)-q(t))\big)\cdot(\dot{q}_n(t)-\dot{q}(t))] = o\left(\mathrm{TV}[\dot{q}_n-\dot{q}]\right)$. We calculate using Lemma 2.49:

$$\begin{aligned}\mathrm{TV}\left[\big((HF)|_{q(t)}(q_n(t)-q(t))\big)\cdot(\dot{q}_n(t)-\dot{q}(t))\right] \leq \\ \mathrm{TV}\left[HF|_{q(t)}\right]\ |q_n-q|_\infty\,|\dot{q}_n-\dot{q}|_\infty + \\ \mathrm{TV}\left[q_n-q\right]\ \left|HF|_{q(t)}\right|_\infty\,|\dot{q}_n-\dot{q}|_\infty + \\ \mathrm{TV}\left[\dot{q}_n-\dot{q}\right]\ \left|HF|_{q(t)}\right|_\infty\,|q_n-q|_\infty\end{aligned}$$

Each of these three terms has a factor which is bounded by $||q_n-q||$ and a different factor which is $o(1)$.

We show $\mathrm{TV}[G_n(t)\cdot\dot{q}(t)] = o\left(\mathrm{TV}[\dot{q}_n-\dot{q}]\right)$. Again, we calculate:

$$\mathrm{TV}[G_n(t)\cdot\dot{q}(t)] \leq \mathrm{TV}[G_n]|\dot{q}|_\infty + \mathrm{TV}[\dot{q}]\,|G_n|_\infty\,.$$

Since there are no unbounded factors, $|G_n| \leq \mathrm{TV}[G_n]$, and $\mathrm{TV}[G_n] = o(||q_n-q||)$, it follows that this term as well is $o\left(||q_n-q||\right)$. We are done.

Now suppose that $d_2 > 1$. The function f may be decomposed into f_k, $k = 1, 2, \cdots, d_2$, such that $f_k := \pi_k \circ f$. The above argument applies to each of the f_k, and it is straightforward to see that the result then extends to f. □

We also need a version of the preceding lemma for the continuous case:

LEMMA 2.54. *Suppose that $F : U \to \mathbb{R}^{d_2}$ is a C^2 mapping, where U is an open subset of $\mathbb{R}^d$. Suppose $q \in C([t_0,t_1],\mathbb{R}^d)$ such that $q(t) \in U$ for all $t \in [t_0,t_1]$. Suppose (q_n) is a sequence of functions in $C([t_0,t_1],\mathbb{R}^d)$ such that $q_n(t) \in U$ for all $t \in [t_0,t_1]$ and*

$q_n(t_0) = q(t_0)$ and $q_n(t_1) = q(t_1)$ for all n. Suppose further that (q_n) tends to q in the uniform sense. Then

$$DF|_{q(t)}(q_n(t) - q(t)) = F(q_n(t)) - F(q(t)) + E_n(t),$$

where functions $E_n : [t_0, t_1] \to \mathbb{R}^{d_2}$ satisfy

$$\lim_{n\to\infty} \frac{E_n(t)}{|q_n - q|_\infty} = 0,$$

where the limit is meant in the uniform sense:

$$|E_n|_\infty = o(|q_n - q|_\infty).$$

Proof. First, we remark that all the limits in this proof are to be meant in the uniform sense. Now we show the result. Because F is C^2, we may bound $E_n(t)$ as follows:

$$|E_n(t)| \leq \kappa \, |q_n(t) - q(t)|^2,$$

where $\kappa > 0$ may be chosen independently of n and t. Accordingly,

$$\lim \frac{|E_n|}{|q_n - q|_\infty} \leq \lim_{n\to\infty} \frac{\kappa |q_n - q|^2}{|q_n - q|_\infty} \leq \kappa \lim_{n\to\infty} |q_n - q| = 0,$$

where we have used the fact that $\frac{|q_n - q|}{|q_n - q|_\infty} \leq 1$ and also the fact that $|q_n - q|$ converges uniformly to 0. □

Continuity Results

The next two lemmas give conditions under which we can transform a convergent sequence in one AC-BV space to a convergent sequence in another AC-BV space.

Both results have continuous versions as well, which describe when we may transform a convergent sequence in a C space to another C space.

LEMMA 2.55. *Suppose that $F : U \to \mathbb{R}^{d_2}$ is a C^2 function, where U is an open subset of $\mathbb{R}^d$. Then the map*

$$\mathcal{F} : \textit{AC-BV}([t_0, t_1], U) \to \textit{AC-BV}([t_0, t_1], \mathbb{R}^{d_2}),$$

defined by

$$\mathcal{F}[x](t) := F(x(t)), \text{ for all } t \in [t_0, t_1],$$

is well defined.

Moreover, if F is C^3, then $\mathcal{F}$ is continuous with respect to the AC-BV metric. If, additionally, $q \in \textit{AC-BV}([t_0, t_1], U)$, then the time-varying matrix-valued function $M(t)$ defined as

$$M(t) = DF|_{q(t)}$$

is an AC-BV function.

This lemma also admits a continuous version*, with all AC-BV spaces replaced with C spaces and all AC-BV convergence replaced with uniform convergence.*

Proof. Let $x \in \text{AC-BV}([t_0, t_1], \mathbb{R}^d)$. We show that $\mathcal{F}[x]$ is AC-BV. To this end we check to see that the AC-BV norm of $\mathcal{F}[x]$ is finite. This will be the case provided that

$$\text{TV}\left[\frac{d}{dt}\mathcal{F}[x]\right] < \infty.$$

The chain rule applies, and we may write

$$\frac{d}{dt}\mathcal{F}[x] = DF|_{x(t)}\dot{x}.$$

By Lemma 2.49,

$$\mathrm{TV}\left[\frac{d}{dt}\mathcal{F}[x]\right] \leq \mathrm{TV}\left[DF|_{x(t)}\right]\ \mathrm{TV}\left[\dot{x}\right].$$

Since we know that $\dot{x}$ has bounded variation (as x is AC-BV), it follows that $\mathcal{F}[x]$ is AC-BV provided that $\mathrm{TV}\left[DF|_{x(t)}\right] < \infty$.

Claim. Suppose $f : U \to \mathbb{R}^{d_2}$ is C^1 and $z : [t_0, t_1] \to \mathbb{R}^d$ is BV. Then $f \circ z$ is BV.

We are assuming of course that the composition exists, i.e. z is U-valued. It is well known that C^1 functions restricted to compact domains are Lipschitz continuous. Accordingly, define $L > 0$ so that $|f(x) - f(y)| \leq L|x - y|$ holds for all $x, y \in \{z(t) : t \in [t_0, t_1]\}$. Now we calculate $\mathrm{TV}[f \circ z]$ from the definition:

$$\lim \sum |f(z(t_{i+1})) - f(z(t_i))| \leq \lim \sum L|z(t_{i+1}) - z(t_i)| = L\ \mathrm{TV}[z].$$

This is finite since z is BV. *This completes the proof of the claim.*

Now we apply the claim to the case of $f(z) := DF|_z$ and $z(t) := x(t)$. We find that $DF|_{x(t)}$ has bounded variation, and by our previous reasoning, $\mathcal{F}$ is well-defined.

Now the moreover part. Assume F is C^3. We show $\mathcal{F}$ is continous. To this end we show $\mathcal{F}$ maps convergent sequences to convergent sequences, which suffices to show continuity for maps between metric spaces. Accordingly, let (q_n) be a sequence

of AC-BV functions converging to q in the AC-BV sense. We show that

$$||\mathcal{F}[q_n] - \mathcal{F}[q]|| \to 0 \text{ as } n \to \infty.$$

Appealing to the definition of the AC-BV norm,

$$||\mathcal{F}[q_n] - \mathcal{F}[q]|| = |F(q_n(t_0)) - F(q(t_0))| + \text{TV}\left[DF|_{q_n(t)}\dot{q}_n - DF|_{q(t)}\dot{q}\right].$$

The first term on the right-hand side clearly tends to zero. We manipulate the second term to obtain the following:

$$\begin{aligned} \text{TV}\left[DF|_{q_n(t)}\dot{q_n} - DF|_{q(t)}\dot{q}\right] \;\leq\; & \text{TV}\left[(DF|_{q_n(t)} - DF|_{q(t)})\dot{q_n}\right] \\ & + \text{TV}\left[DF|_{q(t)}(\dot{q}_n - \dot{q})\right]. \end{aligned} \tag{2.16}$$

We show that the two terms on the right-hand side tend to zero as $n \to \infty$. First we consider the first term on the right-hand side of (2.16), which is $\text{TV}\left[(DF|_{q_n(t)} - DF|_{q(t)})\dot{q_n}\right]$. By Lemma 2.49,

$$\text{TV}\left[(DF|_{q_n(t)} - DF|_{q(t)})\dot{q_n}\right] \leq \text{TV}\left[DF|_{q_n(t)} - DF|_{q(t)}\right] \text{TV}\left[\dot{q_n}\right].$$

Note that DF is C^2. By Lemma 2.52, we have

$$DF|_{q_n(t)} - DF|_{q(t)} = HF|_{q(t)}(q_n(t) - q(t)) + E_n, \tag{2.17}$$

where HF is C^1 and $\text{TV}[E_n] = o(||q_n - q||)$.

The first term on the right-hand side of (2.17) has total variation tending to zero:

$$\text{TV}[HF|_{q(t)}(q_n(t) - q(t))] \leq \text{TV}[HF|_{q(t)}] \,\text{TV}[q_n(t) - q(t)] \to 0 \text{ as } n \to \infty.$$

The second term on the right hand side of (2.17) has total variation tending to zero, since $\mathrm{TV}[E_n] = o(||q_n - q||)$ and $||q_n - q|| \to 0$ as $n \to \infty$.

We conclude that the first term on the right-hand side of (2.16) tends to zero. Now we consider the second term on the right hand side of (2.16), which is given by $\mathrm{TV}\left[DF|_{q(t)}(\dot{q}_n - \dot{q})\right]$. Observe that

$$\mathrm{TV}\left[DF|_{q(t)}(\dot{q}_n - \dot{q})\right] \leq \mathrm{TV}\left[DF|_{q(t)}\right] \mathrm{TV}\left[\dot{q}_n - \dot{q}\right] \to 0 \text{ as } n \to \infty.$$

Since both terms on the right-hand side of (2.16) tend to zero, it follows that $||\mathcal{F}[q_n] - \mathcal{F}[q]||$ tends to zero. Accordingly, $(\mathcal{F}[q_n])$ converges to $\mathcal{F}[q]$ in the AC-BV sense, and we conclude that $\mathcal{F}$ is continuous.

Now we show the second statement of the moreover part. Since F is now C^3, DF is C^2, and we may apply the result about well-defined mapping between AC-BV spaces to the mapping

$$\mathcal{G}[z] = DF|_z, \text{ for } z \in U.$$

The hypotheses are met and $DF|_{x(t)}$ is AC-BV follows.

Now we prove the continuous version of the lemma. The well-defined part of the lemma follows easily from the well known fact that the composition of continuous functions are again continuous. The continuity part of the lemma goes as follows:

$$|\mathcal{F}[q_n] - \mathcal{F}[q]|_\infty \leq |F(q_n(t)) - F(q(t))|_\infty.$$

Apply Lemma 2.54 to obtain

$$F(q_n(t)) - F(q(t)) = DF|_{q(t)}(q_n(t) - q(t)) + E_n(t),$$

where $|E_n|_\infty = o(|q_n - q|_\infty)$. We find that

$$|\mathcal{F}[q_n] - \mathcal{F}[q]|_\infty \leq |DF|_{q(t)}|_\infty \, |q_n - q|_\infty + o(|q_n - q|_\infty),$$

which (as is straightforward to see) tends to zero as $n \to \infty$. Notice in this step we are using the fact that $DF|_{q(t)}$ is continuous (and hence bounded) as it is a composition of continuous functions. □

LEMMA 2.56. *Suppose that $M : [t_0, t_1] \to Mat(\mathbb{R}^d, \mathbb{R}^{d_2})$ is AC-BV. Then the operator*

$$F : AC\text{-}BV([t_0, t_1], \mathbb{R}^d) \to AC\text{-}BV([t_0, t_1], \mathbb{R}^{d_2})$$

given by

$$F[x](t) = M(t)[x(t)]$$

is continuous. Moreover, a continuous version *of this lemma holds: we may replace the AC-BV spaces with the spaces of continuous functions C, while replacing the AC-BV convergence notions with uniform convergence.*

Remark. Notice that it follows that If M is as in the lemma, and $(q_n) \to q$ in the AC-BV sense, then

$$M \lim_{n\to\infty} q_n = \lim_{n\to\infty} Mq_n,$$

where the limit is meant in the AC-BV sense.

Proof. Consider first the case where $d_2 = 1$. In this case we write $M(t)x = \ell(t) \cdot x$, where $\ell \in \text{AC-BV}([t_0, t_1], \mathbb{R}^d)$.

Our method is as follows. First, notice that it is not immediately clear that F is well-defined, since we must see that F is actually AC-BV valued. However, it is certainly C-valued, and as a function from AC-BV functions to C functions it is straightforward to see that F is linear. In order to see that F does indeed have AC-BV outputs and is continuous, it suffices to show that F is bounded (see, for example, [19]).

$$|\ell(t_0) \cdot x(t_0)| + \text{TV}[\dot{\ell}(t) \cdot x(t) + \ell(t) \cdot \dot{x}(t)]$$

In order to proceed, we need to observe

$$TV[x] \leq (t_1 - t_0) TV[\dot{x}].$$

We calculate using Lemma 2.49:

$$\begin{aligned}
||\ell \cdot x|| &= |\ell(t_0) \cdot x(t_0)| + TV\left[\frac{d}{dt}(\ell \cdot x)\right] \\
&\leq |\ell(t_0) \cdot x(t_0)| + \text{TV}[\dot{\ell}] TV[x] + TV[\ell] TV[\dot{x}] \\
&\leq |\ell(t_0)|\,|x(t_0)| + \left(\text{TV}[\dot{\ell}](t_1 - t_0) + TV[\ell]\right) TV[\dot{x}] \\
&\leq \left(|\ell(t_0)| + \text{TV}[\dot{\ell}](t_1 - t_0) + TV[\ell]\right) (|x(t_0)| + TV[\dot{x}]) \\
&\leq C\,||x||, \text{ for some } C > 0.
\end{aligned}$$

It follows that F is bounded (and hence well defined).

Now suppose $d_2 > 1$. We can decompose F into d_2 parts, $\ell_k = \pi_k \circ F$, $k = 1, 2, \cdots, d_2$, each of which the previous argument applies to. F can then be realized

as a cartesian product of the ℓ_k, and continuity is preserved. The details of this argument are straightforward, and we omit them.

Now we show the the continuous version. The proof is similar, except the proof that F is bounded goes as follows:

$$|Fx|_\infty = |\ell \cdot x|_\infty \leq |\ell|_\infty \, |x|_\infty.$$

□

Tangent Cone Results

Now have have sufficient tools to tackle the main problems. Before continuing, we give a brief outline of what follows. We wish to prove Lemma 2.24 (which is Lemma 2.62 below). A preliminary step is to show that the uniform closure of a tangent cone at a point q in a closed subset M (the assumptions on M we spell out below) of an AC-BV space is the tangent cone at q of the uniform closure of M– that is, $\overline{T_q M} = T_q \bar{M}$ – this is Lemma 2.61. In order to show Lemma 2.61, we first prove a simplified version of it for which the set M is very nice: this is Lemma 2.60. In order to use Lemma 2.60 to show Lemma 2.61, we need to be able to straighten out the set M so that it looks "nice." This cannot be done except locally; we must break time into small chunks and have a different straightening-out transformation for each chunk. To formalize such an argument requires a partition of unity approach; the tool we construct for this purpose is Lemma 2.58. This lemma makes extensive use of the results above (the Taylor theorem and continuity results), and also requires the

ability to relate tangent cones between spaces where one is a time-domain restriction of the other. Lemma 2.57 serves this purpose. We now proceed.

The next lemma gives us a relationship between tangent cones in closed subsets of function spaces and the tangent cones of domain restrictions of those subsets:

LEMMA 2.57. *Suppose $\mathcal{Q}$ is a closed subset of $\mathbb{R}^d$. Suppose $a, b \in \mathbb{R}^d$. Define*

$$M := \{p \in AC\text{-}BV([t_0, t_1], \mathcal{Q}) : p(t_0) = a \text{ and } p(t_1) = b\}.$$

Let $q \in M$. Suppose $U := (s_0, s_1) \subset [t_0, t_1]$. Define

$$\tilde{M} := \{p \in AC\text{-}BV([s_0, s_1], \mathcal{Q}) : p(s_0) = q(s_0) \text{ and } p(s_1) = q(s_0)\}.$$

Suppose that (q_n) is a sequence in M. Suppose for each n that $q_n - q$ is supported on $\bar{U}$. Define $\tilde{q} = q|_{\bar{U}}$ and $\tilde{q}_n = q_n|_{\bar{U}}$. Then we have the following:

1. *The sequence $(\tilde{q}_n)$ and the function $\tilde{q}$ are in $\tilde{M}$.*

2. *The sequence (q_n) converges to q in the AC-BV sense in M if and only if the sequence $(\tilde{q}_n)$ converges to $\tilde{q}$ in the AC-BV sense in $\tilde{M}$.*

3. *We have the following relationship of tangent cones:*

   $$T_{\tilde{q}}\tilde{M} \subset \{\tilde{v} : \tilde{v} = v|_{[s_0,s_1]} \text{ for some } v \in T_qM \text{ such that } v \text{ is supported on } [s_0, s_1]\}.$$

4. *Assume $\mathcal{Q}$ is convex. Suppose $v \in T_qM$, and that v is supported on $[s_0, s_1]$. Suppose further that v satisfies the* inner endpoint conditions*:*

   $$\text{Either } s_0 = t_0, \text{ or else } \lim_{t \to s_0^+} \dot{v}(t) = 0; \text{ and}$$

   $$\text{either } s_1 = t_1, \text{ or else } \lim_{t \to s_1^-} \dot{v}(t) = 0.$$

Then

$$v|_{[s_0,s_1]} \in T_{\tilde{q}}\tilde{M}.$$

Moreover, this lemma admits a continuous version in which all AC-BV spaces are replaced with spaces of continuous functions C, AC-BV convergence is replaced with uniform convergence, and the inner endpoint conditions of (4) are omitted.

Proof. **1** is immediate. **2** follows straightforwardly from the observation that, for all n, $||q_n - q||_M = ||\tilde{q}_n - \tilde{q}||_{\tilde{M}}$.

We show **3**. Suppose that $\tilde{v} \in T_{\tilde{q}}\tilde{M}$. Define

$$v := \begin{cases} \tilde{v} & t \in [s_0, s_1] \\ 0 & \text{otherwise} \end{cases} . \tag{2.18}$$

Clearly v is supported on $[s_0, s_1]$; we show that $v \in T_q M$. Consider first the special case that $\tilde{v}$ is a unit geometric tangent vector. Then there exists a sequence $(\tilde{p}_n)$ in $\tilde{M}$ converging to $\tilde{q}$ in the AC-BV sense such that

$$\tilde{v} = \lim_{n\to\infty} \frac{\tilde{p}_n - \tilde{q}}{||\tilde{p}_n - \tilde{q}||},$$

where this limit and others in this proof are meant in the AC-BV sense.

Extend $\tilde{p}_n$ to $p_n \in M$ via

$$p_n := \begin{cases} \tilde{p}_n & t \in [s_0, s_1] \\ q & \text{otherwise} \end{cases} .$$

By a straightforward pasting argument, (p_n) is a sequence in M converging to q. Now we may apply **2** of this lemma to see that

$$v = \lim_{n\to\infty} \frac{p_n - q}{||\tilde{p}_n - \tilde{q}||},$$

from which it follows that $v \in T_q M$.

For the general case, we proceed as follow. Since $\tilde{v} \in T_{\tilde{q}}\tilde{M}$, it is in the AC-BV closure of the positive span of the unit geometric tangent vectors of $T_{\tilde{q}}\tilde{M}$. Accordingly, choose $\epsilon > 0$ and pick a finite set of unit geometric tangent vectors $\tilde{v}^\gamma$, $\gamma \in F$, as well as non-negative coefficients a^γ such that

$$\left|\left| \tilde{v} - \sum_{\gamma \in F} a^\gamma \tilde{v}^\gamma \right|\right| < \epsilon.$$

Extend $\tilde{v}$ and $\tilde{v}^\gamma$ for $\gamma \in F$ as in (2.18) via having the functions vanish where they were not previously defined. We obtain v and v^γ for $\gamma \in F$. By the special case, we know that each v^γ, $\gamma \in F$ is in $T_q M$. Since $T_q M$ is a cone, it follows that $\sum_{\gamma \in F} a^\gamma v^\gamma \in T_q M$. Now observe

$$\left|\left| q - \sum_{\gamma \in F} a^\gamma v^\gamma \right|\right| = \left|\left| \tilde{v} - \sum_{\gamma \in F} a^\gamma \tilde{v}^\gamma \right|\right| < \epsilon.$$

Accordingly, v is within ϵ distance to the tangent cone $T_q M$ as measured by the AC-BV metric. Since $\epsilon > 0$ was arbitrarily small, and $T_q M$ is closed in the AC-BV topology, it follows that $v \in T_q M$.

We show **4**. Now we may assume $\mathcal{Q}$ is convex. Suppose that $v \in T_q M$ is supported on $[s_0, s_1]$. Also assume the endpoint conditions given in **4**. We show that the restriction $\tilde{v} = v|_{[s_0,s_1]}$ is in $T_{\tilde{q}}\tilde{M}$.

Suppose without loss that v is a unit geometric tangent vector. This is without loss by an argument similar to the one just made. Since v is a unit geometric tangent

vector, there exists a sequence (p_n) in M converging to q such that

$$v = \lim_{n\to\infty} \frac{p_n - q}{||p_n - q||}.$$

We proceed in two steps. For the first step, we show that for any C^∞ bump function ϕ supported on $[s_0, s_1]$, we have $\phi|_{[s_0,s_1]}\tilde{v} \in T_{\tilde{q}}\tilde{M}$. For the second step, we show there exists a sequence of such bump functions ϕ_n supported on $[s_0, s_1]$ such that $\phi_n|_{[s_0,s_1]}\tilde{v}$ converges to $\tilde{v}$ in the AC-BV sense. Since $T_{\tilde{q}}\tilde{M}$ is closed, these two steps imply that $\tilde{v} \in T_{\tilde{q}}$, which proves the converse of **3**.

Step 1. Let $\phi \in C^\infty([t_0, t_1], [0, 1])$ such that ϕ is supported on $[s_0, s_1]$. We show $\phi|_{[s_0,s_1]}\tilde{v} \in T_{\tilde{q}}\tilde{M}$. By Lemma 2.56, we obtain

$$\phi v = \lim_{n\to\infty} \frac{\phi(t)(p_n(t) - q(t))}{||p_n - q||}.$$

Manipulation reveals

$$\phi v = \lim_{n\to\infty} \frac{[\phi(t)p_n(t) + (1 - \phi(t))q(t)] - q(t)}{||p_n - q||}.$$

Since $\phi(t)p_n(t) + (1 - \phi(t))q(t)$ is a convex combination of functions in M, and M is convex (since $\mathcal{Q}$ is convex), it follows that $r_n(t) := \phi(t)p_n(t) + (1 - \phi(t))q(t)$ is a sequence in M tending to q in the AC-BV sense. Let $\tilde{r}_n$ denote the restriction of r_n to $[s_0, s_1]$.

By part **2**,

$$\phi|_{[s_0,s_1]}\tilde{v} = \lim_{n\to\infty} \frac{\tilde{r}_n(t) - \tilde{q}(t)}{||p_n - q||}.$$

Thus $\phi\tilde{v} \in T_{\tilde{q}}\tilde{M}$. *This completes the proof of Step 1.*

Step 2. We show there exists a sequence of functions $\phi_n \in C^\infty([t_0, t_1], [0, 1])$ each supported on $[s_0, s_1]$ such that $\phi_n|_{[s_0,s_1]}\tilde{v}$ converges to $\tilde{v}$ in the AC-BV sense.

There are four possibilities to consider. These possibilities depend on whether $s_0 = t_0$ and whether $s_1 = t_1$. We prove one of these four cases. The other three cases are similar. The case we show is $s_0 > t_0$ and $s_1 = t_1$.

Choose ϕ_n to be a bump function supported on $[s_0, s_1]$ such that $\phi(s_0) = 0$ and ϕ is unity on the interval $[s_0 + \frac{1}{n}, s_1]$. (Take n sufficiently high if necessary.) Arrange so ϕ_n is rising on $[s_0, s_0 + \frac{1}{n}]$.

Now we calculate:

$$||\phi_n\tilde{v} - \tilde{v}|| \quad = \quad ||(\phi_n - 1)\tilde{v}|| = ||(\phi_n - 1)\tilde{v}||_{[s_0,s_0+\frac{1}{n}]}.$$

By the definition of the AC-BV norm, it is straightforward to derive

$$||(\phi_n - 1)\tilde{v}|| = \mathrm{TV}[\dot{\phi}_n\tilde{v} + (\phi_n - 1)\dot{\tilde{v}}].$$

Using Lemma 2.49, in particular the submultiplicative property of the TV norm,

$$\mathrm{TV}[\dot{\phi}_n\tilde{v} + (\phi - 1)\dot{\tilde{v}}] \leq \mathrm{TV}[\dot{\phi}_n]\,\mathrm{TV}[\tilde{v}] + \mathrm{TV}[\phi_n - 1]\,\mathrm{TV}[\dot{\tilde{v}}].$$

On the interval $[s_0, s_0 + \frac{1}{n}]$, ϕ rises from 0 to 1. It is straightforward to see that it can be arranged so that $\mathrm{TV}[\dot{\phi}_n] \leq Cn$, for some $C > 0$. (Indeed, this can be arranged for any $C > 1$.) It is also easy to see that $\mathrm{TV}[\phi_n - 1] = 1$, since $\phi_n - 1$ is a monotonic function whose endpoint vaues are -1 and 0 at $t = s_0$ and $t = s_0 + \frac{1}{n}$ respectively.

The total variations $\mathrm{TV}[\tilde{v}]$ and $\mathrm{TV}[\dot{\tilde{v}}]$ vary as n increases since we mean to take the total variation over the ever-shrinking interval $[s_0, s_0 + \frac{1}{n}]$. We point out that

$$\mathrm{TV}[\tilde{v}] \leq \int_{s_0}^{s_0+\frac{1}{n}} |\dot{\tilde{v}}|\, dt.$$

Since $\lim_{t \to s_0^+} \dot{\tilde{v}} = 0$, it follows that $\int_{s_0}^{s_0+\frac{1}{n}} |\dot{\tilde{v}}|\, dt = o\left(\frac{1}{n}\right)$. Also, because $\dot{\tilde{v}}$ is of bounded variation (and recall that we mean only continuous from the right functions of bounded variation), $\mathrm{TV}[\dot{\tilde{v}}]$ tends to 0 as we let $n \to \infty$, as we mean for the total variation norm to be computed over the shrinking interval $[s_0, s_0 + \frac{1}{n}]$.

Combining these facts, we have

$$||\phi_n \tilde{v} - \tilde{v}|| = o\left(\frac{1}{n}\right),$$

and hence $\phi_n \tilde{v}$ indeed converges to $\tilde{v}$ in the AC-BV topology. This completes the proof of **4**.

Now we prove the continuous version of the lemma. In fact, the above proofs of **1**, **2**, and **3** hold for the continuous version if we replace all instances of AC-BV spaces with C spaces, all AC-BV convergence notions AC-BV and norms with uniform convergence and uniform norms. The proof of **4** also carries over in this fashion, except the proof of Step 2 needs to be replaced with the following text:

Step 2. We show there exists a sequence of functions $\phi_n \in C^\infty([t_0, t_1], [0, 1])$, each supported on $[s_0, s_1]$, such that $\phi_n|_{[s_0,s_1]} \tilde{v}$ converges to $\tilde{v}$ uniformly.

There are four possibilities, depending on whether $s_0 = t_0$ and whether $s_1 = t_1$. These four possibilities lead to four similar proofs; we only show one such possibility, where $s_0 > t_0$ and $s_1 = t_1$. The proofs of the other three cases are similar.

Choose ϕ_n to be a bump function which is 0 at s_0 and unity on $[s_0 + \frac{1}{n}, s_1]$. (Take n sufficiently high if necessary.) Arrange so ϕ_n is rising on $[s_0, s_0 + \frac{1}{n}]$. Hence $|\phi_n|_\infty = 1$.

Now we calculate:

$$|\phi_n \tilde{v} - \tilde{v}|_\infty \quad = \quad |(\phi_n - 1)\tilde{v}|_\infty = |(\phi_n - 1)\tilde{v}|_{\infty, [s_0, s_0 + \frac{1}{n}]}.$$

This leads to

$$|\phi_n \tilde{v} - \tilde{v}|_\infty \leq |\tilde{v}|_{\infty, [s_0, s_0 + \frac{1}{n}]},$$

and since $\tilde{v}$ is continuous and $\tilde{v}(s_0) = 0$, it follows that $|\phi_n \tilde{v} - \tilde{v}|_\infty \to 0^+$ as $n \to \infty$. $\square$

The next lemma is quite complicated. Its essential purpose is to extract facts about the tangent cone $T_q\mathcal{M}$ (as above) by understanding the tangent cones in the set of non-negative scalar functions AC-BV($[t_0, t_1], [0, \infty)$). This requires a "straightening out" coordinate transformation which may only be done locally; hence there is a cumbersome decomposition into N parts using a partition of unity. We mention the "inner endpoint conditions" of **3** below are tailor-made to match the conditions given in point **4** of Lemma 2.57.

LEMMA 2.58. *Suppose $q \in M$, where M is the subset of AC-BV($[t_0, t_1], \mathbb{R}^d$) satisfying the constraints*

$$f_\alpha(q(t)) \geq 0 \text{ for all } \alpha \in J,$$

and also fixed endpoint conditions $q(t_0) = a$ and $q(t_1) = b$, for $a, b \in \mathbb{R}^d$. Assume that the constraints f_α, $\alpha \in J$ are C^3 and for any $\vec{r} \in \mathbb{R}^d$, the set of active constraints $\mathcal{A}$ (those $\alpha \in J$ such that $f_\alpha(\vec{r}) = 0$) has the property that

$$\{\nabla f_\alpha\}_{\alpha \in \mathcal{A}} \text{ is linearly independent.}$$

Then there exist functions $\chi_k(t)$, $k = 1, 2, \cdots, N$, which comprise a C^∞ partition of unity on $[t_0, t_1]$, such that each χ_k is vanishes off from a connected open subspace U_k of $[t_0, t_1]$. Moreover, we may choose this partition of unity so that for each $k = 1, 2, \cdots, N$, there exists a set $\mathcal{A}_k \subset J$, open sets $\mathcal{N}_k, \mathcal{D}_k \subset \mathbb{R}^d$, and a C^3 function $F_k : \mathcal{N}_k \to \mathcal{D}_k$ with a C^3 inverse with the following properties:

1. *For all $t \in U_k$, $q(t) \in \mathcal{N}_k$.*

2. *For all $\vec{r} \in \mathcal{N}_k$, the first $|\mathcal{A}|$ coordinates of $F(\vec{r})$ are given by $f_\alpha(\vec{r})$, for $\alpha \in \mathcal{A}$. (The ordering is immaterial except that it is fixed for each k).*

3. *Suppose an AC-BV function $v : [t_0, t_1] \to \mathbb{R}^d$ is supported on $[u_0, u_1] := \bar{U}_k$. Suppose that v satisfies the* outer endpoint conditions*:*

$$v(t_0) = v(t_1) = 0. \tag{2.19}$$

Suppose also that v satisfies the inner endpoint conditions*:*

$$\begin{aligned} &\textit{Either } \lim_{t \to u_0^+} \dot{v}(t) = 0 \textit{ or else } u_0 = t_0\textit{; and} \\ &\textit{either } \lim_{t \to u_1^-} \dot{v}(t) = 0 \textit{ or else } u_1 = t_1. \end{aligned} \tag{2.20}$$

Then $v \in T_qM$ if and only if for each $\alpha \in \mathcal{A}$, the function $z_\alpha : [t_0, t_1] \to \mathbb{R}$ defined as

$$z_\alpha(t) := \begin{cases} Df_\alpha|_{q(t)}(v(t)) & t \in U_k \\ 0 & otherwise \end{cases} \tag{2.21}$$

is a tangent vector at $w_\alpha(t) := f_\alpha(q(t))$ in the set of functions

$$W_\alpha := \{p \in AC\text{-}BV([t_0, t_1], \mathbb{R}) : p([t_0, t_1]) \geq 0 \text{ and } p(t_0) = f_\alpha(a),\ p(t_1) = f_\alpha(b)\},$$

that is, $z_\alpha \in T_{w_\alpha} W_\alpha$.

4. *A function $x : I \to \mathbb{R}^d$ may be decomposed into a finite sum of functions $x_k : I \to \mathbb{R}^d$ each supported on $\bar{U}_k$ for $k = 1, 2, \cdots N$ as follows:*

$$x(t) = \sum_{k=1}^{N} x_k(t),$$

where

$$x_k(t) := \chi_k(t) x(t).$$

In particular, $x \in T_qM$ if and only if $x_k \in T_qM$ for each $k = 1, 2, \cdots, N$.

Also, this lemma remains admits a continuous version: we may replace all AC-BV spaces with spaces of continuous functions C, replace all AC-BV convergence notions with uniform convergence, and omit the inner endpoint conditions of (3).

Proof. For convenience, denote $I := [t_0, t_1]$. For each $s \in I$, denote

$$\mathcal{A}_s = \{\alpha \in J : f_\alpha(q(s)) = 0\}.$$

By assumption, at any s, the active constraint functions f_α, $\alpha \in \mathcal{A}_s$, have gradients ∇f_α which are linearly independent at $q(s)$. Therefore it must be the case that

$a_s := |\mathcal{A}_s| \leq d$. We seek a coordinate transformation in which the first a_s coordinates are the constraint functions f_α, $\alpha \in \mathcal{A}_s$. To this end, we must also supply $d - a_s$ additional C^3 scalar functions, defined in a neighborhood of $q(s) \in \mathbb{R}^d$, which also have linearly independent gradients (with respect to the ∇f_α, $\alpha \in \mathcal{A}_s$ and themselves) – we call these additional $d - a_s$ coordinate functions $\tilde{q}$, collectively. In particular, we may complete the coordinate transformation by choosing the components of $\tilde{q}$ to be judiciously chosen canonical projections of q. (It is straightforward to see this may be done.) To be clear, we point out that $\tilde{q}$ is a C^3 function from $\mathbb{R}^d$ to $\mathbb{R}^{d-a_s}$. The a_s constraint functions f_α, $\alpha \in \mathcal{A}_s$, together with $\tilde{q}$, describe a C^3 coordinate transformation in a neighborhood of $q(s)$. By the inverse function theorem, there exist open sets $\mathcal{N}_s$ and $\mathcal{D}_s$ (of $\mathbb{R}^d$) such that $q(s) \in \mathcal{N}_s$ $F_s : \mathcal{N}_s \to \mathcal{D}_s$ is a C^3 bijection with a C^3 inverse, where

$$F_s(\vec{r}) := (f_\alpha(\vec{r}) : \alpha \in \mathcal{A}_s) \times \tilde{q}(\vec{r}).$$

For any given $s \in I$, let $U_s \subset I$ be a connected neighborhood (that is an open interval) of $s \in I$ for which only the constraints active at s are active for any $t \in U_s$ (though they need not be), and also $F_s(q(t)) \in D_s$ for any $t \in U_s$. To see such a neighborhood exists, notice that constraints not being active is an open condition. Also, the set of times $\{t : F_s(q(t)) \in D_s\}$ is open and contains s, since D_s is open and F_s and q are continuous. In fact, by normality, we may arrange without loss that the closure of the U_s are contained in neighborhoods V_s such that the only constraints active at $t \in V_s$ are in $\mathcal{A}_s$ and also $F_s(q(t)) \in D_s$ for any $t \in V_s$.

The collection of neighborhoods $\{U_s\}$ provides an open cover of I. Since I is compact, there exists a finite subcollection of $\{U_s\}$ which is also a cover of I: call it $\{U_k\}$, for $k = 1, 2, \cdots, N$. Choose corresponding finite subsets of $\{\mathcal{A}_s\}$, $\{\mathcal{N}_s\}$, $\{\mathcal{D}_s\}$, $\{V_s\}$, and $\{F_s\}$, also to be indexed by $k = 1, 2, \cdots, N$. There exists a *partition of unity* $\chi : I \times \{1, 2, \cdots, N\} \to [0, 1]$ such that $\sum_{k=1}^{N} \chi(t, k) = 1$ for all $t \in I$ and the N functions $\chi(\cdot, k)$ for $k = 1, 2, \cdots, N$ are C^∞ and supported on $\bar{U}_1, \bar{U}_2, \cdots, \bar{U}_N$, respectively. For a proof regarding the existence of partitions of unity, see [26].

Inspection of what we have constructed so far reveals we have shown the conclusion of the lemma apart from **3** and **4**.

We show **3**. Let $k \in \{1, 2, \cdots, N\}$; we omit the k subscripts now. Suppose that $v : I \to \mathbb{R}^d$ is an element of T_qM supported on $\bar{U}$ which satisfies the outer endpoint conditions (2.19) and the inner endpoint conditions (2.20). Let $\alpha \in \mathcal{A}$. We show z_α as defined in **3** is in the tangent cone at $w_\alpha(t) := f_\alpha(q(t))$ of

$$W_\alpha := \{p \in \text{AC-BV}([t_0, t_1], [0, \infty)) : p(t_0) = w_\alpha(t_0) \text{ and } p(t_1) = w_\alpha(t_1)\}.$$

That is, we show $z_\alpha \in T_{w_\alpha} W_\alpha$.

