COUPLED PRINCIPLES FOR COMPUTATIONAL FRICTIONAL CONTACT MECHANICS

BY DANNY M. KAUFMAN

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Computer Science

Written under the direction of

Dinesh K. Pai

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2009

ABSTRACT OF THE DISSERTATION

Coupled Principles for Computational Frictional Contact Mechanics

by Danny M. Kaufman

Dissertation Director: Dinesh K. Pai

Abstract:

Methods for simulating frictional contact response are in high demand in robotics, graphics, biomechanics, structural engineering, and many other fields where the accurate modeling of interactions between solids are required. While techniques for accurately simulating structures and continua have advanced rapidly, methods for simulating contact between solids have lagged behind. This thesis considers the difficulties encountered in designing robust, accurate, and efficient computational methods for simulating frictional contact dynamics. We focus on understanding the fundamental sources of difficulty in frictional contact modeling, elucidating existing structures that can be leveraged to minimize them, and designing robust, accurate and efficient algorithms to simulate challenging frictional contact problems.

In this thesis a Coupled Principles formulation of discrete, time-continuous frictional contact is developed. This is then applied as the basis for deriving novel, time-discrete, variational integrators that pose the discrete frictional contact problem as a system of coupled minimizations. Solutions to these resulting systems are given by points that are simultaneously optimal for both minimizations and avoid some known issues present in existing variational integration approaches for frictional contact. We then consider a specific two-step variant of these variational schemes that generalizes the popular Stewart-Trinkle model for frictional contact

simulation. This is taken as a starting point for investigating encountered sources of difficulties found in solving numerical problems posed by these models. We show that many existing algorithms, that have generally been presumed suitable for solving the resulting contact-related numerical optimization problems, fail entirely for many important examples of frictional contact, and then address these limitations with our Staggered Projections algorithm. Applying a fixed-point scheme, derived from the Coupled Principles Formulation, we show that Staggered Projections efficiently obtains accurate solutions to the optimizations problems for many frictional contact problems that were previously impractical to solve. Finally, we also offer convergence analysis of the Staggered Projections algorithm, as well as simulations and instrumented examples that capture frictional contact behaviors for both rigid and large deformation models

Acknowledgements

I would like to thank my advisor Dinesh Pai for guidance and support during this research. You suggested this beautiful research topic right at the very beginning. At the time I had little idea where this project would take me, but your insight and suggestions have always been there to help me along.

I would also like to thank my other co-authors, Doug James, Shinjiro Sueda, and Timothy Edmunds. Collaboration is one of the great pleasures of research and I feel very fortunate to have worked with such keen, talented, and critical partners.

To my committee, Vijay Kumar, Gerard Richter, and Bill Steiger: many thanks for your help, questions, and feedback.

Many other people have also contributed to this work by providing valuable suggestions, technical help, and thought provoking discussions along the way (as well as much needed distractions in some cases); thanks to Uri Ascher, Jernej Barbič, Robert Bridson, Warren Code, Mclean Edwards, Benjamin Gilles, Paul Kry, Cheryl Lau, Mahkameh Lakzedah, David Levin, Philip Loewen, John Lloyd, Jos Stam, Bill Steiger, Qi Wei, and Sang Hoon Yeo.

To my parents, Roberta and Joseph Kaufman, and my brother, Lev Kaufman: you taught me what intellectual curiosity and creative work are all about. Thank you for your love and support all along the way.

Thanks to my amazing daughter Iris. These first two and half years are unlike anything I could have imagined. You've taught me the true meaning of an all-nighter, how to find balance in a hectic year, and a whole new concept of what's funny. I can't wait to see what comes next.

Most of all, I want to thank Barbara. Thank you for everything! Your patience, love, support, and wisdom brought me through this year.

Dedication

For Barbara and Iris.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis addresses the difficulties encountered in designing robust, accurate, and efficient computational methods for simulating frictional contact dynamics. Such methods are in high demand in robotics, graphics, biomechanics, structural engineering and many other fields where the accurate modeling of interactions between solids are required. To date, while methods for simulating structures and continua have advanced rapidly, methods for simulating the contact between solids, which is often cited as one of the prime motivations for continuum modeling [Belytschko, Liu, & Moran 2000], have lagged behind [Kikuchi & Oden 1988; Stewart 2000; Wasfy & Noor 2003]. In this thesis we focus on understanding the fundamental sources of difficulty in frictional contact modeling, elucidating existing structures that can be leveraged to minimize them, and designing robust, accurate and efficient algorithms to simulate challenging frictional contact problems.

1.1 Discrete Frictional Contact

Despite its ubiquity, frictional contact remains one of the most challenging physical phenomena to model and simulate. At large scales, unilateral contact and friction forces can appear to effectively mediate the majority of interactions between solids; thus, understanding frictional contact behaviors is essential in many problem domains. Modern frictional contact modeling is, however, at its essence, a multi-scale problem. Combining macro-scale friction laws for tribological effects and one-sided nonpenetration conditions for boundary interactions that generally occur at much finer time scales, modern methods intrinsically model phenomena occurring at multiple temporal and spatial scales. Inherently, we can expect that any resulting numerical problem will be in some sense fundamentally ill-posed. Nevertheless, the modeling of interfacial behavior at these scales is a necessary complement to any discrete model of continuum

mechanics.

This thesis starts with a discrete (finite-dimensional) spatial model as a starting point. Once we accept a finite-dimensional model as the basis for modeling frictional contact dynamics, a rich range of mathematical and physical modeling problems are obtained. Despite the potential ill-posedness of the resulting problems we will show that complex frictional-contact behaviors can be accurately captured, and consistent and rigorous models can be developed.

There are a range of complementary perspectives that can be adopted for understanding the frictional contact problem:

Tribology: Tribology [Dawson 1998] studies the underlying mechanisms of frictional contact phenomena. Potentially, this can take us down to the quantum level and is generally, with the exception of our adoption of macroscale friction laws, beyond the scope of this thesis.

Finite-Dimensional Models: Spatially finite-dimensional, time-continuous mathematical models are the starting point of this thesis. The adoption of finite-dimensional contact models was pioneered by Lötstedt [1982a] and is, despite outstanding questions of well posedness, existence, and uniqueness, that effectively date back as far as Painlevé [1895], a powerful simplifying assumption. We will discuss these background issues and relevant work further in the next Section.

Computational Models: Despite outstanding theoretical and practical issues, the numerical modeling of frictional contact is a crucial way (and in some cases, the only way) in which we can investigate frictional contact phenomena. Here the question is how can we accurately reproduce observed behavior using computational models? Issues of understanding and reducing accuracy, error, and computational complexity are all important and will be addressed in this thesis.

Time-Continuous Convergence and Solution Existence: Despite modeling successes, an open question in many cases is the relationship between time-discrete models and time-continuous formulations. In many cases the convergence of a numerical scheme to its time-continuous system is still an open problem. On the other hand, existence proofs have been settled for many time-discrete models; however, this still remains another area

of ongoing research. These issues are all related to the multi-scale nature of frictional contact modeling and, as we will see in later chapters, these issues also have implications for the computational complexity of resolving frictional contact in the the fully discretized setting. Time-continuous convergence and existence, however, are also outside the scope of this thesis.

1.2 Overview and Motivation

Numerical methods for the solution of frictional contact problems are in demand in a wide variety of application areas that include structural engineering, interactive medical simulation, computer animation, biomechanics, granular flow, robotics, and real-time video games.

Contacting systems present unique challenges. Pioneering research in applied mathematics (Signorini, Moreau, Panagiotopoulos, Lötstedt, Stewart, Anitescu), mechanics (Ruina, Goyal, Papadopoulos), robotics (Erdmann, Mason), and graphics (Baraff) have helped us to understand many aspects of these challenges; however, implementing robust, accurate and efficient methods for simulating frictional contact has, in general, been very difficult. Understanding the root causes for these difficulties, and developing methods to bypass them, have remained outstanding research problems.

Addressing and understanding these difficulties is the focus of this thesis, which presents a general approach for formulating numerical methods to accurately and effectively solve discrete, frictionally contacting systems. While earlier approaches were often either restricted to frictionless contact or to particular (and/or simplified) contacting systems, the new methods presented here are broadly applicable to many complex multibody configurations with finite degrees of freedom.

Building on seminal results by Moreau, we describe the dynamics of frictional contact resolution using a pair of coupled variational principles. Leveraging variational integration methods (Veselov, Marsden), we can then pose the discrete problem as a system of coupled minimizations that capture convincing and accurate frictional contact behavior. Solutions to these systems are given by points that are optimal for both of the minimizations. Although more general, this problem is similar in nature to finding the solution of a system of equations.

Formulations based on Linear Complementarity Programming (LCP) are standard for frictional contact problems involving linear systems. Initially introduced for contact simulation by Lötstedt, newer variants, such as the popular Stewart-Trinkle formulation, now guarantee solution existence; however, solving the resulting optimization problems accurately and efficiently has remained difficult. We will show that these difficulties can be explained, in part, by noting that optimization problems, obtained from frictional contact, are equivalent to nonconvex problems in global optimization (which are generally NP-hard). We believe that this is the first time that a complexity argument has been made via the connection to global minimization for the Stewart-Trinkle model.

Consistent with these observations we will show that, in practice, standard optimization methods such as Lemke's algorithm, Projected Gauss-Seidel, and interior point methods, that have generally been presumed suitable for solving these contact-related optimization problems, fail entirely for many important examples of frictional contact. We will address these limitations with our Staggered Projections algorithm. Applying a fixed point scheme, derived from the coupled principles formulation, Staggered Projections efficiently obtains accurate solutions to optimization problems for many frictional contact problems that were previously impractical to solve. We will show that Staggered Projections performs well in practice as evidenced by both our simulations of complex multibody frictional contact of rigid and large deformation models and our extensive testing of a Staggered Projections implementation with other available solvers.