Without loss, assume that v is a unit geometric tangent vector. This is without loss, by the following argument. Since $v \in T_qM$, v admits approximation in the AC-BV sense by finite non-negative combinations of unit geometric tangent vectors of T_qM. Let v^γ, $\gamma \in F$ be unit geometric tangent vectors in T_qM such that

$$\left|\left| v - \sum_{\gamma \in F} a^\gamma v^\gamma \right|\right| \leq \epsilon,$$

where the a^γ are the non-negative coefficients. We use the special case of the result **3** (we are showing that if **3** holds unit geometric tangent vectors then it hold generally) to obtain

$$z_\alpha^\gamma(t) := Df_\alpha|_{q(t)} v^\gamma(t) \in T_{w_\alpha} W_\alpha, \;\; \gamma \in F.$$

By Lemma 2.56, we see that $\sum_{\gamma \in F} a^\gamma z_\alpha^\gamma$ is also an approximation for z_α. Since $T_{w_\alpha} W_\alpha$ is a closed cone, and z_α may be approximated by finite non-negative combinations of elements of $T_{w_\alpha} W_\alpha$, it follows that $z_\alpha \in T_{w_\alpha} W_\alpha$. *This completes the without loss argument.*

Since v is a unit geometric tangent vector, there exists a sequence (q_n) of AC-BV functions in M converging in the sense of the AC-BV metric to q such that

$$v := \lim_{n \to \infty} \frac{q_n - q}{||q_n - q||}.$$

Since f_α is C^3, it follows from Lemma 2.55 that $Df_\alpha|_{q(t)}$ is AC-BV. We may apply Lemma 2.56 in order to see that

$$z_\alpha = Df_\alpha|_{q(t)} v = \lim_{n \to \infty} \frac{Df_\alpha|_{q(t)} \, (q_n - q)}{||q_n - q||}.$$

The application of Lemma 2.53 gives us

$$z_\alpha = \lim_{n \to \infty} \frac{f_\alpha(q_n(t)) - f_\alpha(q(t)) + E_n}{||q_n - q||},$$

where $||E_n|| = o(||q - q_n||)$. See that we may omit E_n without changing the expression, leaving us with

$$z_\alpha = \lim_{n \to \infty} \frac{f_\alpha(q_n(t)) - f_\alpha(q(t))}{||q_n - q||}. \tag{2.22}$$

Now we define

$$w_{\alpha,n}(t) := f_\alpha(q_n(t)), \tag{2.23}$$

and we recall

$$w_\alpha(t) := f_\alpha(q(t)). \tag{2.24}$$

See that $(w_{\alpha,n})$ is a sequence in W_α, since they are non-negative AC-BV functions defined on $[t_0, t_1]$ which satisfy fixed endpoint conditions

$$w_{\alpha,n}(t_0) = f_\alpha(t_0) \text{ and } w_{\alpha,n}(t_1) = f_\alpha(q(t_1)),$$

because for each n, $q_n \in M$, and hence satisfies the fixed endpoint conditions $q_n(t_0) = q(t_0)$ and $q_n(t_1) = q(t_1)$.

It is also straightforward that $w_\alpha \in W_\alpha$. Because the f_α are C^3, and (q_n) converges to q in the AC-BV sense, it follows from (2.23), (2.24), and Lemma 2.55 that the sequence $(w_{\alpha,n})$ converges to w_α in the sense of AC-BV functions.

Because $(w_{\alpha,n})$ converges in W_α to w_α, and because we have (substituting (2.23) and (2.24) into (2.22))

$$z_\alpha = \lim_{n\to\infty} \frac{w_{\alpha,n} - w_\alpha}{||q_n - q||},$$

it follows by Lemma 2.17 that $z_\alpha \in T_{w_\alpha} W_\alpha$.

Now we show the converse of **3**. We assume that v is supported on $[u_0, u_1] = \bar{U}$ is an AC-BV function satisfying the outer endpoint conditions (2.19) and the inner endpoint conditons (2.20). We assume for each $\alpha \in \mathcal{A}$ we have that z_α (as

defined in (2.21)) is in the tangent cone at $w_\alpha(t) := f_\alpha(q(t))$ of $W_\alpha := \{p \in \text{AC-BV}([t_0,t_1],[0,\infty)) : p(t_0) = f_\alpha(q(t_0))$ and $p(t_1) = f_\alpha(q(t_1))\}$. We show $v \in T_qM$.

Define $\tilde{q}$ and $\tilde{v}$ as the restrictions of q and v, respectively, to $\bar{U}$. Define

$$\tilde{M} := \{\tilde{p} \in \text{AC-BV}(\bar{U}, \mathbb{R}^d) : \tilde{p} = p|_{\bar{U}} \text{ for some } p \in M, \text{ and } \tilde{p}|_{\partial\bar{U}} = q|_{\partial\bar{U}}\}.$$

Note that the condition $\tilde{p}|_{\partial\bar{U}} = q|_{\partial\bar{U}}$ are fixed endpoint conditions; in particular we have $\bar{U} =: [u_0, u_1]$, and the condition may be rewritten as $\tilde{p}(u_0) = q(u_0)$ and $\tilde{p}(u_1) = q(u_1)$. Since v is supported on $[u_0,u_1] \subset [t_0,t_1]$ and satisfies the outer endpoint conditions (2.19), it follows that $\tilde{v}$ vanishes on $\partial\bar{U}$ (that is, $\tilde{v}(u_0) = \tilde{v}(u_1) = 0$).

We wish to show $v \in T_qM$. By Lemma 2.57, part **3**, $v \in T_qM$ provided that $\tilde{v} \in T_{\tilde{q}}\tilde{M}$. Therefore, what remains to be shown is that $\tilde{v} \in T_{\tilde{q}}\tilde{M}$.

Define the set

$$\tilde{W} := \{p \in \text{AC-BV}(\bar{U}, [0,\infty)^{|\mathcal{A}|} \times \mathbb{R}^{d-|\mathcal{A}|}) : p|_{\partial\bar{U}} = F(\tilde{q})|_{\partial\bar{U}}\}.$$

Notice that the condition $p|_{\partial\bar{U}} = F(q)|_{\partial\bar{U}}$ gives fixed endpoint conditions: we may rewrite these fixed endpoint conditions as $p(u_0) = F(q(u_0))$ and $p(u_1) = F(q(u_1))$ since $\bar{U} = [u_0,u_1]$.

Define

$$z(t) = \begin{cases} DF|_{q(t)}v(t) & t \in U \\ 0 & \text{otherwise} \end{cases}. \tag{2.25}$$

By Lemma 2.55, $DF|_{q(t)}$ is an AC-BV matrix-valued function. By Lemma 2.56 and a straightforward pasting argument, z is AC-BV. Since $v(t)$ is an AC-BV function

supported on $\bar{U}$ satisfying the outer endpoint conditions (2.19), it follows that $z(t)$ is supported on $\bar{U}$ and also satisfies outer endpoint conditions.

Let $\tilde{z}$ be the restriction of z to $\bar{U}$: $\tilde{z} = z|_{\bar{U}}$. Notice that $\tilde{z}|_{\partial\bar{U}} = 0$; in other words, $\tilde{z}$ is zero at the endpoints u_0 and u_1. It is straightforward to verify that $\tilde{z}$ will satisfy the inner endpoint conditions.

Define $\tilde{w}(t) := F(\tilde{q}(t))$ for $t \in \bar{U}$. Observe that $\tilde{w} \in \tilde{W}$.

Claim. The following holds: $\tilde{z} \in T_{\tilde{w}}\tilde{W}$.

Define

$$\tilde{W}_\alpha := \begin{cases} \{p \in \text{AC-BV}(\bar{U}, [0,\infty)) : p|_{\partial\bar{U}} = F_\alpha(\tilde{q})|_{\partial\bar{U}}\} & \text{when } \alpha \in \mathcal{A} \\ \{p \in \text{AC-BV}(\bar{U}, \mathbb{R}) : p|_{\partial\bar{U}} = F_\alpha(\tilde{q})|_{\partial\bar{U}}\} & \text{otherwise} \end{cases}.$$

Let $\tilde{z}_\alpha$, $\alpha = 1, 2, \cdots, d$ be the component functions of $\tilde{z}$. Equivalently, $\tilde{z}_\alpha$ are the restrictions of z_α to $\bar{U}$. By hypothesis, for each $\alpha \in \mathcal{A}$, z_α is in the tangent cone at $w_\alpha(t) := f_\alpha(q(t))$ of

$$W_\alpha := \{p \in \text{AC-BV}([t_0, t_1], [0,\infty)) : p(t_0) = w_\alpha(t_0) \text{ and } p(t_1) = w_\alpha(t_1)\}.$$

It is straightforward to verify that the inner endpoint condition (2.20) applying to v have been inherited by z_α, so we may apply Lemma 2.57. This reveals that $\tilde{z}_\alpha$ is in the tangent cone at $\tilde{w}_\alpha$ of $\tilde{W}_\alpha$ for all $\alpha \in \mathcal{A}$.

For $\alpha \notin \mathcal{A}$, we know that z_α is AC-BV, is supported on $\bar{U}$, and satisfies fixed endpoint conditions

$$z_\alpha(t_0) = z_\alpha(t_1) = 0.$$

From this it is straightforward that for all $\alpha \notin \mathcal{A}$, we have that $\tilde{z}_\alpha$ is in the tangent cone at $F_\alpha(q(t))$ of $\tilde{W}_\alpha$. Define $\tilde{w}_\alpha := F_\alpha(q(t))$ for all $t \in \bar{U}$, all $\alpha = 1, 2, \cdots, d$. Combining these two cases (both $\alpha \in \mathcal{A}$ and $\alpha \notin \mathcal{A}$, we obtain

$$z_\alpha \in T_{\tilde{w}_\alpha}\tilde{W}_\alpha \text{ for all } \alpha = 1, 2, \cdots, d.$$

Now we make the straightforward observation that, in general, if $\tilde{z}_\alpha \in T_{\tilde{w}_\alpha}\tilde{W}_\alpha$ for $\alpha = 1, 2, \cdots, d$, and $\tilde{z} = \Pi\, \tilde{z}_\alpha$, $\tilde{w} = \Pi \tilde{w}_\alpha$, and $\tilde{W} = \Pi\, \tilde{W}_\alpha$, then it follows that $\tilde{z} \in T_{\tilde{w}}\tilde{W}$. This is precisely the situation we face; the claim is true. *This completes the proof of the claim.*

Our goal now is to use the fact that $\tilde{z} \in T_{\tilde{w}}\tilde{W}$ in order to establish $\tilde{v} \in T_{\tilde{q}}\tilde{M}$. Without loss, assume that $\tilde{z}$ is a unit geometric tangent vector. This is without loss for reasons already discussed. Since $\tilde{z}$ is a unit geometric tangent vector, there exists $(\tilde{w}_n)$ in $\tilde{W}$ converging to $\tilde{w} \in \tilde{W}$ in the AC-BV sense such that

$$\tilde{z} = \lim_{n\to\infty} \frac{\tilde{w}_n - \tilde{w}}{||\tilde{w}_n - \tilde{w}||}.$$

It follows from Equation (2.25) and the definitions of $\tilde{z}$ and $\tilde{w}$ that $\tilde{v}(t) = DF^{-1}|_{\tilde{w}(t)}\tilde{z}(t)$. By Lemma 2.55, $DF^{-1}|_{\tilde{w}(t)}$ is AC-BV. Accordingly, we may use Lemma 2.56 in order to write

$$\tilde{v} = \lim_{n\to\infty} \frac{DF^{-1}|_{\tilde{w}(t)}\,(\tilde{w}_n(t) - \tilde{w}(t))}{||\tilde{w}_n - \tilde{w}||}.$$

By Lemma 2.53,

$$DF^{-1}|_{\tilde{w}(t)}\,(\tilde{w}_n(t) - \tilde{w}(t)) = F^{-1}(\tilde{w}_n(t)) - F^{-1}(\tilde{w}(t)) + E_n,$$

where $||E_n|| = o(||\tilde{w}_n - \tilde{w}||)$. Hence we may write

$$\tilde{v} = \lim_{n\to\infty} \frac{F^{-1}(\tilde{w}_n(t)) - F^{-1}(\tilde{w}(t)) + E_n}{||\tilde{w}_n - \tilde{w}||}. \tag{2.26}$$

Define

$$\tilde{q}_n = F^{-1}(\tilde{w}_n). \tag{2.27}$$

We see that this is well-defined. Notice that F^{-1} is not defined everywhere, so the definition of $\tilde{q}_n$ only makes sense if $\tilde{w}_n$ takes values in $\mathcal{D}$. However, since $\tilde{w}$ only takes values in $\mathcal{D}$, which is open, it follows that the closed path $\{\tilde{w}(t) : t \in \bar{U}\}$ admits an ϵ-neighborhood contained in $\mathcal{D}$ (by a straightforward exercise in topology). Since the $\tilde{w}_n$ converge uniformly (AC-BV convergence implies uniform convergence) to $\tilde{w}$, for sufficiently high n the functions $\tilde{w}_n$ take values in $\mathcal{D}$ only. Hence the $\tilde{q}_n$ are well-defined, at least for sufficiently high n, which is no loss. *This completes the proof of well-definedness.*

Observe from the definition of $\tilde{w}$ that

$$\tilde{q} = F^{-1}(\tilde{w}). \tag{2.28}$$

Substitute (2.27) and (2.28) into (2.26) to obtain

$$\tilde{v} = \lim_{n\to\infty} \frac{\tilde{q}_n(t) - \tilde{q}(t)}{||\tilde{w}_n - \tilde{w}||}, \tag{2.29}$$

where we have dropped the E_n term because it is too small to matter.

Claim. The sequence $(\tilde{q}_n)$ is in $\tilde{M}$ and converges to $\tilde{q} \in \tilde{M}$ in the AC-BV sense.

First we see that $\tilde{q}_n \in \tilde{M}$. In order to show a function p is in $\tilde{M}$ we show three things: that it is AC-BV, that is satisfies the constraints $f_\alpha(p(t)) \geq 0$, and that it satisfies the endpoint conditions $p|_{\partial\bar{U}} = q|_{\partial\bar{U}}$.

We show $\tilde{q}_n$ is AC-BV. By (2.27), $\tilde{q}_n = F^{-1}(\tilde{w}_n)$. By Lemma 2.55, since F^{-1} is C^3 and $\tilde{w}_n$ is AC-BV, it follows that $\tilde{q}_n$ is AC-BV. It also follows from Lemma 2.55 that $(\tilde{q}_n)$ converges to $\tilde{q}$ in the AC-BV sense. Since AC-BV limits of AC-BV functions are again AC-BV, it follows that $\tilde{q}$ is AC-BV.

We verify that $f_\alpha(\tilde{q}_n(t)) \geq 0$ for all $\alpha \in \mathcal{A}$, $t \in \bar{U}$. By the definition of $\tilde{W}$, $f_\alpha(\tilde{q}_n(t)) = \tilde{w}_{n,\alpha}(t) \geq 0$ for all $\alpha \in \mathcal{A}$, $t \in \bar{U}$.

We verify the endpoint conditions. Observe that $\tilde{w}_n(u_0) = \tilde{w}(u_0) = f_\alpha(q(u_0))$ and $\tilde{w}_n(u_1) = \tilde{w}(u_1) = f_\alpha(q(u_1))$ for all n. Accordingly, for $\tilde{q}_n(u_0) = F^{-1}(\tilde{w}_n(u_0)) = F^{-1}(\tilde{w}(u_0)) = \tilde{q}(u_0)$ and $\tilde{q}_n(u_1) = F^{-1}(\tilde{w}_n(u_1)) = F^{-1}(\tilde{w}(u_1)) = \tilde{q}(u_1)$. These are the requires endpoint conditions. *This completes the proof of the claim.*

By the above claim and Equation (2.29), we have shown via Lemma 2.17 that $\tilde{v} \in T_{\tilde{q}}\tilde{M}$. The converse of **3** is shown.

We show **4**. That x can be so decomposed into $\sum x_k$ follows from the properties of χ. Since the partition of unity functions are C^∞, the x_k are AC-BV provided x is and vice-versa. Note that if all of the x_k are in T_qM, then so must be x, since cones are closed under finite sums and T_qM is a cone.

Now the converse of **4**. We show that $x \in T_qM$ implies that each x_k is in T_qM. Choose some $k \in \{1, 2, \cdots, N\}$. Henceforth in this proof omit all k-subscripts, except the one on x, as we still need to distinguish x from x_k. We show that x_k is in T_qM.

Without loss, we may assume that x is a unit geometric tangent vector in T_qM. This is without loss because otherwise we may approximate x arbitrarily well in the AC-BV topology with a finite non-negative combination of unit geometric tangent vectors:

$$||x - \sum_{\gamma \in F} a^\gamma x^\gamma|| < \epsilon.$$

By Lemma 2.56, it follows that

$$||\chi x - \sum_{\gamma \in F} a^\gamma \chi x^\gamma||$$

can be made arbitrarily small as well; in other words we may approximate x_k with finite non-negative combinations of unit geometric tangent vectors in T_qM. Since T_qM is closed, it follows that $x_k \in T_qM$. This completes the without loss argument; henceforth we may assume x is a unit geometric tangent vector.

By **3**, $x_k \in T_qM$ if for each $\alpha \in \mathcal{A}$, the function z_α as defined in (2.21) is an element of the tangent cone at $w_\alpha(t) := f_\alpha(q(t))$ of

$$W_\alpha := \{p \in \text{AC-BV}([t_0, t_1], [0, \infty)) : p(t_0) = f_\alpha(q(t_0)) \text{ and } p(t_1) = f_\alpha(q(t_1))\}.$$

Accordingly, let $\alpha \in \mathcal{A}$. We show $z_\alpha \in T_{w_\alpha} W_\alpha$.

Since $x_k = \chi x$, we may write

$$z_\alpha(t) = \begin{cases} DF_\alpha|_{q(t)}(\chi(t)x(t)) & t \in \bar{U} \\ 0 & \text{otherwise} \end{cases}.$$

By linearity of $DF|_{q(t)}$,

$$z_\alpha(t) = \left(\begin{cases} \chi(t) DF_\alpha|_{q(t)} & t \in \bar{U} \\ 0 & \text{otherwise} \end{cases} \right) x(t).$$

Define the time-dependent linear functional

$$\ell(t) = \begin{cases} \chi(t) DF_\alpha|_{q(t)} & t \in \bar{U} \\ 0 & \text{otherwise} \end{cases}$$

Since x is a unit geometric tangent vector in T_qM, there exists $(q_n) \in M$ tending to q in the AC-BV sense such that

$$x = \lim_{n\to\infty} \frac{q_n - q}{||q_n - q||},$$

where the limit is to be interpreted in the AC-BV sense.

Thus,

$$z_\alpha(t) = \ell(t) \lim_{n\to\infty} \frac{q_n - q}{||q_n - q||}.$$

By Lemma 2.55 and straightforward pasting argument, ℓ is AC-BV. By Lemma 2.56 we may move the linear functional inside the limit to obtain

$$z_\alpha(t) = \lim_{n\to\infty} \frac{\ell(q_n - q)}{||q_n - q||}.$$

Expanding out the definition of ℓ,

$$z_\alpha(t) = \lim_{n\to\infty} \frac{\begin{cases} \chi(t) DF_\alpha|_{q(t)}(q_n - q) & t \in \bar{U} \\ 0 & \text{otherwise} \end{cases}}{||q_n - q||}.$$

Applying Lemma 2.53 to this expression yields

$$z_\alpha(t) = \lim_{n\to\infty} \frac{\begin{cases} \chi(t)(f_\alpha(q_n(t)) - f_\alpha(q(t)) + E_n(t)) & t \in \bar{U} \\ 0 & \text{otherwise} \end{cases}}{||q_n - q||},$$

where $||E_n|| = o(||q_n - q||)$. See that the E_n term makes no difference, and we may drop it. Also note that after dropping E_n, the piecewise definition on $\bar{U}$ may be extended to all of $[t_0, t_1]$, since χ is supported on $\bar{U}$. We then have

$$z_\alpha(t) = \lim_{n\to\infty} \frac{\chi(t)(f_\alpha(q_n(t)) - f_\alpha(q(t)))}{||q_n - q||}.$$

Some algebra give us

$$z_\alpha(t) = \lim_{n\to\infty} \frac{[\chi(t)f_\alpha(q_n(t)) + (1-\chi(t))f_\alpha(q(t))] - f_\alpha(q(t))}{||q_n - q||}.$$

Now define

$$w_{\alpha,n}(t) := \chi(t)f_\alpha(q_n(t)) + (1-\chi(t))f_\alpha(q(t))$$

and

$$w_\alpha(t) := f_\alpha(q(t)).$$

Observe that $(w_{\alpha,n})$ is a sequence of nonnegative AC-BV functions. Moreover, notice that the $w_{\alpha,n}$ satisfy the fixed endpoint conditions for W_α. Thus, $w_{\alpha,n} \in W_\alpha$ for all n, and also $w_\alpha \in W_\alpha$. By Lemma 2.55, $(w_{\alpha,n})$ converges to w_α in the sense of AC-BV functions. Since

$$z_\alpha = \lim_{n\to\infty} \frac{w_{\alpha,n} - w_\alpha}{||q_n - q||},$$

it follows that $z_\alpha \in T_{w_\alpha}W_\alpha$. This is, in fact, what we were required to show. It follows that $x_k \in T_qM$.

Now we prove the continuous version of the lemma. In fact, the above proof holds for the continuous proof verbatim, with the following replacements:

1. All AC-BV spaces are replaced with C spaces.

2. All the AC-BV norms are replaced with sup norms.

3. All the limits are interpreted in the uniform sense.

4. The tangent cones are reckoned in the space of continuous functions with the uniform metric.

5. Every use of Lemma 2.53 is replaced with the use of Lemma 2.54.

6. Every use of Lemmas 2.57, 2.55, and 2.56 now refer to the continuous versions of Lemmas 2.57, 2.55, and 2.56, respectively.

7. All mentions of the inner endpoint conditions are omitted.

□

This next lemma is a straightforward result which allows us to make AC-BV approximations of continuous functions which have properties we will find useful in the remaining proofs.

PROPOSITION 2.59. *Suppose* $x : [t_0, t_1] \to \mathbb{R}$ *is a continuous function. Let* $\epsilon > 0$. *Then there exists a* C^∞ *function* $y : [t_0, t_1] \to \mathbb{R}$ *such that*

1. *Both* x *and* y *share the same values at the endpoints:* $x(t_0) = y(t_0)$ *and* $x(t_1) = y(t_1)$.

2. *We have* $y(t) = 0$ *in a neighborhood of the times for which* $x(t) = 0$.

3. *We have* $y(t) \geq 0$ *whenever* $x(t) \geq 0$.

4. *The functions* x *and* y *are uniformly* ϵ*-close:*

$$|x(t) - y(t)| < \epsilon \text{ for all } t \in [t_0, t_1].$$

Remark. We notice, as a consequence of **2** and **3**, that $y(t) \geq 0$ in a neighborhood of the set of times for which $x(t) \geq 0$. We also notice from **2** that $\dot{y}(t) = 0$ in a neighborhood of those times for which $x(t) = 0$.

Proof. It is well known that continuous functions may be uniformly approximated by smooth functions on compact intervals. Let $p(t)$ be a smooth approximation of $x(t)$ such that $|x - w|_\infty < \epsilon$, $p(t_0) = x(t_0)$, and $p(t_1) = x(t_1)$.

Claim. There exists a function $w(t)$ which is a smooth approximation of $x(t)$ such that $|x - w|_\infty < \epsilon$, $w(t_0) = x(t_0)$, and $w(t_1) = x(t_1)$, and $w(t) = 0$ in a neighborhood of times such that $x(t) = 0$.

Define A to be the closed set of times

$$A := \{t \in [t_0, t_1] : |x(t)| \leq \epsilon\}.$$

Define B to be the closed set of times

$$B := \{t \in [t_0, t_1] : |x(t)| \geq 2\epsilon\}.$$

Define a bump function $\psi \in C^\infty([t_0, t_1], [0, 1])$ such that $\psi|_A = 0$, $\psi|_B = 1$,

$$\psi(t_i) = \begin{cases} 1 & |x(t_i)| > 0 \\ 0 & x(t_i) = 0 \end{cases}, \text{ where } i \in \{0, 1\}.$$

It is straightforward to verify that such a bump function exists provided that either $x(t_0) = 0$ or else $\epsilon < |x(t_0)|$, and also either $x(t_1) = 0$ or else $\epsilon < |x(t_1)|$. We assume $\epsilon > 0$ is sufficiently small so that such a bump function exists.

Define $w(t) := \psi(t)p(t)$. Observe that w obeys the endpoint conditions $w(t_0) = x(t_0)$ and $w(t_1) = x(t_1)$. It is also straightforward to see that

$$|w - x|_\infty < 2\epsilon.$$

Finally, we observe that $w(t) = 0$ for all $t \in A$, which contains a neighborhood (specifically, $\{t \in [t_0, t_1] : |x(t)| < \epsilon\}$) of the set of times for which $x(t) = 0$. *This completes the proof of the claim.*

Now we are ready to construct y.

Define A to be the closed set of times for which $w(t) \leq 0$ and $x(t) \geq 0$. Let B be the closed set of times for which $w(t) \geq \epsilon$ or $x(t) \leq -\epsilon$. See that A and B are disjoint closed sets; let θ be a C^∞ bump function which is 0 on A and 1 on B, and also

$$\theta(t_i) = \left\{ \begin{array}{cc} 1 & |x(t_i)| > 0 \\ 0 & x(t_i) = 0 \end{array} \right. , \text{ where } i \in \{0, 1\}.$$

Now define

$$y(t) := \theta(t)w(t).$$

We show **4**: that y is a uniform approximation of x. A few straightforward calculations lead to the formula

$$|y(t) - w(t)| \leq \left\{ \begin{array}{ll} 0 & \text{for } t \in B, \\ \epsilon & \text{for } t \in A, \\ 3\epsilon & \text{otherwise} \end{array} \right. .$$

Hence $y(t)$ uniformly approximates $w(t)$ within 3ϵ, and in turn uniformly approximates $x(t)$ within 4ϵ.

Now we show **3**: that $x(t) \geq 0$ implies that $y(t) \geq 0$. Let $t \in [t_0, t_1]$ such that $x(t) \geq 0$. Suppose $t \in A$. Then $y(t) = 0$, and it follows $y(t) \geq 0$. Now suppose $t \in B$. Then $w(t) \geq \epsilon$ and $y(t) = w(t)$, hence $y(t) \geq 0$. Now suppose that t is in neither A nor B. Then it follows that $w(t) \in (0, \epsilon)$, and $y(t) = \theta(t)w(t) \geq 0$. We have exhausted the cases; $x(t) \geq 0$ implies that $y(t) \geq 0$.

Now we verify the endpoint conditions **1**. If $\theta(t_0) = 1$, the left endpoint condition holds. Otherwise $\theta(t_0) = 0$ implies that $x(t_0) = 0$, and the left endpoint condition holds. Similarly the right endpoint condition holds.

Since $y(t) = 0$ whenever $w(t) = 0$, and $w(t) = 0$ in a neighborhood of the times for which $x(t) = 0$, **2** holds as well. □

Now we prove the "$\overline{T_qM} = T_q\bar{M}$" result in a very special case:

LEMMA 2.60. *Let X be the Banach space AC-BV$([t_0, t_1], \mathbb{R})$. Let M be the closed subset of X consisting of those functions which are non-negative, and for which we have the fixed endpoint conditions $q(t_0) = a \geq 0$, $q(t_1) = b \geq 0$. Let $q \in M$.*

The tangent cone T_qM (as considered in the AC-BV sense) may be taken as a subset of $C([t_0, t_1], \mathbb{R})$. It has a closure (which we call the uniform closure, since C has the sup norm) in this space, which we write as $\overline{T_qM}$. Similarly, M may be taken as a subset of $C([t_0, t_1], \mathbb{R})$, and admits a closure we call $\bar{M}$. We can then construct the tangent cone $T_q\bar{M}$, where the tangent cone is considered in $C([t_0, t_1], \mathbb{R})$. uniform

closure $\bar{M}$. The closure of the tangent cone T_qM in the space of continuous functions $C([t_0,t_1],\mathbb{R})$ is the same as the tangent cone (reckoned in the uniform sense) at q of $\bar{M}$, the (uniform) closure of M in $C([t_0,t_1]),\mathbb{R})$:

$$\overline{T_qM} = T_q\bar{M}.$$

Proof. To avoid confusion we put a superscript "AC-BV" or "C"on the tangent cone symbol T indicating which space it is to be considered in. Barred sets indicate the uniform closure of the set considered as a subset of the Banach space $C([t_0,t_1],\mathbb{R})$. We wish to show $\overline{T_q^{\text{AC-BV}}M} = T_q^C\bar{M}$. To this end we show inclusion both ways.

We show $\overline{T_q^{\text{AC-BV}}M} \subset T_q^C\bar{M}$. Since $M \subset \bar{M}$, and AC-BV convergence implies uniform convergence, it is clear that $T_q^{\text{AC-BV}}M \subset T_q^C\bar{M}$. Since $T_q^C\bar{M}$ is a tangent cone considered in $C([t_0,t_1],\mathbb{R})$, it follows that $T_q^C\bar{M}$ is closed in $C([t_0,t_1],\mathbb{R})$. It is a straightforward fact from topology (if $A \subset \bar{B}$ then $\bar{A} \subset \bar{B}$) now that the uniform closure of $T_q^{\text{AC-BV}}M$ (which is the closure of $T_q^{\text{AC-BV}}M$ in $C([t_0,t_1],\mathbb{R})$) is contained in $T_q^C\bar{M}$.

Now the other inclusion. We show $T_q^C\bar{M} \subset \overline{T_q^{\text{AC-BV}}M}$. Let $x(t) \in T_q^C\bar{M}$. We show $x(t)$ is in the uniform closure of $T_q^{\text{AC-BV}}M$. To this end, we choose $\epsilon > 0$ and show we may construct an element $y(t)$ of $T_q^{\text{AC-BV}}M$ that is within ϵ distance to $x(t)$ in the uniform metric: $|x-y|_\infty < \epsilon$.

By Proposition 2.59, there exists an AC-BV function $y(t)$ defined on $[t_0,t_1]$ such that

1. We have endpoint conditions $y(t_0) = y(t_1) = 0$;

2. The function y is ϵ-close to x:

$$|x - y|_\infty < \epsilon;$$

3. The inequality $y(t) \geq 0$ holds in a neighborhood of the set of times for which $x(t) \geq 0$.

Since y is an arbitrarily close uniform approximation of x, we may prove that x is in the uniform closure of $T_q^{\text{AC-BV}} M$ by showing that $y \in T_q^{\text{AC-BV}} M$.

Observe $y(t) < 0$ is only possible on a closed set of times V disjoint from when $q(t) = 0$. This is because whenever $q(t) = 0$, we have $x(t) \geq 0$, and $y(t) \geq 0$ in a neighborhood of such times. By the extreme value theorem, q attains a minimal value on V which is greater than 0, call it $q_{\min} > 0$. Similarly, $y(t)$ has a minimum value on V (which, of course, may be negative), which we call $y_{\min}$.

Let (δ_n) be a sequence of positive numbers tending to 0: $\delta_n \to 0$ as $n \to \infty$. Define

$$q_n(t) := q(t) + \delta_n y(t).$$

See that q_n is AC-BV and satisfies the fixed endpoint conditions $q_n(t_0) = q(t_0)$ and $q_n(t_1) = q(t_1)$.

Claim. For sufficiently large n, we have $q_n(t) \geq 0$ for all $t \in [t_0, t_1]$.

Observe that for $t \notin V$, $q_n(t) = q(t) + \delta_n y(t)$, where $q(t) \geq 0$, $\delta_n > 0$, and $y(t) \geq 0$. Observe that for $t \in V$, we have $q_n(t) = q(t) + \delta_n y(t) \geq q_{\min} + \delta_n y_{\min}$. For sufficiently large n, $\delta_n < \left| q_{\min}/y_{\min} \right|$. Since $q_{\min} > 0$ and for sufficiently large

n, $0 < \delta_n < \left|q_{\min}/y_{\min}\right|$, it is straightforward to verify that for sufficiently large n, $q_n(t) \geq 0$ for $t \in V$. *This completes the proof of the claim.*

It follows from the preceding that for sufficiently large n, $q_n \in M$. Therefore (q_n) is a sequence of functions in M tending to q in the AC-BV sense. Also, since $\frac{q_n - q}{\delta_n}$ is the constant sequence $(y, y, y, \cdots)$, we have

$$y = \lim_{n\to\infty} \frac{q_n - q}{\delta_n},$$

where the limit is taken in the AC-BV sense. By Lemma 2.17, $y \in T_q^{\text{AC-BV}} M$. We are done: since y can be made arbitrarily close to x in the uniform metric by choosing $\epsilon > 0$ arbitrarily small, it follows that x is in the uniform closure of $T_q^{\text{AC-BV}} M$. $\square$

This last lemma is the critical ingredient, along with Lemma 2.58, to show the following extremely important generalization of Lemma 2.60:

LEMMA 2.61. *Suppose $q \in M$, where M is the subset of AC-BV$([t_0, t_1], \mathbb{R}^d)$ satisfying the constraints*

$$f_\alpha(q(t)) \geq 0 \text{ for all } \alpha \in J,$$

and also fixed endpoint conditions $q(t_0) = a$ and $q(t_1) = b$, for $a, b \in \mathbb{R}^d$. Suppose $x \in T_q M$. Assume that the constraints f_α, $\alpha \in J$ are C^1 and for any $\vec{r} \in \mathbb{R}^d$, the set of active constraints $\mathcal{A}$ (those $\alpha \in J$ such that $f_\alpha(\vec{r}) = 0$) has the property that

$$\{\nabla f_\alpha\}_{\alpha\in\mathcal{A}} \text{ is linearly independent.}$$

Then the uniform closure of the tangent cone of M at q is the tangent cone of $\bar{M}$ at q in $C([t_0, t_1], \mathbb{R}^d)$:

$$\overline{T_q^{AC\text{-}BV} M} = T_q^C \bar{M}.$$

Proof. We wish to show $\overline{T_q^{\text{AC-BV}} M} = T_q^C \bar{M}$. To this end we show inclusion both ways.

We show $\overline{T_q^{\text{AC-BV}} M} \subset T_q^C \bar{M}$. Since $M \subset \bar{M}$, and AC-BV convergence implies uniform convergence, it is clear that $T_q^{\text{AC-BV}} M \subset T_q^C \bar{M}$. Observe that $T_q^C \bar{M}$ is closed in $C([t_0, t_1], \mathbb{R}^d)$. It is a straightforward fact from topology now that the uniform closure of $T_q^{\text{AC-BV}} M$ is contained in $T_q^C \bar{M}$.

Now the other inclusion. We show $T_q^C \bar{M} \subset \overline{T_q^{\text{AC-BV}} M}$. Let $x(t) \in T_q^C \bar{M}$. We show $x(t)$ is in the uniform closure of $T_q^{\text{AC-BV}} M$. To this end we use the continuous version of Lemma 2.58, point **4**, in order to express $x(t)$ as a sum $\sum x_k$, where for $k = 1, 2, \cdots, N$, $x_k \in T_q^C \bar{M}$, and we gain all the other notational baggage indicated in the lemma.

For each $k = 1, 2, \cdots, N$, we show that x_k is in the uniform closure of $T_q^{\text{AC-BV}} M$. Since the uniform closure of $T_q^{\text{AC-BV}} M$ is a cone (we omit the proof that the closure of a cone is again a cone), it follows that x, being the sum of elements in that cone, is also in the uniform closure of $T_q^{\text{AC-BV}} M$.

Accordingly, choose $k \in \{1, 2, \cdots, N\}$. Define $v := x_k$ for convenience. Using the notation of Lemma 2.58, we have that v is supported on $\bar{U} := \bar{U}_k = [u_0, u_1]$. We

omit all k subscripts in what follows. We show that v is in the uniform closure of $T_q^{\text{AC-BV}}M$.