1.2.1 Outline of Thesis

We begin in Section 1.3 with a brief discussion of the history and literature pertaining to discrete frictional contact modeling and simulation. This is followed by a summary of contributions in Section 1.4. We formulate and define the basic elements of the discrete frictional contact problem in a generalized setting in Chapter 2 and motivate the difficulties of frictional contact simulation with a discussion of Painlevé's Paradox. In Chapter 3 we develop the Coupled Principles formulation of discrete, time-continuous frictional contact. This is then applied, in Chapter 4, as the basis for deriving time-discrete variational integrators for frictional contact. Next, Chapter 5 considers a specific two-step variant of these variational schemes and the Stewart-Trinkle

model for frictional contact simulation. Here we also discuss sources of difficulties in solving these methods, and propose the Staggered Projections algorithm to address them. In Chapter 6 we offer a detailed convergence analysis of the Staggered Projections algorithm; while in Chapter 7 we investigate applications. Finally, we conclude in Chapter 8 with a summary and discussion of future work.

1.3 History and Literature

1.3.1 Friction Models

The observed relationship between frictional resistance and normal loads can be traced back at least as far as Leonardo da Vinci [Dawson 1998]. Leonardo further proposed the idea that a value, i.e., a coefficient of friction, could describe the ratio between the two. It was not until two-hundred years later that this observation was revisited by Amontons [1699]. Leonardo and Amontons, however, both concluded that all materials they investigated effectively had the same constant coefficient of friction [Dawson 1998].

Amontons also made the somewhat perplexing observation, at the time, that resistance due to friction did not, apparently, vary with the size of the contact area. He proposed that this potentially paradoxical behavior could be ascribed to surface roughness at finer scales and examined, in detail, how this might occur in the cases of both rigid *and* elastic asperities.

Other noted researchers, including Leibnitz and Euler[1], also looked into the friction problem during the Renaissance. Over these years and into the Industrial Revolution many new developments followed – some more quickly than others; coefficients of friction were found to vary much more widely than initially proposed, the asperity model was further popularized with a focus on the relation between friction and the shearing of asperities, and, perhaps most importantly, variations of Leonardo and Amontons's friction laws were extended to the dynamic case. Coulomb's role [Coulomb 1785] in this last development is most generally noted. See Dawson [1998] for a comprehensive history of these advances.

It is interesting to note that these early results in friction laws were generally developed independently from consideration of the contact problem. That is, in most of the investigations

[1]Euler is believed to have introduced the use of the symbol μ for the coefficient of friction [Dawson 1998].

of friction described above, tangential motion (or possibly no motion at all) was presumed a priori and little or no consideration was given to how a body, subject to contacts, might be described.

1.3.2 Contact Mechanics

In a parallel fashion initial investigations into contact mechanics also generally ignored frictional phenomenon. Early investigations of rigid-body contact date back at least as far as Galileo [1638]. Newton [1868] introduced the idea of restitution laws and a coefficient of restitution by which to describe it .

The modeling of contact for continuum mechanics perhaps begins with Poisson and Saint-Venant. The first successful approach to the elastic contact problem appears to have been formulated by Hertz [1882], whose elastic half-space model is still applied in various forms today [Johnson 1985].

The modern formulation of the frictionless contact problem, however, most likely begins with Signorini [1933]. Signorini developed a nonlinear, complementarity-based, boundary condition formulation for the static problem of a linear elastic body in contact with a rigid base. Fichera [1964] showed that an equivalent variational inequality formulation of Signorini's problem was possible and used convex minimization results to analyze uniqueness and existence questions for frictionless contact problems. Around the same time Moreau [1963, 1966] analyzed the more general case of a system subject to arbitrary one-sided constraints.

1.3.3 Nonsmooth Formulations

Inspired by these and other modeling problems, Moreau [1968] soon after introduced the concept of superpotentials and elucidated their connection to variational inequalities. When combined with subdifferentials (see Section 2.5.3), superpotentials allow for beautiful and almost transparent generalizations of classical mechanics to many nonsmooth settings beyond unilateral contact constraints, including plasticity and friction [Moreau 1973]. In a complementary

manner, friction can also be viewed as an the application of the Maximal Dissipation Principle [Moreau 1973; Goyal, Ruina, & Papadopoulos 1991]. First introduced as a tool for plasticity, the Maximal Dissipation Principle effectively provides a variational interpretation of the more familiar Coulomb friction model and will be applied throughout this thesis as a useful framework for understanding and modeling frictional behavior.

1.3.4 Frictional Contact Dynamics

The rigorous consideration of the interplay between one-sided non-penetration constraints and Coulomb-type friction forces perhaps begins in earnest with Painlevé [1895]. From the beginning, Painlevé's observation [Painlevé 1895] that some initial value problems involving rigid bodies and Coulomb friction had no solution [2], incited controversy [Klein 1910; Mises 1910, Prandtl 1910]. This potential inconsistency in frictional contact dynamics, dubbed "Painlevé's Paradox", has been the subject of a long and, in some places, ongoing debate. Early observations by Delassus [1905], Lecornu [1905], Beghin [1923], and others, however, suggested that these cases perhaps corresponded to what were, effectively, observed jumps in velocity. The modern resolution of the Painlevé Paradox hinges on formalizing this notion of a velocity jump through the careful application of impulsive (measure-based) forces, differential inclusions, and superpotentials [Moreau 1988; Stewart 2000].

1.3.5 Computational Methods

As discussed above, the accurate numerical modeling of contact for continuum mechanics has long been an active area of research. Computational contact mechanics has been broadly covered, with varying degrees of formalism [Johnson 1985; Kikuchi & Oden 1988; Brogliato 1999; Wriggers 2002]. Yet resolving multiple contacts with friction for both rigid and deformable bodies is still challenging [Stewart 2000]. In particular, flexible multibody systems have received increased attention in the last decade [Wasfy & Noor 2003], but fast and reliable algorithms for impact and friction remain open problems. Also of importance, existing algorithms tend to specialize to either rigid *or* flexible multibody systems, but rarely generalize to all cases.

[2] Painlevé also noted the, possibly less controversial, phenomena that in other case where solutions did exist they were possibly non-unique.

Restricting the frictional contact problem to rigid body models opened the development of acceleration-level LCP solutions [Lötstedt 1981, 1982a,b; Baraff 1991, 1994; Trinkle *et al.* 1995]. Baraff [1993] noted that Painlevé's Paradox implies that these acceleration formulations were not guaranteed to always have a solution and used these inconsistencies to show that the acceleration LCP is NP-hard. This observation led to more recent velocity-level LCP methods [Stewart & Trinkle 1996; Anitescu & Potra 1997]. While velocity-level LCPs can be shown to always have a solution [Anitescu & Potra 1997], Anitescu & Hart [2004] demonstrate examples where the solution set is non-convex and, based on this observation, suggest that velocity-level LCPs may also be difficult, in general, to solve. LCP approaches have also been extended to many elastic formulations, including quasi-rigid [Song & Kumar 2003; Pauly, Pai, & Guibas 2004], linear modal [Stewart 2001; Hauser, Shen, & O'Brien 2003] and Finite Element models (FEM) [Klarbring 1986; Baraff & Witkin 1992; Jourdan, Alart, & Jean 1998; Jean 1998; Duriez *et al.* 2006; Otaduy *et al.* 2007]. As with the rigid case, these formulations can also potentially suffer from inconsistency and non-uniqueness [Klarbring 1990, 1993], which again implies the likelihood of computational difficulties. Alternate formulations [Kikuchi & Oden 1988] with tight bounds on allowable coefficients of friction or even nonlocal friction models have also been proposed to side-step some of these issues.

Kane *et al.* [1999] have proposed a nonsmooth extension of variational integration methods [Marsden & West 2001] for frictionless contact that offers an alternate approach to resolving contact constraints in the discrete setting. While the specific methods proposed in Kane *et al.* [1999] have some known issues [Stewart 2000], this general approach is exciting in that it offers a natural bridge between the implied optimization interpretation of Discrete Variational Mechanics [Marsden & West 2001] and the modern tools of Mathematical Programming necessary for resolving contact. A natural and desirable extension would be to rigorously incorporate friction into this formulation; however, recent approaches have had limited success due to restricting assumptions [Pandolfi *et al.* 2002; Pandolfi & Ortiz 2007] and/or ad-hoc simplifications [Cirak & West 2005]. One bar to applying discrete variational techniques has been understanding what an appropriate Langrangian function for frictional contact should look like. Instead, friction has generally been incorporated as a Rayleigh-type damping force – with limited success. We will discuss these background issues and the potential difficulties with existing

variational methods for both friction and contact further in Chapter 4. In the same chapter we will also propose a new discrete variational integration scheme that resolves many of these difficulties and then, later in Chapter 5, clarify the relationship between the resulting variational methods and existing complementarity approaches.

In general, proving the convergence of numerical integration schemes to their corresponding time-continuous systems has been plagued with difficulties. A variety of methods in recent years, however, have been shown to converge [Monteiro-Marques 1993; Stewart 1999; Pang *et al.* 2006; Gavrea *et al.* 2008]. In general, Painlevé Paradox-type conditions often present the technical crux for these problems. A complementary question is that of time-continuous solution existence and uniqueness. Here too, convergence results can also be applied fruitfully. While many outstanding questions of existence in the time-continuous setting still remain, general solution existence for a broad class of problems, in the fully discrete setting, has recently been obtained [Pang & Stewart 1999], placing discrete frictional contact problems on a firmer footing.

Despite these outstanding theoretical questions, LCP integration methods and related formulations can now be considered as rigorously posed mathematical problems. Then, given an appropriately formulated LCP method with guaranteed solution existence, the technical question becomes how (and whether) we can obtain accurate numerical solutions. Generally, it has been widely assumed that, when fully accurate and robust solutions were desired, a direct, pivoting-based LCP algorithm, such as Lemke's [Cottle, Pang, & Stone 1992], could be applied, provided enough computation time was available. It has been recently noted, however, that direct LCP solvers do not, in practice, scale. They are not, in fact, currently able to return solutions for non-trivial frictionally contacting problems beyond relatively small-scale examples [Anitescu & Hart 2004; Erleben 2007]. In Chapter 5 we will discuss these difficulties further and investigate the reasons why these methods often fail.