Define

$$z(t) := \begin{cases} DF|_{q(t)}(v(t)) & \text{whenever } t \in U \\ 0 & \text{otherwise} \end{cases}. \tag{2.30}$$

Claim. The components z_α of z, for $\alpha \in \mathcal{A}$ satisfy

$$z_\alpha(t) = \begin{cases} Df_\alpha|_{q(t)}(v(t)) & \text{whenever } t \in U \\ 0 & \text{otherwise} \end{cases}. \tag{2.31}$$

Note that $f_\alpha(\vec{r}) = \pi_\alpha \circ F(\vec{r})$ for all $\vec{r} \in \mathcal{N}$ (recall $\mathcal{N}$ is the domain of F, as we are using the notation from Lemma 2.58). Differentiate this to obtain $Df_\alpha = D\pi_\alpha\, DF = \pi_\alpha \circ DF = DF_\alpha$. Applying π_α to (2.30) and using these relations verifies the claim. *This completes the proof of the claim.*

Since $v \in T_q^C M$ and also because of (2.31), we may appeal to point **3** of the continuous version of Lemma 2.58 (which does not require inner endpoint conditions) in order to see that for each $\alpha \in \mathcal{A}$, the function z_α is in the tangent cone $T_{w_\alpha}^C \bar{W}_\alpha$, where $w_\alpha := f_\alpha(q(t))$ and

$$\bar{W}_\alpha := \{p \in C([t_0, t_1], [0, \infty)) : p(t_0) = w_\alpha(t_0) \text{ and } p(t_1) = w_\alpha(t_1)\}.$$

For $\alpha \in \mathcal{A}$, define

$$W_\alpha := \{p \in \text{AC-BV}([t_0, t_1], [0, \infty)) : p(t_0) = w_\alpha(t_0) \text{ and } p(t_1) = w_\alpha(t_1)\}.$$

It is straightforward to verify, for every $\alpha \in \mathcal{A}$, that $\bar{W}_\alpha$ is the uniform closure of W_α in the Banach space $C([t_0, t_1], \mathbb{R})$. By Lemma 2.60, for each $\alpha \in \mathcal{A}$,

$$z_\alpha \in \overline{T_{w_\alpha}^{\text{AC-BV}} W_\alpha}.$$

Claim: Suppose $\epsilon > 0$. For each $\alpha \in \mathcal{A}$, there exists a function $\tilde{z}_\alpha$ satisfying the following:

1. The function $\tilde{z}_\alpha$ is an ϵ-close uniform approximation of z_α:

$$|\tilde{z}_\alpha - z_\alpha|_\infty < \epsilon.$$

2. We have $\tilde{z}_\alpha \in T^{AC-BV}_{w_\alpha} W_\alpha$.

3. The function $\tilde{z}_\alpha$ obeys the *inner endpoint conditions*:

$$\text{Either } \lim_{t \to u_0^+} \dot{\tilde{z}}_\alpha(t) = 0 \text{ or else } u_0 = t_0; \text{ and} \tag{2.32}$$

$$\text{either } \lim_{t \to u_1^-} \dot{\tilde{z}}_\alpha(t) = 0 \text{ or else } u_1 = t_1.$$

Apart from **3**, the claim holds because z_α is in the uniform closure of $T_{w_\alpha} W_\alpha$ for $\alpha \in \mathcal{A}$. This enables us to approximate the z_α uniformly with elements of $T_{w_\alpha} W_\alpha$. To see that **3** can be satisfied as well, we examine the proof of Lemma 2.60 and the remark of Proposition 2.59. *This completes the proof of the claim.*

Let $\epsilon > 0$ and construct $\tilde{z}_\alpha$ for $\alpha \in \mathcal{A}$ as in the above claim. For all remaining $d - |\mathcal{A}|$ indices α, we construct $\tilde{z}_\alpha \in C^\infty([t_0, t_1], \mathbb{R})$ to be an ϵ-close uniform approximation of $z_\alpha(t)$ supported on $\bar{U}$, obeying $\tilde{z}_\alpha(t_0) = z_\alpha(t_0)$, $\tilde{z}_\alpha(t_1) = z_\alpha(t_1)$, and also satisfying the inner endpoint conditions (2.32). Proposition 2.59 shows that this may be done.

Let $\tilde{z} \in \text{AC-BV}([t_0, t_1], \mathbb{R}^d)$ be the concatenation (cartesian product) of the functions $\tilde{z}_\alpha \in \text{AC-BV}([t_0, t_1], \mathbb{R})$, $\alpha = 1, 2, \cdots, d$. See that $\tilde{z}$ is supported on $\bar{U}$ and

satisfies the inner endpoint conditions. Now define

$$\tilde{v}(t) := \begin{cases} DF^{-1}|_{F(q(t))}\tilde{z}(t) & \text{whenever } t \in U \\ 0 & \text{otherwise} \end{cases}. \tag{2.33}$$

By Lemma 2.55, it follows from the fact that F^{-1} is C^3 that $DF^{-1}|_{q(t)}$ is AC-BV. By Lemma 2.56 and a straightforward pasting argument, $\tilde{v}$ is AC-BV. Also, see that $\tilde{v}$ is supported on $\bar{U}$ and satisfies the inner endpoint conditions (which it inherits from $\tilde{z}$ in a straightforward manner):

$$\text{Either } \lim_{t\to u_0^+} \dot{\tilde{v}}(t) = 0 \text{ or else } u_0 = t_0; \text{ and} \tag{2.34}$$

$$\text{either } \lim_{t\to u_1^-} \dot{\tilde{v}}(t) = 0 \text{ or else } u_1 = t_1.$$

Claim. We have the following: For each $\alpha \in \mathcal{A}$,

$$\tilde{z}_\alpha = \begin{cases} Df_\alpha|_{q(t)}\tilde{v}(t) & \text{whenever } t \in U \\ 0 & \text{otherwise} \end{cases}. \tag{2.35}$$

Observe that $Df_\alpha|_{q(t)} = DF_\alpha|_{q(t)} = \pi_\alpha \circ DF|_{q(t)}$. Hence, for any $t \in U$, we have

$$Df_\alpha|_{q(t)}\tilde{v}(t) = Df_\alpha|_{q(t)}DF^{-1}|_{F(q(t))}\tilde{z}(t) = \pi_\alpha DF|_{q(t)}DF^{-1}|_{F(q(t))}\tilde{z}(t) = \pi_\alpha\tilde{z}(t) = \tilde{z}_\alpha(t).$$

For $t \notin U$ the calculation is trivial. *This completes the proof of the claim.*

By the AC-BV version of Lemma 2.58, point **3**, it follows from (2.34), (2.35), the fact that $\tilde{v}$ is AC-BV and supported on $\bar{U}$, and the fact $\tilde{z}_\alpha \in T_{w_\alpha}W_\alpha$ for all $\alpha \in \mathcal{A}$ that

$$\tilde{v} \in T_q^{\text{AC-BV}}M.$$

It follows straightforwardly from (2.30) that

$$v(t) = \begin{cases} DF^{-1}|_{F(q(t))}z(t) & \text{whenever } t \in U \\ 0 & \text{otherwise} \end{cases}. \tag{2.36}$$

We wish to show that $\tilde{v}$ approximates v. From (2.33) and (2.36) we may obtain

$$\tilde{v} - v = \begin{cases} DF^{-1}|_{F(q(t))}(\tilde{z} - z) & t \in \bar{U} \\ 0 & \text{otherwise} \end{cases}.$$

By Lemma 2.55, $DF^{-1}|_{F(q(t))}$ is in AC-BV$([u_0, u_1], \mathbb{R}^{d^2})$. By Lemma 2.56, the mapping $\mathcal{F}$ given by $(\mathcal{F}[u])(t) := DF^{-1}|_{F(q(t))}u(t)$ is a bounded linear operator on AC-BV$([u_0, u_1], \mathbb{R}^d)$. It follows that

$$|\tilde{v} - v|_\infty \leq ||\mathcal{F}||\ |\tilde{z} - z|_\infty.$$

Since $|\tilde{z} - z|_\infty < \epsilon$, and ϵ may be made arbitrarily small, it follows that $\tilde{v}$ may be made arbitrarily close to v in the uniform metric as we decrease $\epsilon > 0$.

In other words, we have approximated v arbitrarily closely with an element of $T_q^{\text{AC-BV}}M$. We conclude that v is in the uniform closure of $T_q^{\text{AC-BV}}M$. The proof is complete. □

We are now finally ready to prove Lemma 2.24, which we repeat here with all notation and hypotheses spelled out:

LEMMA 2.62 (Proof of 2.24). *Define M to be the subset of AC-BV$([t_0, t_1], \mathbb{R}^d)$ satisfying the constraints*

$$f_\alpha(q(t)) \geq 0 \text{ for all } \alpha \in J,$$

and also fixed endpoint conditions $q(t_0) = a$ and $q(t_1) = b$, for $a, b \in \mathbb{R}^d$. Suppose $q \in M$. Assume that the constraints f_α, $\alpha \in J$ are C^3 and for any $\vec{r} \in \mathbb{R}^d$, the set of active constraints $\mathcal{A}$ (those $\alpha \in J$ such that $f_\alpha(\vec{r}) = 0$) has the property that

$$\{\nabla f_\alpha\}_{\alpha \in \mathcal{A}} \text{ is linearly independent.}$$

Define the following collection of bounded linear functionals on $C([t_0, t_1], \mathbb{R}^d)$*:*

$$\begin{aligned}\mathcal{C}_q := \quad & \{h \mapsto Df_\alpha|_{q(t)}h(t) : \alpha \in Z \text{ and } t \in [t_0, t_1] \text{ such that } f_\alpha \circ q(t) = 0\} \\ \cup \ & \{h \mapsto \pm\hat{e}_i \cdot h(t_0) : i = 1, 2, \cdots d\} \\ \cup \ & \{h \mapsto \pm\hat{e}_i \cdot h(t_1) : i = 1, 2, \cdots d\}.\end{aligned}$$

The cone of $\mathcal{C}_q$ *is precisely the uniform closure of the tangent cone* T_qM*:*

$$\overline{T_qM} = \text{ cone of } \mathcal{C}_q.$$

Proof. By Lemma 2.61, the uniform closure of T_qM is $T_q\bar{M}$. We show $T_q\bar{M}$ is the cone of $\mathcal{C}_q$. We prove set equality by showing inclusion both ways.

Part 1. We show that $T_q\bar{M}$ is contained in the cone of $\mathcal{C}_q$.

We show that every geometric tangent vector in $T_q\bar{M}$ is contained in the cone of $\mathcal{C}_q$. Since the cone of $\mathcal{C}_q$ is closed and convex in $C([t_0, t_1], \mathbb{R}^d)$, it follows that $T_q\bar{M}$ is contained in the cone of $\mathcal{C}_q$.

Let x be a geometric tangent vector in $T_q\bar{M}$. We show x is in the cone of $\mathcal{C}_q$. Since x is a geometric tangent vector in $T_q\bar{M}$, there exists a sequence of continuous functions (q_n) in $\bar{M}$ converging uniformly to q such that

$$x = \lim_{n\to\infty} \frac{q_n - q}{|q_n - q|_\infty},$$

where the limit is to be understood in the uniform sense.

In order to show that x is in the cone of $\mathcal{C}_q$, it suffices to show that for every linear functional ℓ in $\mathcal{C}_q$, $\ell(x) \geq 0$.

Since the (q_n) all share the same endpoint values, it follows that $x(t_0) = x(t_1) = 0$. Accordingly, $\ell(x) = 0$ for $\ell : h \mapsto \pm\hat{e}_i \cdot x(t^*)$, $i = 1, 2, \cdots, n$, and $t^* = t_0, t_1$.

All remaining linear functionals in $\mathcal{C}_q$ may be identified by a pair (α, s), such that $\alpha \in Z$ is a constraint index and $s \in [t_0, t_1]$ such that $f_\alpha(q(s)) = 0$. Then $\ell : x \mapsto \nabla f_\alpha(q(s)) \cdot x(s)$ is a linear functional in $\mathcal{C}_q$. All that remains is to show $\ell(x) \geq 0$ for this case.

Since the linear functional ℓ is bounded, it is continuous [19]. Hence we may calculate

$$\ell(x) = \lim_{n\to\infty} \frac{\ell(q_n) - \ell(q)}{|q_n - q|_\infty} = \lim_{n\to\infty} \frac{\nabla f_\alpha|_{q(s)} \cdot q_n(s) - \nabla f_\alpha|_{q(s)} \cdot q(s)}{|q_n - q|_\infty}.$$

Since $f_\alpha(q(s)) = 0$ and $f_\alpha(q_n(s)) \geq 0$, it follows that

$$f_\alpha(q_n(s)) - f_\alpha(q(s)) \geq 0. \tag{2.37}$$

Since f_α is differentiable,

$$f(q_n(s)) - f(q(s)) = \nabla f_\alpha(q(s)) \cdot (q_n(s) - q(s)) + E_n, \tag{2.38}$$

where, by the definition of differentiability, the error E_n satisfies

$$\lim_{n\to\infty} \frac{E_n}{|q_n(s) - q(s)|} = 0.$$

We may now arrive at the expression

$$\ell(x) = \lim_{n\to\infty} \frac{f_\alpha(q_n(s)) - f_\alpha(q(s))}{|q_n - q|_\infty}.$$

Notice that $f_\alpha(q(s)) = 0$, or else (α, s) doesn't correspond to a functional ℓ in $\mathcal{C}_q$. Note that $f_\alpha(q_n(s)) \geq 0$ for all n, since the q_n are in $\bar{M}$. It follows that $\ell(x)$ may be

realized as the limit of a non-negative sequence of real numbers, hence $\ell(x) \geq 0$. We have finished Part 1.

Part 2. We show that the cone of $\mathcal{C}_q$ is contained in $T_q\bar{M}$.

Let $x(t)$ be in the cone of $\mathcal{C}_q$. We show x is in $T_q\bar{M}$.

By point **4** of the continuous version of Lemma 2.58, we see that x is in $T_q\bar{M}$ provided that

$$v(t) := \chi_k(t)x(t), \text{ for } t \in [t_0, t_1],$$

is in $T_q\bar{M}$ for each $k = 1, 2, \cdots, N$. Here, $\{\chi_k\}$ is the partition of unity in Lemma 2.58; we borrow all pertinent notation and objects. In particular, define all the objects in the scope of Lemma 2.58 as applied to $q \in M$.

Let $k \in \{1, 2, \cdots, N\}$. We show $v \in T_q\bar{M}$. From here on out, we omit k-subscripts. We show $v \in T_q\bar{M}$. By point **3** of Lemma 2.58 applied to the continuous case, this will be true if for each $\alpha \in \mathcal{A}$, the function

$$z_\alpha = \begin{cases} Df_\alpha|_{q(t)}x(t) & \text{whenever } t \in U \\ 0 & \text{otherwise} \end{cases} \tag{2.39}$$

is in the tangent cone at $w_\alpha(t) := f_\alpha(q(t))$ of

$$\bar{W}_\alpha := \{p \in C([t_0, t_1], [0, \infty)) : p(t_0) = f_\alpha(q(t_0)) \text{ and } p(t_1) = f_\alpha(q(t_1))\}.$$

Claim. Suppose $z \in C([t_0, t_1], \mathbb{R})$ such that $z(t_0) = z(t_1) = 0$ and $z(t) \geq 0$ in a neighborhood of the set of $\{t \in [t_0, t_1] : f_\alpha(q(t)) = 0\}$. Then

$$z \in T_{w_\alpha}\bar{W}_\alpha.$$

Let z be is as in the claim. We show that $z \in T_{w_\alpha}\bar{W}_\alpha$. A straightforward consequence of the hypothesis of the claim is that there exists a closed set $V \subset [t_0, t_1]$ such that $w_\alpha(t) > 0$ for $t \in V$ and $z(t) \geq 0$ for $t \notin V$.

By the extreme value theorem, let $w_{\min} > 0$ be the minimum value attained by $w_\alpha(t)$ for $t \in V$ and let $z_{\min}$ (which may be negative) be the minimum value attained by $z(t)$ for $t \in V$. Choose a sequence (δ_n) of positive reals tending to zero. Define

$$w_{\alpha,n}(t) = w_\alpha(t) + \delta_n z(t).$$

We show that $w_{\alpha,n}(t) \geq 0$ for all $t \in [t_0, t_1]$. Suppose first that $t \notin V$. Then $z(t) \geq 0$. Consequently, $w_{\alpha,n}(t) \geq 0$. Now suppose that $t \in V$. Then for sufficiently large n, $\delta_n < |\frac{w_{\min}}{z_{\min}}|$, and it is straightforward to show that

$$w_{\alpha,n}(t) = w_\alpha(t) + \delta_n z(t) \geq w_{\min} + \delta_n z_{\min} \geq 0.$$

It follows from $z(t_0) = z(t_1) = 0$ that $w_{\alpha,n}(t_0) = w_\alpha(t_0)$, $w_{\alpha,n}(t_1) = w_\alpha(t_1)$. Hence (w_n) is a sequence in $\bar{W}$. It is clear that (w_n) converges uniformly to w. We conclude that $z \in T_{w_\alpha}\bar{W}_\alpha$. *This completes the proof of the claim.*

Now we use our claim in order to show that $z_\alpha \in T_{w_\alpha}W_\alpha$ which will complete the proof.

Observe that since x is in the cone of $\mathcal{C}_q$, it follows that v is in the cone of $\mathcal{C}_q$:

$$\text{For each } \ell \in \mathcal{C}_q,\ \ell(v) = \ell(\chi x) = \chi(s)\nabla f_\alpha|_{q(s)} \cdot x(s) \geq 0.$$

It follows from this fact, (2.39), and the definition of $\mathcal{C}_q$ that for each $\alpha \in \mathcal{A}$, for all $t \in [t_0, t_1]$, we have $z_\alpha(t) \geq 0$ whenever $w_\alpha(t) = f_\alpha(q(t)) = 0$.

Let $\epsilon > 0$. Define $\phi : \mathbb{R} \to \mathbb{R}$ to be a continuous function satisfying $|\phi(s) - s| < \epsilon$ for all $s \in \mathbb{R}$ and also $\phi(s) = 0$ for all s sufficiently close to 0.

Define $z(t) := \phi(z_\alpha(t))$. It is clear that z is continuous, is within ϵ distance to z_α in the uniform metric, and also that $z(t_0) = z(t_1) = 0$.

If we can apply the last claim to z, then we will have shown $z \in T_{w_\alpha}\bar{W}_\alpha$. Since z may be constructed to be arbitrarily close to z_α in the uniform metric by letting $\epsilon \to 0^+$, it follows that z_α is in the uniform closure of $T_{w_\alpha}\bar{W}_\alpha$. But $T_{w_\alpha}\bar{W}_\alpha$ is already a closed set, so we will have shown $z_\alpha \in T_{w_\alpha}\bar{W}_\alpha$ and the lemma will be shown. Hence, all that remains to verify the conditions of the previous claim for $z(t)$. The only condition left to verify is the following:

Claim. The inequality $z(t) \geq 0$ holds in a neighborhood of $\{t \in [t_0, t_1] : f_\alpha(q(t)) = 0\}$.

Let $V = \{t \in [t_0, t_1] : f_\alpha(q(t)) = 0\}$. We show $z(t) \geq 0$ in a neighborhood of V. Recall that $z_\alpha(t) \geq 0$ for all $t \in V$. Let $\delta > 0$ such that $\phi(s) = 0$ whenever $|s| < \delta$. Define A to be the set of times

$$A := \{t \in [t_0, t_1] : |z_\alpha(t)| < \delta\}.$$

Observe that A is an open subset of $[t_0, t_1]$ since it is the preimage of $(-\delta, \delta)$ through the continuous function z_α. Note that A contains V. For all $t \in A$, $z(t) = \phi(z_\alpha(t)) = 0$. Hence $z(t) = 0$ in a neighborhood of V, which suffices. *This completes the proof of the claim.* □

Summary

We have derived the measure-theoretic Euler-Lagrange equations by stipulating a generalized version of the usual Principle of Stationary Action. The generalization is analogous to the generalization one makes when writing optimality conditions for constrained problems rather than unconstrained problems.

Related Work

Using a variational principle directly with constrained function spaces has been done before; see [11]. However, the principle given is the usual one, and consequently, the impacts are conservative. Our method allows for a wider class of motions; for example, our principle could allow inelastic collisions.

Our principle does not give unique motions. It leaves much freedom for impacts. One would need to supplement this theory with a so-called *constitutive law* describing the impacts to even begin to hope for uniqueness of solutions. Even when this is done, we refer the reader to examples of Schatzman and Ballard in [34], [35], and [6] to see that uniqueness might not be obtained.

We did not address the problem of existence for the measure-theoretic Euler-Lagrange equations. Along these lines, we mention the work of Moreau [24], who did not give existence theory, but did invent the concept of a measure differential inclusion (his framework for the measure-theoretic constructs necessary in problems with unilateral constraints) and gave a numerical method for solving it – the so-called

sweeping process and *catching up algorithm* [18]. This numerical method would later prove to be an important tool for both numerical techniques and existence proofs in related work; see [38], also [9]. The first rigorous existence results for such problems as ours were given Monteiro Marques [23]. Later work was done by Mabrouk [22] and Ballard [6].

Open Questions

In closing the chapter, we consider a few interesting questions. First, we ask if constitutive laws for impacts may be incorporated into our scheme. For example, problems in the calculus of variations often come with so-called *subsidiary conditions* (see, for example, [13]). Is there a way to phrase a constitutive law like elasticity, inelasticity, or a coefficient of restitution law as a subsidiary condition? Secondly, we ask whether or not it is possible to use this result in order to obtain existence results. Existence results for measure differential inclusions have been obtained by analyzing the convergence of numerical schemes (see [18]). Can the variational approach lead to non-constructive existence proofs?

CHAPTER 3

LINEAR COMPLEMENTARITY PROBLEMS AND FRICTIONLESS, COMPLETELY INELASTIC UNILATERAL CONSTRAINTS

Introduction

In the current chapter, we consider a modification of Hamiltonian dynamics designed to accommodate frictionless unilateral constraints. This modification allows for the phenomena of impact and reactive forces in the directions normal to the contact surfaces. Additionally, we demand a so-called *constitutive law*: impacts must be *inelastic*.

Inelastic impacts are energetically maximally dissipative: the impact is selected among candidates such that the energy immediately after the time of impact ($H(t_i^+)$, where H is the Hamiltonian and t^+ indicates a right-handed limit) is minimized. Indeed, we may take this to be the definition of inelastic. However, there is an equivalent manner of specifying the inelastic constitutive law, based on the so-called *complementarity formulation*, which we will employ instead. We show that maximal dissipation is equivalent to the satisfaction of complementarity conditions which essentially assert that no impact involves an impulsive force at a contact which consequently releases as a result of the impact.

Given the specification of a unilaterally constrained Hamiltonian system, will be able to find unique reactive forces and impacts at any time. Moreover, we will see that

the determination of impact or force at a given time is a well-posed problem. Well-posedness in the entire evolution problem, however, does not follow. In particular, the possibility of *right accumulations of impacts* allows for pathological non-uniqueness examples. We give such an example, due to Ballard [6].

This chapter relates to the rest of the thesis as follows. Like the analysis of Chapter 2, we deal with a dynamical system in which we have inserted unilateral constraints which are to be frictionless. Unlike Chapter 2, we make no attempt to derive our dynamical rules from an underlying principle: we simply postulate them. What we postulate is similar to the measure-theoretic Euler-Lagrange equations of Chapter 2, except it is translated to the case of a Hamiltonian system and we introduce an additional condition on the impulsive forces. Namely, we assume that singular part of the measures are supported on a countable set of times[1]. Also, we have a constitutive law (the inelastic assumption) which was a missing ingredient in Chapter 2. This allows us to solve for reactive forces and impulses. In order to do this, we introduce a tool, the *linear complementarity problem*, or LCP, which finds application in many optimization problems. We also see in Chapter 4 that the LCP is used in the frictional contact algorithms in the literature.

The Linear Complementarity Problem.

We introduce a vital tool: the *linear complementarity problem.*

[1]Prohibiting, for example, the Cantor measure

Motivation

In the study of optimization problems we are very familiar with the usual idea: set the derivative of the cost functional equal to zero and solve. The situation is complicated by constraints. When we constrain an optimization problem, we have to worry about local extrema on the boundary of the feasible set. Accordingly, we have to generalize our notion of "setting the derivative equal to zero." The correct generalization is to assert that the derivative of the cost functional (which is vector-valued) is in *the positive span of active constraint normals.*

In practical problems (such as linear programming or quadratic programming) the constraints are often linear; in any case one associates to every constraint in such a problem a *Lagrange multiplier* $s_k \geq 0$. The multiplier is to act as a coefficient when one writes the derivative of the optimization cost functional as a linear combination of constraint normals. In particular, a Lagrange multiplier may be non-zero only when the corresponding constraint is "active."

For example, we may have an optimization problem with constraints such that the kth constraint is given by $x_k \geq 0$. We say this constraint is active when $x_k = 0$. It is this case, when the kth constraint is active, that may we have $s_k > 0$. Otherwise, we must have $s_k = 0$. The relationship between x_k and s_k is called *complementarity* and we write

$$0 \leq x_k \perp s_k \geq 0.$$

Often we want an entire vector $x \in \mathbb{R}^n$ to be complementary to a vector $y \in \mathbb{R}^n$. For each $k = 1, 2, \cdots, n$, we have $0 \leq x_k \perp y_k \geq 0$. However, this may be conveniently expressed as $x, y \geq 0$, $x^T y = 0$. Note that when we say $z \geq 0$ for some vector z, we mean each component of z is non-negative.

Example. One wishes to minimize $c^T x$ for $x \in K$, where K is the set given by

$$K := \{x \in \mathbb{R}^n : Ax + b \geq 0, \text{ and } x \geq 0\}.$$

Here A is $k \times n$ matrix, b is a k-vector, and c is an n-vector.

It turns out, using the optimality ideas just discussed, (see [28], for example) that $x \in \mathbb{R}^n$ is an optimizer in K if and only if there exists $s \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}^k$ such that

$$Ax + b = y, \tag{3.1}$$

$$A^T \lambda + s, = c \tag{3.2}$$

$$x, s \geq 0 \quad \text{and} \quad x^T s = 0. \tag{3.3}$$

$$y, \lambda \geq 0 \quad \text{and} \quad y^T \lambda = 0 \tag{3.4}$$

Such a list of conditions are often called the KKT conditions after Karush, Kuhn, and Tucker [28]. These equations may be written as

$$\begin{bmatrix} y \\ s \end{bmatrix} - \begin{bmatrix} 0 & A \\ -A^T & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ x \end{bmatrix} = \begin{bmatrix} b \\ c \end{bmatrix}.$$

If we write the matrix

$$M := \begin{bmatrix} 0 & A \\ -A^T & 0 \end{bmatrix},$$

and define

$$X := \begin{bmatrix} \lambda \\ x \end{bmatrix}, \text{ and } Y := \begin{bmatrix} y \\ s \end{bmatrix},$$

and also

$$Q := \begin{bmatrix} b \\ c \end{bmatrix},$$

the above becomes $Y - MX = Q$, $X, Y \geq 0$, $X^T Y = 0$.

DEFINITION 3.1. *Given a linear operator M on $\mathbb{R}^n$ and a vector Q in $\mathbb{R}^n$, the* linear complementarity problem*, LCP(M, Q), is the task of finding $x, y \in \mathbb{R}^n$ such that*

$$x - My = Q, \quad x \geq 0, \quad y \geq 0, \quad \textit{and } x^T y = 0.$$

Numerical Solutions

Solving an LCP is possible, at least when the kernel M is *copositive*, by the application of Lemke's algorithm [20]. Lemke's algorithm is a popular pivot-based method. Asymptotically better methods exist, however, known as *interior point methods.* The reader is referred to [27] and [28] for more details.

Applications

We motivated the LCP earlier through a linear program; in fact they also have application in quadratic programming. Indeed, the LCP unifies the two problems of linear and quadratic programming into a single, coherent framework.

Our current applications of the LCP are to resolve force and impact problems arising in mechanical problems with unilateral constraints that are frictionless and

have completely inelastic impacts. The LCP structure arises naturally from the complementarity conditions between the forces at contacts and the normal accelerations at contacts; we cannot push if the contact is accelerating off the constraint. An analogous situation exists for the case of impact. Here, the impulses at contacts and resultant normal velocity are the complementary variables. We will see shortly how both of these problems may be formulated as LCP's.

Hamiltonian Dynamics with Inelastic Unilateral Constraints

We consider the Hamiltonian system

$$H(t,p,q) = \frac{1}{2}p^T A(t,q)p + b^T(t,q)p + c(t,q). \tag{3.5}$$

Here, A, b, and c are differentiable matrix, vector, and scalar valued (respectively) functions of the time variable t and the position variables $q \in \mathbb{R}^d$. We assume that A is a $d \times d$ symmetric positive definite (SPD) matrix, that b is $\mathbb{R}^d$-valued, and c is scalar-valued. We refer to $p \in \mathbb{R}^d$ as the momentum, $q \in \mathbb{R}^d$ as the position, and a pair (p,q) we may refer to as a *state of the system.* We take the dynamical system to have $2d$ variables; d position coordinates and d momentum coordinates.

In addition to the Hamiltonian, we have a collection of constraints $\{f^k\}_{k\in J}$, and also $\{g^k\}_{k\in K}$ for some index sets J and K. Here, each f^k, $k \in J$ is a scalar-valued C^2 function on $\mathbb{R}^d$. Similarly for the g^k. The functions of the collection $\{f^k\}_{k\in J}$ encode unilateral constraints. The functions of the collection $\{g^k\}_{k\in K}$ encode bilateral constraints. In particular, the allowed positions which are said to *obey the constraints*

are those points $q \in \mathbb{R}^d$ such that

$$f^k(q) \geq 0 \text{ for all } k \in J, \text{ and also } g^k(q) = 0 \text{ for all } k \in K.$$

The equations of motion of a Hamiltonian system subjected to external forces F are given by

$$\dot{q} = \frac{\partial H}{\partial p} \qquad \dot{p} = F - \frac{\partial H}{\partial q}. \tag{3.6}$$

The external forces F are reactive forces and impulses. We are to interpret (3.6) measure-theoretically, choosing the reactive forces and impulses F to obey the constitutive law. We now formulate this precisely.

Precise Formulation

We now define what we mean by saying a motion $q : (t_i, t_f) \to \mathbb{R}^d$ is a solution to *inelastic dynamics*, given by Hamiltonian H as in Equation (3.5), unilateral constraints $f^k(q) \geq 0$ for $k \in J$, and bilateral constraints $g^k(q) = 0$, $k \in K$. Here, the constraints $\{f^k\}$ and $\{g^k\}$ are both assumed to be C^2 functions from $\mathbb{R}^d$ to $\mathbb{R}$. Further, we assume that the bilateral constraints g^k have linearly independent gradients.

Definition 3.2. *We say that $q : (t_i, t_f) \to \mathbb{R}^d$ is a* solution to the inelastic dynamics *provided the following hold:*

1. *The function q satisfies the unilateral and bilateral constraints for all $t \in (t_i, t_f)$:*

$$f^k(q(t)) \geq 0 \text{ for all } k \in J, \text{ and also } g^k(q(t)) = 0 \text{ for all } k \in K.$$

2. *The function q is absolutely continuous and admits a derivative of bounded variation v, which we take to be continuous from the right. Note this implies that the momentum $p(t)$, which is given by*

$$p(t) = A^{-1}(t, q(t))v(t) - b(t, q(t)),$$

is of bounded variation and continuous from the right as well.

3. *The Borel-Stieltjes measure dp satisfies the measure-theoretic equation*

$$dp = F - \frac{\partial H}{\partial q}\, dt,$$

where dt is Lebesgue measure and F is a vector-valued finite measure.

4. *The vector-valued finite measure F is of the form*

$$F = \sum_{k=1}^{n_1} \nabla f_k(q(t))\mu_k + \sum_{k=1}^{n_2} \nabla g_k(q(t))\nu_k,$$

where the μ_k are finite non-negative measures supported on the times which $f_k(q(t)) = 0$, and the ν_k are finite signed measures supported on the times which $g_k(q(t)) = 0$.

5. *The measures μ_k and ν_k may each be decomposed into an absolutely continuous part and a countable sum of point-masses (Dirac delta functions):*

$$a(t)dt + \sum a_i \delta(t - t_i),$$

where $a(t)$ is integrable.

6. *The measure μ_k may not have mass at time t unless $f_k(q(t)) = 0$ and $\nabla f_k(q(t)) \cdot v(t^+) = 0$, where $v(t^+) := \lim_{s \to t^+} v(s)$.*

7. *The measure ν_k may not have mass at time t unless $g_k(q(t)) = 0$ and $\nabla g_k(q(t)) \cdot v(t^+) = 0$, where $v(t^+) := \lim_{s \to t^+} v(s)$.*

We understand the previous definition as follows. Condition (1) enforces the unilateral and bilateral constraints. Conditions (2) and (3) generalize the Hamiltonian dynamics to the unilaterally constrained setting by allowing measure-theoretic forces. Conditions (4) and (5) tell us the specific form of the reactive forces from the constraints we allow. Conditions (6) and (7) enforce the concept of *complementarity*, which prohibits impacts which would release the impacting constraint. Later, we will see that this is what causes the solution to be energetically maximally dissipative (inelastic).

Impact Determination using LCP's

The notation in the following two sections involves many indices. The Einstein summation convention (as in the Ricci Calculus) is in force: repeated indices imply summation.

Consider a Hamiltonian system as described above constrained by the relations $f^k(q) \geq 0$ and $g^k(q) = 0$. We say that the system is *in a state of impact at time t_0* when for some $k \in J$,

$$f^k(q(t_0)) = 0 \text{ and } \left(\frac{d}{dt} f^k(q(t_0))\right)^- = \frac{\partial f^k}{\partial q^\mu} \left(\frac{dq^\mu}{dt}\right)^- = \frac{\partial f^k}{\partial q^\mu} \left(\frac{\partial H}{\partial p^\mu}\right)^- < 0.$$

Notice we have suppressed writing some of the arguments for convenience. The "super-script minus" notation means *in the left-handed sense.* Note that since $q(t)$ has a derivative of bounded variation, this notion always exists. We must use this sense, since in the right-handed sense the impact has already occurred and the impact has been "resolved." That an impact must occur should be clear, for if $q(t)$ is to continue to obey the constraint then a velocity discontinuity is required.

We wish to find an impact at time t_0 which satisfies the requirements of Definition 3.2. To this end, we posit the following rules:

Impact Determination Rules.

1. We take change of state to be a change in momentum coordinates only, in fact of the form

$$\Delta p = (Df)^T s + (Dg)^T \lambda.$$

2. The final value of $\dot{f}$ is in the positive orthant; the final value of $\dot{g}$ is the zero vector.

3. We assume $s \geq 0$.

4. We assume that whenever s^k is strictly greater than zero, the resulting value of $\dot{f}^k$ must be zero..

Justification of Rules (1)-(4). We justify these rules by appealing to Definition 3.2. In particular, impulses deal with the "Dirac deltas" one finds in this definition. We point out the correspondence between the Lagrange multipliers s_k and the coefficients

to the Dirac deltas in the measure μ_k; similarly we point out the correspondence between λ_k and ν_k. From these correspondences, (1)-(4) follow. We now comment on the interpretation of these rules: The first assumption states that an impulse must be driven by impulsive forces normal to the constraints. This corresponds to the frictionless assumption. The second assumption is necessary to prevent constraint violation in the state resulting from impact. The third assumption states that the unilateral constraints may only push away, they may never pull in. They are not sticky! The fourth assumption, which we call complementarity of s and $\dot{f}$, encodes the notion of inelastic impact. It says that a unilateral constraint only pushes as hard as it needs to in order to prevent the constraint from being violated, but no harder.

Now that we have justified (1)-(4), we use these rules to formulate an LCP which an impact specified by the dynamics of Definition 3.2 must be a solution to.

Because of the nature of impacts, we have a discontinuity of the momentum p at a fixed time. We write the initial (or left-handed limit) momentum as $p := p(t_0^-)$, and we write the final (or right-handed limit) momentum as $p + \Delta p := p(t_0^+)$.

We begin with the following expressions for $\dot{f}^k$ and $\dot{g}^k$:

$$\dot{f}^k = \frac{\partial f^k}{\partial q^\mu}\frac{\partial H}{\partial p_\mu} \text{ and } \dot{g}^k = \frac{\partial g^k}{\partial q^\mu}\frac{\partial H}{\partial p_\mu}. \tag{3.7}$$

These equations hold whether interpreted in the left-hand or right-hand sense. Note that if we interpret them in the left-handed sense, we mean that H_p is being evaluated at $(t_0, p(t_0^-), q(t_0))$, but if we interpret it in the right-handed sense H_p is being evaluated at $(t_0, p(t_0^+), q(t_0)) = (t_0, p(t_0^-) + \Delta p, q(t_0))$.

Define Q^k to be the left-handed limit of $\dot{f}^k$:

$$Q^k = \frac{\partial f^k}{\partial q^\mu}\frac{\partial H}{\partial p_\mu}\bigg|_{(t_0,p(t_0^-),q(t_0))}. \tag{3.8}$$

Take $\dot{f}^k$ and $\dot{g}^k$ to be defined in the right-handed sense. We express them in terms of Δp. By Equation 3.5, we have

$$\frac{\partial H}{\partial p^\mu} = [A(p+\Delta p)+b]_\mu.$$

Substitution of this into Equation 3.7, usage of Equation 3.8, and writing Δp according to Impact Rule (1) yields

$$\dot{f}^k = \frac{\partial f^k}{\partial q^\mu}A_{\mu\nu}(\frac{\partial f^j}{\partial q^\nu}s_j + \frac{\partial g^j}{\partial q^\nu}\lambda_j) + Q^k$$

and

$$\dot{g}^k = \frac{\partial g^k}{\partial q^\mu}A_{\mu\nu}(\frac{\partial f^j}{\partial q^\nu}s_j + \frac{\partial g^j}{\partial q^\nu}\lambda_j) = 0.$$

We make some notational abbreviations. We concatenate all of the $\dot{f}^k$ into a single vector $\dot{f}$, and similarly we define $\dot{g}$ and Q. We also define the four matrices

$$(M_{\alpha\beta})_{jk} := \frac{\partial \alpha^j}{\partial q^\mu}\frac{\partial^2 H}{\partial p^\mu \partial p^\nu}\frac{\partial \beta^k}{\partial q^\nu},$$

where the symbols α and β are to be either f or g.