Alternate, sometimes ad-hoc, approaches for frictional contact resolution have long been standard in many fields as well. Among the most popular alternatives, penalty-based methods, are well known in mechanics [Kikuchi & Oden 1988], and first introduced contact resolution problems to the graphics community [Hahn 1988; Bridson, Fedkiw, & Anderson 2002; Hauser, Shen, & O'Brien 2003], but suffer from stiffness and stability issues despite recent

advances [Spillmann, Becker, & Teschner 2007]. Optimization-based approaches in graphics [Baraff 1989, 1991, 1994; Redon, Kheddar, & Coquillart 2002; Raghupathi & Faure 2006], which are generally equivalent to LCP formulations, introduced improved accuracy in exchange for greater computational cost and motivated several extensions using local [Milenkovic & Schmidl 2001] and nonlocal [Kaufman, Edmunds, & Pai 2005] approximations of friction. In some cases the frictional contact problem has been simplified by reducing the contact inequality constraints to equalities [Irving, Schroeder, & Fedkiw 2007; Harmon *et al.* 2008]; in others it is mollified by treating individual constraints sequentially rather than simultaneously [Moore & Wilhelms 1988; Mirtich & Canny 1995; Cirak & West 2005].

In general, however, with the exception of LCP-based approaches, which we will discuss further below, existing methods for frictional contact resolution generally suffer from instabilities, often exhibited by unnatural vibrations at contacts, constraint drift (i.e., elements sinking into one another over time), popping artifacts (created by stabilization methods applied to combat drift), energy gain (although contact resolution should be dissipative) and other issues which have so far made the accurate simulation of many frictional contact phenomena out of reach. In the following discussion, as well as above, we use the term "robustness" to denote that these problems do not manifest, regardless of the accuracy of the obtained solution.

In part, to address the perceived high computational cost of direct LCP algorithms, recent attention has been focused on iterative LCP solution methods [Murty 1988; Cottle, Pang, & Stone 1992] customized for contacting systems [Moreau 1994; Jourdan, Alart, & Jean 1998; Jean 1998; Guendelman, Bridson, & Fedkiw 2003; Duriez *et al.* 2006; Erleben 2007]. These methods, generally based on Gauss-Seidel variants, can potentially be quite fast, and, for LCP problems obtained from frictionless contacting systems, have reasonably nice convergence guarantees [Cottle, Pang, & Stone 1992]. When friction is also modeled, however, Gauss-Seidel-based methods are no longer guaranteed to converge for the resulting LCPs and, as we will show in Chapters 5 and 7, often fail for LCPs obtained from complex problems. The inherent lack of accuracy, stability and robustness this causes generally requires hand tuning and large amounts of non-physical constraint stabilization [Erleben 2007] to make complex multibody simulations work. Effectively, iterative LCP methods often reintroduce many of the same errors and artifacts, discussed above, that LCP methods were designed to avoid in the first

place.

In general, none of the above negative results suggests that solutions to these LCP problems would not generate effective simulations of frictional contact. It just has not been practical, in many cases, to actually examine the behavior of numerical solutions to large and/or complex LCP frictional-contact problems. Part of the difficulty in solving these LCPs, which we will discuss in detail in Chapter 5, can be ascribed to an asymmetry in the LCP problem that is imposed by friction constraints (see Section 2.4 for details on friction constraints) and made worse as friction increases.

In an effort to address the asymmetry problem, Anitescu & Hart [2004] have proposed a relaxation algorithm for the frictional contact problem by symmetrizing the LCP formulation. By formulating these relaxations as an iterative sequence, the proposed algorithm can be shown to converge to the true LCP solution, provided the coefficient is below an (unknown) threshold. In this approach, however, contact and friction constraints are folded together, which effectively moves the contact constraint directions away from the surface normals, as the coefficients of friction increase, and thus changing the nature of the contact problem when convergence is not obtained. While their approach is convergent for very small coefficients of friction ($\mu \simeq 0.05$), the authors report that it fails for coefficient values where frictional behavior becomes pronounced, e.g., $\mu = 0.2$.

1.4 Contributions

This thesis offers the following contributions:

- We show how the conditions for the so-called Painlevé Paradox can be generalized to define an important metric of frictional contact problem difficulty, on par with the coefficient of friction.

- We reformulate existing time-continous models [Moreau 1988; Stewart 2000] to explicitly partition the frictional contact problem into two naturally convex subproblems. This leads to discretizations and algorithms that naturally leverage the existing frictional contact problem's structure.

- Combining this model with discrete variational integration methods [Marsden & West

2001; Hairer, Lubich, & Wanner 2002], we then derive new variational integrators that pose the discrete frictional contact problem as a system of coupled minimizations. Solutions to these systems are given by points that are optimal for both of the minimizations [3] and avoid known issues with existing variational integration approaches for friction and contact.

- We show how a specific two-step variant of these variational schemes [Kaufman *et al.* 2008] generalizes the popular Stewart-Trinkle model [Stewart & Trinkle 1996] and essentially requires the solution of the same underlying Linear Complementarity Problem (LCP).

- We then show that difficulties specifically encountered in solving these LCP systems can be explained, in part, by noting that optimization problems obtained from these frictional contact formulations are equivalent to nonconvex problems in *global* optimization (which are generally NP-hard). We believe that this is the first time that a complexity argument has been made via the connection to global minimization for the Stewart-Trinkle model.

- We show that existing solution methods that have generally been presumed suitable for solving these contact-related optimization problems fail entirely for many important examples of frictional contact, and then address these limitations with our Staggered Projections algorithm. Applying a fixed-point scheme, derived from our coupled principles work, Staggered Projections efficiently obtains accurate solutions to optimization problems for many frictional contact problems that were previously impractical to solve.

- We show that Staggered Projections performs well in practice by offering simulations and instrumented examples that capture convincing and accurate frictional contact behaviors for both rigid and large deformation models, as well as tests of the Staggered Projections implementation with other available popular frictional contact solvers.

- Finally, we offer a detailed convergence analysis of the Staggered Projections algorithm in which we discuss its behavior, as well as its suitability for practical applications and its potential limitations.

[3] Although more general, this problem is similar in nature to finding the solution of a system of equations.

Chapter 2

Discrete Frictional Contact

In this chapter we formulate and define the basic elements of the discrete frictional contact problem in a generalized setting. These will essentially define the fundamental elements of the contact problems that we will be considering throughout the remainder of this thesis. We will then, for completeness, quickly introduce some useful tools from convex and variational analysis that we will apply in later chapters. Finally, with our definitions, assumptions, and structures in place, we will discuss and motivate the difficulties involved in discrete frictional contact a bit more rigorously.

2.1 Semi-Discrete and Reduced Systems

In this thesis we restrict our attention to finite degree of freedom systems with generalized coordinates given by a vector q, mass matrices denoted by M, and a convex, potential energy function given by $V(\mathsf{q})$. Systems covered potentially include Finite Element Models (FEM), reduced modal models, rigid body models, reduced mechanical models and many other formulations that either directly have discrete coordinates or have had a spatial discretization scheme applied to them.

2.2 Frictional-Contact Response

Along with system state (i.e., $\mathsf{q}, \dot{\mathsf{q}}, \ddot{\mathsf{q}}$) frictional contact problems have an additional unknown: the frictional contact response, which we will denote by r. This frictional contact response can then be decomposed into a contact (non-penetration) component, c, and a frictional response component, f. In most cases it will be sufficient to consider these terms as forces. However, for full rigor, left and right limits of these terms must be handled carefully (see Section 2.8.2 for a

brief discussion and Stewart [2000] for a full coverage of the related issues). To limit the scope of the following presentation, beyond Section 2.8.2, we will not discuss these issues further, except to note that in the following sections we will first treat these terms as forces, while from Chapter 4 onward we will overload the notation by using superscripted versions of r, c, and f to describe discrete *impulses* [1]. Within each of the following sections we will clarify the local usage of these terms.

Also note that in this thesis, to simplify our discussion, we will reserve the word "contact" to specify terms (e.g., forces, constraints, impulses) related to the enforcement of non-penetration. Similarly, we use the word "friction" to specify terms related to frictional phenomena. The second usage may be obvious, but we find that this division of terms is helpful in clarifying the discussion.

2.3 Material Domain and Contacts

We presume that the material domain of the system, potentially composed of many different bodies, is given by $\Omega \subset \mathbb{R}^3$. Each point belonging to the system is identified by a material coordinate $\boldsymbol{X} \in \Omega$, with local spatial coordinates given by

$$\boldsymbol{x} = \Phi(\boldsymbol{X}, \mathsf{q}) \in \mathbb{R}^3. \tag{2.1}$$

2.3.1 Contact Set

We also define the set of all possible contact pairs, $\mathscr{C}$. Generally speaking, we will rarely compute this set explicitly; however, this is a useful convention in the foregoing discussion.

[1] i.e., integrals of force with respect to time-step

2.3.2 Relative Jacobian

Arbitrary contacts, indexed by $k \in \mathscr{C}$ and occurring at $\boldsymbol{x}_k \in \mathbb{R}^3$, between points $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ belonging to the boundary[2] of the material domain, $\partial\Omega$, have a relative velocity given by

$$\dot{\boldsymbol{x}}_k = \nabla_{\mathsf{q}}\Big(\Phi(\boldsymbol{X}_i, \mathsf{q}) - \Phi(\boldsymbol{X}_j, \mathsf{q})\Big)\dot{\mathsf{q}}. \tag{2.2}$$

To simplify notation in much of the following, we specify the relative Jacobian, employed above, using

$$\Gamma_k \stackrel{\text{def}}{=} \nabla_{\mathsf{q}}\Big(\Phi(\boldsymbol{X}_i, \mathsf{q}) - \Phi(\boldsymbol{X}_j, \mathsf{q})\Big). \tag{2.3}$$

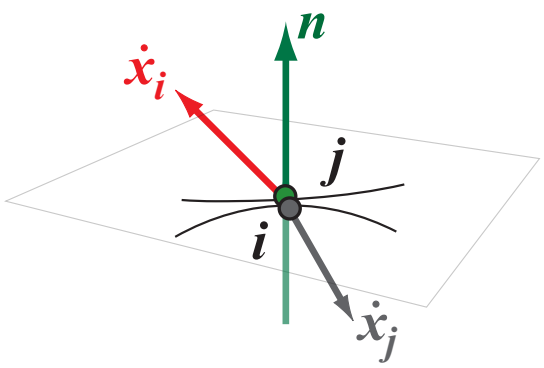


Figure 2.1: **Relative Velocity:** we track the relative velocity at a contact k between two points i and j using the relative jacobian, Γ_k.