Using these abbreviations, we may write the previous expressions as

$$\dot{f} = M_{ff}\,s + M_{fg}\,\lambda + Q \geq 0. \tag{3.9}$$

$$\dot{g} = M_{gf}\,s + M_{gg}\,\lambda = 0. \tag{3.10}$$

From Equation 3.10, we may solve for λ:

$$\lambda = -M_{gg}^{-1} M_{gf}\, s. \tag{3.11}$$

Substituting Equation 3.11 into Equation 3.9 yields

$$(M_{ff} - M_{fg} M_{gg}^{-1} M_{gf}) s + Q \geq 0.$$

Define

$$M := M_{ff} - M_{fg} M_{gg}^{-1} M_{gf}. \tag{3.12}$$

By Impact Rule (4), there is a complementarity condition between $\dot{f}$ and s. By Impact Rules (2) and (3), $\dot{f}$ and s are non-negative. We have

$$\dot{f} - Ms = Q, \quad \dot{f}, s \geq 0 \quad \dot{f}^T s = 0.$$

This is the linear complementarity problem LCP(M, Q). We call this *the LCP for impact determination.*

We have now accomplished the goal of this section, but before we move on to the analogous problem of determining reactive forces (rather than impulses), we stop to prove our claim that the LCP solution we obtain (which was based on the complementarity assumption) really is equivalent to maximizing dissipation due to the impact:

PROPOSITION 3.3. *Let (p, q) be some momentum-position state satisfying unilateral and bilateral constraints as given above. Let $\Omega \subset \mathbb{R}^d$ be the closed, convex set of*

impulses which may be obtained via

$$\sum_{f^k(q)=0} \nabla f^k|_q s_k, \text{ where } s_k \geq 0 \text{ for all } k.$$

Let $\phi : \mathbb{R}^d \to \mathbb{R}$ *be the scalar valued function such that*

$$\phi(\Delta p) := H(t, p + \Delta p, q).$$

Then there is a unique minimizer Δp *of* ϕ *on* Ω*. Moreover, given any solution* $(\dot{f}, s)$ *of the linear complementarity problem* $LCP(M, Q)$*, we have the formula*

$$\Delta p = (Df)^T s - (Dg)^T M_{gg}^{-1} M_{gf} s.$$

Proof. That ϕ admits a unique minimizer follows from the fact that $H(t, p + \Delta p, q)$ is a quadratic functional with a symmetric positive definite kernel, and the feasible set Ω of the optimization problem is convex. In particular (redefining ϕ by adding a constant, which does not affect the optimizers) we may write

$$\phi(\Delta p) = \frac{1}{2}(\Delta p)^T A(\Delta p) + (\Delta p)^T(Ap + b).$$

We show that if $(\dot{f}, s)$ is a solution to $LCP(M, Q)$, then

$$\Delta p := (Df)^T s - (Dg)^T M_{gg}^{-1} M_{gf} s$$

satisfies the optimality conditions for ϕ on Ω.

The gradient of ϕ is given by

$$\nabla\phi = A\Delta p + (Ap + b).$$

We know that $s \geq 0$ and $\dot{f} - Ms = Q$. Hence

$$\nabla\phi|_{\Delta p} = A\left((Df)^T - (Dg)^T M_{gg}^{-1} M_{gf}\right) s + (Ap + b).$$

Notice that $Ap + b$ represents the initial velocity, $\dot{q}^-$. The optimality condition for ϕ in Ω states that $\nabla\phi$ is contained in the positive span of constraint normals. The constraint normal corresponding to $s_k = 0$ is ∇f^k. Hence, the optimality condition for ϕ is that

$$s^T(Df)\nabla\phi = 0, \text{ where } s \geq 0 \text{ and } (Df)\nabla\phi \geq 0.$$

Observe that $(Df)\nabla\phi = Ms + Q$, where we use the fact that $Q = (Df)\dot{q}^-$. Thus, we see that the optimality condition is indeed satisfied: since $(\dot{f}, s)$ is a solution to LCP(M, Q) it follows that $(Df)\nabla\phi = \dot{f}$, and $\dot{f}$ is complementary to s. □

Force Determination using LCP's

Now we determine the reactive forces of the constraints which act over time. There is some ambiguity in this problem; the reactive forces which act over time are only defined almost everywhere. Indeed, any almost-everywhere equal representative of the reactive forces will do. We use this fact to our advantage: since contact breaking happens on a set of times of measure zero, we can may compute whatever force we'd like at such times without being erroneous. However, for aesthetic reasons, we find such an approach slightly less than satisfactory, and we instead try to compute the force when contacts break by imposing complementarity conditions between normal acceleration and the normal contact force. We will again obtain an LCP.

Assume a Hamiltonian system as described above that is not in a state of impact. We will say that the kth unilateral constraint is active iff for the corresponding index k, we have that $f^k = \dot{f}^k = 0$. All bilateral constraints are said to be active at all times.

For the remainder of this section, for notational convenience, we will take every unilateral constraint to be active for notational convenience. We will also, as in the last section, concatenate contacts into vectors f, g, $\dot{f}$, and $\dot{g}$.

Since the system is not in a state of impact, we know that $f = \dot{f} = 0$. We also know that $g = \dot{g} = 0$. We maintain these constraints as prescribed in Definition 3.2 by forcing the Hamiltonian dynamics with extra terms:

$$\dot{q} = \frac{\partial H}{\partial p} \qquad \dot{p} = (Df)^T s + (Dg)^T \lambda - \frac{\partial H}{\partial q}.$$

We interpret the components of s to be the measure of force corresponding to each of the unilateral constraints; similarly the components of λ correspond to the forces maintaining the equality constraints. These are both functions of bounded variation, rather than measures. Indeed $s\,dt$ and λdt correspond to the absolutely continuous part of the measures of Definition 3.2.

To determine the force is to determine what s and λ should be. We impose the following on s and λ:

Force Determination Rules.

1. $\ddot{g} = 0$.

2. $\ddot{f} \geq 0$.

3. $s \geq 0$.

4. Whenever a component of $\ddot{f}$ is greater than 0, the corresponding component of s should be 0.

Justification of Assumptions (1)-(4). We justify these rules by appealing to Definition 3.2. In particular, forces represent the absolutely continuous parts of the measures μ_k and ν_k in this definition. We point out the correspondence between the Lagrange multipliers s_k and the density part of μ_k; similarly we point out the correspondence between λ_k and the absolutely continuous part of ν_k. Rule (1) follows straightforwardly from the fact that $g(t) = 0$ for all t. Rule (2) follows from noting that if $f_k(t) = \dot{f}_k(t) = 0$ and $\ddot{f}_k(t) < 0$, then $f_k < 0$ for sufficiently small $s > t$, which would violate a constraint. Rule (3) follows from the correspondence between s_k and the absolutely continuous part of μ_k, which must be a non-negative function for μ_k to be a non-negative finite measure. Rule (4) is more difficult. Observe that this condition must hold almost always, or else the complementarity in Definition 3.2 would be violated. Now choose an almost-everywhere equal representative of the density part of μ_k which obeys the complementarity condition everywhere, and note that we may let s correspond to that representative. Now we give a more intuitive description of what these rules mean. The first assumption follows from the demand that $g(t) = 0$ for all time, and hence the second derivative must be zero as well. The

second assumption, meant only to apply to unilateral constraints which are active, follows from the fact that $f(t_0) = \dot{f}(t_0) = 0$ and $\ddot{f}(t_0) < 0$ implies $f(t) < 0$ for sufficiently small $t > t_0$. The third assumption states that the unilateral constraints may only push away in the normal direction (into the set of admissible positions), they may never pull towards. The fourth assumption, which we call complementarity of s and $\ddot{f}$, simply says that we never need to worry about trying to compute a situation where we are applying a normal force which breaks the contact supplying that force.

Now we use these Force Rules to find an LCP which will determine $\dot{p}$ when it exists. The derivation is quite similar to the impact case, albeit with some extra terms due to second derivatives.

$$\ddot{f}^k = \frac{\partial H}{\partial p^\mu}\frac{\partial^2 f^k}{\partial q^\mu \partial q^\nu}\frac{\partial H}{\partial p^\nu} + \frac{\partial f^k}{\partial q^\mu}\frac{\partial^2 H}{\partial p^\mu \partial q^\nu}\frac{\partial H}{\partial p^\nu} + \frac{\partial f^k}{\partial q^\mu}\frac{\partial^2 H}{\partial p^\mu \partial p^\nu}\left(\frac{\partial f^\sigma}{\partial q_\nu}s^\sigma + \frac{\partial g^\rho}{\partial q^\nu}\lambda^\rho - \frac{\partial H}{\partial q^\nu}\right) \geq 0.$$

$$\ddot{g}^k = \frac{\partial H}{\partial p^\mu}\frac{\partial^2 g^k}{\partial q^\mu \partial q^\nu}\frac{\partial H}{\partial p^\nu} + \frac{\partial g^k}{\partial q^\mu}\frac{\partial^2 H}{\partial p^\mu \partial q^\nu}\frac{\partial H}{\partial p^\nu} + \frac{\partial g^k}{\partial q^\mu}\frac{\partial^2 H}{\partial p^\mu \partial p^\nu}\left(\frac{\partial f^\sigma}{\partial q_\nu}s^\sigma + \frac{\partial g^\rho}{\partial q^\nu}\lambda^\rho - \frac{\partial H}{\partial q^\nu}\right) = 0.$$

We use the second equation to find λ in terms of s. First we make some abbreviations. The symbols α and β are to be either f or g:

$$(M_{\alpha\beta})_{jk} := \frac{\partial \alpha^j}{\partial q^\mu}\frac{\partial^2 H}{\partial p^\mu \partial p^\nu}\frac{\partial \beta^k}{\partial q^\nu}.$$

$$(Q_\alpha)^k := \frac{\partial H}{\partial p^\mu}\frac{\partial^2 \alpha^k}{\partial q^\mu \partial q^\nu}\frac{\partial H}{\partial p^\nu} + \frac{\partial \alpha^k}{\partial q^\mu}\frac{\partial^2 H}{\partial p^\mu \partial q^\nu}\frac{\partial H}{\partial p^\nu} - \frac{\partial \alpha^k}{\partial q^\mu}\frac{\partial^2 H}{\partial p^\mu \partial p^\nu}\frac{\partial H}{\partial q^\nu}$$

Using this notation and by Force Rules (1) and (2) we write

$$\ddot{g} = Q_g + M_{gg}\lambda + M_{gf}s = 0,$$

and also

$$\ddot{f} = Q_f + M_{ff}s + M_{fg}\lambda \geq 0.$$

For a given s, a solution for λ is given by

$$\lambda = -M_{gg}^{-1}(Q_g + M_{gf}s) \tag{3.13}$$

Substitution into the $\ddot{f} \geq 0$ equation yields

$$\ddot{f} = Q_f + M_{ff}s - M_{fg}M_{gg}^{-1}M_{gf}s - M_{fg}M_{gg}^{-1}Q_g \geq 0.$$

Now define

$$M := M_{ff} - M_{fg}M_{gg}^{-1}M_{gf}, \text{ and also } Q := Q_f - M_{fg}M_{gg}^{-1}Q_g. \tag{3.14}$$

Now we have, recalling Force Rules (1) and (3),

$$\ddot{f} = Ms + Q, \text{ subject to } \ddot{f} \geq 0 \text{ and } s \geq 0.$$

Along with Force Rule (4), our problem becomes the linear complementarity problem LCP(M,Q). We call this *the LCP for force determination.* Given the solution to this LCP, one has determined s. With this value of s, we may use Equation (3.13) to obtain λ as well. With these quantities determined, we have

$$\dot{p} = (Df)^T s + (Dg)^T \lambda - \frac{\partial H}{\partial q},$$

and the force problem is solved.

Remark. In the case that there do not exist bilateral constraints, or alternatively there do not exist unilateral constraints, it is straightforward to determine the appropriate modifications to the above analysis. In particular it is simply a matter of omitting the appropriate terms. When there are no unilateral constraints, the LCP step is unnecessary.

Well-posedness of Force and Impact Determination

In both the force and impulse cases studied above, we found that we may compute them by using a linear complementarity problem (LCP). It turns out that although linear complementarity problems in general may have multiple solutions, for our current case all of the LCP solutions produce the same force or impulse on the system. Moreover, the force or impact has continuous dependence on the problem data. This property is called *well-posedness*.

Schur Complement

In this subsection, we show that matrices constructed as M (see (3.14) and (3.12); they are the same) above in the force and impact determination sections are symmetric positive semidefinite (SPSD). The construction here is that of the so-called *Schur complement*, which arises when trying to invert block matrices.

LEMMA 3.4. *Let A be an $n \times n$ symmetric PSD matrix. Let P be an $n \times n$ symmetric idempotent matrix, i.e. $P^2 = P$. Define $Q := \boldsymbol{id} - P$. Then*

$$\textit{Range } PAQ \subset \textit{Range } PAP.$$

Proof. Let x be in the range of PAQ. We show x is in the range of PAP.

Since PAP is symmetric, $\mathbb{R}^n$ is the direct sum of the range of PAP and the null space of PAP. Let $x = y+z$ be the unique decomposition of x so that $y \in$ Range PAP and $z \in$ Null PAP. These two spaces are orthogonal, so $y^T z = 0$.

First we show $Qz = 0$. Observe that $Qx = Qy + Qz$. Since x is in the range of PAQ and $PQ = 0$, $Qx = 0$. Similarly, since y is in the range of PAP, $Qy = 0$. It follows that $Qz = 0$.

Now we show that $z = 0$. Since x is in the Range of PAQ, there exists u such that $x = PAQu$. Let α be any real number. Then

$$(Qu + \alpha z)^T A(Qu + \alpha z) = u^T QAQu + 2\alpha z^T x.$$

Notice the $\alpha^2 z^T Az$ term is zero, since $z = Pz$ and z is in the null space of PAP, and hence $z^T Az = z^T PAPz = 0$.

Positive semidefiniteness of A requires that $(Qu+\alpha z)^T A(Qu+\alpha z) \geq 0$. However, since α may range freely, this is only possible if $z^T x = 0$. Observe that $x = y + z$, so $z^T x = z^T y + z^T z$. Recalling $z^T y = 0$, we arrive at $z^T z = 0$, which implies $z = 0$.

Hence $x = y$, so that x is in the range of PAP. The inclusion is shown. □

In what follows, we denote the pseudoinverse of a matrix A by A^+. See [17] for a definition.

LEMMA 3.5. *Let A be an $n \times m$ matrix. Let r, s be nonnegative integers. Then*

$$\begin{bmatrix} A & 0_{n\times s} \\ 0_{r\times m} & 0_{r\times s} \end{bmatrix}^+ = \begin{bmatrix} A^+ & 0_{m\times r} \\ 0_{s\times n} & 0_{s\times r} \end{bmatrix},$$

where the 0's represent zero matrices of the indicated sizes.

Proof. It is straightforward to verify that the Penrose properties hold. See [17]. This guarantees that the right hand side is indeed the unique pseudoinverse of the left hand side. □

LEMMA 3.6. *Suppose that A is $n \times m$, U is a unitary $n \times n$ matrix, and V is a unitary $m \times m$ matrix.*

$$V(AV)^+ = A^+ = (UA)^+U.$$

Proof. A admits a singular value decomposition $A = XDY$, so $V(AV)^+ = V(XDYV)^+ = VV^TY^TD^+X^T = A^+$. The other part is similar. □

LEMMA 3.7. *Let A be an $n \times n$ symmetric PSD matrix. Let G be an $n \times d$ matrix. Then*

$$M := A - AG(G^TAG)^+G^TA$$

is symmetric PSD.

Proof. Symmetry is obvious; it is straightforward to see the expression is invariant under the transpose given that A is symmetric.

We show M is PSD. The matrix G admits a singular value decomposition $G = UDV$, where U is an $n \times n$ unitary matrix, V is a $d \times d$ unitary matrix, and D is an $n \times d$ matrix whose only nonzero entries are positive and lie on the diagonal. Substituting,

$$M = A - AUDV(V^T D^T U^T AUDV)^+ V^T D^T A.$$

By Lemma 3.6,

$$M = A - AUD(D^T U^T AUD)^+ D^T U^T A.$$

Defining $\tilde{A} := U^T AU$, we see that

$$U^T MU = \tilde{A} - \tilde{A}D(D^T \tilde{A} D)^+ D^T \tilde{A}.$$

Since U is unitary, we see that M is PSD if and only if $U^T MU$ is PSD. Note also that $\tilde{A}$ is symmetric PSD. Observe now that we have reduced the problem to the case where G is an $n \times d$ diagonal matrix with nonnegative entries. We may therefore take $G = D$ and $A = \tilde{A}$ without loss; we will favor the names A and D and we seek to show that $A - AD(D^T AD)^+ D^T A$ is PSD.

Without loss, take the diagonal entries of D to be arranged in descending order, so that $D_{11} \geq D_{22} \geq \cdots$. This is without loss since we have freedom to choose the singular value decomposition of G above in this way. Let s denote the number of nonzero singular values of G (or D). In other words, let s be the maximal index such

that $D_{ss} > 0$. We have that

$$D = \begin{bmatrix} S & 0_{s\times(d-s)} \\ 0_{(n-s)\times s} & 0_{(n-s)\times(d-s)} \end{bmatrix}$$

for some nonsingular $s \times s$ symmetric matrix S. Write the $n \times n$ matrix A in a similar block manner, so that

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

The matrix A_{11} is to be $s \times s$, A_{12} is $s \times (n-s)$, etc. We have by Lemma 3.5 that

$$(D^T A D)^+ = \begin{bmatrix} SA_{11}S & 0_{s\times(d-s)} \\ 0_{(d-s)\times s} & 0_{(d-s)\times(d-s)} \end{bmatrix}^+ = \begin{bmatrix} (SA_{11}S)^+ & 0_{s\times(d-s)} \\ 0_{(d-s)\times s} & 0_{(d-s)\times(d-s)} \end{bmatrix}.$$

From here on out, we drop the dimension of the zero matrices for convenience and economy of space. This will make the matrices for D and D^T appear to be the same, though they are not in general:

$$AD(D^T AD)^+ D^T A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} (SA_{11}S)^+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} =$$

$$\begin{bmatrix} A_{11}S(SA_{11}S)^+SA_{11} & A_{11}S(SA_{11}S)^+SA_{12} \\ A_{21}S(SA_{11}S)^+SA_{11} & A_{21}S(SA_{11}S)^+SA_{12} \end{bmatrix}.$$

It is an application of Lemma 3.4 that the range of A_{12} is contained in the range of A_{11}. In particular, we choose P and Q in the lemma to be the projections onto the subspaces corresponding to the block decomposition of A.

Because the range of A_{12} is contained in the range of A_{11}, there must exist a matrix η such that $A_{12} = A_{11}\eta$. Since A is symmetric, $A_{21} = \eta^T A_{11}$. We can rewrite the above matrix as

$$\begin{bmatrix} S^{-1}(SA_{11}S)(SA_{11}S)^{+}(SA_{11}S)S^{-1} & S^{-1}(SA_{11}S)(SA_{11}S)^{+}(SA_{11}S)S^{-1}\eta \\ \eta^T S^{-1}(SA_{11}S)(SA_{11}S)^{+}(SA_{11}S)S^{-1} & \eta^T S^{-1}(SA_{11}S)(SA_{11}S)^{+}(SA_{11}S)S^{-1}\eta \end{bmatrix}.$$

Applying the Penrose pseudoinverse identity $XX^{+}X = X$ (see [17]) for $X = SA_{11}S$, we can simplify this to

$$\begin{bmatrix} S^{-1}(SA_{11}S)S^{-1} & S^{-1}(SA_{11}S)S^{-1}\eta \\ \eta^T S^{-1}(SA_{11}S)S^{-1} & \eta^T S^{-1}(SA_{11}S)S^{-1}\eta \end{bmatrix} = \begin{bmatrix} A_{11} & A_{11}\eta \\ \eta^T A_{11} & \eta^T A_{11}\eta \end{bmatrix} =$$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & \eta^T A_{11}A_{11}^{+}A_{11}\eta \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{21}A_{11}^{+}A_{12} \end{bmatrix}.$$

Finally, we have that

$$A - AD(D^T AD)^{+}DA = \begin{bmatrix} 0 & 0 \\ 0 & A_{22} - A_{21}A_{11}^{+}A_{12} \end{bmatrix}.$$

This matrix is PSD iff $A_{22} - A_{21}A_{11}^{+}A_{12}$ is PSD. We show this now. We know that A is PSD, so we know that

$$\begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \geq 0,$$

$$x^T A_{11}x + x^T A_{12}y + y^T A_{21}x + y^T A_{22}y \geq 0.$$

This holds for any x, y in the appropriate spaces. In particular it holds for the choice $x = -A_{11}^{+}A_{12}y$. Substituting this choice, we obtain

$$y^T A_{21}A_{11}^{+}A_{11}A_{11}^{+}A_{12}y - y^T A_{21}A_{11}^{+}A_{12}y - y^T A_{21}A_{11}^{+}A_{12}y + y^T A_{22}y \geq 0.$$

By the Penrose pseudoinverse property $X^{+}XX^{+} = X^{+}$, the first two terms cancel, leaving us with $y^T A_{22}y - y^T A_{21}A_{11}^{+}A_{12}y \geq 0$, which proves that $A_{22} - A_{21}A_{11}^{+}A_{12}$ is PSD. The proof is complete. □

Well-Posedness of LCP Solutions

THEOREM 3.8. *Suppose $x, y \geq 0$ are solutions to the linear complementarity problem LCP(M, Q), so that $x - My = Q$. Assume that M is symmetric positive semidefinite. Then x is uniquely determined by M and Q. Moreover, the dependence of the solution to LCP(M, Q) on M and Q is locally Lipschitz continuous.*

Proof. Suppose that $x - My = Q$ and $u - Mv = Q$ with complementarity conditions $x^T y = u^T v = 0$ for $x, y, u, v \geq 0$. We have that $(x - u) - M(y - v) = 0$, and also that $(x - u)^T (y - v) = -u^T y - x^T v \leq 0$. Multiplying, we see that $(y - v)^T (x - u) = (y - v)^T M(y - v) \leq 0$. But since M is PSD, we must have that $(y - v)^T M(y - v) = 0$. Since M is symmetric, it admits a decomposition $M = U^T DU$ which can be used to show that $(y - v)$ is necessarily in the null space of M. Hence $My = Mv$ from which it follows $x = u$.

Now the moreover part. Define the scalar-valued function $\phi : \mathbb{R}^d \to \mathbb{R}$ as

$$\phi(z) := \frac{z^T Mz}{2} + Q^T z.$$

Since M is SPSD, this is a convex functional. Consider the optimization problem of minimizing ϕ on the feasible set $[0, \infty)^d$. By convexity of the cost functional and the feasible set, there exists a convex set of optimizers consisting precisely of those points $y \in \mathbb{R}^d$ such that $\nabla\phi|_y$ is non-negative and complementary to y. Since $\nabla\phi|_y = My + Q$, we may rephrase this, and say that the optimizers are those points $y \in \mathbb{R}^d$ such that (x, y) is a solution to LCP(M, Q) for some $x \in \mathbb{R}^d$. Indeed, a unique choice of x, by

the first part. Thus, all the optimizers of ϕ in $[0,\infty)^d$ differ only by displacements in the null space of M.

In light of this, we consider the optimization problem (again a quadratic program) of minimizing ϕ on the projection of $[0,\infty)$ onto the range of M. This quadratic program now has a symmetric positive definite kernel. It is a straightforward to see that the optimizer of such a quadratic program is locally Lipschitz continuous with respect to the data. It is also straightforward to see that x may be recovered from this solution in a locally Lipschitz continuous manner; in particular, the same formula $x = Mz + Q$ applies. □

COROLLARY 3.9. *The LCP formulations for impulse and force determination uniquely determine the impulse or force. Moreover, the impulse and force depend on the data in a locally Lipschitz continuous fashion.*

Proof. It is straightforward to verify that the LCP kernel M for the impact and force problems is constructed in a manner satisfying the assumptions of Lemma 3.7. Accordingly, the kernel of the impact and force LCP problems is SPSD. Now we obtain the result as a consequence of Lemma 3.8. □

Existence and Uniqueness

While it may seem from the preceding that we have effectively solved the problem of constrained Hamiltonian dynamics with frictionless, completely inelastic constraints, we have in fact done nothing of the sort. Rather, we have provided a means

of calculating forces on the system at a given time, or impulses at a given time. But whether or not solutions actually *exist* for which these calculations actually apply is a question so far left open. Also, though we have shown uniqueness and well-posedness *for our calculations of forces and impulses at a given moment*, whether or not this yields uniqueness of the entire problem we have also left unanswered. In this section we indicate the answers to these questions, or at least where they might be found.

Existence

Existence is non-trivial. The mathematical framework for the problem was given by Moreau [24]; this led to practical numerical methods rather than an existence proof. It was Monteiro Marques [23] who first gave an existence result. Mabrouk [22] extended the work of Monteiro Marques to handle impacts which were not completely inelastic, but also elastic or partially elastic. Ballard [6] gives an existence result when the underlying configuration space is a Riemannian manifold rather than Euclidean space.

Uniqueness

Given our ability to resolve impacts and determine reactive forces, we have the possibility for the following algorithm for computing solution trajectories to inelastic dynamics:

Proposed Solution Method: We begin with an initial condition: if it is impacting, we find the correct impulse via the LCP for impact determination. We then integrate

the results of the force determination LCP for at least some small amount of time, until an impact occurs or a contact releases. At such a moment we again turn to the impact determination LCP, and then go back to integration. This is to be continued indefinitely.

If this method indeed did construct a solution, it is straightforward to see that it is unique, since we have an ordinary differential equation with a locally Lipschitz vector field. There are discontinuities, but they are uniquely determined, well-posed, and never accumulate from the right.

A right accumulation of impacts is a sequence of times (t_n) tending to t_0 in the right-handed sense[2] such that a discontinuity of momentum of the inelastic dynamics solution occurs for each t_n. That the impacts do not accumulate from the right is the key ingredient for this solution construction method to work: if there is a right-hand accumulation of impacts, then at the time of accumulation we cannot use the method to extend the solution any further! If there is a right accumulation of impacts at t_0, then we cannot apply the "integrate the forces" step of our proposed solution construction method. This step requires that we extend the solution over some finite (non-zero) time interval, and yet that is impossible since every interval $[t_0, t_0 + \epsilon)$, $\epsilon > 0$ contains infinitely many impacts, which we are not allowed to integrate over.

In fact it is possible to have C^∞ data for an inelastic dynamics evolution problem for which all solutions involve right accumulations of impacts. That is, if we want

[2]that is, $(t_n) \to t_0$ and $t_n > t_0$ for a tail end of the sequence (t_n)

to have an existence result which works for all C^∞ data, we have to accept that this proposed solution method may not compute a solution.

The failure of this proposed solution method to work in situations with right-hand accumulations opens the door for non-uniqueness. With *right accumulations of impacts*, non-uniqueness is possible despite the impact determination and the force determination being (essentially) unique and well-posed.

Ballard [6] has given an example to show that these right-hand accumulations of impacts can in fact be responsible for non-unique solutions, even in a one dimensional mechanical system with a point particle and the single constraint $q \geq 0$. We give his example:.

Example. *(Ballard.)* Consider the one dimensional dynamics of a point particle with unit mass that lives on the nonnegative real number line $[0, \infty)$: $q \geq 0$. We suppose that this constraint is inelastic, and that the particle's motion is due to the inelastic reactive forces/impulses of the unilateral constraint $q \geq 0$ and also an external forcing function $f(t)$. Take initial conditions to be $q(0) = \dot{q} = 0$. We consider solutions such that $\dot{q}$ has bounded variation, and we take $\dot{q}$ to be the continuous from the right (in particular the initial condition $\dot{q}(0) = 0$ thus implies $\lim_{t \to 0^+} \dot{q}(t) = 0$). Since functions of bounded variation have distributional derivatives which are measures, we may write the measure-theoretic equation

$$d\dot{q} = R + f(t)dt,$$

where R is a finite nonnegative measure supported on $\{t \in [0,1] : q(t) = 0\}$ which represents the action of the constraint on the particle. The inelastic assumption may be expressed as $\dot{q} = 0$ whenever $q(t) = 0$.

Ballard was able to produce an external forcing function $f(t)$ in the C^∞ class for which two distinct solutions to the dynamics exist. We give such a forcing function and demonstrate two distinct solutions. For what follows, take $\rho(t)$ to be a *bump function*: a nonnegative C^∞ function supported on $[0,1]$ such that $\int_0^1 \rho(t)dt = 1$.

Define

$$f(t) := \begin{cases} A_n \rho\left(\frac{t-\frac{1}{n+1}}{\tau_n^1}\right) & \text{whenever } t \in \left[\frac{1}{n+1}, \frac{1}{n+1} + \tau_n^1\right) \\ 0 & \text{whenever } t \in \left[\frac{1}{n+1} + \tau_n^1, \frac{1}{n} - \tau_n^2\right) \\ -B_n \rho\left(\frac{t-\frac{1}{n}+\tau_n^2}{\tau_n^2}\right) & \text{whenever } t \in \left[\frac{1}{n+1} - \tau_n^2, \frac{1}{n}\right) \end{cases} .$$

Here, (A_n), (B_n), (τ_n^1), and (τ_n^2) are nonnegative sequences to be determined. Note however that the definition of f implies $\tau_n^1 + \tau_n^2 \leq \frac{1}{n} - \frac{1}{n+1}$. In fact we will choose τ_n^1 and τ_n^2 so they are each less than half the size of the interval $[\frac{1}{n+1}, \frac{1}{n})$. We may visualize f on each interval $\left[\frac{1}{n+1}, \frac{1}{n}\right)$ as consisting of an initial upward bump, a flat region, and then an downward bump. We have yet to determine the width (τ_n^1 and τ_n^2) and height (A_n and B_n) of these C^∞ bumps.

In order to discover what we could make these undetermined coefficients in order to produce a non-uniqueness, we stipulate that there are two solutions, $x(t)$ and $y(t)$, whose values and derivatives we will specify on the times $t_n = (\frac{1}{n})$.

When n is an even positive integer, we stipulate

$$x\left(\frac{1}{n}\right) = 0 \qquad \dot{x}\left(\frac{1}{n}\right) = 0$$

and

$$y\left(\frac{1}{n}\right) = q_n \qquad \dot{y}\left(\frac{1}{n}\right) = -v_n.$$

When n is an odd positive integer, we stipulate

$$x\left(\frac{1}{n}\right) = q_n \qquad \dot{x}\left(\frac{1}{n}\right) = -v_n$$

and

$$y\left(\frac{1}{n}\right) = 0 \qquad \dot{y}\left(\frac{1}{n}\right) = 0.$$

Here, (v_n) and (q_n) are positive sequences of real numbers to be determined. Notice that $x(t)$ and $y(t)$ are certainly distinct if they obey these relationships.

We wish to arrange, for each positive integer $n \geq 2$, that a particle at position $q_{n+1} > 0$ with velocity $-v_{n+1} < 0$ at time $\frac{1}{n+1}$ will reach the constraint $q = 0$ at the end of the time interval $\left[\frac{1}{n+1}, \frac{1}{n+1} + \tau_n^1\right)$ without experiencing any reactive forces or impulses. Integration reveals

$$0 = q_{n+1} - \tau_n^1 v_n + \int_0^{\tau_n^1} \int_0^t A_n \rho\left(\frac{s}{\tau_n^1}\right) ds\, dt.$$

The integral may be manipulated to give

$$0 = q_{n+1} - \tau_n^1 v_n + A_n (\tau_n^1)^2 \int_0^1 \int_0^t \rho(s) ds\, dt.$$

We have chosen our bump function ρ so that $\int_0^1 \int_0^t \rho(s) ds\, dt = \frac{1}{2}$, hence

$$0 = q_{n+1} - \tau_n^1 v_n + \frac{1}{2} A_n (\tau_n^1)^2.$$

From the quadratic formula we discover

$$\tau_n^1 = \frac{v_n - \sqrt{v_n^2 - 2A_n q_{n+1}}}{A_n}. \tag{3.15}$$

We wish to arrange, for each positive integer $n \geq 2$, that a particle at position 0 and with velocity 0 at time $t = \frac{1}{n}$ should be at position q_n with velocity $-v_n$ at time $t = \frac{1}{n+1}$.

Single integration reveals

$$-v_n = \int_{\frac{1}{n+1}}^{\frac{1}{n}} f(t)dt = A_n\tau_n^1 - B_n\tau_n^2, \tag{3.16}$$

while double integration reveals

$$q_n = \frac{1}{2}A_n\left(\tau_n^1\right)^2 + A_n\tau_n^1(\frac{1}{n} - (\frac{1}{n+1} + \tau_n^1)) - \frac{1}{2}B_n\left(\tau_n^2\right)^2. \tag{3.17}$$

We may use Equation 3.16 to rewrite Equation 3.17 as

$$q_n = \frac{1}{2}A_n\left(\tau_n^1\right)^2 + A_n\tau_n^1\left(\frac{1}{n} - (\frac{1}{n+1} + \tau_n^1)\right) - \frac{1}{2}(A_n\tau_n^1 + v_n)\left(\tau_n^2\right), \tag{3.18}$$

which in turn simplifies to

$$q_n = -\frac{1}{2}A_n\left(\tau_n^1\right)^2 + A_n\tau_n^1\frac{1}{n(n+1)} - \frac{1}{2}A_n\tau_n^1\tau_n^2 - \frac{1}{2}v_n\tau_n^2 \tag{3.19}$$

Now we are ready to make concrete choices. We choose

$$q_n = \frac{1}{n^4 2^n}, \quad v_n = \frac{1}{2^n}, \quad A_n = \frac{n^3}{2^n}. \tag{3.20}$$

Through the choice of Equation (3.20) we can determine τ_n^1 via Equation 3.15:

$$\tau_n^1 = \frac{1}{n^3}\left(1 - \sqrt{1 - \frac{n^3}{(n+1)^4}}\right) \sim \frac{1}{2n^4}. \tag{3.21}$$

We can use Equation (3.19) to solve for τ_n^2, obtaining

$$\tau_n^2 = \frac{\frac{2n\tau_n^1}{n+1} - \frac{2}{n^4} - n^3\left(\tau_n^1\right)^2}{1 + n^3\tau_n^1} \sim \frac{2}{n^3}. \tag{3.22}$$

We can use Equation (3.16) to solve for B_n, obtaining

$$B_n = \frac{1}{2^n}\frac{n^3\tau_n^1 + 1}{\tau_n^2} \sim \frac{n^3}{2^{n+1}}. \tag{3.23}$$

For sufficiently high n, we see that we have chosen τ_n^1 and τ_n^2 less than half the width of the interval $[\frac{1}{n+1}, \frac{1}{n}]$. Choose $T > 0$ sufficiently small so that $\frac{1}{n} < T$ implies that τ_n^1 and τ_n^2 are each less than half the width of the interval $[\frac{1}{n+1}, \frac{1}{n}]$.

The reader is referred to Ballard [6] for a proof of the following:

PROPOSITION 3.10. *The function $f(t)$, defined on the interval $[0, T)$, is C^∞.*

We now define the two functions $u(t)$ and $v(t)$.

$$u(t) = \begin{cases} -v_{n+1} + A_n \int_{\frac{1}{n+1}}^t \rho\left(\frac{s - \frac{1}{n+1}}{\tau_n^1}\right) ds & \text{whenever } t \in \left[\frac{1}{n+1}, \frac{1}{n+1} + \tau_n^1\right), \\ 0 & \text{otherwise.} \end{cases} \tag{3.24}$$

$$v(t) = \begin{cases} A_n \int_{\frac{1}{n+1}}^t \rho\left(\frac{s - \frac{1}{n+1}}{\tau_n^1}\right) ds & \text{whenever } t \in \left[\frac{1}{n+1}, \frac{1}{n+1} + \tau_n^1\right), \\ A_n\tau_n^1 & \text{whenever } t \in \left[\frac{1}{n+1} + \tau_n^1, \frac{1}{n} - \tau_n^2\right), \\ A_n\tau_n^1 - B_n \int_{\frac{1}{n} - \tau_n^2}^t \rho\left(\frac{s - \frac{1}{n} + \tau_n^2}{\tau_n^2}\right) ds & \text{whenever } t \in \left[\frac{1}{n} - \tau_n^2, \frac{1}{n}\right), \end{cases} \tag{3.25}$$

Now we construct a solution we call $\dot{x}(t)$. Define $\dot{x}(0) = 0$, and

$$\dot{x}(t) = \begin{cases} u(t) & \text{whenever } t \in \left[\frac{1}{2n+1}, \frac{1}{2n}\right), \\ v(t) & \text{whenever } t \in \left[\frac{1}{2n}, \frac{1}{2n-1}\right). \end{cases} \tag{3.26}$$

Now we construct a solution we call $\dot{y}(t)$. Define $\dot{y}(0) = 0$, and

$$\dot{y}(t) = \begin{cases} v(t) & \text{whenever } t \in \left[\frac{1}{2n+1}, \frac{1}{2n}\right), \\ u(t) & \text{whenever } t \in \left[\frac{1}{2n}, \frac{1}{2n-1}\right). \end{cases} \tag{3.27}$$

The reader is referred to Ballard [6] for a proof of the following (or he may do the necessary computations himself):

PROPOSITION 3.11. *The function*

$$x(t) = \int_0^t \dot{x}(s)\, ds$$

is a solution to inelastic dynamics with forcing function $f(t)$ and unilateral constraint $x(t) \geq 0$.

PROPOSITION 3.12. *The function*

$$y(t) = \int_0^t \dot{y}(s)\, ds$$

is a solution to inelastic dynamics with forcing function $f(t)$ and unilateral constraint $y(t) \geq 0$.

Since $x(t)$ and $y(t)$ are both solutions, and they are not the same, we have given a non-uniqueness example. Since $f \in C^\infty$, this shows that even very nice data can lead to non-uniqueness, even with inelastic impacts.

Non-uniqueness of solutions is not a phenomenon exclusive to inelastic impact laws. Schatzman ([34], [35]) has given an example of non-uniqueness in the totally elastic case using C^∞ data. Again, the technique is based on right-accumulations of impacts.

An attempt has been made to circumvent the problem of right-accumulations of impacts. One specifies that all problem data is *real analytic*. It turns out that one can still get left-accumulations of impacts, but not right accumulations. One can then formulate a uniqueness result. This approach was pioneered by Percivale ([31], [32]), and later taken up by Schatzman [35]. Ballard gives the most general result of this type available in [6].

Numerics

We have given a means of computing a solution to a problem in inelastic dynamics, provided that there exists a solution which does not contain right accumulations of impact times. However, we also know that it is possible that the only existing solutions do have such right-handed accumulations. We would like to have a numerical technique which may compute such solutions.

In fact, in some sense, we already do. On any real computer, small enough impacts are not detectable. Hence, practically, there are no right-hand accumulations of impacts – at least not any impacts strong enough to overcome tolerance limits set in the program. Hence the "*proposed solution method*", in a practical implementation, would happily time-step over infinitely many tiny impulses, none the wiser. What of the computed solution? Is it valid? As we increase precision, does it converge to a solution of inelastic dynamics?

Such questions have indeed been tackled in the literature. One accepts that any practical algorithm for computing solutions with right accumulations of impacts must

have the ability to time-step over infinitely many impacts. Accordingly, we become interested in algorithms which treat the forces and impacts on equal footing; over the time-step, we are no longer interested in whether the net change in momentum was due to impulses or integrated forces. This is the key notion of a so-called *time-stepping algorithm.*

Moreau [25] appears to have initiated the study of such algorithms in the context of his so-called *sweeping processes* [18]. The key notion (for inelastic dynamics) is to use an implicit method to evaluate the impulse Δp over a time-step such that Δp is in the non-negative span of the active contact surface normals *at the end of the time-step.* This prevents releasing constraints from "pushing," and hence enforcing inelasticity.

This reasoning leads to time-stepping schemes where an LCP is solved at each time step. It was the analysis of such schemes that led to the existence theorems mentioned earlier. The first such result was due to Monteiro Marques [23]. We also mention the work of Lötstedt, [21], who appears to have been the first to apply the LCP approach to the problem of unilateral constraints, though his existence results are weaker – they require analyticity or else an assumption of that the active set is constant for small time intervals (essentially the assumptions which would make our *proposed solution method* work).

Summary.

In this chapter we formulated a frictionless unilaterally constrained dynamical problem and investigated how LCP's could be used to solve it. Despite showing that LCP's are very well behaved, it turns out that we may still get non-uniqueness of solutions due to anomalies which are caused by the possibility of right-accumulations of impacts. An example due to Ballard was used to prove this point. We discussed numerical strategies. Of particular importance was the time-stepping method, which also plays a central role in the study of modern Coulomb friction, the subject of the next chapter.

CHAPTER 4

COULOMB FRICTION

Introduction

In this chapter, we begin our study of frictional contact. We review the notion of Coulomb friction, and discuss the relevant literature.

Although experimentally successful after its inception, it was soon realized that there were serious well-posedness problems in the Coulomb friction model. In particular, Painlevé [29] discovered examples of simple mechanical problems in which the Coulomb model admits no solutions.

Recently, modern methods have been devised to circumvent this so-called *Painlevé paradox* by the usage of impulsive forces when they are ostensibly not needed: so-called tangential impacts. This resolution was pioneered by David Stewart [37]. His formulation of the problem requires a dynamical framework which can accommodate the discontinuities due to impacts. To this end *measure-differential inclusions*, a generalization of *differential inclusions*, is invoked. These were introduced by Moreau [24], whose investigations led to the so-called *time-stepping* schemes. These algorithms are the basis for numerical methods in the modern friction methods developed by Stewart, Trinkle, Anitescu, Potra, and others – see [38] for a survey.

Although Stewart has shown that the modern model for Coulomb friction (which admits tangential shocks) has an existence theorem for single-contact cases [37], for multiple-contact cases this appears to be unknown. Despite this, computational schemes have been devised [3]. Stewart remarks in his recent survey [38]:

> *In many ways it is easier to write down a numerical method for rigid-body dynamics than say what the method is trying to compute.*

Even though it is only known that the time-stepping formulations for solving frictional contact problems are known to converge to a solution of the continuous formulation of Coulomb friction (with the tangential impulse resolution to Painlevé's paradox) only for single-contact problems, it is expected that the essential idea is correct and convergence for the multiple-contact case will someday be shown as well.

Specification of Coulomb Friction Model.

The usual model of frictional contact studied is known as *Coulomb Friction.* This model of friction exclusively applies to three-dimensional mechanical systems. Coulomb friction models prevent interpenetration of unilateral constraints by allowing both normal and tangential forces to push the contacts. Contacts are categorized into two categories: *static* (or sometimes called *rolling*) or *kinetic* (or sometimes called *sliding*). The categorization is based on whether or not the contacts have non-zero velocity relative to the constraint. The constraint forces at the contacts are chosen under the following rules:

- A sliding contact always experiences a tangential force retrograde to its motion, and the magnitude of this force is a fixed constant μ_k times the magnitude of the normal force. The constant μ_k is called the *coefficient of kinetic friction.*
- A static contact may experience any force within the so-called *friction cone.* which is the set of forces such that the tangential force on the contact is less than or equal to a fixed constant μ_s times the normal force. The constant μ_s is called the *coefficient of static friction.*
- The combination of forces experienced on all contacts does not lead to inter-penetration of the constraints.

The sliding contact rule may be modified for situations with anisotropic friction. Instead of demanding that the frictional force at the contact is precisely in the opposite direction of the sliding, the force at a contact should be chosen in such a way that the rate of dissipation at the contact is maximized, the so-called *maximal dissipation principle.* This determines the direction of the force. If there are multiple contacts, we establish the direction of each individually before deciding on the magnitudes.

Example. A four-legged table is sliding across flat concrete which has a coefficient of friction $\mu > 0$. The frictional contact between the table leg and concrete is isotropic. Then the tangential frictional force on each leg is exactly retrograde to the sliding velocity of the leg's contact to the concrete.

Remark. In a later chapter, we develop a different theory of friction which also has a maximal dissipation principle. However, it differs in that we attempt to choose

frictional forces so that dissipation is maximized in a *global sense*. This will be discussed in Chapter 6. In Coulomb friction, we do not have a single, global dissipation maximization problem. Rather, for each contact we have a two-dimensional optimization (assuming we are working with contacts in a three-dimensional geometry) in the plane of the contact surface. This optimization problem is to minimize $\langle v_s, T \rangle$, where v_s is the sliding velocity and T is an acceptable tangential force due to friction. For the isotropic case, we have $|T| \leq \mu |N|$, where N is the normal force. In the anisotropic case, the allowed values of T may be an ellipse or some other convex figure. It is straightforward that for the isotropic case the frictional forces will always be retrograde to the sliding velocity of the contact: in this case, the allowed values of T comprise a two-dimensional ball.

Painlevé's Paradox.

The specification of Coulomb Friction given in the previous section does not always yield solutions. One situation in which it does not yield a solution is expected: when an impact occurs. No force can prevent interpenetration; only an *impulse* can stop it. Barring impacts, one could hope that the Coulomb model always admits some force solution.

Unfortunately, as Painlevé discovered in [29], it does not. He produced an example of a simple rigid bar in frictional contact with a flat surface that was not in a state of impact for which the Coulomb prescription yields no possible forces.

We spell out his example:

Example. *(Painlevé)* Painlevé's example is a rigid bar of mass m, rotational inertia J, and length ℓ in frictional contact with a flat surface (which we envision as the x-axis of the coordinate plane). The bar makes an angle $\theta < \frac{\pi}{2}$ with the surface, and is instantaneously not rotating: $\dot{\theta}(0) = 0$. However the bar is moving via translation at a speed $|v|$ directly left, as indicated in Figure 1.

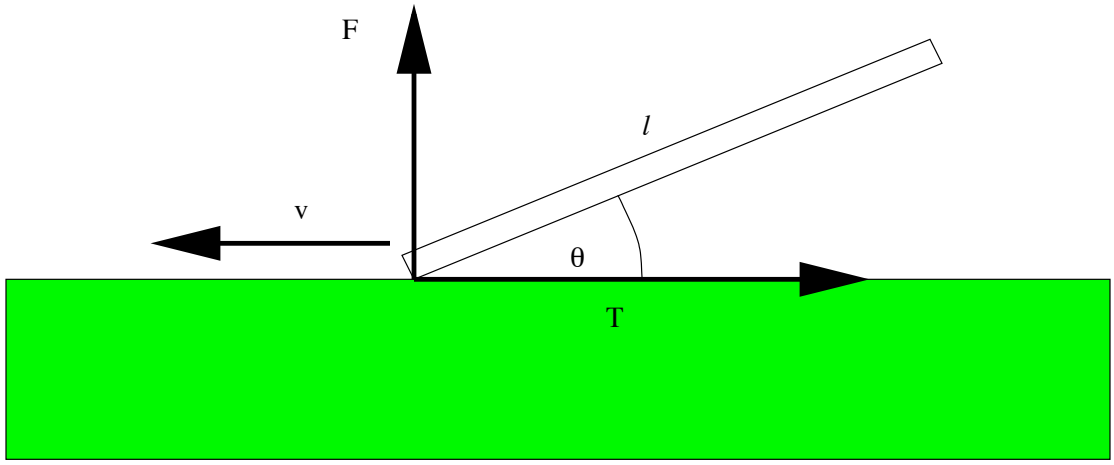

Figure 1. Painlevé's Paradox.

According to Coulomb Friction, there must be a tangential force T pushing the bar at its point of contact directly right (the positive x direction) with a magnitude $T = \mu N$, where N is the normal force which pushes directly up at the contact (the positive y direction). Some force is necessary to prevent interpenetrating in the presence of a gravitational force $-mg\hat{j}$ at the center of mass of the bar. The following equations indicate how the dynamics respond to the forces:

$$m\ddot{x} = \mu N \tag{4.1}$$

$$m\ddot{y} = N - mg \tag{4.2}$$

$$J\ddot{\theta} = \frac{\ell}{2}\left(-N\cos\theta + \mu N\sin\theta\right). \tag{4.3}$$

We are interested in the interpenetration problem, so we want to know that acceleration of the y-value of the contacting point. The y-coordinate of the contact has the formula

$$y_c = y - \frac{1}{2}\ell\sin\theta,$$

which we differentiate twice to obtain

$$\ddot{y}_c = \ddot{y} - \frac{\ell}{2}(\cos\theta)\ddot{\theta} + \frac{\ell}{2}(\sin\theta)\dot{\theta}^2.$$

Since we have chosen $\dot{\theta} = 0$ for our initial condition, we have

$$\ddot{y}_c = \frac{N}{m} - g - \frac{\ell^2}{4J}(\cos\theta)(-N\cos\theta + \mu N\sin\theta). \tag{4.4}$$

We must have either $N \geq 0$ and $\ddot{y}_c \geq 0$, since only nonnegative normal forces are allowed (no sticking) and the contact may not accelerate into the surface. Also, we cannot have both N and $\ddot{y}_c$ simultaneously positive, by complementarity – this was discussed in an analogous situation in Chapter 3. This is commonly written as

$$0 \leq \ddot{y}_c \perp N \geq 0.$$

If we choose $N = 0$, then $\ddot{y}_c = -g < 0$, which will violate the constraint. Hence we know that $N > 0$, which forces $\ddot{y}_c = 0$. Substituting this into Equation (4.4), we can solve for N:

$$N = \frac{mg}{1 - \frac{m\ell^2}{4J}\cos\theta(\mu\sin\theta - \cos\theta)}. \tag{4.5}$$

The paradox occurs when $c := \frac{m\ell^2}{4J}$ and μ are both large. In such a case, the denominator of the right-hand side of Equation (4.5) can become negative, and $N > 0$ is violated for the only Coulomb force abiding solution which does not break the contact. In particular, a paradox will occur when

$$1 \leq c\cos\theta(\mu\sin\theta - \cos\theta).$$

Stewart's Resolution to the Painlevé Paradox.

David Stewart [38] has championed the following resolution to this paradox: he proposes to allow impulses for ostensibly non-impacting situations. These are called *tangential impacts.* We illustrate with an example how a tangential impact remedies the problem:

Example. *(Stewart)* Suppose we allow ourselves the usage of an impulsive force F which may change the velocity of the system discontinuously. Rather than being forced to use an impulse in the direction $\mu\hat{i} + \hat{j}$, we instead may use other directions – provided that we put the system into an "end state" where those directions become

allowed. By "end state" we mean the position and velocity configuration (which we call the state) "after" the impact; mathematically, we mean the right-handed limit of the state. For example, we would be able to use forces in a direction $\alpha\mu\hat{i} + \hat{j}$, $\alpha \in [-1, 1]$ *provided the selection of impulse stops the contact from sliding.*

The following equations tell us what the resulting velocity of the system will be after experiencing an impulse $F = T\hat{i} + N\hat{j}$ at the contact:

$$m\dot{x} = mv + T, \tag{4.6}$$

$$m\dot{y} = N, \tag{4.7}$$

$$J\dot{\theta} = \frac{\ell}{2}\left(-N\cos\theta + T\sin\theta\right). \tag{4.8}$$

We want to find T and N so that the velocity of the contact will be zero after the impulse. To this end we need formulas for $\dot{x}_c$ and $\dot{y}_c$ after the impulse. Here they are:

$$\begin{aligned} \dot{x}_c &= v + \frac{T}{m} + \frac{\ell}{2}(\sin\theta)\dot{\theta}, \\ \dot{y}_c &= \frac{N}{m} - \frac{\ell}{2}(\cos\theta)\dot{\theta}. \end{aligned}$$

Substituting in the expression for $\dot{\theta}$ and also $\dot{x}_c = \dot{y}_c = 0$,

$$\begin{aligned} 0 &= v + \frac{T}{m} + \frac{\ell^2}{4J}(\sin\theta)\left(-N\cos\theta + T\sin\theta\right), \\ 0 &= \frac{N}{m} - \frac{\ell^2}{4J}(\cos\theta)\left(-N\cos\theta + T\sin\theta\right). \end{aligned}$$

Writing $c = \frac{m\ell^2}{4J}$ and $p = mv$, we write this system in matrix form as:

$$\begin{bmatrix} 1 + c\sin^2\theta & -c\sin\theta\cos\theta \\ -c\sin\theta\cos\theta & 1 + c\cos^2\theta \end{bmatrix} \begin{bmatrix} T \\ N \end{bmatrix} = \begin{bmatrix} -p \\ 0 \end{bmatrix}.$$

The solution to this, as one could verify, is

$$\begin{bmatrix} T \\ N \end{bmatrix} = \frac{-p}{1+c} \begin{bmatrix} 1 + c\cos^2\theta \\ c\sin\theta\cos\theta \end{bmatrix}.$$

We ask: is this solution within the allowed friction cone? Is $|T| \leq \mu N$? The solution is in the friction cone provided

$$1 + c\cos^2\theta \leq \mu c\sin\theta\cos\theta,$$

or, equivalently,

$$1 \leq c\cos\theta(\mu\sin\theta - \cos\theta).$$

But this is precisely the condition that must be satisfied for a Painlevé paradox above! Hence, a tangential impulse will exist precisely when it is needed.

The Modern Coulomb Model

Stewart made the theory of tangential impacts rigorous by casting the problem as a *measure differential inclusion.* He managed to obtain a single-contact existence result [37], but at present there is no known multiple contact existence theorem. In this section we (very informally) describe the mathematical framework of the modern Coulomb model and the numerical schemes for finding solutions.

Measure Differential Inclusions

The mathematical framework of the modern Coulomb friction model invokes the so-called *measure differential inclusions*.

DEFINITION 4.1. *A* measure differential inclusion *(MDI) is the problem of finding an absolutely continuous function $x(t)$ and a function of bounded variation $v(t)$ such that*

$$\frac{dx}{dt} = f(t, x, v), \tag{4.9}$$

$$\frac{dv}{dt} \in F(t, x), \tag{4.10}$$

where f is continuous in its arguments, and $F(t, x)$ is a set-valued function upper-semicontinuous in its arguments, and has nonempty, closed, convex values.

This definition doesn't say it all, however. Although the first equation, $\dot{x} = f(t, x, v)$, is easy enough to understand – it is meant to be satisfied for almost all t (rather than everywhere) – it is difficult to understand what the inclusion means. This is because $v(t)$ is assumed only to be of bounded variation, and need not have a derivative, or even a Radon-Nikodym derivative. However, being of bounded variation, it does admit a *distributional derivative*, and in fact its distributional derivative is a signed finite measure we call Dv. In particular we may define Dv via the Riemann-Stieltjes approach so that

$$\int_a^b \phi(t) Dv = \lim \sum \phi(t_i)(v(t_{i+1}) - v(t_i)),$$

where the limit is over finer partitions $a = t_1 < t_2 < \cdots < t_N = b$.

It is known that every signed finite measure may be decomposed into two parts that are absolutely continuous and singular, respectively, with respect to some other measure; this is known as the Lebesgue Decomposition Theorem (see [8]). In particular we apply this theorem to Dv and the standard Lebesgue measure on the real line to obtain

$$Dv = a(t)dt + \mu_s,$$

where $a(t)dt$ is the absolutely continuous part of Dv and μ_s is the singular part of Dv. Now we are in a position to indicate what Equation (4.10) actually means:

$$a(t) \in F(t,x) \tag{4.11}$$

$$\frac{d\mu_s}{d|\mu_s|} \in F_\infty(t,x), \tag{4.12}$$

where the first inclusion holds Lebesgue almost-everywhere, and the second inclusion holds $|\mu_s|$-almost everywhere. The notation F_∞ means the asymptotic cone, or regression cone, or cone of unbounded rays:

$$F_\infty := \{v \in \mathbb{R}^d : F + \kappa v \subset F \text{ for all } \kappa \geq 0\}.$$

Inclusion (4.11) makes plenty of sense; it is, after all, precisely how we would interpret (4.10) if $v(t)$ were absolutely continuous and $a(t)$ were its Radon-Nikodym derivative. Inclusion (4.12) is still, perhaps, counterintuitive. To remedy this confusion, we consider the case where $v(t)$ is absolutely continuous except for a single jump discontinuity at $t = 0$. In this case, $|\mu_s|$ is a measure that is supported only on 0;

it is proportional to a so-called *delta function.* The Radon-Nikodym derivative $\frac{d\mu_s}{d|\mu_s|}$ is then the equivalence class of functions that agree at $t = 0$, and at $t = 0$ have the value which is the unit vector in the direction $\mu_s(\{0\})$ – which is just $\frac{\Delta v}{|\Delta v|}$. Inclusion (4.12), then, ultimately asserts that Δv is in the regression cone F_∞.

Intuitively, we think of it as follows. When $v(t)$ is absolutely continuous, the "force" which compels the velocity to change is in the non-empty closed convex set F. Occasionally, v may have discontinuities. We think of these discontinuities as corresponding to choosing "infinite" values in F; we formalize this concept by saying that the singular part of Dv "takes values in the unbounded ray cone" of F – and this in turn is formalized by constructing the Radon-Nikodym derivative $\frac{d\mu_s}{d|\mu_s|}$.

Now, we might wish to revise Equation (4.10) so that there is v-dependence:

$$\frac{dv}{dt} \in F(t, x, v). \tag{4.13}$$

In fact, such a revision is necessary in order to apply the MDI approach to non-trivial physical problems. However, when v undergoes a discontinuity, it is ambiguous what Equation (4.13) means. In Coulomb friction with inelastic impacts, however, this ambiguity is resolved by demanding that $v(t)$ be defined as its right-hand limit everywhere, so we would have

$$\frac{dv}{dt} \in F(t, x, v^+).$$

We do not specifically study MDIs nor do we indicate precisely how one formulates the modern Coulomb friction model using them. Rather, we give a brief account of

the literature concerning numerical methods, pointing out that the schemes originated from the work done by Moreau [24], who first formulated the MDI.

Numerical Methods

The MDI framework sets the stage for numerical methods called *time-stepping methods*, which resulted from the work of Moreau [24].

The key computational machinery invoked to calculate the time-steps is the linear complementarity problem, or LCP. Complementarity arises in Coulomb friction models in two ways. First, we demand maximal dissipation of energy at the contact. Optimization problems give rise to complementarity conditions between *Lagrange multipliers* and constraint functions. Indeed, both linear and quadratic programs are easily cast as LCP's. See, for example, Murty [27]. The second way that complementarity arises is through the stipulation that forces and impacts cannot occur if they result in breaking the very contacts responsible for those forces or impulses.

The time-stepping method breaks time into steps, and over each step considers the integral of the forces and impulses. A linear complementarity problem (complementing integrals of forces to contact breaks and optimizing for maximal dissipation at each contact) is solved at each time step. The LCP essentially asks the question: how may the contacts push so that each pushing contact remains active at the end of the time step, each frictional force is in an allowed direction which maximizes dissipation *with respect to the final velocity after the time step*, and the constraints are not violated at the end of the time step?

We phrased such a question in Chapter 3, with frictionless constraints, except we did not have to worry about any dissipation. In fact, we arrived at a very nicely behaved LCP with a symmetric positive semidefinite "kernel" matrix.

However, in the rigid-body time-stepping schemes with friction, one does not get such a well-behaved LCP. But the situation is not impossible: Anitescu and Potra [3] give a time-stepping scheme such that the LCP has solutions for each time step.

At least for the single contact case, one may consider a sequence of numerical solutions as one decreases step-size. Stewart [37] showed that they converge to a solution of an MDI with suitable extra conditions encoding ingredients such as maximal dissipation, non-violation of constraints, and the friction law. One gets convergence to a solution for single contact Coulomb friction problems.

Time-stepping methods using complementarity formulations have been studied in a number of papers by Stewart, Trinkle, Pang, Anitescu, Potra, and others; see [30], [40], [3], [4], [2], [1], [38]. We mention also the work of Baraff, who considered modifications to LCP algorithms for the case of force computaions [7]. A book by Pfeiffer and Glocker [33] develops an ODE approach (rather than a time-stepping approach) which keeps strong account of the impacts.

Summary

In this chapter we have introduced the notion of Coulomb friction. We showed that if one does not allow for tangential shocks in a Coulomb friction model, then

we may obtain non-existence examples called Painlevé paradoxes. Subsequently, we saw how allowing tangential shocks resolved the paradox. The MDI framework was introduced. It is a framework which allows for the discontinuities required in impact mechanics. Existence results were discussed. We discussed practical methods for solving MDI's, known as time-stepping methods. We discussed the importance of the LCP as the crucial computational tool in implementing these time-stepping schemes for modern Coulomb friction.

We noticed in this chapter that no multiple-contact existence result is currently available for modern Coulomb friction models. In a later chapter, we construct a novel model of friction (applying only to the case of persistent contact) for which we prove existence for multiple contacts. This model differs from Coulomb friction in some important respects.

CHAPTER 5

DIFFERENTIAL INCLUSIONS AND THE FEEDBACK PROBLEM

Introduction

In this chapter we consider a purely mathematical problem, which we will refer to as the Feedback Problem (FP). These FP's are a special case of differential inclusions (DI) which involve dynamically changing convex programs (CP). The FP will provide the theoretical framework for a model of friction we present later. Our current goal is to formulate the feedback problem it and discuss its properties.

Unfortunately, at the time of writing the well-posedness properties of FP's are not entirely clear. An existence result is known (due to Filippov's work on DI's [12]), but uniqueness and dependence of initial conditions and perturbations in data remain topics with open questions. We present the existence result and some partial results regarding uniqueness and well-posedness in special cases.

Although our motivation for studying FP's will not be entirely clear immediately, we can give an illustrative and familiar example which demonstrates a mechanical situation involving persistent frictional contact and a differential inclusion.

Example. *(Inclined Plane.)* Imagine a brick sliding on a inclined ramp. For convenience, we consider the "brick" to be a point particle of mass m. Gravity is in effect, applying a force $-mg\hat{j}$. Suppose the ramp is at an angle θ from the ground, and the

coefficient of friction (both static and dynamic) is μ. We can describe the force on the particle due to contact with the plane as having a normal component N and a tangential component T. Assume that $T > 0$ corresponds to pushing up the ramp. See Figure 2.

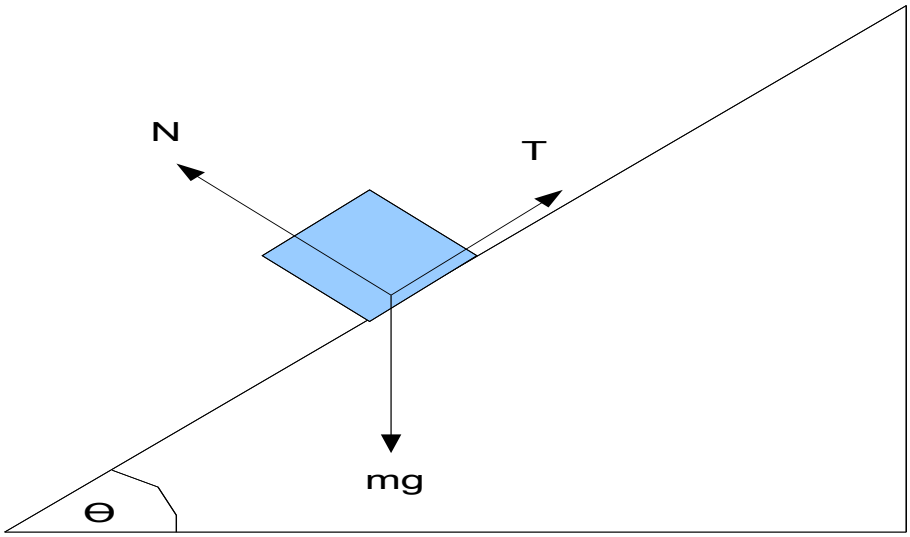

Figure 2. Inclined Plane Example.

To prevent the particle from accelerating into or off from the inclined plane, we must have

$$N = mg\cos\theta.$$

The tangential force T should then be chosen to be $T = -\mu N$ if the particle is sliding up the ramp, $T = +\mu N$ if the particle is sliding down the ramp, and $T \in [-\mu N, +\mu N]$ if the particle is at rest of the ramp. All three of these cases may be subsumed into the single statement:

$$T \in \mu N \cdot \mathrm{SOL}([-1, 1], v_s),$$

where v_s is the sliding velocity ($v_s > 0$ for sliding up the ramp, $v_s < 0$ for sliding down the ramp, $v_s = 0$ for at rest on the ramp) and $\mathrm{SOL}([-1,1], v_s)$ is the set of optimizers of the linear program with feasible set $[-1,1]$ and cost functional $f^{\mathrm{cost}}(z) = v_s \cdot z$. When $v_s = 0$, all the points in $[-1,1]$ are optimizers. When $v_s > 0$, only -1 is an optimizer. When $v_s < 0$, only $+1$ is an optimizer.

Now we write

$$\dot{v}_s = -mg\sin\theta + T \in -mg\sin\theta + \mu mg\cos\theta\ \mathrm{SOL}([-1,1], v_s). \tag{5.1}$$

Equation (5.1) is not an equation: it is an inclusion. Nevertheless, we will refer to inclusions as equations freely as a matter of practice. It is of the form

$$\dot{x} \in F(t,x),$$

where $F(x)$ is a *set-valued function*. This set-valued function is closed, bounded, and convex valued for all selections $t, x \in \mathbb{R}$. It also has a property we call *upper semi-continuity*, which is a generalization of continuity to set-valued functions. Because of these properties, it can be shown [12] that solutions exist.

We will find in the next chapter than we may use differential inclusions for many frictional contact problems, if we choose our friction law in a certain manner. In the current chapter, we study the differential inclusion. We pay special attention to differential inclusions that arise, as in this example, from solutions to optimization problems on convex sets.

The organization of the Chapter is as follows. In the first section, we introduce the concept of differential inclusions, and indicate what is meant by a solution. We present the celebrated[1] existence result due to Filippov [12]. In the second section, we discuss convex programs. We will only require a special case of convex programs, particularly, those with linear cost functionals[2]. Finally, in the third section, we pioneer a new type of problem which is a differential inclusion whose right hand side is the convex set of solutions to a convex program. This problem is the aforementioned feedback problem. We give conditions which guarantee the existence of solutions to the feedback problem. We show uniqueness and well-posedness of the feedback problem for special cases. We indicate open questions concerning the feedback problem.

Differential Inclusions

In this section we define differential inclusions and give conditions which guarantee existence. The existence result for differential inclusions was developed by Filippov in the 1960's [12]. We do not present a comprehensive account of differential inclusions, so we refer the interested reader to Aubin and Cellina [5].

Set-Valued Functions

We begin by defining the concept of a *set-valued function*:

[1] celebrated by nerds, at least

[2] Linear Programming, more or less, but with more general feasible sets than linear polytopes

DEFINITION 5.1. *A function $F : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^n)$ is said to be a* set-valued function. *The notation $\mathcal{P}(S)$, for some set S, denotes the so-called* power set*, or set of all subsets, of S. Thus, F is a set-valued function if its values are subsets of $\mathbb{R}^n$.*

We will need to be able to assert some regularity properties on set-valued functions. Much of this regularity may be obtained by asserting regularity of the values of a set-valued function F. This leads us to the following three definitions (condensed into one):

DEFINITION 5.2. *A set-valued function $F : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^n)$ is said to have (*closed, convex, bounded*)* values *if for each $x \in \mathbb{R}^d$, $F(x)$ is a (closed, convex, bounded) subset of $\mathbb{R}^n$.*

We also need to be able to discuss regularity of F in terms of how its values change as we vary its arguments. We use the following definitions:

DEFINITION 5.3. *A set-valued function $F : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^n)$ is said to be* upper semicontinuous *if it has a closed graph:*

$$\bigcup_{x \in \mathbb{R}^d} \{x\} \times F(x) \text{ is closed in } \mathbb{R}^d \times \mathbb{R}^n.$$

Equivalently, F is upper semicontinuous provided that for every convergent sequence (x_n) in $\mathbb{R}^d$ converging to x_0 and every convergent sequence (z_n) in $\mathbb{R}^n$ converging to z_0 such that $z_n \in F(x_n)$ for all n, we have that $z_0 \in F(x_0)$. Given a subset $U \subset \mathbb{R}^d$, we say that F is upper semicontinuous on U if $F|_U$ is upper semicontinuous.

The proof of the equivalence of these definitions is straightforward.

DEFINITION 5.4. *A set-valued function $F : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^n)$ is said to be* lower semicontinuous *if for every sequence (x_n) in $\mathbb{R}^d$ tending to x_0 and for every point $z_0 \in F(x_0)$, there exists a sequence (z_n) in $\mathbb{R}^n$ tending to z_0 such that $z_n \in F(x_n)$ for each n. Given a subset $U \subset \mathbb{R}^d$, we say that F is lower semicontinuous on U if $F|_U$ is lower semicontinuous.*

DEFINITION 5.5. *A set-valued function is said to be* continuous *if it is both upper semicontinuous and lower semicontinuous.*

Statement

A differential inclusion (DI) is a generalization of an ODE interpreted *in the sense of Carathéodory* (the almost always sense) where the right-hand side is allowed to be multi-valued. In particular, one has a set-valued function $F(t, x) : \mathbb{R} \times \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$, and an initial condition (t_0, x_0), and wishes to find an absolutely continuous function $x(t)$ such that $x(t_0) = x_0$ and

$$\dot{x}(t) \in F(t, x)$$

for almost all t, in the sense of Lebesgue measure. Such a function is called a *solution* to the differential inclusion with initial value condition $x(t_0) = x_0$.

Existence

DEFINITION 5.6. *We say a set-valued function* $F : \mathbb{R} \times \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ *satisfies the* basic conditions *on a set* $U \subset \mathbb{R} \times \mathbb{R}^n$ *whenever:*

$$\begin{cases} \bullet \text{ For every } (t,x) \in U, \text{ the set } F(t,x) \text{ is non-empty, closed, bounded, and convex.} \\ \bullet \text{ The set-valued function } F \text{ is upper-semicontinuous on } U. \end{cases}$$

THEOREM 5.7 (Filippov. [12]). *Let* U *be a neighborhood of an initial condition* $(t_0, x_0) \in \mathbb{R} \times \mathbb{R}^d$. *Suppose that a set-valued function* $F : \mathbb{R} \times \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^d)$ *satisfies the* basic conditions *on* U. *Then the differential inclusion* $\dot{x} \in F(t,x)$ *with initial condition* $x(t_0) = x_0$ *admits a solution on some interval* $t \in [t_0, t_0 + \epsilon)$, *for some* $\epsilon > 0$. *Moreover, any such solution may be continued until it reaches the boundary of* U.

Numerics

Work has been done on the numerical solution of differential inclusions. This work appears to have been initiated by Taubert [39]. A survey of such literature has been given by Dontchev and Lempio [10]. Under certain conditions guaranteeing uniqueness, high order methods have been given by Kastner-Maresch [16] and Stewart [36].

Differential Inclusions on Submanifolds

We will have occasion to use a slightly generalized version of Theorem 5.7 for which F is known only to satisfy basic conditions on some C^1 submanifold $\mathcal{S} \subset \mathcal{R}^n$. We supplement the basic conditions with a *tangency condition* which states that every

element of $F(t,x)$ is tangent to the submanifold $\mathcal{S}$. Then we may use a flattener argument in order to reduce it to the previously known case.

THEOREM 5.8. *Suppose that $\mathcal{S}$ is a C^1 submanifold of $\mathbb{R}^n$. Suppose that $F(t,x) : \mathbb{R}\times\mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ is a set-valued function that satisfies the basic conditions on $\mathbb{R}\times\mathcal{S}$. Further suppose the* tangency condition*: for every $(t,x)\in\mathbb{R}\times\mathcal{S}$,*

$$F(t,x) \subset T_x\mathcal{S}.$$

Then for any initial condition $(t_0,x_0)\in\mathbb{R}\times\mathcal{S}$ the differential inclusion $\dot{x}\in F(t,x)$ admits a solution for sufficiently small time.

Proof. Let $(t_0,x_0)\in\mathcal{S}$. We show a solution to the DI exists. First, we will need to introduce some notions regarding the submanifold.

Since $\mathcal{S}$ is a submanifold of $\mathbb{R}^n$, there exists a neighborhood U (open in $\mathbb{R}^n$) about $x_0\in\mathcal{S}$ which comes equipped with a flattener $\psi : U\to\mathbb{R}^n$. See [15] for more details on submanifolds; in particular, the flattener ψ is C^1, has a C^1 inverse, and satisfies

$$\psi(U\cap\mathcal{S}) \subset \mathbb{R}^k\times\{0\}^{n-k}. \tag{5.2}$$

The derivative of the flattener is a map between tangent spaces: For each $p\in U$,

$$D\psi|_p : T_pU \to T_{\psi(p)}\mathbb{R}^n = \mathbb{R}^n.$$

For convenience, we identify the tangent space of the Euclidean space $\mathbb{R}^n$ with $\mathbb{R}^n$ itself.

It is straightforward from (5.2) that the tangent directions of $\mathcal{S}$ map through $D\psi$ to tangent directions of $\mathbb{R}^k \times \{0\}^{n-k}$: For each $p \in U \cap \mathcal{S}$,

$$D\psi|_p : T_p\mathcal{S} \to \mathbb{R}^k \times \{0\}^{n-k}. \tag{5.3}$$

Now that we have introduced the flattener map we are ready to continue. Our proof proceeds in three steps.

Step 1. Define the set-valued function $G(t, u) : \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}^k$ so that

$$G(t, u) \times \{0\}^{n-k} = D\psi|_x \left(F(t, x)\right), \tag{5.4}$$

where

$$x = \psi^{-1}(\{u\} \times \{0\}^{n-k}).$$

We show G is well defined. By (5.3), observe that the tangent directions in $\mathcal{S}$ at some point $x \in \mathcal{S}$ are mapped by $D\psi|_x$ to $\mathbb{R}^k \times \{0\}^{n-k}$. By the tangency condition, $F(t, x)$ consists only of elements which are tangent to $\mathcal{S}$ at $x \in \mathcal{S}$. It follows that $D\psi|_x F(t, x) \subset \mathbb{R}^k \times \{0\}^{n-k}$, so G is well-defined.

Step 2. We show that G satisfies the *basic conditions* of Definition 5.6.

To see that G is nonempty, closed, convex, and bounded, we consider (5.4) and note that for each $x \in \mathcal{S}$, $D\psi|_x$ is a linear map between finite dimensional spaces and also that $F(t, x)$ is nonempty, closed, convex, and bounded. That G is the projection of the image through a linear map of a nonempty, closed, convex, and bounded set makes it nonempty, closed, convex, and bounded itself.

Now we show that $G(t,u)$ is upper-semicontinuous in (t,u). We show that if $((t_n,u_n))$ is a sequence in $\mathbb{R}\times\mathbb{R}^k$ tending to (t_0,u_0) and (g_n) is a sequence in $\mathbb{R}^k$ tending to g_0 such that for each n,

$$g_n \in G(t_n,u_n),$$

then $g_0 \in G(t_0,u_0)$.

Define

$$z_n := D\psi^{-1}|_{u_n\times\{0\}^{n-k}}(g_n\times\{0\}^{n-k}).$$

Since ψ^{-1} is C^1, the linear map $D\psi_p^{-1}$ is continuous for all $p\in\psi(\mathcal{S})$. It follows that z_n converges to

$$z_0 := D\psi^{-1}|_{u_0\times\{0\}^{n-k}}(g_0\times\{0\}^{n-k}).$$

We define $x_n := \psi^{-1}(u_n\times\{0\}^{n-k})$. Observe that $x_n\in\mathcal{S}$. The sequence (x_n) converges to $x_0 := \psi^{-1}(u_0\times\{0\}^{n-k})\in\mathcal{S}$ by continuity of ψ^{-1}.