2.4 Coulomb-Type Friction Laws

Coulomb-type friction laws constrain friction forces locally, at each possible contact, to lie in an origin centered, convex subset, scaled by the magnitude of the applied normal force. In the following, for each possible contact, k, we will let $\boldsymbol{f}_k$ denote the local friction force and α_k the local normal force magnitude. Friction laws can then be specified in either an inclusion form or an inequality constraint form. We'll discuss both variations here.

---

[2]The ∂ operator applied to Ω will be used to define the boundary of the domain, however, elsewhere, when applied to nonsmooth mappings, it will be used to indicate the application of the generalized gradient operator.

2.4.1 Inclusion Form

Friction laws can be enforced using set valued mappings, $\boldsymbol{F}_k(\alpha_k, \mathsf{q})$. These mappings generate scaled, origin centered, convex subsets in which the current friction force must lie, so that, at each contact, we can define the local friction law via the inclusion

$$\boldsymbol{f}_k \in \boldsymbol{F}_k(\alpha_k, \mathsf{q}). \tag{2.4}$$

2.4.2 Inequality Form

Alternately, letting $\boldsymbol{T}_k$ denote an arbitrary, generally orthogonal, span of the minimal subspace containing $\boldsymbol{F}_k$, we can define friction laws by pairs of equalities and inequalities. Local friction forces are constrained to the spanning subspace by requiring

$$\boldsymbol{f}_k = \boldsymbol{T}_k \beta_k, \tag{2.5}$$

where β_k is a corresponding vector of friction force magnitudes. Friction is then further constrained to the correct subset of $\boldsymbol{T}_k$, via inequality constraint functions, $\breve{\boldsymbol{F}}_k(\alpha_k, \beta_k, \mathsf{q})$. These functions are constructed so that the friction law is satisfied if and only if

$$\breve{\boldsymbol{F}}_k(\alpha_k, \beta_k, \mathsf{q}) \geq 0, \tag{2.6}$$

for all possible contacts k. In most cases of interest we can further decompose the inequality constraint so that we have

$$\breve{\boldsymbol{F}}_k(\alpha_k, \beta_k, \mathsf{q}) = a_k(\alpha_k) - b_k(\beta_k). \tag{2.7}$$

2.4.3 Example1: Isotropic and Anisotropic Coulomb Laws

For the classic form of the Coulomb-Amontons friction law, $\boldsymbol{T}_k$ is given by the two-dimensional span of a local tangent plane defined as common to both contacting points at $\boldsymbol{x}_k$. The constraint function, $\breve{F}_k$ is then given by $a_k(\alpha_k) = \mu_k \alpha_k$, while $b_k(\beta_k) = \| \beta_k \|$ for the isotropic model and $b_k(\beta_k) = \sqrt{c_1 \beta_{k_1} + c_2 \beta_{k_2}}$, where $c_1, c_2 > 0$, for the anisotropic model.

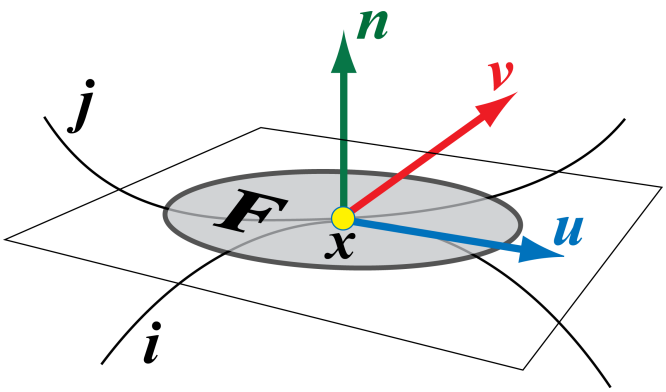


Figure 2.2: **Coulomb Friction Disk:** Coulomb's Law defines a contact centered disk of possible friction forces lying in the tangent plane.

2.4.4 Example2: Linearized Coulomb Laws

In many formulations a linearized, isotropic Coulomb Law [Stewart 2000] is employed. We can consider this law as a particular instance of an anisotropic Coulomb law variant or, alternately, we can view this law as a linear approximation of the nonlinear isotropic Coulomb Law. Effectively, each isotropic Coulomb friction disk is approximated by an inscribed polygon so that, at each contact k, we uniformly sample an arbitrarily large, *symmetric* set of unit length vectors from the tangent plane. We then redefine $\boldsymbol{T}_k$ to be the matrix composed column-wise of these samples so that a friction force, $\boldsymbol{f}_k \in \mathbb{R}^3$, applied at a contact k, will lie in the span of $\boldsymbol{T}_k$. As before, we continue to enforce the side constraint $\boldsymbol{f}_k = \boldsymbol{T}_k \boldsymbol{\beta}_k$; however, the benefits of linearization come with the cost of an additional constraint. Now that we are using symmetric samples (rather than two orthogonal samples), we additionally constrain the friction force magnitudes, β_k, to be nonnegative, i.e.,

$$\boldsymbol{\beta}_k \geq 0. \tag{2.8}$$

We then obtain the linearized Coulomb law friction subset

$$\boldsymbol{F}_k(\alpha_k, \mathsf{q}) = \{\boldsymbol{T}_k \beta : \boldsymbol{e}^T \beta_k \leq \mu_k \alpha_k, \beta_k \geq 0\}, \tag{2.9}$$

where $\boldsymbol{e} = (1....1)^T$.

Similarly, the constraint function, $\breve{\boldsymbol{F}}_k$, in this case, is given by $a_k(\alpha_k) = \mu_k \alpha_k$ and $b_k(\beta_k) =$

$e^T\beta_k$. so that we enforce

$$\breve{F}(\alpha_k,\beta_k) = \mu_k\alpha_k - e^T\beta_k \geq 0. \tag{2.10}$$

Note that, for multiple points of contact, we can concatenate all of the respective μ_k values into a single vector μ and correspondingly define a vector that concatenates the friction force magnitudes for all of the contacts

$$\beta = (\beta_1^T ... \beta_{|\mathscr{C}|}^T)^T. \tag{2.11}$$

Then the linearized Coulomb constraint, for all contacts, can be reduced to the simple form of the two inequalities

$$E^T\beta \leq \text{diag}(\mu)\alpha, \quad \beta \geq 0. \tag{2.12}$$

Here $\text{diag}(\mu)$ constructs the diagonal matrix composed of the μ_k values along the diagonal and zeros elsewhere. The matrix E corresponds to the vector β such that each column k of E has ones in rows that correspond to entries in the subvector $\beta_k \in \beta$ and zeros in all other rows.

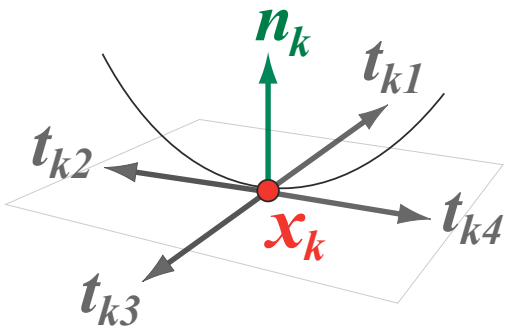


Figure 2.3: The tangent plane is symmetrically sampled for the linearized Coulomb Law.

2.5 A Variational Toolkit

In the interest of making this dissertation mostly self-contained we will briefly introduce a small set of tools from Variational and Convex analysis that will be useful in the following discussion.

2.5.1 Indicator Functions

Given an arbitrary set S, the extended value indicator function is defined as

$$I_{\mathsf{S}}(\mathsf{x}) = \begin{cases} 0, \ \mathsf{x} \in \mathsf{S} \\ \infty, \ \mathsf{x} \notin \mathsf{S}. \end{cases} \tag{2.13}$$

This nonsmooth, extended value function will, in the following sections, be applied to construct nonsmooth, dissipative, (super) potentials for both contact and friction.

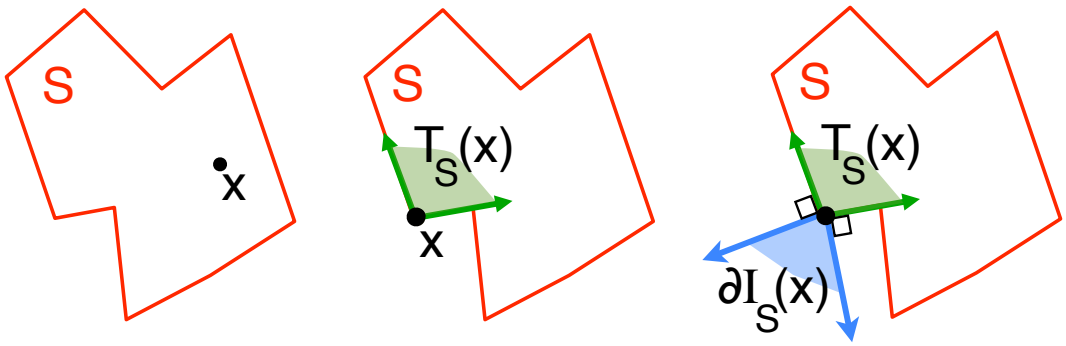


Figure 2.4: The geometric picture for the extended value indicator function.