By (5.4), observe that $g_n\times\{0\}^{n-k} = D\psi|_{x_n}F(t_n,x_n)$ for all $n=0,1,2,\cdots$. Accordingly,

$$z_n = D\psi^{-1}|_{u_n\times\{0\}^{n-k}}(g_n\times\{0\}^{n-k})\in F(t_n,x_n)$$

for each n. Since F is upper semicontinuous, it follows that $z_0\in F(t_0,x_0)$. Applying the map $D\psi|_{x_0}$, we find

$$D\psi|_{x_0}z_0 \in D\psi|_{x_0}F(t_0,x_0),$$

from which we conclude $g_0\in G(t_0,x_0)$. We have shown G is upper semicontinuous.

Step 3. We now construct a solution.

It follows from Theorem 5.7 that there is a solution to the DI

$$\dot{u} \in G(t, u) \tag{5.5}$$

with the initial condition $u_0 = u(t_0) = \psi(x_0)$. Let $u(t)$ denote such a solution, defined on the time domain $[t_0, t_0 + T)$.

Define, for $t \in [t_0, t_0 + T)$,

$$x(t) := \psi^{-1}(\{u(t)\} \times \{0\}^{n-k}). \tag{5.6}$$

We show that $x(t)$ is a solution to the DI $\dot{x} \in F(t, x)$, with initial condition $x(t_0) = x_0$.

Since $x(t) \in \mathcal{S}$ for all $t \in [t_0, t_0+T)$, it follows straightforwardly that $\frac{dx}{dt}(t) \in T_{x(t)}\mathcal{S}$ for all $t \in [t_0, t_0 + T)$. Differentiating Equation (5.6) with respect to time, obtain

$$\frac{dx}{dt}(t) = D\psi^{-1}|_{(\{u(t)\}\times\{0\}^{n-k})}(\{\dot{u}(t)\} \times \{0\}^{n-k}). \tag{5.7}$$

Continuing,

$$D\psi|_{x(t)} \left(\frac{dx}{dt}(t) \right) = \{\dot{u}(t)\} \times \{0\}^{n-k}. \tag{5.8}$$

Substituting Equation (5.5) into Equation (5.8), we have

$$D\psi|_{x(t)} \left(\frac{dx}{dt}(t) \right) \in G(t, u) \times \{0\}^{n-k}. \tag{5.9}$$

Comparing Equation (5.9) to Equation (5.4) and noting that for all $x \in U \cap \mathcal{S}$, $D\psi|_x$ is a bijection from $T_x\mathcal{S}$ to $\mathbb{R}^k \times \{0\}^{n-k}$, we conclude that

$$\dot{x} \in F(t, x)$$

almost always. The initial condition follows from $x_0 = \psi^{-1}\left(\{u(t_0)\} \times \{0\}^{n-k}\right)$. $\square$

Convex Programs

Convex programs (CP's) are a type of optimization problem. One is given a non-empty, closed convex subset K of a Euclidean space $\mathbb{R}^n$, and a cost functional $f^{\text{cost}}(x)$ which is again convex, and wishes to produce the subset of K consisting of the minimizers of $f^{\text{cost}}(x)$. In this thesis, we restrict our attention to linear cost functionals

$$f^{\text{cost}}(\cdot) = \langle c, \cdot \rangle \,,$$

where $c \in \mathbb{R}^n$. Here, c is known as a *cost vector*. This specialization yields optimality problems similar to those of linear programming, except one is allowed any convex feasible set rather than only linear polytopes.

Statement and Notation

DEFINITION 5.9. *Given a closed, bounded, convex subset K of $\mathbb{R}^n$ and a so-called* cost vector $c \in \mathbb{R}^n$*, we use the terminology $SOL(K, c)$ to refer to the set of vectors v for which $v \in K$ and $c^T v$ is minimized under this constraint. That is,*

$$SOL(K, c) := \{z \in K : \langle c, z \rangle = \min_{w \in K} \langle c, w \rangle\}.$$

Numerics

We will not worry too much about the numerical methods for actually finding the optimal set. We do mention that in the case where K is a linear polytope and the cost functional in linear, one may employ an LCP approach, and use Lemke's algorithm.

In fact, this works as well for a quadratic cost functional. For more information, consult [27], [28], or [17].

Preliminary Results

LEMMA 5.10. *Suppose that $K(t,x)$ is a set-valued function which is satisfies the following conditions:*

$$\begin{cases} K(t,x) \text{ is nonempty, closed, convex, and bounded for all } (t,x). \\ \\ K(t,x) \text{ is continuous in } (x,t). \end{cases}$$

Suppose further that $c(t,x)$ is an $\mathbb{R}^n$-valued function that is continuous in (t,x). Then the set-valued function

$$F(t,x) := SOL(K(t,x), c(t,x))$$

satisfies the basic conditions *of Definition 5.6.*

Remark. To avoid confusion, we explicitly point out that $c(t,x)$ is a vector which acts as a cost functional through the formula $\langle c(t,x), z\rangle$, for $z \in K$. It does not act on x as if we meant $c(t,x) = \langle c(t), x\rangle$.

Proof. **Step 1.** First we show that $F(t,x)$ is always closed, convex, bounded, and non-empty.

To see that F is closed-valued, note that a convergent sequence of optimizers in K must converge to an optimizer since

$$\lim_{n\to\infty} \langle c, x_n\rangle = \langle c, x\rangle \text{ whenever } \lim_{n\to\infty} x_n = x.$$

Since K itself is closed, the limit of optimizers must be in K as well, and also an optimizer in K. Hence the set of optimizers in K is closed.

To see that F is convex-valued, suppose that x and y are optimizers in K. Let $\alpha \in [0,1]$ and define $z = \alpha x + (1-\alpha)y$. Then

$$\langle c, z\rangle = \langle c, \alpha x + (1-\alpha)y\rangle = \alpha \langle c, x\rangle + (1-\alpha)\langle c, y\rangle .$$

Since both $\langle c, x\rangle$ and $\langle c, y\rangle$ are the same optimal value, it follows that a convex combination of them is that optimal value as well. Hence $z = \alpha x + (1-\alpha)y$ is an optimizer of K provided it is in K; it is, since K is convex and z is a convex combination of elements of K.

To see that $F(t,x)$ is non-empty, we invoke the extreme value theorem: knowing that K is a closed and bounded subset implies that it is compact; hence the extreme value theorem applies and we know that the function $f : K \to \mathbb{R}$ such that $f(z) = \langle c, z\rangle$ admits a minimizer. Such a minimizer is an element of $F(t,x)$.

To see that $F(t,x)$ is bounded, observe that it is a subset of $K(t,x)$ which is assumed to be bounded.

Step 2. We show that $F(t,x)$ is upper-semicontinuous in (t,x).

We show upper semicontinuity at (t_0, x_0). Let (x_n) be sequence of points converging to x_0. Let (t_n) be a sequence of real numbers converging to t_0. Suppose that (z_n) is a sequence converging to z_0 such that for each n, $z_n \in F(t_n, x_n)$. We show $z_0 \in F(t_0, x_0)$.

Since $z_n \in F(t_n, x_n)$, we know that $\langle c(t_n, x_n), z_n \rangle$ is the optimal value in $K(t_n, x_n)$ with cost functional $c(t_n, x_n)$ for each positive integer n. Since c is continuous,

$$\lim_{n\to\infty} \langle c(t_n, x_n), z_n \rangle = \langle c(t_0, x_0), z_0 \rangle .$$

Since K is upper semicontinuous and $z_n \in K(t_n, x_n)$, it follows that $z_0 \in K(t_0, x_0)$. Accordingly, to show that $z_0 \in F(t_0, x_0)$, it suffices to show that z_0 is an optimizer of $\langle c(t_0, x_0), z \rangle$ for $z \in K(t_0, x_0)$.

Suppose to the contrary: assume that z_0 is not optimal in $K(t_0, x_0)$ with cost functional $c(t_0, x_0)$. Let w_0 be an optimizer in $K(t_0, x_0)$ with cost functional $c(t_0, x_0)$. By optimality of w_0,

$$\langle c(t_0, x_0), w_0 \rangle < \langle c(t_0, x_0), z_0 \rangle . \tag{5.10}$$

Define the sequence (w_n) to consist of the unique closest points in the nonempty, closed, convex sets $K(t_n, x_n)$ to the point w_0. Because K is lower semicontinuous, the sequence (w_n) must converge to w. Since c is continuous,

$$\lim_{n\to\infty} \langle c(t_n, x_n), w_n \rangle = \langle c(t_0, x_0), w_0 \rangle .$$

Now we consider the squeeze theorem of limits. For each positive integer n, we know by optimality of z_n in $K(t_n, x_n)$ that

$$\langle c(t_n, x_n), w_n \rangle \geq \langle c(t_n, x_n, z_n) \rangle .$$

From the preceding we may observe

$$\langle c(t_0, x_0), w_0 \rangle = \lim_{n\to\infty} \langle c(t_n, x_n), w_n \rangle \geq \lim_{n\to\infty} \langle c(t_n, x_n, z_n) \rangle = \langle c(t_0, x_0), z_0 \rangle . \tag{5.11}$$

Comparing Equations (5.10) and (5.11) we have contradiction. This contradiction arose through assuming the non-optimality of z_0 in $K(t_0, x_0)$. It follows that $z_0 \in F(t_0, x_0)$, and we have shown upper-semicontinuity of F. □

Feedback Problems

Statement

We now state the basic equations of the feedback problem:

$$\dot{u} = f(t, u, x) \tag{5.12}$$

$$\dot{x} \in \text{SOL}(K(t, u, x); c(t, u, x)) + g(t, u, x). \tag{5.13}$$

Here, the solutions $u(t)$ and $x(t)$ are understood to be absolutely continuous functions of time, and equations (5.12-5.13) may be violated on a set of measure zero (they hold almost everywhere).

We define a *feedback problem* to be finding an absolutely continuous pair of functions $(u(t), x(t))$ satisfying initial conditions and also equations (5.12-5.13) almost-everywhere. Also, we specialize to the case where f, K, c and g satisfy certain assumptions.

Remark. Since (5.12-5.13) are the basis for the model of friction of the next chapter, we take a moment to compare it to (4.9-4.10), which are the framework for the modern Coulomb model. In particular, (5.12-5.13) are a differential inclusion (DI) whereas (4.9-4.10) are a measure differential inclusion (MDI). The inability for the feedback problem to handle discontinuities prevents it from dealing with a theory with impacts.

Accordingly, the theory we present in the next chapter deals only with the case of persistent contact and no impacts. On the other hand, (5.12-5.13) imposes much more structure. The details of the modern Coulomb friction model are not visible in (4.9-4.10), but rather are inserted as auxiliary conditions. However, in the upcoming model (in Chapter 6), (5.12-5.13) will give a complete formulation of our novel model for persistent frictional contact.

Existence.

DEFINITION 5.11 (Hypothesis H.). *Let n, k be positive integers. Suppose that $f : \mathbb{R} \times \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R}^k$ (which we write $f(t, u, x)$) is continuous. Suppose that $g, c : \mathbb{R} \times \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R}^n$ (which we write $g(t, u, x)$ and $c(t, u, x)$, respectively) are continuous. Assume that $K(t, u, x)$ is a set-valued function from $\mathbb{R} \times \mathbb{R}^k \times \mathbb{R}^n$ to $\mathcal{P}(\mathbb{R}^n)$ which is continuous in (t, u, x). Further suppose that K has non-empty, closed, convex, and bounded values.*

We define the set-valued function

$$F(t, (u, x)) := \{f(t, u, x)\} \times (\mathrm{SOL}(K(t, u, x), c(t, u, x)) + \{g(t, u, x)\}) . \tag{5.14}$$

We have the following straightforward observation:

PROPOSITION 5.12. *The feedback equation (5.12-5.13) with the initial condition (t_0, u_0, x_0) and the differential inclusion*

$$\dot{w} \in F(t, w), \tag{5.15}$$

with initial condition $(t_0, w_0) = (t_0, u_0 \times x_0)$ *where* F *is defined as in Equation 5.14 are identical; i.e. a solution of either one is a solution to the other.*

PROPOSITION 5.13. *Under Hypothesis H, the set-valued function* $F(t, w)$ *of equation 5.14 satisfies the basic conditions of Definition 5.6.*

Proof. Define $\tilde{K}(t, w) := K(t, u, x)$ and see that it, along with $\tilde{c}(t, w) := c(t, u, x)$ satisfy the assumptions of Lemma 5.10, where we write $w := (u, x)$. Hence the set-valued function

$$G(t, w) := \mathrm{SOL}(K(t, u, x), c(t, u, x))$$

satisfies the basic conditions. Observe also that f and g may be regarded as singleton-set-valued functions which satisfy the basic conditions. Now we apply two straightforward observations:

1. The sum of set-valued functions satisfying the basic conditions again satisfies the basic conditions.

2. The cartesian product of set-valued functions satisfying the basic conditions again satisfies the basic conditions.

Since $F(t, w)$ of Equation (5.14) can be constructed from set-valued functions satisfying the basic conditions using these two rules of construction, it follows that F satisfies the basic conditions. □

THEOREM 5.14. *Under Hypothesis H, the Feedback Problem given by Equations (5.12-5.13) admits a solution for sufficiently short time.*

Proof. This follows from Propositions 5.12, 5.13, and Theorem 5.7. These show that the Feedback Problem under Hypothesis H is a differential inclusion satisfying basic conditions; hence we may use the existence theorem for differential inclusions. □

We will need an existence result for Feedback Problems on submanifolds:

THEOREM 5.15. *Suppose that $\mathcal{S}$ is a C^1 submanifold of $\mathbb{R}^k \times \mathbb{R}^n$. Suppose that the Feedback Problem given by Equations (5.12-5.13) satisfies Hypothesis H when restricted to $\mathbb{R} \times \mathcal{S}$. Take F to be defined as in Equation 5.14. Suppose that for every $t \in \mathbb{R}$, every $(u, x) \in \mathcal{S}$, we have that every element of $F(t, (u, x))$ is tangent to $\mathcal{S}$ at (u, x):*

$$F(t, (u, x)) \subset T_{(u,x)}\mathcal{S}. \tag{5.16}$$

Then for any initial condition $(t_0, u_0, x_0) \in \mathbb{R} \times \mathcal{S}$ the Feedback Problem given by Equations (5.12-5.13) admits a solution on $\mathbb{R} \times \mathcal{S}$ satisfying the initial condition for sufficiently small time.

Remark. It is straightforward to see this theorem also applies when the conditions on F only hold on an open subspace $U \subset \mathbb{R} \times \mathcal{S}$, provided the initial condition $(t_0, u_0, x_0) \in U$.

Proof. This follows from Propositions 5.12, 5.13, and Theorem 5.8. These show that the Feedback Problem satisfying Hypothesis H (on $\mathcal{S}$) is a differential inclusion satisfying basic conditions on $\mathcal{S}$; hence we may use the existence theorem for differential inclusions on submanifolds. □

Negative Feedback Problems

We have shown that FP's satisfying Hypothesis H admit solutions. We have not addressed whether the solutions are unique, or whether or not they vary continuously with respect to the problem data (well-posedness).

In this section, we give addition assumptions which guarantee uniqueness. A result due to Filippov [12] shows that uniqueness implies well-posedness in the context of DI's.

Negative Feedback Condition

We wish to find a condition which leads to uniqueness, or at least gets us part of the way there. To this end, we consider a simple uniqueness counterexample. We analyze how non-uniqueness arises, and then stipulate a condition which prevents non-uniqueness in a class of FP's, as we see in the next subsection.

Example. Suppose $K(t, u, x)$ is the constant set $[-1, 1]$. Choose the dimension $k = 0$, and $n = 1$. Choose $c(t, u, x) = -x$, and $g(t, u, x) = 0$. Observe that we obtain the following FP satisfying Hypothesis H:

$$\dot{x} \in \mathrm{SOL}([-1, 1], -x).$$

Take an initial time $t_0 = 0$ and an initial condition $x(0) = x_0 = 0$. It is straightforward to verify that the following three functions (which all satisfy the initial conditions) are each solutions to the feedback initial value problem: $x^1(t) = 0$, $x^2(t) = t$, and $x^3(t) = -t$.

We verify the solution $x^1(t) = 0$. Observe that $0 \in \text{SOL}([-1, 1], 0)$ always. Hence $x^1(t)$ is a solution.

We verify the solution $x^2(t) = t$. Observe that $1 \in \text{SOL}([-1, 1], -t)$ for all $t > 0$. Hence $x^2(t)$ is a solution.

We verify the solution $x^3(t) = -t$. Observe that $-1 \in \text{SOL}([-1, 1], +t)$ for all $t > 0$. Hence $x^3(t)$ is a solution.

The crux of this example is straightforward: if perturbing c causes a positive feedback effect (that is, the perturbation ends up reinforcing itself, like perturbing from an unstable equilibrium), then it is possible to have non-uniqueness whenever the convex program is degenerate. We wish to devise a condition which counteracts this. Notice in the above example that if we replaced $-x$ with $+x$, then we obtain uniqueness.

In more general cases, we have higher dimension for x, and a cost vector $c(t, u, x)$ which may be more general than $\pm x$. Speaking informally, we suspect that if we characterize what it means for c to be "like $+x$," then we may impose a condition on it which will lead to uniqueness (or at least prevent certain kinds of non-uniqueness). With c depending on x in this "positive fashion" the optimization problem will depend on x in a "negative fashion," and in turn $\dot{x}$ will depend on changes in x in a "negative fashion." Also, of course, for uniqueness we demand that data be Lipschitz continuous. This intuition leads to the following definition, which attempts to make sense of this informal thought:

DEFINITION 5.16. *We state the* negative feedback condition *on FP's satisfying Hypothesis H: the data (f, g, c) are now assumed to be Lipschitz continuous, and the vector-valued function $c(t, u, x)$ is additionally assumed to be C^2 and have the following property: For all $(t, u, x) \in \mathbb{R} \times \mathbb{R}^k \times \mathbb{R}^n$, the $n \times n$ Jacobian matrix $[c_x]_{ij} := \frac{\partial c_i}{\partial x_j}$ is symmetric positive definite. We refer to a feedback problem satisfying Hypothesis H and the negative feedback condition* a negative feedback problem*, or NFP.*

A Special Class of NFP's with Uniqueness

We have a condition which we hope may prevent non-uniqueness in FP's. Although we will see that this hope fails (some other ingredient is required other than the negative feedback condition), we do find uniqueness for a special class of NFP's, as the following theorem asserts.

THEOREM 5.17. *Let (K, f, g, c) be as in an NFP. Assume further that $K(t, u, x) = K(t)$; that is, K depends only on t, not on u or x. Then the NFP admits a unique solution for any initial condition.*

Proof. Suppose not. Let $(u(t), x(t))$ and $(v(t), y(t))$ be distinct solutions sharing the initial condition (u_0, x_0) at time $t = t_0$.

Define $a(t) = \dot{x}(t) - g(t, u(t), x(t))$. Observe that $a(t) \in \mathrm{SOL}(K(t), c(t, u(t), x(t))$ for all t in the solution domain. Similarly define $b(t) = \dot{y}(t) - g(t, v(t), y(t))$ and observe that $a(t) \in \mathrm{SOL}(K(t), c(t, v(t), y(t))$ for all t in the solution domain.

Define

$$M(t) = \int_0^1 c_x(t, u(t), (1-s)x(t) + sy(t))ds.$$

Since the integrand is SPD valued, if follows that $M(t)$ is symmetric positive definite for every t in the solution domain. By the Fundamental Theorem of Calculus,

$$M(t)(x(t) - y(t)) = c(t, u(x), x(t)) - c(t, u(t), y(t)).$$

Claim. For some $C > 0$, the following inequality holds:

$$\langle x(t) - y(t), M(t)(\dot{x} - \dot{y})\rangle \leq C|x(t) - y(t)|^2.$$

Observe that

$$\begin{aligned} \langle M(t)(x(t) - y(t)), \dot{x}(t) - \dot{y}(t)\rangle = \\ \langle M(t)(x(t) - y(t)), g(t, u(t), x(t)) - g(t, v(t), y(t))\rangle \\ + \langle M(t)(x(t) - y(t)), a(t) - b(t)\rangle . \end{aligned} \tag{5.17}$$

By the Cauchy-Schwarz inequality, the boundedness of $M(t)$, and Lipschitz continuity of g, the first term on the right hand side of (5.17) is bounded by $C|x(t) - y(t)|\left(|x(t) + y(t)| + |u(t) + v(t)|\right)$, for some $C > 0$.

The second term on the right-hand side of (5.17) is equal to

$$\begin{aligned} \langle c(t, u(t), x(t)) - c(t, u(t), y(t)), a(t) - b(t)\rangle = \\ \langle c(t, u(t), x(t)), a(t) - b(t)\rangle \quad + \langle c(t, u(t), y(t)), b(t) - a(t)\rangle . \end{aligned} \tag{5.18}$$

The first term on the right-hand side of (5.18) is non-positive, since $a(t)$ is an optimizer in $K(t)$ with the cost functional $c(t, u(t), x(t))$, and $b(t) \in K(t)$.

Since c is Lipschitz continuous, the second term on the right-hand side of (5.18) satisfies the inequality

$$\langle c(t,u(t),y(t)),b(t)-a(t)\rangle \leq \langle c(t,v(t),y(t)),b(t)-a(t)\rangle + |a(t)-b(t)||u(t)-v(t)|. \tag{5.19}$$

The first term on the right-hand side of (5.19) is non-positive, since $b(t)$ is an optimizer in $K(t)$ with the cost functional $c(t,v(t),y(t))$, and $a(t) \in K(t)$.

The second term on the right-hand side of (5.19) is bounded by $C|u(t)-v(t)|$ for some $C>0$. In particular we may choose C to be the diameter of $\cup_{t\in[t_0,t_0+\epsilon]}K(t)$, which is finite since K has bounded values and is upper-semicontinuous in its argument t.

Taking this all together, we get the inequality

$$\langle x(t)-y(t),M(t)(\dot{x}-\dot{y})\rangle \leq C|x(t)-y(t)|\left(|x(t)-y(t)|+|u(t)-v(t)|\right). \tag{5.20}$$

By the Lipschitz property of f, $|\dot{u}-\dot{v}| = |f(t,u(t),x(t))-f(t,v(t),y(t))| \leq C(|u(t)-v(t)|+|x(t)-y(t)|)$ for some $C>0$. Notice

$$\frac{d}{dt}|u(t)-v(t)| \leq |\dot{u}(t)-\dot{v}(t)|,$$

and hence

$$\frac{d}{dt}|u(t)-v(t)| \leq C(|u(t)-v(t)|+|x(t)-y(t)|), \text{ some } C>0.$$

Since $u(t_0)=v(t_0)$, it follows from Gronwall's Lemma that for any sufficiently small $\epsilon>0$, there exists $C>0$ such that for $t\in[t_0,t_0+\epsilon]$,

$$|u(t)-v(t)| \leq C|x(t)-y(t)|.$$

Substituting this into (5.20), we have the desired result. *This completes the proof of the claim.*

Now we consider the function $\phi(t) := \frac{1}{2}\langle x(t) - y(t), M(t)(x(t) - y(t))\rangle$. Since $x(t) \neq y(t)$ for all t, it follows that $\phi(t) > 0$ for some t. We show this is impossible:

$$\frac{d}{dt}\phi(t) = \frac{1}{2}\left\langle x(t) - y(t), \dot{M}(t)(x(t) - y(t))\right\rangle + \left\langle x(t) - y(t), \dot{M}(t)(\dot{x}(t) - \dot{y}(t))\right\rangle.$$

By the boundedness of $\dot{M}$, the Cauchy-Schwarz inequality, and the claim, we may now write

$$\frac{d}{dt}\phi(t) \leq C|x(t) - y(t)|^2.$$

Let $\kappa > 0$ be the smallest eigenvalue of $M(t)$ for any $t \in [t_0, t_0 + \epsilon]$. Then for $t \in [t_0, t_0 + \epsilon]$,

$$|x(t) - y(t)|^2 \leq \frac{1}{2\kappa}\langle x(t) - y(t), M(t)(x(t) - y(t))\rangle.$$

It follows that $\dot{\phi} \leq C\phi$, and also $\phi \geq 0$, and $\phi(t_0) = 0$. Gronwall's Lemma shows that $\phi(t) = 0$ for all $t \in [t_0, t_0 + \epsilon]$. This shows that $x(t) = y(t)$ for all $t \in [t_0, t_0 + \epsilon]$, from which uniqueness follows. □

A Uniqueness Counterexample for NFP's

Do all NFP's have uniqueness? The answer is no, even for the class for which K depends only on u, as the following example shows.

Example. Let (K, f, g, c, k, d) be the data for an NFP, with the following definitions. We take $k = 2$ and $d = 2$. (This makes x and u two-dimensional,

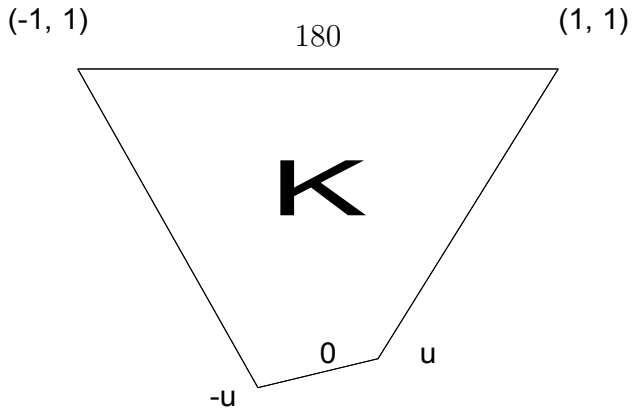

Figure 3. Feasible set $K(u)$ for uniqueness counterexample.

and K a convex subset of the plane.) Let $K(u)$ be the convex hull of the points $\{(1,1),(-1,1),u,-u\} \subset \mathbb{R}^2$. Our example has an initital condition $u_0 = (0,0)$, so K is initially a triangle. As u varies, the bottom vertex of this triangle may split into two vertices. See Figure 3. Choose $c(t,u,x) := x$. Choose $g(t,u,x) := 0$. Choose

$$f(t,u,x) := \begin{bmatrix} 0 & 0 \\ -4 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Take the initial condition to be $(u_0, x_0) = ((0,0),(0,1))$. We give two distinct solutions to this NFP. (There is also a third solution, but it is similar to the first.)

The manner in which these solutions can be determined is the following (we indicate this so it does not appear that we are pulling a rabbit out from a hat). As the problem evolves, we consider three possible behaviors that could emerge. In the first behavior, the point $u \in K$ is optimal for sufficiently small $t > 0$. In the second behavior, $0 \in K$ is optimal for sufficiently small $t > 0$. In the third behavior, $-u \in K$ is optimal for sufficiently small $t > 0$.

We consider these behaviors in turn. For the first behavior we would have to have

$$\begin{cases} \dot{u}_1 &= 1 \\ \dot{u}_2 &= -4x_1 \\ \dot{x}_1 &= u_1 \\ \dot{x}_2 &= u_2 \end{cases}.$$

We can solve this by first solving for u_1, then x_1, then u_2, and finally x_2. We arrive at

$$u(t) = \left(t, \frac{-2t^3}{3}\right) \text{ and } x(t) = \left(\frac{t^2}{2}, 1 - \frac{t^4}{6}\right).$$

To show that this candidate solution really is a solution of the FP it is necessary (and sufficient) to show that $u(t)$ is optimal in $K(u(t))$ with cost vector $x(t)$ for $t > 0$ sufficiently small. The points $(1, 1)$ and $(-1, 1)$ are clearly non-optimal, and it suffices to show that $\langle x(t), u(t)\rangle < \langle x(t), -u(t)\rangle$, that is, that u is more optimal than $-u$. This simplifies to proving that $\langle x(t), u(t)\rangle < 0$, which holds due to the calculation

$$u(t) \cdot x(t) = \frac{t^3}{2} - \frac{2t^3}{3} + \frac{t^7}{9} < 0 \text{ for sufficiently small } t > 0.$$

The second behavior involves $0 \in K$ being optimal for sufficiently small $t > 0$, and accordingly we try to find a candidate solution such that $\dot{x} = 0$. It is straightforward to see that such a candidate solution must be given by

$$u(t) = (t, 0) \text{ and } x(t) = (0, 1), \text{ for sufficiently small } t > 0.$$

To show that this candidate solution is a solution of the FP, observe that as $x(t) = (0, 1)$, and $u_2(t) = 0$, the point $(0, 0)$ is optimal in $K(u(t))$ with the cost vector $x(t)$ and hence $\dot{x}(t) = (0, 0)$ is allowed.

The third solution is determined similarly as the first two, but from the third possible behavior described, such that $-u$ is optimal in K for sufficiently short time.

Conjectures and Future Work

The counterexample of the previous subsection shows that uniqueness questions regarding NFP's are non-trivial. Despite this, there is hope that many important classes of NFP's will submit to a uniqueness result. In particular, we conjecture, but have not shown, that the NFP's arising in the work of the next chapter have unique solutions that vary continuously with the data. See the discussion in Chapter 6 for more details.

Another idea concerning FP's is that we should allow for the same extension to measure-theoretic solutions as is done for DI's. In particular, if we allow the set $K(t)$ to not have bounded values, then it is possible that $\mathrm{SOL}(K, c)$ is empty. The solution of the optimization problem is *an unbounded ray.* It appears possible to take this "solution set" of unbounded rays as "unbounded values" for the derivative $\dot{x}$, analogous to the discussion in Chapter 4 regarding MDI's. We discuss this more at the end of Chapter 6.

Summary

In this section we studied differential inclusions and an existence result due to Filippov [12]. We also considered a special type of differential inclusion involving a convex program, which we named the feedback problem. We gave conditions for

existence to the feedback problem. We gave conditions for uniqueness for a special class of feedback problems obeying the so-called negative feedback condition. We showed that non-uniqueness issues persist even for negative feedback problems, but also indicated that there is hope for a useful class of such problems to admit a well-posedness result. The possibility of measure-theoretic feedback problems was also suggested.

CHAPTER 6

A NOVEL MODEL OF FRICTION

Introduction

In this chapter, we use the mathematical formalism of the Feedback Problem developed in Chapter 5 in order to frame a new model of frictional contact. The model we construct admits a multi-contact existence theorem for the case of persistent contact. Persistent contact means that the active (touching) set of contacts remains constant for a sufficiently small amount of time.

Our friction model departs from the Coulomb friction model in two ways. First, we alter the maximal dissipation principle, in effect taking it more literally. At each contact, we stipulate that there is an allowed friction set. The frictional force at the contact is chosen from this allowed friction set. The choices of forces are restricted so that the bilateral constraints are not violated. In the resulting collection of possible forces on the contacts, we choose a frictional force which maximizes the rate of energy lost due to friction. This differs from the Coulomb model's maximal dissipation principle: in Coulomb friction, we determine the directions of forces at each contact, in effect a "local" version of maximal dissipation. In the model of this chapter, we maximize dissipation *globally*. Hence we call this assumption the *global maximal dissipation principle.*

We will see that this friction law results in a linear optimization problem on some convex set K. This is the problem of determining the frictional force $F \in K$. We describe later how K is derived from the constraints, dynamics, and admissible reactive forces. For now, note that K is closed, convex, and (in cases of interest) nonempty. (If the set is in fact empty, we interpret this to mean that a contact must break; we do not have a persistent contact solution. We discuss this point later.) The cost vector v of this optimization problem is the velocity of the system $\dot{q}$. This is because, as one remembers from elementary physics, power is force times velocity (and more generally, the inner product $F \cdot v$). We derive this in the context of Hamiltonian systems shortly.

In order for $\mathrm{SOL}(K, v)$ to be non-empty, we must also know that K is *bounded*; if K is not bounded, then there may not be an optimizer. In order to assure that K is bounded, we depart from the Coulomb model again, and assume that the set of possible frictional forces at the kth contact, $\mathcal{F}_k$, is a non-empty, closed, convex set *whose only unbounded ray is in the normal direction to the constraint.* With this assumption, the possible frictional forces at a contact do not comprise a cone, as they do in Coulomb friction, unless the contact is frictionless. Because of this, we will refer to the possibilities as the *friction set* rather than friction cone. In particular, we must have, at a contact, that the tangential friction F_t and the normal force F_n satisfy

$$\frac{|F_t|}{|F_n|} \to 0 \text{ as } |F_n| \to \infty.$$

Because of this, we call this the *tapered friction set assumption.* See Figure 4 for a simple example.

With these two assumptions, (1) global maximal dissipation principle and (2) the tapered friction set assumption, we find that the frictional force is the solution to a convex optimization problem. We may cast the dynamics as a feedback problem, and invoke the existence theorem proved in the last chapter. We will also consider well-posedness questions, and questions of physicality.

New Friction Law

As discussed in Chapter 4, a formulation of Coulomb friction that does not allow for tangential impulses is plagued by non-existence problems. This is illustrated by the so-called Painlevé Paradox. These problems are remedied, at least for the single contact case (and presumably for multiple contacts, but a proof is lacking), by allowing tangential impulses and formulating the frictional contact problem as a measure differential inclusion.

In this chapter, we do not resolve Painlevé's paradox by allowing tangential impacts. Instead, we consider a different type of friction which is not Coulomb friction. We call it our *novel model of friction,* and we give a precise description below. We consider only the case of persistent contact. We do not allow for impacts of any kind, tangential or not. To allow this to be possible, the novel friction model has the *tapered friction set* assumption. The reason for this assumption is that it renders

the Painlevé Paradox solvable, as we see shortly. This removes the necessity of a framework which may handle tangential impacts.

In other aspects, the new friction law we propose is very similar to Coulomb's law. We determine the frictional force according to the following principles:

1. For each contact k, we assign a *friction set* $\mathcal{F}_k$ which represents the set of possible reactive forces from the constraint on the kth contact.

2. The net reactive force F from all constraints may be expressed as

$$F = \sum F_i, \text{ where } F_k \in \mathcal{F}_k \text{ for each contact } k.$$

3. The chosen force does not result in imminent interpenetration or contact breaking.

4. The force F chosen among possible candidates maximizes energy dissipation due to friction.

We hasten to make a few remarks regarding this prescription. First, it does not, for a fixed time, yield a unique force. There may be an entire set of possibilities for F which are optimal in the sense of maximizing dissipation. Secondly, one may wish to regard this as an optimization problem. The sum of the sets $\mathcal{F}_k$ yields a large set $\mathcal{F}$. We will find that preventing interpenetration and contact breaking amounts to introducing some linear bilateral constraints. We show later that intersecting $\mathcal{F}$ with these bilateral constraints yields a closed, bounded convex set of possible forces

which we call K. The set K are the "possible candidates" of forces mentioned in (3). Accordingly, our choice criteria is $F \in \mathrm{SOL}(K, v)$.

There is, of course, the possibility that K is empty: this indicates non-existence of solution continuations *undergoing persistent contact.* In other words, it signals that a contact break is necessary. This is different from Painlevé's paradox, which cannot be resolved by contact breaking.

Our goal in this chapter is to frame an existence result for this novel model. We define a set of "good" initial conditions for which we can assure short-time existence of a persistent contact solution. What this "good" set is will be called a *persistent contact submanifold*, which we study below.

Painlevé's Paradox, Revisited.

With the principles of our novel friction model in hand, we seek to resolve the paradox of Painlevé without resorting to impulses.

With the Coulomb Law of Chapter 2, we require $T = \pm\mu N$. With the new law, we need to specify a tapered friction set. We will choose

$$|T| \leq \mu\sqrt{|N|}.$$

See Figure 4.

Now we consider the equation at the heart of Painlevé's paradox,

$$\ddot{y}_c = \frac{N}{m} - g - \frac{\ell^2}{4J}(\cos\theta)(-N\cos\theta + T\sin\theta). \tag{6.1}$$

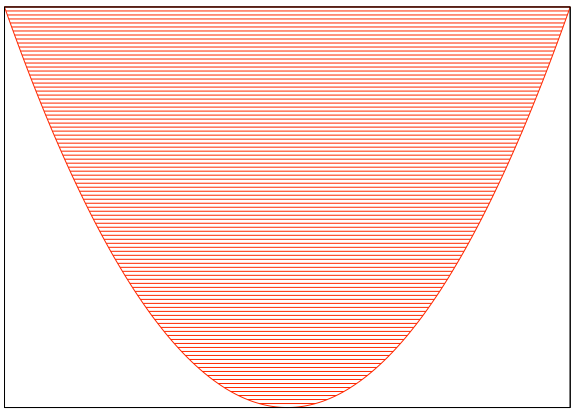

Figure 4. A Tapered Friction Set.

We have $0 \leq \ddot{y}_c \perp N \geq 0$ by complementarity.

If $\frac{\ell}{2}\dot{\theta}^2 \cos\theta - g \geq 0$, then we may choose $N = 0$ and see that we have an acceptable force solution. Otherwise, we need $N > 0$ or else $\ddot{y}_c < 0$. If we choose $N > 0$, then we must have $\ddot{y}_c = 0$. We may then write:

$$0 = \frac{N}{m} - g - \frac{\ell^2}{4J}(\cos\theta)(-N\cos\theta + T\sin\theta). \tag{6.2}$$

This provides a linear constraint on T and N; knowledge of one specifies the other. Since we demand $N \geq 0$ and $|T| < \mu\sqrt{N}$, and since the line of (T, N) solutions to Equation (6.2) is not in the direction of the unbounded ray of the convex set $|T| < \mu\sqrt{N}$, it follows that the solutions to Equation (6.2) satisfying $|T| < \mu\sqrt{N}$ comprise a line segment.