2.5.2 Generalized Gradient

Because we will be working with nonsmooth functions (such as the indicator function described above) we will need to occasionally apply a generalized notion of calculus. A fundamental tool for nonsmooth analysis is the generalized gradient, which, at any differentiable point, reduces to the classical notion of a gradient. For nonsmooth points, however, the generalized gradient at x is, loosely speaking, given by the convex hull of the gradients at all of the closest differentiable points to x. More formally, we can define the generalized gradient of a locally Lipschitz function, f, as

$$\partial f(\mathsf{x}) = \mathrm{conv}\Big\{ \lim_{i \to \infty} \nabla f(\mathsf{x}_i) : \mathsf{x}_i \to x, \ \mathsf{x}_i \notin U, \ \mathsf{x}_i \notin V \Big\}, \tag{2.14}$$

were U is the set of non-differentiable points in the neighborhood of x and V is any set of measure zero [Clarke 1983].

2.5.3 Subdifferentials

For arbitrary *convex* functions, f, the generalized gradient operator, ∂, reduces to the subdifferential [Rockafellar 1970] of convex analysis which can be defined as

$$\partial f(\mathsf{x}) = \{\mathsf{y} : f(\mathsf{z}) \geq f(\mathsf{x}) + \mathsf{y}^T(\mathsf{z}-\mathsf{x}), \forall \mathsf{z}\}. \tag{2.15}$$

Basic calculus rules such as

$$\partial(-f)(\mathsf{x}) = -\partial f(\mathsf{x}). \tag{2.16}$$

can be then be inferred and stationary points follow inuitively; however, because of convexity, generalized stationarity, i.e., $0 \in \partial f(x^*)$, specifically implies that x^* is a *minimum* of f.

2.5.4 Tangent Cones

A tangent cone at an arbitrary point $\mathsf{x} \in \mathsf{S}$, effectively describes the set of all possible infinitesimal motions that can taken from x, while still remaining in the set S.

More formally, in the convex setting, the tangent cone is given by

$$T_{\mathsf{S}}(\mathsf{x}) = \Big\{\lambda(\mathsf{z}-\mathsf{x}) : \lambda \geq 0, \mathsf{z} \in \mathsf{S}\Big\}. \tag{2.17}$$

Taking advantage of convexity, this simply gives us the set of all vectors that, starting from x, are directed towards other points in S.

The generalized definition of the tangent cone, for the nonconvex case, effectively defines a similar, but localized, version of the tangent cone, constructed via a limiting process similar to the construction of the generalized differential. For the sake of brevity, however, we refer the interested reader to Clarke [1983] for details.

2.5.5 Normal Cones

The normal cone to S at x, then similarly generalizes the idea of an outward pointing normal direction at a nonsmooth point. In particular, the normal cone for both the convex and nonconvex cases is given by

$$N_{\mathsf{S}}(\mathsf{x}) = \Big\{\mathsf{y} : \mathsf{y}^T\mathsf{z} \leq 0, \forall \mathsf{z} \in T_{\mathsf{S}}(\mathsf{x})\Big\} \tag{2.18}$$

The normal cone is also given by applying the generalized gradient to the indicator function on S. That is, we always have

$$\partial I_{\mathsf{S}}(\mathsf{x}) = N_{\mathsf{S}}(\mathsf{x}). \tag{2.19}$$

For intuition building, the convex case can be helpful here. Applying the definition of the tangent cone from above, the normal cone must then be given by

$$N_{\mathsf{S}}(\mathsf{x}) = \left\{ \mathsf{y} : 0 \geq \mathsf{y}^T(\mathsf{z}-\mathsf{x}), \forall\, \mathsf{z} \in \mathsf{S} \right\}. \tag{2.20}$$

Then, for the particular case where we have an indicator function defined on a *convex* set S, the subdifferential of the indicator is

$$\partial I_{\mathsf{S}}(\mathsf{x}) = \{ \mathsf{y} : I_{\mathsf{S}}(z) \geq I_{\mathsf{S}}(\mathsf{x}) + \mathsf{y}^T(\mathsf{z}-\mathsf{x}), \forall\, \mathsf{z} \} \tag{2.21}$$

so that

$$\partial I_{\mathsf{S}}(\mathsf{x}) = \begin{cases} \{ \mathsf{y} : 0 \geq \mathsf{y}^T(\mathsf{z}-\mathsf{x}), \forall\, \mathsf{z} \in \mathsf{S} \} = N_{\mathsf{S}}(x), \text{ if } \mathsf{x} \in \mathsf{S}, \\ \emptyset, \text{ if } \mathsf{x} \notin \mathsf{S}. \end{cases} \tag{2.22}$$

We also note, in particular, that if $\mathsf{x} \in \text{interior}\,\mathsf{S}$, $\partial I_{\mathsf{S}}(\mathsf{x}) = N_{\mathsf{S}}(x) = \{0\}$.

2.5.6 Legendre-Fenchel Conjugates

The Legendre-Fenchel transform [Rockafellar 1970], of an arbitrary function, f, effectively generalizes the Legendre Transform to nonsmooth systems and is given by

$$f^*(\mathsf{x}) = \sup_{\mathsf{y}} (\mathsf{y}^T\mathsf{x} - f(\mathsf{y})). \tag{2.23}$$

Applying the Fenchel Transform to the case where $f = I_{\mathsf{S}}$ we then obtain the conjugate function

$$I^*_{\mathsf{S}}(\mathsf{x}) = \sup_{\mathsf{y} \in \mathsf{S}} \mathsf{y}^T\mathsf{x}. \tag{2.24}$$

The Legendre-Fenchel Transform will be helpful, in the next chapter, for defining an appropriate (super) potential function for friction.

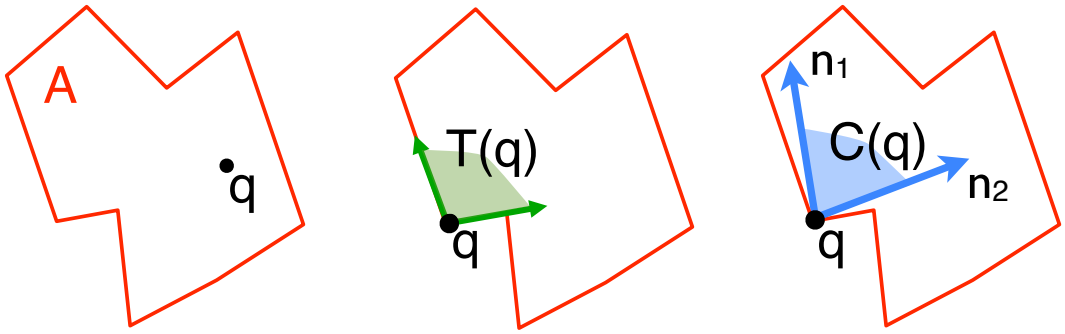


Figure 2.5: **Feasible Set:** Contact constraints describe a feasible set A, which the contacting system must lie in. For each configuration q, we can describe an associated tangent cone $\mathsf{T}(\mathsf{q})$, and normal cone $\mathsf{N}(\mathsf{q})$.

2.6 Feasible Set

In the generalized discrete contact setting we will find it useful to define the set of all acceptable *non-penetrating* configurations using the notion of a feasible set (or acceptable set) A. In general, A can be both nonconvex and nonsmooth.

We also presume, without loss of generality, that the feasible set A can always be described by a set of constraint functions,

$$\mathsf{G}(\mathsf{q}) = (g_0(\mathsf{q}), ..., g_m(\mathsf{q}))^T \geq 0. \tag{2.25}$$

2.7 Normal and Tangent Cones Via Constraint Gradients

For the specific context of discrete contact, we will also find it useful to define the tangent and normal cone to the feasible set, A, using the gradients of the constraint functions, G.

We first let N denote the matrix composed of the gradients of the zero-valued constraints at q, i.e.,

$$\mathsf{N}(\mathsf{q}) \stackrel{\text{def}}{=} \{\nabla g_i(\mathsf{q}) : \ g_i(\mathsf{q}) = 0, \ i \in [0, m]\}. \tag{2.26}$$

The inward pointing normal cone, the negation of the outward pointing normal cone, N_{A}, defined above, can then equivalently be given by

$$\mathsf{C}(\mathsf{q}) \stackrel{\text{def}}{=} \{\mathsf{N}\mathsf{x} : \mathsf{x} \geq 0\}. \tag{2.27}$$

Similarly, as hinted above, the set of feasible directions, along which infinitesimal motion is locally permissible, will be described by the tangent cone to A at q, which can also be specified

as

$$\mathsf{T}(\mathsf{q}) \stackrel{\text{def}}{=} \{\mathsf{y} : \mathsf{N}^T\mathsf{y} \geq 0\}. \tag{2.28}$$

Note that throughout this thesis, to simplify notation, the tangent cone to the feasible set, A, will be written, as above, *without* a subscript.

2.8 The Coupling Problem

With the preliminary problem formulation in place we can now discuss some of the fundamental challenges of frictional contact with greater detail.

2.8.1 Frictional Contact Challenges

Much of the difficulty in frictional contact resolution algorithms comes from the feedback between friction impulses (or forces) and contact (non-penetration) impulses (forces). Many ad-hoc contact resolution algorithms propose variants on a two pass strategy in which contact constraints are processed in a first pass, followed by a second pass in which friction impulses are applied. For deformable models, with low stiffness, this can often work reasonably well. As stiffness increases, however, friction impulses can begin to apply significant global changes in the system's velocity that, in turn, can create new contact constraint violations. In a similar way, contact correction impulses can effect global changes in the system's sliding velocities and the Coulomb constraints, which must affect friction impulse calculations.