On this line segment, we wish to choose (T, N) to maximize dissipation. This is achieved, since the velocity of the contact is negative, by picking T as large as

possible. This is done by choosing

$$T = \mu\sqrt{N}.$$

Writing $c := \frac{m\ell^2}{4J}$,

$$0 = N - mg - c(\cos\theta)(-N\cos\theta + \mu\sqrt{N}\sin\theta). \tag{6.3}$$

Equation (6.3) is quadratic in $\sqrt{N}$, with solution

$$\sqrt{N} = \frac{\mu c\sin\theta\cos\theta \pm \sqrt{\mu^2c^2\sin^2\theta\cos^2\theta + 4mg(1 + c\cos^2\theta)}}{1 + c\cos^2\theta}.$$

Observe that the $\pm$ must in fact be $+$, since otherwise we arrive at a negative value for $\sqrt{N}$. Hence we arrive at the unique force prescribed by our model.

Precise Formulation of Model

Model Specification.

For matrix and vector valued functions A, b, c to be described below, we define a Hamiltonian function $H : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, such that

$$H(t, p, q) = \frac{1}{2}p^T A(q)p + p^T b(q) + c(t, q). \tag{6.4}$$

Here, p and q are position and momentum variables, each in $\mathbb{R}^n$. We make the following assumptions about A, b, and c:

- A is an $n \times n$ matrix such that A_q exists and is continuous.

- A is everywhere symmetric positive definite.
- b is an $n \times 1$ matrix such that b_q exists and is continuous.
- c is a scalar function such that c_q and c_t exist and are continuous.

We have disallowed time dependence for simplicity in everything except c, where we felt it would be too restrictive a theory without allowing c to have time dependence. This is because external driving forces require c to have t dependence. That is, if one wanted to apply an external force $F_{\text{ext}}(t)$ to the system, that could be arranged by adding $F_{\text{ext}}(t) \cdot q$ to the $c(t, q)$ term in the Hamiltonian above.

We suppose we are also supplied with finitely many constraints $\{f^k\}_{k \in J}$, each constraint f being unilateral and time-independent: they constrain the allowed positions via the inequalities

$$\mathcal{Q} := \{q : f^k(q) \geq 0 \text{ for all } k \in J\},$$

where we are defining $\mathcal{Q}$ to be the set of admissible (constraint-respecting) positions. An admissible position-momentum pair (q, p) is an element of $\mathcal{Q} \times \mathbb{R}^n$; we do not constraint the momentum variables.

For a fixed position $q \in \mathcal{Q}$, we define the *active set at* q,

$$\mathcal{A}_q := \{k \in J : f^k(q) = 0\}.$$

We suppose

- For each $k \in J$, the function f^k admits first and second derivatives f^k_q and f^k_{qq} which are both continuous.

- For each $k \in J$, constraint gradient f_q^k is non-zero in a neighborhood of $\{q : f^k(q) = 0\}$.
- For each $q \in \mathcal{Q}$, the active constraint gradients $\{f_q^k : k \in \mathcal{A}_q\}$ are linearly independent.

In addition to the Hamiltonian H and the constraints $\{f^k\}$, we have the so-called *friction sets* $\mathcal{F}_k$. There is a friction set for each possible constraint, and they satisfy the following:

- For each $k \in J$, the set $\mathcal{F}_k(q)$ is closed and convex.
- For each $k \in J$, we have that $f^k(q) > 0$ implies

$$\mathcal{F}_k(q) = \emptyset.$$

- For each $k \in J$, the set $\mathcal{F}_k$ is nonempty whenever $f^k(q) = 0$. In this situation it contains precisely one unbounded ray, and specifically,

$$\{\alpha f_q^k(q) : \alpha \geq 0\} \subset \mathcal{F}_k.$$

- The set-valued function $\mathcal{F}_k(q)$ is continuous on the submanifold defined by $f^k(q) = 0$.
- The *total friction set*

$$\mathcal{F}(q) := \sum_{f^k(q)=0} \mathcal{F}_k(q)$$

is of constant dimension when restricted to any of the so-called $\mathcal{A}$-*contact submanifolds*

$$\mathcal{Q}^{\mathcal{A}} := \mathcal{Q} \cap \{q \in \mathbb{R}^n : f^k(q) = 0 \text{ for all } k \in \mathcal{A}\},$$

where $\mathcal{A} \subset J$.

DEFINITION 6.1. *We call data* $(H, A, b, c, J, \{f^k\}_{k\in J}, \{\mathcal{F}_k\}_{k\in J}, \mathcal{F}, \mathcal{A}, \{\mathcal{Q}^{\mathcal{A}} : \mathcal{A} \subset J\})$ *satisfying the above a* model specification.

Remark. We will often call the triple (t, p, q) of time, momentum, and position an *state*. However, other times we will simply refer to the pair (p, q) as a state. Which meaning we use is will be clear from context on each occasion.

Equations of Motion

Now that we have specified the data of our framework, we consider a problem pertaining to it. Given a model specification and an initial condition (p_0, q_0), we formulate what we mean by a *short time persistent contact solution to the initial value problem of the friction model*, henceforth just *solution*.

DEFINITION 6.2. *A* friction problem with persistent contact *consists of the model specification, an initial time* t_0*, and an initial condition* $(p_0, q_0) \in \mathbb{R}^n \times \mathbb{R}^n$*. A* solution *to the friction problem with persistent contact is a pair of absolutely continuous functions* $(q(t), p(t))$*. Given such a pair, we define* $F(t) := \dot{p} + H_q(t, p(t), q(t))$*, which we call the reactive force. We demand the following of a solution:*

1. *The position* q*, the momentum* p*, and the reactive force* F *are all* $\mathbb{R}^n$*-valued functions defined on a time interval* $[t_0, t_0 + \epsilon)$*, some* $\epsilon > 0$*.*

2. *The functions* p, q*, and* F *obey Hamilton's equations on their domain of definition* $[t_0, t_0 + \epsilon)$*:*

$$\dot{q} = H_p \tag{6.5}$$

$$\dot{p} = -H_q + F \tag{6.6}$$

The equations are meant here in the Lebesgue almost-everywhere sense. (Notice that (6.6) is trivial, as it is the definition of F.)

3. *Initial conditions are satisfied: $p(t_0) = p_0$ and $q(t_0) = q_0$.*

4. *The reactive force F satisfies*

$$F(t) \in \mathcal{F}(q(t)).$$

5. *For all $t \in [t_0, t_0 + \epsilon)$, $\mathcal{A}_{q(t)} = \mathcal{A}_{q(t_0)} =: \mathcal{A}$. This is the persistent contact condition; it prohibits any contacts from releasing or impacting.*

6. *The reactive force F satisfies the* maximal dissipation principle*, to be described below.*

Persistent Contact Equations.

Definition 6.2, point (5), requires that we find a force solution which manages to keep all intially active contacts remaining active. This is to be done by preventing acceleration into or out of each constraint which we mean to be persistently active. That is, for each initally (and hence persistently) active constraint k, we are constrained to choose the force solution F in order to enforce

$$\ddot{f}^k(q) = 0.$$

It turns out these constraints yield a linear constraint on F, as we now establish.

By the chain rule of multivariable calculus and substitution of the expressions (6.5 - 6.6) , we derive expressions for $\dot{f}^k_q$ and $\ddot{f}^k(q)$:

$$\dot{f}^k(q) = f^k_q H_p.$$

$$\ddot{f}^k(q) = H_p^T f^k_{qq} H_p + f^k_q (H_{pq} H_p + H_{pp}(F - H_q)).$$

Noting linearity in F, we conclude that setting $\ddot{f}^k(q) = 0$ for each active contact $k \in \mathcal{A}$ is equivalent to the following constraint on F:

$$E(q)F + d(t, p, q) = 0. \tag{6.7}$$

We wish to write convenient expressions for E and d. To this end, we define the $m \times n$ matrix Z whose rows are the constraint normals f^k_q for all $k \in \mathcal{A}$, or, to be more explicit,

$$Z_{ij} = \frac{\partial f^i}{\partial q^j}.$$

With this notation, we can give expressions for E and d in Equation (6.7):

$$E(q) := Z H_{pp} \tag{6.8}$$

$$d(t, p, q) := Z(H_{pq} H_p - H_{pp} H_q) + \sum_{k \in \mathcal{A}} \left(H_p^T f^k_{qq} H_p \right) \hat{e}_k. \tag{6.9}$$

Notice that d has t-dependence since it involves the term H_q, and that Z has full row rank because the constraint gradients $\{f^k_q : k \in \mathcal{A}_q\}$ are linearly independent by

hypothesis. Since H_{pp} is positive definite and Z has full row rank, it follows that E has full range and (6.7) admits solutions in $\mathbb{R}^n$.

DEFINITION 6.3. *We define* $\mathcal{E}(t,p,q) := \{z : E(q)z + d(t,p,q) = 0\}$. *Technically we should write* $\mathcal{E}^{\mathcal{A}}$, *since its definition depends on the choice of the active set, but we omit this notation. We take the domain of* $\mathcal{E}$ *to be the states* $(t,p,q) \in \mathbb{R} \times \mathbb{R}^n \times \mathcal{Q}^{\mathcal{A}}$.

LEMMA 6.4. *Let* $\mathcal{A} \subset J$. *The set-valued function* $\mathcal{E} : \mathbb{R} \times \mathbb{R}^n \times \mathcal{Q}^{\mathcal{A}} \to \mathcal{P}(\mathbb{R}^n)$ *is a continuous, affine-subspace valued function of constant dimension.*

Proof. It is clearly affine-subspace valued. To see that it is of constant dimension we point out that E is a continuous matrix of constant size that has full range for all $q \in \mathcal{Q}^{\mathcal{A}}$.

To see that $\mathcal{E}$ is upper-semicontinuous, observe that $\{(z,t,p,q) : E(q)z+d(t,p,q) = 0\}$ is closed. The reason this set is closed is because is may be realized as the preimage of the closed set $\{0\}$ through the continuous map $\phi(z,t,p,q) := E(q)z + d(t,p,q)$.

Now we show that $\mathcal{E}$ is lower-semicontinous. To see this, assume that $z_0 \in \mathcal{E}(x_0)$ for some $x_0 = (t_0, p_0, q_0)$ and z_0. Then

$$E(q_0)z_0 + d(t_0,p_0,q_0) = 0.$$

Let $\tilde{E}$ and $\tilde{d}$ be perturbed versions of $E(q_0)$ and $d(t_0,p_0,q_0)$. Assume the perturbation is small enough so that $\tilde{E}$ continues to have full rank. It follows that $\tilde{E}z + \tilde{d} = 0$ must have a solution arbitrarily close to z_0 for arbitrarily small perturbations. In particular, for $z = z_0 + \Delta z$, we are solving $\tilde{E}z_0 + \tilde{E}\Delta z_0 + \tilde{d} = 0$. See that $\tilde{E}z_0 =$

$Ez_0 + (\Delta E)z_0 = \Delta E z_0 - d$. Hence

$$\tilde{E}\Delta z_0 + (\Delta E)z_0 + \Delta d = 0.$$

For small perturbations, $\Delta E z_0 + \Delta d$ is very small; the small solution for Δz_0 may be found. □

The Admissible Contact Force Set

DEFINITION 6.5 (Admissible Contact Forces.). *Define $K(t,p,q)$ to be the set of forces satisfying (6.7) that are also in the total friction set $\mathcal{F}(q)$. We may call K the* set of admissible contact forces*. We have*

$$K(t,p,q) = \mathcal{E}(t,p,q) \cap \mathcal{F}(q).$$

Remark. In the special case where

$$H(t,p,q) := \frac{1}{2}p^T M^{-1} p + V(t,q),$$

where M^{-1} is an SPD constant matrix we recover Newtonian mechanics:

$$M\ddot{q} = F - \nabla V.$$

Specializing further to the case of linear constraints $\{f^k\}$, so that f^k_{qq} vanishes, we find that E and d above depend *only on the time and position coordinates t and q.* In this case, the admissible contact forces $K(t,p,q)$ can be written $K(t,q)$. In this special case, the admissible contact forces depend only on position.

PROPOSITION 6.6. *The set-valued funtion $K(t,p,q)$ is closed, convex, and bounded.*

Proof. Since K is the sum of finitely many closed, convex sets intersected with the closed, convex set of solutions to (6.7), it must be closed and convex. Now we show that K must be bounded. Suppose otherwise. Then there exists, for each $M > 0$, a force $F = \sum F_i$, $F_i \in \mathcal{F}_i$, such that F satisfies Equation (6.7) and $|F| > M$. For each F_i, we can break it into two parts: a part in the direction of the unbounded ray that $\mathcal{F}_i$ possesses, and the remainder. Do so, obtaining $F_i = F_i^n + F_i^t$. By the tapered friction set assumption, we see that as $|F| \to \infty$, then

$$\sum |F_i^t| = o\left(\sum |F_i^n|\right).$$

Since the rows of the matrix Z are the unbounded directions of the sets $\mathcal{F}_i$, we can write $F = Z^T x + \epsilon$, where ϵ is a vector whose length is arbitrarily small compared to the length of F. Substituting this into Equation (6.7),

$$ZH_{pp}Z^T x + ZH_{pp}\epsilon + d = 0.$$

The matrix $ZH_{pp}Z^T$ is symmetric positive definite since H_{pp} is symmetric positive definite and Z has full range. Let $\delta > 0$ be its smallest eigenvalue. Then we see that

$$|x| \leq \delta |ZH_{pp}\epsilon + d| \leq \delta |\epsilon||ZH_{pp}| + \delta |d|.$$

Since $|\epsilon| < \delta^{-1}|ZH_{pp}|^{-1}|x|$ for sufficiently large $|F|$, we can conclude

$$|x| \leq 2\delta |d|,$$

which bounds the size of F, a contradiction. □

Maximal Dissipation Selection Principle

We postponed the formulation of one of the requirements in our definition of a *solution* above. This was the assumption which requires us to maximize the dissipation of energy due to friction. We of course do not mean this entirely literally, as maximizing the total dissipation of energy due to friction might result in very bizarre solutions. In such a bizarre solution, one may have a frictional surface craftily manuever its contacts into places with higher friction as if it had a mind of its own. No; we mean that friction has maximal dissipation in a greedy sense: at each time, the friction tries to maximize dissipation at that time only – not in the future. This corresponds to maximizing the dissipation *power*, where power is defined as the time derivative of energy.

This indicates that we want to formulate what it means to maximize power dissipation at a given time. To this end, we must first compute the dissipation as a function of the chosen force F. This is a simple application of the chain rule:

PROPOSITION 6.7. *The rate of energy dissipation, $-\dot{H}$, may be found via the formula*

$$\dot{H} = H_t + H_q^T H_p + H_p^T (F - H_q). \tag{6.10}$$

Proof. This is obtained from applying (6.5-6.6) to the formula for $\dot{H}$ given by the chain rule:

$$\dot{H} = H_t + H_p \cdot \dot{p} + H_q \cdot \dot{q}.$$

□

COROLLARY 6.8. *Energy dissipation is maximized by choosing F to minimize*

$$\langle H_p, F \rangle .$$

Proof. Energy dissipation, by definition, is maximized by minimizing $\dot{H}$. The only term involving F (and hence controllable by a choice of frictional force) in (6.10) is $H_p \cdot F$. □

Notice that this is a linear function of F. This turns force selection into a particularly nice optimization problem: a linear functional being optimized on the convex set K above.

PROPOSITION 6.9. *(Maximal Dissipation Rule) The force F prescribed by our new friction model satisfies*

$$F \in SOL(K(t,p,q), \dot{q}). \tag{6.11}$$

Proof. In order to satisfy the assumptions of the friction model given above, $F = \sum F_i$, with $F_i \in \mathcal{F}_i$ for all $i \in \mathcal{A}$. Also, F satisfies Equation (6.7). By the definition of K, this means $F \in K$. In order to maximize dissipation, we wish to minimize $F^T H_p$. By the Hamiltonian equations (6.5-6.6) this is equivalent to minimizing $F^T \dot{q}$. Thus F must solve the variational inequality $\mathrm{SOL}(K, \dot{q})$. □

Also, we point out that the cost functional for this optimization problem is H_p, which is just $\dot{q}$, or the *generalized velocity* of the Hamiltonian system. Power, which

is the rate of energy transfer, is usually given by the formula "force times velocity". Here, the power loss through the contacts is given by the dot product of the generalized velocity with the force due to the contacts. This is befitting.

Formulation as a Feedback Problem

We have so far given a precise statement (Definition 6.2, Definition 6.5, and Proposition 6.11) of what is meant by a persistent frictional contact problem and its solution. The purpose of the current section is to obtain a convenient formulation of the persistent frictional contact problem by casting it as a "Feedback Problem" of the last chapter. However, in order to do this we need to analyze the underlying spaces.

Contact Submanifolds

In this section, we identify sets of states (t, p, q) which we call the *persistent contact submanifolds*. Restricted to one of these sets, the set-valued function K will satisfy regularity conditions sufficient to apply the existence theory of the previous chapter.

The section is long and tedious: the important results are Proposition 6.11 and Lemma 6.21.

DEFINITION 6.10. *For a given active set $\mathcal{A}$, we define the $\mathcal{A}$-contact submanifold to be the set of states (p, q) in the set*

$$\mathcal{S}^{\mathcal{A}} := \{(p, q) : f^k = 0 \text{ and } f_q^k \cdot v(p, q) = 0 \text{ for } k \in \mathcal{A}, \text{ and } f^k > 0 \text{ for } k \notin \mathcal{A}.\},$$

where $v(p,q) := H_p(p,q)$ is the velocity $\dot{q}$ corresponding to the state (p,q).

Remark. Notice that $v = H_p$ is time independent, even though H is not.

PROPOSITION 6.11. *For a given active set $\mathcal{A}$, the $\mathcal{A}$-contact submanifold $\mathcal{S}^{\mathcal{A}}$ is a C^1 submanifold of $\mathbb{R}^n \times \mathbb{R}^n$.*

Proof. Denote by m the cardinality of $\mathcal{A}$, so $m \leq n$. (Otherwise, it would be impossible for $\{f_q^k : k \in \mathcal{A}\}$ to linearly independent.) Construct the function

$$g(p,q) := \{f\} \times \{f_q \cdot v\}.$$

Observe that g is a differentiable map from $\mathcal{M} := \mathbb{R}^n \times \mathbb{R}^n$ to $\mathcal{N} := \mathbb{R}^m \times \mathbb{R}^m$.

With this notation, we see that

$$\mathcal{S}^{\mathcal{A}} := g^{-1}(\{0\}^m \times \{0\}^m).$$

By the Regular Value Theorem (see [15]), $\mathcal{S}^{\mathcal{A}}$ is a C^1 submanifold of $\mathbb{R}^n \times \mathbb{R}^n$ provided that $\{0\}^m \times \{0\}^m$, which we will abbreviate hereafter as 0, is a regular value of g.

We show that $g(t,p,q)$ has full rank about any point (t_0, p_0, q_0) for which we have $g(t_0, p_0, q_0) = 0$. This demonstrates that 0 is a regular value of g; see [15]. This amounts to showing the Jacobian Dg, evaluated at (t_0, p_0, q_0) has full rank.

$$Dg \begin{bmatrix} dp \\ dq \end{bmatrix} = \begin{bmatrix} 0 & f_q \\ f_q \cdot v_p & f_{qq} \cdot v + f_q \cdot v_q \end{bmatrix} \begin{bmatrix} dp \\ dq \end{bmatrix}. \tag{6.12}$$

To show the matrix Dg has full rank, we apply row operations which do not effect the dimension of the range. When we arrive at a matrix that has full rank, we can

conclude, having used moves that preserved the dimension of the range, that range of Dg is full as well.

The $k \times n$ matrix f_q has full rank, hence we can use row operations to eliminate the second block-column:

$$\begin{bmatrix} 0 & f_q \\ f_q \cdot v_p & 0 \end{bmatrix}$$

We can rearrange rows:

$$\begin{bmatrix} f_q \cdot v_p & 0 \\ 0 & f_q \end{bmatrix}$$

Any block diagonal matrix (whose blocks need not even be square) has full rank provided its diagonal blocks have full rank. The $k \times n$ matrix $f_q v_p$, being the product of a $k \times n$ matrix f_q with full rank and an $n \times n$ matrix v_p that is non-singular must have full rank. Also the $k \times n$ matrix f_q has full rank. We have thus reduced Dg with row operations to a matrix which we see has full rank. We conclude that Dg has full rank.

Hence we have shown that 0 is a regular value of g, and thus $\mathcal{S}^{\mathcal{A}}$ is a C^1 submanifold of $\mathbb{R}^n \times \mathbb{R}^n$. □

Definition 6.12. *Given a convex set A, we define the* tangent space of A *to be the subspace*

$$T_A := \{v \in \mathbb{R}^n : v = \kappa(z_1 - z_2), \text{ where } \kappa \in \mathbb{R},\ z_1, z_2 \in A\}.$$

Given a point $z \in A$, we say that z is in the relative interior of A *if z is topologically interior to A with respect to the subspace topology on $T_A + \{z\}$.*

If A is a convex-valued function, then we take T_A to be the subspace-valued function such that $T_A(x)$ is the tangent space of $A(x)$ for all x in the domain of A.

DEFINITION 6.13. *We say that a given state (t, p, q) is* typical *provided that there exists an element $z \in K(t, p, q)$ such that z is in the relative interior of $\mathcal{F}$. We denote the set of typical states as $\mathcal{T}$.*

Remark. We prove later (Proposition 6.22) that the concept of typicality in $\mathbb{R} \times \mathcal{S}^{\mathcal{A}}$ is an open one. The next definition uses this fact.

DEFINITION 6.14. *For a given active set $\mathcal{A}$, and for a fixed time t^*, we define the $\mathcal{A}$-persistent contact submanifold $\mathcal{S}_p^{\mathcal{A}}(t^*)$ to be*

$$\mathcal{S}_p^{\mathcal{A}}(t^*) := \{(p, q) \in \mathbb{R}^n \times \mathbb{R}^n : (p, q) \in \mathcal{S}^{\mathcal{A}} \text{ and } (t^*, p, q) \in \mathcal{T}\}.$$

We define the time-dependent persistent-contact submanifold *to be*

$$\mathcal{P} := \{\{t\} \times \mathcal{S}_p^{\mathcal{A}}(t) : t \in \mathbb{R}\}.$$

Notice that the $\mathcal{A}$ dependence of $\mathcal{P}$ is not explicitly noted in the notation.

Remark. Since typicality is an open concept, $\mathcal{S}_p^{\mathcal{A}}(t^*)$ is a submanifold of the same dimension as $\mathcal{S}^{\mathcal{A}}$. Similarly, $\mathcal{P} = (\mathbb{R} \times \mathcal{S}^{\mathcal{A}}) \cap \mathcal{T}$ is an open submanifold of $\mathbb{R} \times \mathcal{S}^{\mathcal{A}}$. The states $(t, p, q) \in \mathcal{P}$ will be the allowed initial conditions for an upcoming existence theorem for the novel model of friction.

Continuity Proofs

In the following section, we wish to show that the set-valued function $K(t, p, q)$ is continuous when restricted to the time-dependent persistent-contact submanifold $\mathcal{P}$. This is finally obtained in Lemma 6.21. Another result of this section, Proposition 6.22, has already been used in Definition 6.14. This is the result that proves that typicality is an open concept.

LEMMA 6.15. *Suppose $A(x)$ is a continuous set-valued function with convex values such that z_0 is in the topological interior of $A(x_0)$. Then there exists $\epsilon > 0$ such that $z \in A(x)$ for all x, z such that $|x - x_0| < \epsilon$ and $|z - z_0| < \epsilon$.*

Proof. The topological boundary $\partial A(x)$ is also a continuous set-valued function in a neighborhood of x_0. Define the function

$$\phi(x, z) := d_H(z, \partial A(x)).$$

By continuity, the preimage $\phi^{-1}((0, \infty)$ is open. Since z_0 is in the topological interior of $A(x_0)$, we have $\phi(x_0, z_0) > 0$. Hence, (x_0, z_0) is in $\phi^{-1}((0, \infty)$, which is open. Hence there is an entire open ball B about (x_0, z_0) contained in $\phi^{-1}((0, \infty)$. Since each element in the open ball B is in the same connected component of $\phi^{-1}((0, \infty)$ as (x_0, z_0), it follows that every element (x, z) in B satisfies the criteria that $z \in A(x)$. The conclusion of the lemma is now satisfied. □

LEMMA 6.16. *Suppose that $F : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^m)$ is a continuous set-valued function which has nonempty, closed, convex values. Let $x_0 \in \mathbb{R}^n$ and $y_0 \in F(x_0) \subset \mathbb{R}^m$. Then*

there exists a continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$ *such that* $f(x_0) = y_0$ *and* $f(x) \in F(x)$ *for all* $x \in \mathbb{R}^n$.

Proof. Define $f : \mathbb{R}^n \to \mathbb{R}^m$ such that $f(x)$ is defined to be the (unique) closest point to y_0 in $F(x)$. Notice that $f(x_0) = y_0$ since $y_0 \in F(x_0)$. Notice that $f(x) \in F(x)$ always. All that remains is to show f is continuous.

To this end, consider a sequence (x_n) tending to some point x. We show that $f(x_n)$ converges to $f(x)$. Define y_n to be the closest point in $F(x_n)$ to $f(x)$. Observe, since $f(x_n)$ is the closest point to y_0 in $F(x_n)$ (and, in particular, closer to y_0 than y_n), that

$$|f(x_n) - y_0| \leq |y_n - y_0| \leq |f(x) - y_n| + |f(x) - y_0|,$$

where we have used the triangle inequality. Since $\lim |f(x) - y_n| = 0$ by lower-semicontinuity of F at x, we see that

$$\limsup_{n\to\infty} |f(x_n) - y_0| \leq |f(x) - y_0|. \tag{6.13}$$

Since F is upper-semicontinuous at x, observe that every limit point of $(f(x_n))$ is in $F(x)$. Since $f(x) \in F(x)$ is the unique (by uniqueness of closest points on closed convex sets) optimizer of $c(z) := |z - y_0|$, and every limit point of $(f(x_n))$ must be at least as optimal as $f(x)$ (by 6.13), it follows that every limit point of $(f(x_n))$ is $f(x)$, i.e. $(f(x_n))$ converges to $f(x)$. □

LEMMA 6.17. *Suppose* $A : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^N)$ *is a continuous set-valued function with non-empty, closed, convex values of fixed dimension* d_2. *Then the tangent space of*

A, T_A, is a continuous subspace-valued function. Moreover, if instead A is upper semicontinuous but also affine-subspace-valued, then T_A is upper semicontinuous as well.

Proof. We prove the moreover part first. We assume that A is upper semicontinuous and affine-subspace-valued and show that T_A is upper semicontinuous. To this end, assume that (x_n) is a sequence of points tending to $x_0 \in \mathbb{R}^d$ and (z_n) is a sequence of points tending to $z_0 \in \mathbb{R}^N$ such that $z_n \in T_A(x_n)$ for all n. We show that $z_0 \in T_A(x_0)$. For each n, choose points a_n and b_n in $A(x_n)$ such that $z_n = a_n - b_n$. This is possible because $A(x_n)$ is an affine subspace. Observe that we can choose the sequences (a_n) and (b_n) to be bounded, without loss. (For example, choose a_n to be the minimal distance point to the origin on $A(x_n)$ and choose $b_n = a_n - z_n$.) Consider a subsequence such that (a_n) and (b_n) both converge to a_0 and b_0 respectively. Since $A(x)$ is upper semicontinuous, $a_0 \in A(x_0)$ and $b_0 \in A(x_0)$. It follows that $z_0 = a_0 - b_0$ is in the tangent space of $A(x_0)$. This completes the moreover proof.

Now we assume that A is continuous (but not necessarily affine-subspace-valued) and has constant dimension. We show T_A is continuous at $x_0 \in \mathbb{R}^d$. Choose $d_2 + 1$ points in $A(x_0)$ that are in *general position*. We remark that $d_2 + 1$ points are said to be in general position in $\mathbb{R}^n$ if their convex hull has dimension d_2. Call these points z_0^k, for $k = 1, 2, \cdots, d_2 + 1$. See that $\{z_0^k - z_0^{d_2+1} : k = 1, 2, \cdots, d_2\}$ is a linearly independent set which comprises a basis for $T_A(x_0)$.

By Lemma 6.16, we can extend the z_0^k into continuous functions $z^k(x)$ such that $z^k(x_0) = z_0^k$ and $z^k(x) \in A(x)$ for x sufficiently close to x_0.

Since the functions $\{z^k\}$ are continuous, the points $\{z^k(x) : k = 1, 2, \cdots, d_2 + 1\}$ will remain in general position for x sufficiently close to x_0. Observe that $\{z_0^k - z_0^{d_2+1} : k = 1, 2, \cdots, d_2\}$ is a linearly independent set which must span a d_2-dimensional subspace of $T_A(x)$. But $T_A(x)$ is of fixed dimension, so $T_A(x)$ must simply be the span of $\{z_0^k - z_0^{d_2+1} : k = 1, 2, \cdots, d_2\}$.

Now that we see that there is a continuous set of basis vectors which span the constant dimensional space $T_A(x)$, the arguments for upper and lower semicontinuity are straightforward. In particular, we simply decompose everything into the basis. □

LEMMA 6.18. *Suppose $A, B : \mathbb{R}^m \to \mathcal{P}(\mathbb{R}^N)$ are continuous set-valued functions. Suppose that A and B have values that are affine subspaces of $\mathbb{R}^N$. Suppose $x_0 \in \mathbb{R}^d$. Assume that the dimension of $A(x) \cap B(x)$ is constant in a neighborhood of x_0. Then $A(x) \cap B(x)$ is continuous in a neighborhood of x_0.*

Proof. Upper semicontinuity follows because the intersection of upper semi-continuous functions is again upper-semicontinuous. This follows from the definition of upper-semicontinuous functions as multi-valued functions with closed graphs.

Now we show lower semicontinuity. Define $C(x) := A(x) \cap B(x)$. Observe that for x sufficiently close to x_0, $C(x)$ is an upper-semicontinuous set-valued functions with values that are d-dimensional affine subspaces of $\mathbb{R}^N$. We show $C(x)$ is lower semicontinuous.

It suffices to show lower semi-continuity at x_0 (since x_0 is not special compared to some neighborhood of itself with respect to the hypotheses of the lemma). Let $z_0 \in C(x_0)$. Let (x_n) be a sequence of points tending to $x_0 \in \mathbb{R}^m$. We show such there exists a sequence (z_n) converging to z_0 that satisfies $z_n \in C(x_n)$ for sufficiently high n.

Define (z_n) to be the unique closest point in $C(x_n)$ to z_0. We show that (z_n) converges to z_0. Suppose not. Define $v_n := z_n - z_0$. Since (z_n) does not converge to z_0, it follows that (v_n) admits a subsequence bounded away from 0. Therefore there is a subsequence (v_{n_k}) converging to some non-zero vector v_0. By upper semicontinuity of $C(x)$, the corresponding subsequence (z_{n_k}) converges to a point $z^* \in C(x_0)$. Since v_n is the displacement between z_0 and the closest point to z_0 on the affine subspace $C(x_n)$, it follows that $v_n \in T^{\perp}_{C(x_n)}$.

Claim. The set-valued function $T^{\perp}_{C(x)}$ is upper semicontinuous for x sufficiently close to x_0.

By Lemma 6.17, $T_{C(x)}$ is upper semicontinuous for x sufficiently close to x_0. We show that if there is a sequence (w_n) such that for each n, $w_n \in T^{\perp}_{C(x_n)}$, and also (w_n) converges to w_0, then $w_0 \in T^{\perp}_{C(x_0)}$.

For each n, choose an orthonormal basis $\{b^k_n : k = 1, 2, \cdots, d\}$ of $T_{C(x_n)}$. Restrict attention to a subsequence in which the basis converges to an orthonormal set $\{b^k_0 : k = 1, 2, \cdots, d\}$. By upper semicontinuity of $T_{C(x_n)}$, the basis vectors $\{b^k_0 : k =$

$1, 2, \cdots, d\}$ span a d-dimensional subset of $T_{C(x_0)}$. Since the dimension of $T_{C(x_0)}$ is d, $\{b_0^k : k = 1, 2, \cdots, d\}$ is an orthonormal basis for $T_{C(x_0)}$.

For each n, choose an orthonormal basis $\{c_n^k : k = 1, 2, \cdots, N-d\}$ of $T_{C(x_n)}^{\perp}$. Restrict attention to a subsequence in which the basis converges to an orthonormal set $\{c_0^k : k = 1, 2, \cdots, N-d\}$. Observe that for all $j \in \{1, 2, \cdots, d\}$ and $k \in \{1, 2, \cdots, N-d\}$,

$$\left\langle b_0^j, c_0^k \right\rangle = \lim_{n\to\infty} \left\langle b_n^j, c_n^k \right\rangle = \lim_{n\to\infty} 0 = 0.$$

It follows that $\{c_0^k : k = 1, 2, \cdots, N-d\}$ is an orthonormal basis for $T_{C(x_0)}^{\perp}$. We may write, for each n in the subsequence,

$$w_n = \sum_{k=1}^{N-d} a_n^k c_n^k,$$

where a_n^k are coefficients. Since the (w_n) are bounded, these coefficients are bounded, and again, we may choose a convergent subsequence (so that the (a_n^k) each converge to a_0^k) and we have

$$w_0 = \sum_{k=1}^{N-d} a_0^k c_0^k,$$

from which it follows that $w_0 \in T_{C(x_0)}^{\perp}$. *This completes the proof of the claim.*

Since $T_{C(x)}^{\perp}$ is upper semicontinuous for x sufficiently close to x_0, and $v_n \in T_{C(x_n)}^{\perp}$ for all n, it follows that $v_0 \in T_{C(x_0)}^{\perp}$. On the other hand, $v_0 = z^* - z_0$. But both z^* and z_0 are in $C(x_0)$, and hence $v_0 \in T_{C(x_0)}$. It follows that $v_0 = 0$. This is a contradiction; we conclude that (z_n) converges to z_0. $\square$

LEMMA 6.19. *Let $A : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^n)$ be a continuous set-valued function. Suppose that $P : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^n)$ is a continuous subspace-valued function. Then the set-valued function obtained by projecting $A(x)$ onto $P(x)$, given by*

$$F(x) := \Pi_{P(x)} A(x),$$

is a continuous set-valued function.

Proof. We show that if (z_n) converges to z_0 in $\mathbb{R}^n$, then $(\Pi_{P(x_n)} z_n)$ converges to $\Pi_{P(x_0)} z_0$. Both lower and upper semi-continuity follow rather straightforwardly after we have this fact.

Notice that we may consider $Q_n = \Pi_{P(x_n)}$ to be a square projection matrix for each n.

Claim. The sequence of matrices (Q_n) converges to $Q_0 := \Pi_{P(x_0)}$.

This can be established by considering an orthogonal basis V_n of $P(x_n)$ which converges to a basis V_0 of $P(x_0)$ as $n \to \infty$. Here we mean for V_n to have columns spanning $P(x_n)$ which are of unit length and are orthogonal. Observe that V_n may be extended into an orthogonal matrix U_n by completing the basis, and similarly V_0 to U_0, while maintaining the convergence: (U_n) tends to U_0. Then we have

$$Q_n = U_n D U_n^T, \text{ for } n = 0, 1, 2, \cdots .$$

It then follows quickly that (Q_n) converges to Q_0. *This completes the proof of the claim.*

Since (z_n) converges to z_0, and (Q_n) converges to Q_0, it follows that

$$\lim_{n\to\infty} Q_n z_n = \lim_{n\to\infty} Q_n \lim_{n\to\infty} z_n = Q_0 z_0.$$

We have shown the fact we wanted to prove; the lower and upper semicontinuity of the projected set now follows readily from the lower and upper semicontinuity of the original set using standard arguments. □

LEMMA 6.20. *Suppose $A, B : \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^n)$ are continuous set-valued functions with convex values. Suppose for some neighborhood of $x_0 \in \mathbb{R}^d$ we have the following: for each $x \in U$,*

1. *$A(x) \cap B(x)$ contains a point $p(x)$ in the relative interior of both A and B.*
2. *The dimension of both $T_A(x)$ and $T_B(x)$ remain fixed.*
3. *The vectors in $T_A(x)$ and $T_B(x)$ together span $\mathbb{R}^n$:*

$$\text{span } (T_A(x) \cup T_B(x)) = \mathbb{R}^n \text{ for all } x \in \mathbb{R}^d.$$

Then $A \cap B$ is continuous in a neighborhood of x_0.

Proof. First we show upper-semicontinuity. Since A and B are both upper semicontinuous, so is their intersection, as the intersection of two closed graphs $\{x \times A(x) : x \in \mathbb{R}^d\} \cap \{x \times B(x) : x \in \mathbb{R}^d\}$ is again closed.

Now we show lower-semicontinuity. To prove lower-semicontinuity it suffices to show lower-semicontinuity on a dense subset. We use this fact to prove $A(x) \cap B(x)$ is lower semicontinuous.

Claim: Define $D(x) \subset A(x) \cap B(x)$ to be the set of points which are simultaneously in the relative interior of both $A(x)$ and $B(x)$. Then $D(x)$ is dense in $A(x) \cap B(x)$.