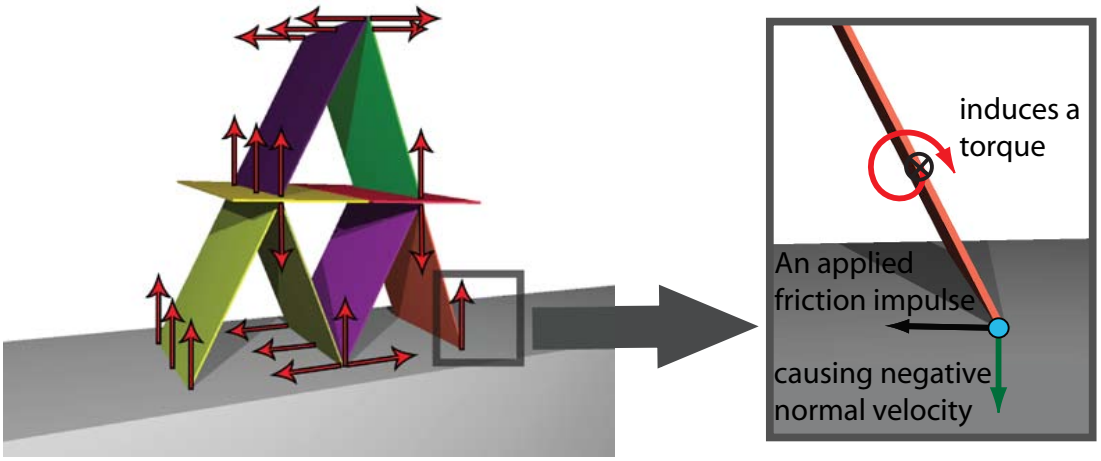


Figure 2.6: **Rigid Card House:** Constraint and friction force coupling.

Rigid body systems, which are essentially deformable bodies with infinite stiffness, pose an extreme version of this problem. Consider the card house example where the structure is composed of thin, rigid planks. A tangential friction impulse applied at the base of a card, to oppose outward sliding, will also induce a torque on the card. The torque, in turn, will cause a negative velocity along the normal *at the same point* where the friction impulse was originally applied. Similarly, a normal impulse, applied at the same contact point, would generate a faster outward sliding velocity (see Figure 2.6). These effects are related to the Painlevé Paradox [Painlevé 1895], and the observation by Erdmann [1994] that generalized normal and friction impulses, for rigid bodies, are not, in general, orthogonal. They can, in fact, oppose or even reinforce one another. We also note that, while this effect is most marked for rigid bodies, we can extend the same observations to deformable models.

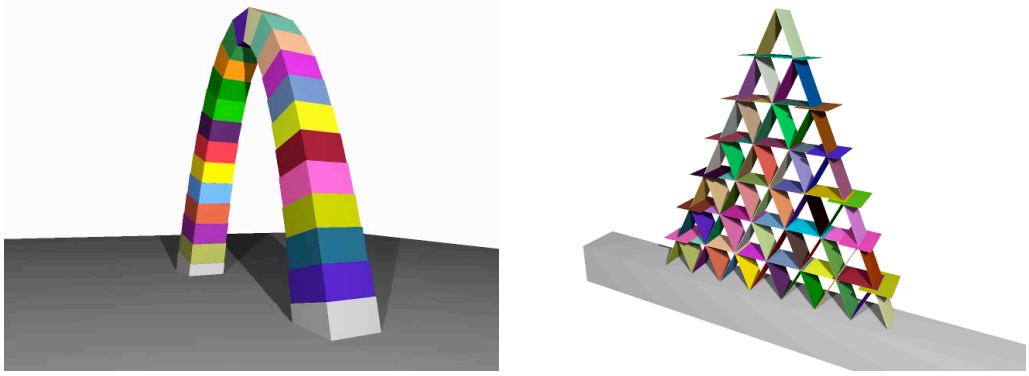


Figure 2.7: **Friction Dependent Masonry:** Stable masonry structures can be built without adhesives using materials that have a suitably high coefficient of friction. **Left:** In this simulated example a stable catenary arch is constructed using rigid body block models with a high coefficient of friction ($\mu = 0.6$). **Right:** In this simulated example a stable house of rigid cards is constructed using rigid body card models with a high coefficient of friction ($\mu = 0.8$). See Chapter 7 for details on both of these simulation examples.

There has recently been a focus on stable structured stacking in gaming and graphics [Erleben 2007]. While these examples require accurate contact constraint enforcement they require very little accuracy from friction. Structures like the rigid card house in Figure 2.7 right, masonry arches *without* adhesives (see Figure 2.7 left), or woven elastic-frictional composites (see Figure 7.1), on the other hand, depend on very accurate coupling between friction and contact impulse calculations. Non-orthogonal friction and contact impulses must be accurately balanced, at each active contact point, to prevent incremental errors from crashing or breaking

apart the structures. Because of this, inaccuracies imposed by two-pass, penalty, iterative LCP, and other methods have previously made the stable simulation of such frictionally dependent structures impractical.

2.8.2 The Painlevé Paradox

Painlevé [1895] famously observed that some initial value problems involving rigid bodies and Coulomb friction appeared to have no solution. For obvious reasons this has long stirred controversy.

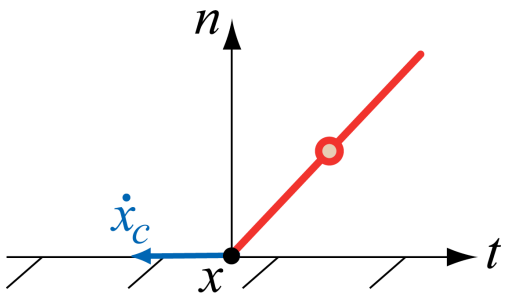


Figure 2.8: **Painlevé Rod Example**

The now classic example given by Painlevé is that of a rigid rod sliding along a flat surface, with the rod at an angle to the horizontal (see Figure 2.8). The Painlevé rod example is generally formulated using the local coordinates of the contacting point (see Stewart [2000] for example). Here, we will reformulate the problem in a generalized coordinate formulation to make explicit some properties of the Painlevé Paradox that we will want to revisit in later chapters.

Generalized Coordinate Painlevé Example

In this simple example we have a one dimensional rod of length 2 with a single point of contact at $\boldsymbol{x}_c$ (Figure 2.8). The rod has the inertia matrix M and the magnitude of gravity is g. The contacting surface is fixed with a normal direction $\boldsymbol{n}$ and a tangent $\boldsymbol{t}$. Here the degrees of freedom of the system are given by

$$\mathsf{q} = (a, \boldsymbol{p})^T \in SE(2), \tag{2.29}$$

where a gives the angle of rotation about the center of mass, and $\boldsymbol{p}$ indicates the $\mathbb{R}^2$ coordinates of the center of mass [3]The generalized velocity is then

$$\dot{\mathsf{q}} = (\dot{a}, \dot{\boldsymbol{p}})^T \in se(2). \tag{2.30}$$

The body frame coordinates of the contact point could then be given by

$$\hat{\boldsymbol{x}}_c = (0, -1)^T. \tag{2.31}$$

The world coordinates of the contacting point are the given by

$$\boldsymbol{x}_c(\mathsf{q}) = R(a)\hat{\boldsymbol{x}} + \boldsymbol{p}, \tag{2.32}$$

Where $R(a)$ is the rotation specified by a; i.e.,

$$R(a) = \begin{pmatrix} \cos(a) & -\sin(a) \\ \sin(a) & \cos(a) \end{pmatrix}. \tag{2.33}$$

Then, we have

$$\dot{\boldsymbol{x}}_c = \begin{pmatrix} \cos(a) & 1 & 0 \\ \sin(a) & 0 & 1 \end{pmatrix} \dot{\mathsf{q}}, \tag{2.34}$$

so that the jacobian for contact point $\boldsymbol{x}_c$ is

$$\Gamma_c = \begin{pmatrix} \cos(a) & 1 & 0 \\ \sin(a) & 0 & 1 \end{pmatrix}, \tag{2.35}$$

with $\dot{\boldsymbol{x}}_c = \Gamma_c \dot{\mathsf{q}}$.

To enforce non-penetration, we enforce the constraint that acceleration of the contact point be nonnegative along the normal, i.e.,

$$\boldsymbol{n}^T \ddot{\boldsymbol{x}}_c \geq 0. \tag{2.36}$$

Then contact forces, required to enforce this constraint, are applied with a magnitude of $\alpha \geq 0$ along the direction of the normal. The corresponding generalized force, if this contact force, $\boldsymbol{n}\alpha$, is applied at $\boldsymbol{x}_c$ is then

$$\mathsf{c} = \Gamma_c^T \boldsymbol{n}\alpha \in se^*(2). \tag{2.37}$$

---

[3]For more details on the Special Euclidean groups for two and three dimensional rigid motion (i.e., $SE(2)$ and $SE(3)$) and their correspodning Lie Algebras ($se(2)$ and $se(3)$) we refer the interested reader to Murray, Li, & Sastry [1994]

Similarly, friction is applied as a magnitude β along the tangent direction $\boldsymbol{t}$. The corresponding generalized force if this friction force, $\boldsymbol{t}\beta$, is applied at $\boldsymbol{x}_c$ is then

$$\mathsf{f} = \Gamma_c^T \boldsymbol{t}\beta \in se^*(2). \tag{2.38}$$

Intuitively, these three-dimensional, generalized forces make sense in the context of a contact in $\mathbb{R}^2$. The last two entries in each generalized force correspond to the linear force applied at $\boldsymbol{x}_c$, while the first entry in each generalized force corresponds to the torque that is generated.

The equations of motion for this contacting system are then

$$\mathsf{M}\ddot{\mathsf{q}} = \mathsf{c} + \mathsf{f} + \mathsf{M}(0, -g, 0)^T. \tag{2.39}$$

If contact is breaking, i.e., $\boldsymbol{n}^T \ddot{\boldsymbol{x}}_c > 0$, then, since we do not consider adhesion, the contact force must be zero. Similarly, if the contact force is greater than zero, i.e., $\alpha > 0$, the rod must stay in contact, implying $\boldsymbol{n}^T \ddot{\boldsymbol{x}}_c = 0$. Combining these observations together we get a simple variant of the Signorini condition [Signorini 1933],

$$0 \leq \alpha \perp \boldsymbol{n}^T \ddot{\boldsymbol{x}}_c \geq 0, \tag{2.40}$$

where the $\perp$ notation simply indicates $\alpha \boldsymbol{n}^T \ddot{\boldsymbol{x}}_c = 0$.

We now consider the case of zero angular velocity ($\dot{a} = 0$), zero normal velocity ($\dot{\boldsymbol{p}}_2 = 0$), and negative tangential velocity ($\dot{\boldsymbol{p}}_1 < 0$). This, in turn, implies that the contact point must be sliding to the left, i.e, $\boldsymbol{t}^T \dot{\boldsymbol{x}}_c < 0$. Because of this sliding, Coulomb's law specifically requires that magnitude of the friction force take the maximum value

$$\beta = \mu \alpha. \tag{2.41}$$

Expanding the contact constraint from Equation (2.36) we have

$$\boldsymbol{n}^T \left(\frac{d}{dt} \left(\Gamma_c \dot{\mathsf{q}} \right) \right) = \boldsymbol{n}^T \Gamma_c \ddot{\mathsf{q}} + \boldsymbol{n}^T \dot{\Gamma}_c \dot{\mathsf{q}} \geq 0. \tag{2.42}$$

Then, noting that, in this context, we have

$$\dot{\Gamma}_c = \begin{pmatrix} -\sin(a)\dot{a} & 0 & 0 \\ \cos(a)\dot{a} & 0 & 0 \end{pmatrix}, \tag{2.43}$$

the $\mathsf{n}^T\dot{\Gamma}_c\dot{\mathsf{q}}$ term drops out of the above inequality (since $\dot{a}=0$) and we are left with the constraint

$$\boldsymbol{n}^T\Gamma_c\ddot{\mathsf{q}} \geq 0. \tag{2.44}$$

Then substituting in Equations (2.39),(2.41), (2.37), and (2.38) we obtain

$$\boldsymbol{n}^T\Gamma_c\mathsf{M}^{-1}\Gamma_c^T\boldsymbol{n}\alpha + \boldsymbol{n}^T\Gamma_c\mathsf{M}^{-1}\Gamma_c^T\boldsymbol{t}(\mu\alpha) - \mathsf{g} \geq 0, \tag{2.45}$$

or

$$\left(\boldsymbol{n}^T\Gamma_c\mathsf{M}^{-1}\Gamma_c^T\boldsymbol{n} + \mu\boldsymbol{n}^T\Gamma_c\mathsf{M}^{-1}\Gamma_c^T\boldsymbol{t}\right)\alpha - \mathsf{g} \geq 0. \tag{2.46}$$

Recalling that $g > 0$ and noting that $\boldsymbol{n}^T\Gamma_c\mathsf{M}^{-1}\Gamma_c^T\boldsymbol{n} > 0$, we see that if $\boldsymbol{n}^T\Gamma_c\mathsf{M}^{-1}\Gamma_c^T\boldsymbol{t} < 0$ and μ is sufficiently large, there does not exist any contact force $\alpha \geq 0$ that can satisfy the contact constraint. This is Painlevé's Paradox. Effectively, in these Painlevé cases, the applied forces will induce a torque that rotates the contacting point *into* the constraint surfaces and, due to friction, no feasible contact force exists that can prevent this from happening.

The modern resolution of the so-called paradox hinges on the observation that the entire analysis above depends on the assumption that there are no impulsive forces involved. The moment we presume that the sliding velocity can be driven to zero *instantaneously*, the paradox disappears. To rigorously include impulsive forces, however, requires a measure-based formulation that carefully handles right and left limits at jumps. We refer the interested reader to Moreau [1988] and Stewart [2000] for further details.

2.8.3 Painlevé Conditions and Non-Orthogonality

Considering the above generalized force definitions in Equations (2.37), and (2.38), we see that they effectively define a generalized normal direction $\mathsf{n} = \Gamma_c^T\boldsymbol{n}$ and a generalized tangent direction $\mathsf{d} = \Gamma_c^T\boldsymbol{t}$.

In this framework, it is also often helpful to consider the kinetic metric given by the system's mass and inertia. That is, for arbitrary covariant vectors a and b, the kinetic metric is given by

$$< \mathsf{a}, \mathsf{b} >_{\mathsf{M}^{-1}} = \mathsf{a}^T\mathsf{M}^{-1}\mathsf{b}. \tag{2.47}$$

From this perspective, conditions for the Painlevé Paradox, as given above in Equation (2.46), can occur whenever we have

$$< \mathsf{n}, \mathsf{d} >_{\mathsf{M}^{-1}} > 0. \tag{2.48}$$

The problem is then that, although $\boldsymbol{n}$ and $\boldsymbol{t}$ are orthogonal in the contact space, their corresponding generalized force space analogues, n and d, no longer necessarily inherit orthogonality [Erdmann 1994].

Equation (2.46) suggests that as μ increases, the higher the likelihood of conditions being suitable for the Painlevé Paradox to occur. On the other hand, however, we also see that the angle between n and d, is also important. As this angle decreases, the likelihood of conditions suitable for the Painlevé Paradox also correspondingly increases.

As we will show in later chapters, conditions that can cause the Painlevé Paradox are not only of theoretical interest. We will show that in combination, these two criteria (μ and the generalized, inertia-weighted angle) are responsible, in large part, for much of the difficulty involved in solving computational problems related to frictional contact problems.

Given the modern resolution of the Painlevé Paradox, we will, in the rest of this thesis, simply refer to the conditions that give rise to these troublesome Painlevé Paradox-type conditions as "Painlevé Conditions".

In the generalized, multi-point setting these issues are of even greater importance. For these cases we will introduce a simple, useful, and intuitive generalization of the Painlevé Condition and will show its direct connection with the difficulty of computational frictional contact problems. We will show that, as is generally well appreciated, increasing the coefficient of friction, μ, is certainly a crucial indicator of potential difficulties, but that a generalized notion of angle between friction and contact directions, is also of equal importance.

2.8.4 Two-Way Coupling in this Thesis

In Section 2.8.1 we motivated how coupling between friction forces and contact forces could be troublesome for accurate frictional contact algorithms. In Sections 2.8.2 and 2.8.3 we introduced the Painlevé Paradox and discussed how non-orthogonality and the coefficient of friction play a role in somehow making frictional contact problems hard.

What we have shown and will continue to discuss throughout this thesis is that friction is, in essence, a two-way coupled problem. The first coupling, as we saw above, comes from the interaction between generalized friction forces and generalized contact forces. If these two forces are orthogonal, then this form of coupling is removed. As we saw in the last three sections non-orthogonal cases encode important global changes that can be imposed by local contact and friction forces. The other source of coupling comes from Coulomb-type friction subsets. These subsets are scaled by the magnitude of the contact force; this scaling causes a second form of feedback between friction and contact forces. If either form of coupling is removed we will show, in Chapters 5 and 6, that the frictional contact problem generally becomes much easier. With this coupling in place, however, frictional contact imposes unique challenges that must be addressed for accurate frictional contact modeling.

Chapter 3

Coupled Variational Principles

In this chapter we will develop a coupled model of frictional contact. We will show that discrete frictional contact, in the time-continuous setting, can be posed as a system of coupled variational principles. From this perspective we will see that frictional contact is, essentially, a multi-physics problem. we will then show that *both* of the obtained principles are effectively convex minimization problems and so lead (in Chapter 4) to partitioned discretizations that leverage the frictional contact problem's inherent structure. Morever, the coupled variational formulation we adopt here allows us to construct natural Lagrangians that, in turn, will be applied, using discrete variational techniques, to construct novel numerical integration schemes for frictional contact in Chapter 4.

We first begin by noting that frictional contact response is, by definition, a purely dissipative phenomena. Similar to Rayleigh-type damping it is velocity dependent. Unlike Rayleigh-type damping, however, frictional-contact response can not, in general, be obtained from the direct substitution of a single potential gradient into the Euler-Lagrange equations. Instead, frictional-contact response terms can be obtained from the gradients of two separate, nonsmooth potentials, which are mutually dependent.

We will highlight this coupling by showing that, leveraging seminal results by Moreau [1966, 1973, 1988] a variational interpretation of frictional contact can be formulated as a pair of two, independently convex, variational principles, in which the friction and contact potentials are each embedded in separate variational principles. As we will show, these two potentials are intrinsically coupled in both directions. Due to this coupling, each variational principle can not, in general, be solved independently of the other and will generally be found to be the chief source of unavoidable computational difficulty in solving related frictional contact problems accurately.

3.1 A Non-Smooth Hamilton's Stationarity Principle

To begin our development we start with Hamilton's Stationarity Principle subject to constraints.

3.1.1 The Basic Form

We can implicitly include constraints in the Hamiltonian setting by augmenting the classical action functional with an indicator function of the acceptable set, I_{A}, to construct an extended value action

$$\int_0^T \frac{1}{2}\,\dot{\mathsf{q}}^T \mathsf{M}\dot{\mathsf{q}} - V(\mathsf{q}) - I_{\mathsf{A}}(\mathsf{q})\; dt. \tag{3.1}$$

Note that all terms here are the same as in the classical action, except for the addition of the indicator function. This extra term infinitely penalizes all paths that leave the feasible set A. Thus extremizing this action leads to a generalized form of the Euler-Lagrange equations that naturally include constraint enforcing forces [Clarke 1983; Kane *et al.* 1999]. We will see more on this in a moment.