By assumption, $D(x)$ is non-empty for all x in some neighborhood of x_0. We verify that the existence of just a single such point in $A(x) \cap B(x)$ implies that $D(x)$ is dense in $A(x) \cap B(x)$. Suppose $z^* \in A(x) \cap B(x)$ is not in $D(x)$. Let $z_0 \in D(x) \subset A(x) \cap B(x)$. Observe that $\alpha z_0 + (1 - \alpha) z^*$ is in the relative interior of both $A(x)$ and $B(x)$ for all $\alpha \in [0, 1)$. This implies that z^* is the limit of points in $D(x)$, and hence $D(x)$ is dense in $A(x) \cap B(x)$. *This completes the proof of the claim.*

Take $z_0 \in A(x_0) \cap B(x_0)$ to be a point that is in the relative interior of both $A(x_0)$ and $B(x_0)$. Assume (x_n) is a sequence tending to x_0. We show there exists a sequence (z_n) converging to z_0 such that for sufficiently high n, $z_n \in A(x_n) \cap B(x_n)$. This will prove that $A(x) \cap B(x)$ is lower semicontinuous. (Actually, this proves lower semicontinuity at x_0. But there is nothing special about x_0 in U, so this suffices.)

We will use the notation $X^\perp$ to indicate the orthogonal complement of a finite dimensional subspace X of $\mathbb{R}^n$. Using this notation, we note that $T_A^\perp \cap T_B^\perp = \emptyset$, since otherwise T_A and T_B would not together span $\mathbb{R}^n$. Because of this, we may decompose $\mathbb{R}^n$ into three subspace-valued functions $U(x), V(x)$, and $W(x)$ such that for all x in a neighborhood of x_0,

1.

$$\mathbb{R}^n = U(x) \oplus V(x) \oplus W(x).$$

2.

$$U(x) := T_A(x) \cap T_B^{\perp}(x).$$

3.

$$V(x) := T_A^{\perp}(x) \cap T_B(x).$$

4.

$$W(x) = T_A(x) \cap T_B(x).$$

It is straightforward to see this direct sum decomposition is possible.

Claim. The subspace-valued functions U, V, and W are continuous.

Because A and B are continuous, closed, convex valued functions with constant dimensionality, it follows by Lemma 6.17 that T_A and T_B are continuous subspace-valued functions. (Of course, we mean in a neighborhood of x_0.) Let a and b denote the dimension of T_A and T_B respectively. By the assumption that for all x sufficiently close to x_0 that $T_A(x)$ and $T_B(x)$ together span $\mathbb{R}^n$, we see that $a + b \geq n$. Since $a + b \geq n$, it follows that $T_A^{\perp} \subset T_B$ and $T_B^{\perp} \subset T_A$ or else $T_A + T_B = \mathbb{R}^n$ would be impossible. Consequently, $U = T_B^{\perp}$ and $V = T_A^{\perp}$ have dimension $n - b$ and $n - a$ respectively. Since $\mathbb{R}^n$ admits a direct sum decomposition into U, V, and W, it follows that the sum of the dimensions of U, V, and W is n. Accordingly, the dimension of W is $a + b - n$.

We have determined that U, V, and W are of constant dimension in a neighborhood of x_0. By Lemma 6.18, it follows that U, V, and W are continuous subspace-valued functions of fixed dimension for a neighborhood of x_0. *This completes the proof of the claim.*

Consider x close to x_0. Let $\ell \in U(x)$. Then $\ell \in T_B^{\perp}(x)$, and it follows that $\langle \ell, B(x) \rangle$ is a singleton set. To be clear,

$$\langle \ell, B(x) \rangle := \{ \langle \ell, z \rangle : z \in B(x) \}.$$

From this observation, it may be determined that there exists a unique vector $u \in U(x)$ such that $\langle \rho, B(x) \rangle = \langle \rho, u \rangle$ for all $\rho \in U(x)$. It also follows that the projection of $B(x)$ into the subspace $U(x)$ consists of a single point, $\{u\}$. See that u depends on x: we write $u(x)$. Notice that we must have

$$B(x) \subset u(x) \times V(x) \times W(x).$$

Claim. The function $u(x)$ is a continuous function of x.

Note that $\langle \ell, B(x) \rangle = \langle \ell, u(x) \rangle$ for all $\ell \in U(x)$ implies that the projection of $B(x)$ onto the subspace $U(x)$ results in the singleton set $\{u(x)\}$. This observation defines $u(x)$. To see that $u(x)$ is continuous, we show that the projection $\Pi_{U(x)} B(x)$ is a continuous set-valued function. This is accomplished by Lemma 6.19. Since, in addition, $\Pi_{U(x)} B(x)$ is always singleton-valued (it is $\{u(x)\}$), it follows that $u(x)$ is continuous. *This completes the proof of the claim.*

Similarly to how we defined $u(x)$, define $v(x)$ to be a continuous function defined in a neighborhood of x_0 such that for each x in that neighborhood of x_0, we have $\langle \ell, A(x) \rangle = \langle \ell, v(x) \rangle$ for all $\ell \in V(x)$. Accordingly, we have that

$$A(x) \subset U(x) \times v(x) \times W(x).$$

Claim. The set-valued function $\tilde{A}(x) := A(x) \cap (u(x) \times v(x) \times W(x))$ is continuous.

Upper semi-continuity is straightforward as we are considering the intersection of $A(x)$ with a continuous affine-subspace valued function. In particular both have closed graphs, and thus so does their intersection.

Now we show lower semi-continuity. Suppose (x_n) is a sequence tending to x_0. Suppose $z_0 \in \tilde{A}(x)$. We show there exists a sequence (z_n) tending to $z_0 = (u(x_0), v(x_0), w_0)$ such that $z_n \in \tilde{A}(x_n)$ for sufficiently high n. The projection of $A(x)$ onto the U and W spaces results in a continuous convex-valued function containing z_0 in its topological interior. By Lemma 6.15, there exists $\epsilon > 0$ such that for sufficiently high n, we have $(u, v(x_n), w) \in \tilde{A}(x_n)$ for all u, w such that $|u - u(x_0)| < \epsilon$ and $|w - w_0| < \epsilon$. Note that $|u(x_n) - u(x_0)| < \epsilon$ for n sufficiently high. Choose $z_n = (u(x_n), v(x_n), w_0)$ and observe that $z_n \in \tilde{A}(x)$ for sufficiently high n. *This completes the proof of the claim.*

Now we can project $\tilde{A}(x)$ into the $W(x)$ space, where it is a continuous convex set with topological interior for all x in a neighborhood of x_0. Similarly we can construct $\tilde{B}(x)$, which we may project into the W space as well. We refer to Lemma 6.19 to see that the projections of $\tilde{A}(x)$ and $\tilde{B}(x)$ remain continuous. We may simultaneously

rotate these projections into $\mathbb{R}^{a+b-n} \times \{0\}^{2n-(a+b)}$ while preserving continuity. This may be accomplished using an orthogonal matrix $M(x)$ which brings $W(x)$ into the canonical position: $M(x)W(x) = \mathbb{R}^{a+b-n} \times \{0\}^{2n-(a+b)}$. Since W is continuous, it is straightforward to show that M may be made continuous, at least in a neighborhood of x_0. In particular, one could use Lemma 6.16 on a set of points of $W(x)$ in general position in order to create a continuous basis $\beta(x)$ for $W(x)$. This basis $\beta(x)$ could then be extended into an orthogonal basis $U(x)$ for $\mathbb{R}^n$ in a neighborhood of x_0 without sacrificing continuity; observe that any ordinary scheme (Gram-Schmidt, for example) for extending the basis will have continuous output in a neighborhood of the basis at x_0. Then define $M(x) := U^T(x)$, which is continuous as it is the matrix transpose of a continuous orthogonal matrix. See that $M(x)$ has the desired properties, in particular since $U^T = U^{-1}$ is a matrix which brings $W(x)$ into canonical position.

Now we define the functions

$$A^*(x) := M(x)\Pi_{W(x)}\tilde{A}(x)$$

and

$$B^*(x) := M(x)\Pi_{W(x)}\tilde{B}(x)$$

as continuous set-valued functions with values in $\mathbb{R}^{a+b-n}$. To see that A^* and B^* are continuous is straightforward. Observe that A^* and B^* have topological interior; importantly, observe that $w_0^* := M(x_0)w_0$ is in the topological interior of both $A^*(x_0)$ and $B^*(x_0)$.

We now will see that $A^*(x) \cap B^*(x)$ is lower-semicontinuous at $(x_0; w_0^*)$ only if $A(x) \cap B(x)$ is lower semi-continuous (x_0, z_0). This is because if we find a sequence (w_n) converging to w_0 such that $w_n^* := M(x_n)w_n \in A^*(x_n) \cap B^*(x_n)$ for all n, there is also a sequence

$$z_n := (u(x_n), v(x_n), w_n),$$

which converges to $z_0 = (u(x_0), v(x_0), w_0)$ and satisfies $z_n \in A(x_n) \cap B(x_n)$ for all n.

Accordingly, we show $A^*(x) \cap B^*(x)$ is lower-semicontinuous at $(x_0; w_0^*)$.

Observe that w_0 is interior to $A^*(x_0)$. By Lemma 6.15, there exists $\rho_A > 0$ such that $w^* \in A^*(x)$ whenever $|x - x_0| < \rho_A$ and $|w^* - w_0^*| < \rho_A$.

Observe that w_0 is interior to $B^*(x_0)$. By Lemma 6.15, there exists $\rho_B > 0$ such that $w^* \in B^*(x)$ whenever $|x - x_0| < \rho_B$ and $|w^* - w_0^*| < \rho_B$.

Choose $\rho = \min\{\rho_A, \rho_B\}$. Combining the last two results, we have $w^* \in A^*(x) \cap B^*(x)$ whenever $|x - x_0| < \rho$ and $|w^* - w_0^*| < \rho$. In particular, we may choose

$$w_n^* := w_0^*.$$

By the last result we have that for sufficiently high n, $w_n^* \in A^*(x_n) \cap B^*(y_n)$. It follows that $A^*(x) \cap B^*(x)$ is lower semi-continuous at $(x_0; w_0^*)$. The lemma is shown. □

LEMMA 6.21. *The set-valued function $K(t, p, q)$ is continuous when restricted to the time-dependent persistent-contact submanifold*

$$\mathcal{P} := \{\{t\} \times \mathcal{S}_p^{\mathcal{A}}(t) : t \in \mathbb{R}\},$$

where $\mathcal{A}$ is some active set of contacts.

Proof. We show K is continuous. Observe that $K = \mathcal{F} \cap \mathcal{E}$. We verify the conditions of Lemma 6.20:

1. $\mathcal{F}$ is continuous on $\mathcal{S}^{\mathcal{A}}$. This follows since for $x \in \mathcal{S}^{\mathcal{A}}$,

$$\mathcal{F}(x) = \sum_{k \in \mathcal{A}} \mathcal{F}_k,$$

and hence, being a finite sum of continuous set-valued functions, $\mathcal{F}$ is again continuous.

2. $\mathcal{E}$ is continuous. The set $\mathcal{E}(t, x)$ for $(t, x) \in \mathbb{R} \times \mathcal{S}^{\mathcal{A}}$ is defined by the persistent contact equations:

$$E(x)z + d(t, x) = 0,$$

where E is a full-rank matrix and both E and d are continuous. The solutions to this equation $\mathcal{E}(t, x)$ are thus continuous by Lemma 6.4.

3. For every $(t, x) \in \mathcal{P}$, by the definition of the persistent contact submanifold, $K(t, x)$ admits a point that is in the relative interior of $\mathcal{F}(x)$. Every point in $K(t, x)$ is in the relative interior of $\mathcal{E}(t, x)$, since $\mathcal{E}(t, x)$ is a affine subspace. Hence for every $(t, x) \in \mathcal{P}$, there exists a point $p \in K(t, x)$ which is in the relative interior of both $\mathcal{F}(x)$ and $\mathcal{E}(t, x)$.

4. By Lemma 6.4, $\mathcal{E}(t, x)$ is of constant dimension. By a hypothesis of the model, $\mathcal{F}(x)$ is of fixed dimension for $x \in \mathcal{S}^{\mathcal{A}}$.

5. The tangent space of $\mathcal{E}(t,x)$ contains every direction which is perpendicular to the constraint gradients f_q^k, $k \in \mathcal{A}(x)$. The tangent space of $\mathcal{F}(x)$ contains the span of the constraint gradients f_q^k, $k \in \mathcal{A}(x)$. Therefore, the vectors in the tangent space of $\mathcal{E}(t,x)$ and the vectors in the tangent space $\mathcal{F}(x)$ span $\mathbb{R}^n$.

Thus, Lemma 6.20 applies and we conclude that K is continuous. □

PROPOSITION 6.22. *The concept of typicality in $\mathbb{R} \times \mathcal{S}^{\mathcal{A}}$ is an open one. That is, if $(t_0, x_0) \in \mathbb{R} \times \mathcal{S}^{\mathcal{A}}$ is typical, where $x_0 = (p_0, q_0)$, then there exists $\epsilon > 0$ such that $B_\epsilon((t_0, x_0)) \cap \left(\mathbb{R} \times \mathcal{S}^{\mathcal{A}}\right)$ consists only of typical states (t,x).*

Proof. Without loss, assume $0 \in \mathcal{F}(x)$ for all x in a neighborhood of x_0. This is without loss since otherwise, we could redefine both $\mathcal{F}$ and $\mathcal{E}$ (for our purposes in this proof) to be translated by a continuous function $f(x)$ ($\mathcal{F}'(x) = \mathcal{F}(x) - f(x)$ and $\mathcal{E}'(t,x) = \mathcal{E}(t,x) - f(x)$) in order to make $0 \in \mathcal{F}$. The function $f(x) \in \mathcal{F}(x)$ may be constructed via Lemma 6.16. Notice that this redefinition preserves the notion of typicality. This completes the without loss argument.

Observe that $\mathcal{F}$ is a continuous set-valued function when restricted to $\mathcal{S}^{\mathcal{A}}$. Moreover, it has convex values of constant dimension; call this dimension d. It is possible to construct a continuous orthogonal matrix $U(x)$ for all $x \in \mathbb{R}^n \times \mathcal{Q}$ such that for all x in a neighborhood of x_0,

$$\tilde{\mathcal{F}}(x) := U(x)\mathcal{F}(x)$$

is a continuous, convex-valued function whose values are subsets of $\Omega := \mathbb{R}^d \times \{0\}^{n-d}$. See Lemma 6.17 and its proof to see this.

Claim. The dimension of $\tilde{\mathcal{E}}(t,x)$ is fixed in a neighborhood of (t_0, x_0).

As seen in the proof of Lemma 6.21, the tangent spaces of $\mathcal{E}(t,x)$ and $\mathcal{F}(x)$ suffice to span $\mathbb{R}^n$. The tangent space of $\mathcal{E}(t,x)$ is of dimension $n-|\mathcal{A}|$, for all $x \in \mathcal{S}^{\mathcal{A}}$. Since $U(x) \circ \mathcal{E}(t,x)$ and $\tilde{\mathcal{F}}(x)$ have tangent spaces which together span the space, it must follow that the tangent space of $U(x) \circ \mathcal{E}(t,x)$ contains every direction in $\Omega^{\perp}$. It is now straightforward to see that $U(x) \circ \mathcal{E}(t,x) \cap \Omega$ has dimension $n-|\mathcal{A}|-(n-d) = d-|\mathcal{A}|$. *This completes the proof of the claim.*

By Lemma 6.18, $\tilde{\mathcal{E}}(t,x) := (U(x) \circ \mathcal{E}(t,x)) \cap \Omega$ is continuous.

Now consider $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{F}}$ to have values that are subsets of $\mathbb{R}^d$, rather than $\Omega := \mathbb{R}^d \times \{0\}^{n-d}$. The following fact is straightforward: a convex set *which contains the origin* has topological interior when considered in the subspace topology of its tangent space (which it is contained in). Since $\tilde{\mathcal{F}}$ contains the origin, it has topological interior. Indeed, the relative interior of $\mathcal{F}$ has been transformed one-to-one into the topological interior of $\tilde{\mathcal{F}}$ by our construction.

Since $\mathcal{E}$ intersects $\mathcal{F}$ in its relative interior, the intersection of $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{F}}$ contains a point in the topological interior of $\tilde{\mathcal{F}}$.

Let z_0 be a point of intersection of $\tilde{\mathcal{E}}(t_0, x_0)$ and $\tilde{\mathcal{F}}(x_0)$ that is in the topological interior of $\tilde{\mathcal{F}}(x_0)$. By Lemma 6.15, there exists $\epsilon > 0$ such that if $((t,x),z)$ is ϵ-close to $((t_0,x_0),z_0)$, then $z \in \tilde{\mathcal{F}}(t,x)$. Since $\tilde{\mathcal{E}}(t,x)$ is continuous and $z_0 \in \tilde{\mathcal{E}}(t_0,x_0)$, it

is straightforward to see that $\tilde{\mathcal{E}}(t,x)$ contains a point in $B_\epsilon(z_0)$ for (t,x) sufficiently close to (t_0,x_0). It follows from the previous two sentences that for (t,x) sufficiently close to (t_0,x_0), $\mathcal{E}(t,x)$ intersects $\mathcal{F}(t,x)$ in its interior. Accordingly, typicality is an open concept in $\mathcal{S}^{\mathcal{A}}$. □

Associated Feedback Problem

We define

$$v(p,q) := H_p(p,q) = A(q)p + b(q). \tag{6.14}$$

Also define

$$g(t,p,q) = H_q(t,p,q). \tag{6.15}$$

Note that $v = H_p = \dot{q}$. We chose the letter v for "velocity".

We now substitute (6.11) and (6.14) into (6.5-6.6) to obtain

$$\dot{q} = v(p,q) \tag{6.16}$$

$$\dot{p} \in \mathrm{SOL}(K(t,p,q), v(p,q)) + g(t,p,q) \tag{6.17}$$

Proposition 6.23. *The equation/inclusion (6.16 - 6.17) is a Feedback Problem satisfying Hypothesis H and the tangency condition (5.16) when constrained to the time-dependent persistent contact submanifold*

$$\mathcal{P} := \{\{t\} \times \mathcal{S}_p^{\mathcal{A}}(t) : t \in \mathbb{R}\}.$$

Proof. By Proposition 6.11, $\mathcal{S}_p^{\mathcal{A}}$ is a submanifold.

We show the tangency condition holds. To this end, we show that for a neighborhood of (t_0, p_0, q_0) on $\mathcal{P}$ that the following holds:

$$F(t,p,q) \subset T_{(p,q)}\mathcal{S}_p^{\mathcal{A}}(t),$$

where

$$F(t,p,q) := v(p,q) \times \left(\mathrm{SOL}(K(t,p,q), v(p,q)) + g(t,p,q)\right).$$

Since $F(t,p,q) \subset K(t,p,q)$, it suffices to show that for a neighborhood of (t_0, p_0, q_0) on $\mathcal{S}$ that

$$K(t,p,q) \subset T_{(p,q)}\mathcal{S}_p^{\mathcal{A}}(t).$$

To this end, we show that the elements of K are vectors which are tangent to both $\mathcal{S}_1$ and $\mathcal{S}_2$, where we have defined

$$\mathcal{S}_1 := \{(p,q) : f^k(q) = 0 \text{ for all } k \in \mathcal{A}\} \tag{6.18}$$

$$\mathcal{S}_2 := \{(p,q) : f_q^k \cdot v(p,q) = 0 \text{ for all } k \in \mathcal{A}\} \tag{6.19}$$

This suffices since $\mathcal{S}_p^{\mathcal{A}}(t)$ is an open submanifold of $\mathcal{S}^{\mathcal{A}}$ which in turn is an open submanifold of $\mathcal{S}_1 \cap \mathcal{S}_2$.

To see tangency to $\mathcal{S}_1$, we show that for all $(p,q) \in \mathcal{S}_p^{\mathcal{A}}(t)$,

$$\left.\frac{d}{ds}\right|_{s=0} f^k(q + v(p,q)s) = 0$$

Since f^k is differentiable, this is satisfied provided $f_q^k(q) \cdot v(p,q) = 0$ for all $k \in K(t,p,q)$. Since $(p,q) \in \mathcal{S}_p^{\mathcal{A}}(t)$, we also have $(p,q) \in \mathcal{S}_2$. tangency to $\mathcal{S}_1$ follows.

Now we show tangency to $\mathcal{S}_2$. It suffices to show

$$\frac{d}{ds}\bigg|_{s=0} \Big(f_q^k(q + v(p,q)s) \cdot v(p + \vec{k}s + g(t,p,q)s, q + v(p,q)s) \Big) = 0 \text{ for all } \vec{k} \in K(t,p,q).$$

Since f and v are differentiable, we may write this condition as

$$\big\langle v(p,q), f_{qq}^k(q)v(p,q) \big\rangle + f_q^k(q) \cdot \Big(v_p \cdot (\vec{k} + g(t,p,q)) + v_q \cdot v(p,q) \Big) \text{ for all } \vec{k} \in K(t,p,q).$$

Substitute (6.14) and (6.15) to rewrite the condition as

$$H_p^T f_{qq}^k H_p + f_q^k (H_{pq} H_p + H_{pp}(\vec{k} - H_q)) = 0 \text{ for all } \vec{k} \in K(t,p,q).$$

Since $K(t,p,q) \subset \mathcal{E}(t,p,q)$, this condition follows straightforwardly from the persistent contact equations (6.7), (6.8), and (6.9) which define $\mathcal{E}$. We have tangency to $\mathcal{S}_2$.

Now we show Hypothesis H is satisfied. By Lemma 6.21, $K(t,p,q)$ is non-empty and continuous on $\mathcal{P}$. By Lemma 6.6, K has closed, convex, and bounded values. Note next that v and g are continuous.

We have shown (6.16-6.17) is a Feedback problem satisfying Hypothesis H and the tangency condition when constrained to the submanifold $\mathcal{P}$. □

We of course need to know that the solutions to the Feedback Problem indeed give us solutions as we defined them earlier:

PROPOSITION 6.24. *Any solution $(p(t), q(t))$ to the Feedback Problem (6.16-6.17) with initial condition $(t_0; p_0, q_0)$ which stays in the persistent contact submanifold $\mathcal{P}$ is a solution to the friction problem with persistent contact as defined in Definition 6.2.*

Proof. Observe that Hamilton's equations are obeyed, and the obtained force $F(t) := \dot{p}(t) + H_q(p(t), q(t))$ is indeed a maximally dissipative choice within the allowed set $\mathcal{F} \cap \mathcal{E}$. Clearly, F is integrable. Since the solutions to Feedback Problems are AC, we have that p and q are AC. Since we assumed the contact remained in the time-dependent persistent-contact submanifold $\mathcal{P}$, the persistent contact conditions holds as well. Hence all of the conditions of Definition 6.2 are satisfied. □

Existence Of Solutions.

In this section we show that the Feedback Problem (6.16-6.17) modeling persistent frictional contact admits solutions for some short time provided that we choose an initial condition in the time-dependent persistent contact submanifold $\mathcal{P}$.

THEOREM 6.25. *Given an initial condition $(t_0, p_0, q_0) \in \mathcal{P}$, for some active constraint set $\mathcal{A}$, the Feedback Problem (6.16 - 6.17) with initial condition (t_0, p_0, q_0) admits a solution for some short span of time $[t_0, t_0 + \epsilon)$, $\epsilon > 0$.*

Proof. By Proposition 6.23 and Theorem 5.15, this result follows. □

This is the best sort of existence theorem we can hope for, since we know that many configurations are in a state of "imminent release". If we interpret $K(t, p, q)$ being empty as meaning "release is imminent", then we see that Theorem 6.25 yields existence of a persistent contact solution provided that the initial condition is bounded away from states where contact release is necessary. In the absence of a encompassing

theory which deals not only with persistent contact, but also with impacts and contact releases, this is the best we can obtain.

Unfortunately there are some non-generic cases one might be able to construct, where there really is a persistent contact solution, yet our existence result is unable to guarantee it. This will arise when a solution exists on the boundary of the persistent contact submanifold. An example of this is furnished by the one-dimensional problem with the single constraint $q \geq 0$, a Hamiltonian $H = \dot{q}^2$, and an initial condition $q(0) = \dot{q} = 0$. Here, there is a solution $q(t) = 0$, yet Theorem 6.25 does not guarantee it. See that in this example $K = \{0\}$ is non-empty, but does not contain any points in the relative interior of $\mathcal{F} = [0, \infty)$. We may informally describe these situations of persistent contact for which are not covered by Theorem 6.25 as those where a contact is "skimming the surface." Although we make no attempt to rigorously define and justify this characterization, it should be apparent that these situations are somewhat contrived (not generic). We save the complication of worrying about existence theory for these initial conditions for a future theory which allows contact releases.

If for an initial condition (t_0, p_0, q_0) we have that $K(t, p, q)$ is empty, then the existence result Theorem 6.25 does not apply. This does not concern us, since we do not expect a persistent contact solution when $K = \emptyset$. In this situation, we expect a contact break, which we have made no attempt to model. One might ask if this interpretation is valid. Does the emptyness of K imply there is a contact breaking solution? Although this question cannot really be answered without developing a

more complete theory beyond that of persistent contact, we point out that one could apply the theory of Chapter 3 and find the LCP solution to the force problem for frictionless unilateral constraints. This yields some unique answer; if there is a persistent contact solution for the frictionless problem, then K is nonempty. Therefore, when K is empty, then the closely related frictionless problem has a contact-breaking solution. Accordingly, it is not unreasonable to expect that a more complete theory allowing contact breaks will admit an existence theorem in the case where $K(t_0, p_0, q_0) = \emptyset$.

A word on uniqueness. We provide no uniqueness result, but we mention that it is certainly possible. We leave this as an open problem. If there is a uniqueness result, we also have well-posedness by a theorem of Filippov [12].

Physicality

Although we have managed to give an existence result for multiple contacts without contact breaking, one might critique the two assumptions leading to the model. Consider first the *global maximal dissipation* assumption. This is to be contrasted with the *local* version of the assumption, used by standard Coulomb friction.

Consider the following objection: "The local maximal dissipation assumption is more physical than the global maximal dissipation assumption." We can think of two objections to global maximal dissipation supporting this statement. First, global maximal dissipation appears to require that the contacts communicate to each other instantaneously, thus *communicating faster than the speed of light.* Second, it appears

possible that one could find a situation in which the direction of the frictional force at some contact was within ninety degrees of the sliding velocity of the contact. This would cause kinetic energy to be "absorbed" from the surface. Maximal dissipation can lead to this behavior if it happens that there are greater losses of energy through contacts elsewhere, which this paradoxical absorption somehow helped incur. This would violate *the second law of thermodynamics*, at least in the neighborhood of the questionable contact.

To respond to the first objection, regarding faster-than-light communication, we can simply shrug and say "If you don't like violations in special relativity, don't use rigid body approximations." For after all, even with the local maximal dissipation prescription, only the directions of the frictional force is determined. The relative magnitudes of the frictional forces are still communicated instantly among the contacts. That is, this objection applies equally well to the usual Coulomb model.

The second objection to the global maximal dissipation assumption is far more compelling. We give no example, but we indicate that such thermodynamically unsound solutions do exist. However, the severity of such problems can be made very small if one makes the friction sets $\mathcal{F}_k$ *velocity dependent*. Our assumptions demand that the only unbounded ray of $\mathcal{F}_k$ be in the direction ∇f_k, but other than this we are free to distort the shape of $\mathcal{F}_k$ so that it does not contain elements which have a large inner product with the sliding velocity of the corresponding contact. We do not describe such a construction any further, but only mention that such a technique

may prevent the solution obtained from violating the second law of thermodynamics to any great extent.

Our second assumption – the *tapered set assumption* – said that the tangential component of the friction compared to the normal component of friction became arbitrarily small in comparison for very large magnitudes of force. One could claim that this was not physical. However, for any desired bound of friction strength, we have arbitrary control over the shape of the frictional forces available within that bound (except that they must be convex). Yet examples are known [14] in modern Coulomb friction where the reactive force of a contact surface is integrable but not bounded.

Accordingly, we should want to allow a more interesting unbounded ray cone for the frictional sets. This is fair enough; but it opens the door for Painlevé's paradox, which requires tangential impulses to resolve. Tangential impulses cannot exist in the context of FP's, and our theory breaks down.

There is a possibility of developing an extension of feedback problems similar to the extension made of differential inclusions to measure differential incluions. We could then use "measure feedback problems" and attempt to tackle the Painlevé type paradoxes we encounter. This could also allow for a theory involving contact breaking and impacts. How to construct such a theory and give an existence result we leave an open question.

Summary

In this section we modified the Coulomb law of friction in two ways. The first way was to choose frictional forces to maximally dissipate energy in a global fashion, rather than locally determining the frictional force directions. The second way was to "taper" the friction sets and allow them to have unbounded rays only in the direction of the constraint normals.

After these modifications, we were able to cast the persistent contact frictional model with multiple contacts into the Feedback problem of the previous chapter, and apply the existence result. Our existence result only applies to initial conditions in the so-called persistent contact submanifold. We discussed the possibility of initial conditions elsewhere.

Finally, we discussed the physicality of the assumptions that went into the model. We indicated that if we extended our theoretical framework to a measure-theoretic version of FP's, it might be possible to make a theory with impacts, contact releases, and tangential shocks.

CHAPTER 7

CONCLUSION

Overview of Thesis

In this thesis we have surveyed many topics concerning unilateral constraints and frictional models.

Chapter 2 introduced unilateral constraints into the calculus of variations. By generalizing the principle of stationary action to a new principle reflecting optimality conditions in a constrained environment, we were able to derive the measure-theoretic Euler-Lagrange equations. We discussed related work and future directions for this idea.

In Chapter 3 we considered a specific frictionless impact model with the inelastic constitutive law. We found that it could be cast into a so-called *complementarity formulation*, for both force and impact situations, and that the complementarity problem obtained was well-behaved under perturbations of problem data. Despite this, we saw from an example of Ballard that uniqueness is not guaranteed, even with C^∞ data. The culprit appears to be right accumulations of impacts, and real analytic data eliminates the non-uniqueness.

Chapter 4 reviewed the modern theory of Coulomb friction and associated work.

In Chapter 5 we introduced the so-called Feedback Problem, which we proved to be a special case of a differential inclusion satisfying the *basic conditions* given by Filippov. Because of this, we were able to give an existence result for it. We also considered the questions of uniqueness and well-posedness. To this end we proved uniqueness for a special case and gave a non-uniqueness example for another well-behaved case.

Chapter 6 unveiled a novel model for frictional contact. Although we only developed it for the case of persistent contact, it is the only model of frictional contact we are aware of that has an existence result for both multiple contacts and large frictional coefficients. There is also hope that it is well-posed, although this is largely because we disallow impacting behavior. However, the model is fundamentally different from Coulomb models.

Open Questions

History shows that the formulation of frictional contact problems is a delicate matter, as Coulomb friction is an inconsistent theory without the usage of tangential impacts. When one goes to multiple contacts, existence results have yet to be shown. Nevertheless, we have given an alternative theory which yields existence for the persistent contact case without appealing to the usage of tangential impacts.

The following questions are raised and left open:

1. If one adopts the "tapered friction set assumption" and uses the local maximal dissipation rule, then does one obtain a frictional contact theory which is immune to the Painlevé paradox and more closely resembles Coulomb friction?

2. What other applications are there for the so-called Feedback Problem?

3. What is required to obtain uniqueness in the Negative Feedback Problem?

4. Can one combine the notions of measure differential inclusions with that of the Feedback Problem by interpreting unbounded ray solutions of convex programs to correspond to allowed discontinuities in the solution? If so, can then one relax the tapered friction set assumption in the above work? Also, do uniqueness results for special classes of Negative Feedback Problems carry over into the measure-theoretic version? If not, is it because of right accumulations of impacts? If so, and if we assume real-analytic data, could we get uniqueness for a class of measure feedback problems?

REFERENCES CITED

[1] M. Anitescu and D. Hart, *A fixed-point iteration approach for multibody dynamics with contact and small friction*, Math. Program, Ser. B, (2004).

[2] M. Anitescu and G. D. Hart, *Solving nonconvex problems of multibody dynamics with joints, contact and small friction by successive convex relaxation.*

[3] M. Anitescu and F. A. Potra, *Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems*, ASME J. Nonlinear Dynamics, 14 (1997), pp. 231–247.

[4] M. Anitescu, F. A. Potra, and D. E. Stewart, *Time-stepping for three-dimensional rigid body dynamics*, Comput. Methods Appl. Mech. Engrg., 177 (1999), pp. 183–197.

[5] J. P. Aubin and A. Cellina, *Differential Inclusions*, Springer Verlag, Berlin, 1994.

[6] P. Ballard, *The dynamics of discrete mechanical systems with perfect unilateral constraints*, Archives for Rational Mechanics and Analysis, 154 (2000), pp. 199–274.

[7] D. Baraff, *Fast contact force computation for non-penetrating rigid bodies*, Computer Graphics (Proc. SIGGRAPH), 28 (1994), pp. 23–34.

[8] R. Bartle, *The Elements of Integration and Lebesgue Measure*, John Wiley and Sons, New York, 1966.

[9] B. Brogliato, *Nonsmooth Impact Mechanics: Models, Dynamics, and Control*, vol. 220 of Lecture Notes in Control and Inform. Sci., Springer-Verlag, New York, 1996.

[10] A. Dontchev and F. Lempio, *Difference methods for differential inclusions: A survey*, SIAM Review, 34 (1992), pp. 263–294.

[11] R. C. Fetecau, J. E. Marsden, M. Ortiz, and M. West, *Nonsmooth Lagrangian mechanics and variational collision integrators*, SIAM J. Applied Dynamical Systems, 2 (2003), pp. 381–416.

[12] A. F. Filippov, *Differential equations with discontinuous right-hand side*, AMS Trans., 42 (1964), pp. 199–231.

[13] I. M. Gelfand and S. B. Fomin, *Calculus of Variations*, Dover, New York, 1963.

[14] F. Génot and B. Brogliato, *New results on Painlevé paradoxes*, Eur. J. Mech. A/Solids, 18 (1999), pp. 653–677.

[15] K. Jänich, *Vector Analysis*, Springer, New York, 2001.

[16] A. E. Kastner-Maresch, *Implicit Runge-Kutta methods for differential inclusions*, Numer. Funct. Anal. Optim., 11 (1990), pp. 937–958.

[17] D. Kincaid and W. Cheney, *Numerical Analysis: Mathematics of Scientific Computing*, Brooks/Cole, Pacific Grove, CA, 2001.

[18] M. Kunze and M. D. P. M. Marqués, *An introduction to Moreau's sweeping process*, Lecture Notes in Physics, 551 (2000), pp. 1–60.

[19] P. D. Lax, *Functional Analysis*, John WIley and Sons, New York, 2002.

[20] C. E. Lemke and J. T. Howson, *Equilibrium points of bimatrix games*, SIAM Journal of Applied Mathematics, 12 (1964), pp. 413–423.

[21] P. Lötstedt, *Mechanical systems of rigid bodies subject to unilateral constraints*, SIAM Journal of Applied Mathematics, 42(2) (1982), pp. 281–296.

[22] M. Mabrouk, *A unified variational model for the dynamics of perfect unilateral constraints.*, Eur J Mech A/Solids, 17 (1998), pp. 819–842.

[23] M. D. P. Monteiro Marques, *Chocs inélastiques standards: Un reśultat d'existence*, Séminaire d'Analyse Convexe, Montpellier 15 (1985), p. Exposé 4.

[24] J. Moreau, *Liaisons unilatérales sans frottement et chocs inélastiques*, C R Acad Sci Paris, Sér II, 296 (1983), pp. 1473–1476.

[25] ——, *Unilateral contact and dry friction in finite freedom dynamics*, In Nonsmooth Mechanics and Applications (J.J. Moreau and P.D. Panagiotopoulos, editors), International Center for Mechanical Sciences, Springer-Verlag, 302 (1988), pp. 1–82.

[26] J. Munkres, *Analysis on Manifolds*, Westview Press, Boulder, Colorado, 1991.

[27] K. G. Murty, *Linear and Nonlinear Programming*, Heldermann, Available Online, 1988.

[28] J. Nocedal and S. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.

[29] P. Painléve, *Sur le lois du frottement de glissemment*, C. R. Acad. Sci. Paris, 121 (1895), pp. 112–115.

[30] J. Pang and J. Trinkle, *Complementarity formulations and existence of solutions of dynamic multi-rigid -body contact problems with Coulomb friction*, Math. Programming., 73 (1996), pp. 199–226.

[31] D. Percivale, *Uniqueness in the elastic bounce problem I.*, J Differential Equations, 56 (1985), pp. 206–215.

[32] ——, *Uniqueness in the elastic bounce problem II.*, J Differential Equations, 90 (1991), pp. 304–315.

[33] F. Pfeiffer and C. Glocker, *Multibody Dynamics with Unilateral Contacts*, John Wiley and Sons, New York, 1996.

[34] M. Schatzman, *A class of nonlinear differential equations of second order in time*, Nonlinear Anal, 2 (1978), pp. 355–373.

[35] ——, *Uniqueness and continuous dependence on data for one dimensional impact problems*, Math and Computational Modeling, 28 (1998), pp. 1–18.

[36] D. E. Stewart, *A high accuracy method for solving ode's with discontinuous right-hand side*, Numer. Math., 58 (1990), pp. 299–328.

[37] D. E. Stewart, *Convergence of a time-stepping scheme for rigid-body dynamics and resolution of Painlevé's problem*, Archives for Rational Mechanics and Analysis, 145 (1998), pp. 215–260.

[38] D. E. Stewart, *Rigid-body dynamics with friction and impact*, SIAM Review, 42 (2000), pp. 3–39.

[39] K. Taubert, *Converging multistep methods for initial value problems involving multivalued maps*, Computing, 27 (1981), pp. 123–136.

[40] J. Trinkle, J.-S. Pang, S. Sudarsky, and G. Lo, *On dynamic multi-rigid-body contact problems with Coulomb friction*, Z. Agnew. Math. Mech., 77 (1997), pp. 267–279.