3.1.2 A Frictional External-Force/Dissipation Term

Our formulation begins with a relatively simple addition. With the assumption that local coordinates for q exist, that are compatible with the generalized friction force, f, we can add an extra term, $\mathsf{f}^T\mathsf{q}$, to the action. This additional function describes the dissipation performed by friction. Adding this term to the action then gives us the augmented functional

$$\int_0^T \frac{1}{2}\,\dot{\mathsf{q}}^T \mathsf{M}\dot{\mathsf{q}} - V(\mathsf{q}) - I_{\mathsf{A}}(\mathsf{q}) + \mathsf{f}^T\mathsf{q}\; dt. \tag{3.2}$$

Then, applying Hamilton's Stationarity Principle, we can extremize the action in this nonsmooth setting as,

$$\delta_{\mathsf{q}} \int_0^T \frac{1}{2}\,\dot{\mathsf{q}}^T \mathsf{M}\dot{\mathsf{q}} - V(\mathsf{q}) - I_{\mathsf{A}}(\mathsf{q}) + \mathsf{f}^T\mathsf{q}\; dt \ni 0. \tag{3.3}$$

3.1.3 Integral Lagrange-d'Alembert Principle for Friction

Alternately, if we do not wish to apply a local coordinate argument, we can invoke the integral Lagrange-d'Alembert principle [Marsden & Ratiu 2003] in the nonsmooth setting,

$$\delta_{\mathsf{q}} \int_0^T \frac{1}{2}\,\dot{\mathsf{q}}^T \mathsf{M}\dot{\mathsf{q}} - V(\mathsf{q}) - I_{\mathsf{A}}(\mathsf{q})\; dt + \int_0^T \mathsf{f}\cdot\delta\mathsf{q}\; dt \ni 0, \tag{3.4}$$

to formulate our augmented system. In adopting this perspective, we effectively treat friction as an externally prescribed force; however, since friction is essentially path dependent, the inclusion of this term in the action here, as above, also requires friction to be maximally dissipative.

3.1.4 Stationarity

By either route, applying Nonsmooth Analysis [Rockafellar & Wets 1998], we obtain a multivalued stationarity condition,

$$\mathsf{M}\ddot{\mathsf{q}} - \mathsf{f} - \mathsf{g} \in -\partial I_{\mathsf{A}}(\mathsf{q}). \tag{3.5}$$

This is a differential inclusion (DI) form of the Euler-Lagrange equations of motion with

$$\mathsf{g} \stackrel{\text{def}}{=} -\nabla_{\mathsf{q}} V(\mathsf{q}).$$

3.1.5 Discussion: Generalized Gradients, Normal Forces

By definition, the generalized gradient of the indicator function returns 0 whenever q lies in the interior of A. In this case the DI in Equation (3.5) reduces to the familiar Euler-Lagrange equation, $\mathsf{M}\ddot{\mathsf{q}} = \mathsf{g}$. While, on the boundary of A, it reduces to the outward pointing normal cone. Thus we have $-\partial I_{\mathsf{A}}(\mathsf{q}) = \mathsf{C}(\mathsf{q})$.

Then, Equation (3.5) implies that all corrective forces, i.e., those that enforce the feasibility or *contact* constraints, must be applied positively along the constraint gradient directions, giving us

$$\mathsf{N}\alpha = \mathsf{M}\ddot{\mathsf{q}} - \mathsf{f} - \mathsf{g} \in -\partial I_{\mathsf{A}}(\mathsf{q}), \quad \alpha \geq 0. \tag{3.6}$$

From this perspective, the above α terms can thus be interpreted as force magnitudes, along the normal directions of system contacts.

3.1.6 A Lagrangian Dual Formulation

Equivalently, from an optimization perspective, we can view the α terms as Lagrange multipiers for enforcing the constraints, G. To explicitly expose these multipliers we can also reformulate the non-smooth action using Lagrangian duality

$$\int_0^T \frac{1}{2}\dot{\mathsf{q}}^T\mathsf{M}\dot{\mathsf{q}} - V(\mathsf{q}) - \alpha\mathsf{G}(\mathsf{q}) - I_{\mathbb{R}^m_+}(\alpha) + \mathsf{f}^T\mathsf{q}\; dt, \tag{3.7}$$

Extremizing with respect to both q and α, an equivalent stationarity condition is then given by the pair

$$\mathsf{M}\ddot{\mathsf{q}} = \mathsf{g} + \mathsf{f} + \mathsf{N}\alpha, \tag{3.8}$$

$$\mathsf{G}(\mathsf{q}) \in -\partial I_{\mathbb{R}^m_+}(\alpha). \tag{3.9}$$

3.1.7 Kinematic Equivalence, Local Convexity and Optimality

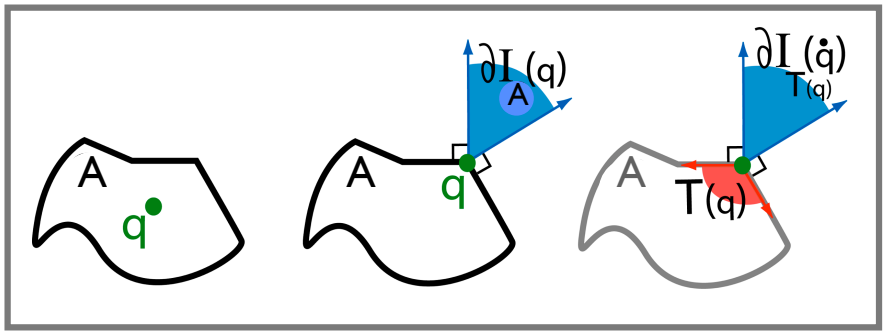


Figure 3.1: At the boundary of the feasible set, the relationship between the normal cone and the tangent cone allow us to transform constraints locally from configuration space to velocity space.

Locally we have an equivalence between the constraining the system's position to the feasible set, $\mathsf{q} \in \mathsf{A}$, and constraining velocities to the tangent cone, $\dot{\mathsf{q}} \in \mathsf{T}(\mathsf{q})$. This observation can be extended [Moreau 1988] to compose an equivalent, but *convex*, form of Equation (3.5), via the equivalence

$$\mathsf{M}\ddot{\mathsf{q}} - \mathsf{f} - \mathsf{g} \in -\partial I_{\mathsf{A}}(\mathsf{q}) \iff \mathsf{M}\ddot{\mathsf{q}} - \mathsf{f} - \mathsf{g} \in -\partial I_{\mathsf{T}(\mathsf{q})}(\dot{\mathsf{q}}). \tag{3.10}$$

See Figure 3.1 for intuition. This equivalence exposes the underlying local convexity of the stationarity criterion, since $\mathsf{T}(\mathsf{q})$, and correspondingly $I_{\mathsf{T}(\mathsf{q})}$, are both convex for all configurations, even when A is not. In the following we will refer to this equivalence as the "Moreau Transform". We now can compactly describe the contact response optimality using the convex formulation

$$-\mathsf{N}\alpha \in \partial I_{\mathsf{T}(\mathsf{q})}(\dot{\mathsf{q}}), \tag{3.11}$$

so that ∂ explicitly indicates the subdifferential operator.

3.2 A Maximal Dissipation Principle

Moreau [1973] introduced a rate-based Maximal Dissipation Principle to describe single point frictional contact. This approach reinterprets Coulomb friction as maximizing dissipation over all feasible friction forces at a contact [Goyal, Ruina, & Papadopoulos 1991]. Here we show how this approach can be generalized to construct a variational principle for friction in the multi-contact setting.

3.2.1 Local Dissipation Rate

First we need to define the rate of dissipation with respect to the generalized velocity The rate of dissipation performed by a friction force, $\boldsymbol{f}_k$, acting locally at a contacting point, $\boldsymbol{x}_k$, with an associated relative velocity, $\dot{\boldsymbol{x}}_k$, is given by

$$-\boldsymbol{f}_k^T \dot{\boldsymbol{x}}_k = -\boldsymbol{f}_k^T \Gamma_k \dot{\mathsf{q}}. \tag{3.12}$$

3.2.2 A Global Formulation of Dissipation

Then summing the local dissipation rates over all potential contacts and rearranging the terms, we obtain a generalized, global form of the total dissipation rate,

$$-\Big(\sum_k \Gamma_k^T \boldsymbol{f}_k\Big)^T \dot{\mathsf{q}}. \tag{3.13}$$

Thus the generalized friction force, generated at contacting point k, is given by

$$\mathsf{f}_k = \Gamma_k^T \boldsymbol{f}_k, \tag{3.14}$$

while, the corresponding full generalized friction term is then given by their sum,

$$\mathsf{f} = \sum_k \mathsf{f}_k. \tag{3.15}$$

3.2.3 Generalized Friction Constraints

Above we've shown how to generate the generalized friction forces. Here we will correspondingly derive generalized friction sets.

As discussed in Section 2.4, we can enforce Coulomb-type friction laws at each possible contact, k, using the inclusion constraints

$$\boldsymbol{f}_k \in \boldsymbol{F}_k(\alpha_k, \mathsf{q}). \tag{3.16}$$

Here we merely require that each $\boldsymbol{F}_k$ be convex for all[1] α_k.

Applying Equation (3.14) we can generate a corresponding generalized friction constraint for each potential contact,

$$\mathsf{f}_k \in \mathsf{F}_k(\alpha_k, \mathsf{q}) \stackrel{\text{def}}{=} \Gamma_k^T \boldsymbol{F}_k(\alpha_k, \mathsf{q}). \tag{3.17}$$

This, in turn, defines a single, global generalized friction set to which we constrain the global, generalized friction force

$$\mathsf{f} \in \mathsf{F}(\alpha, \mathsf{q}) \stackrel{\text{def}}{=} \bigoplus_k \mathsf{F}_k(\alpha_k, \mathsf{q}). \tag{3.18}$$

Note that because the generalized friction set, $\mathsf{F}(\alpha, \mathsf{q})$, is generated from this direct sum [2] of convex sets, it too is convex.

3.2.4 Dissipation Rate Functional

In the last section we reduced the friction constraints to a single generalized friction set in Equation (3.18). Using this friction set we can now construct a constrained work rate functional for friction,

$$\int_0^T \dot{\mathsf{q}}^T \mathsf{f} + I_{\mathsf{F}(\alpha,\mathsf{q})}(\mathsf{f})\, dt, \tag{3.19}$$

which we will call the Friction Functional. By construction, this functional is convex, so that by minimizing it, we maximize the rate of dissipation *globally* over the system's path, subject to the selected friction law.

3.2.5 Maximal Dissipation Principle: Variational Form

Extremizing the Friction Functional,

$$\delta_{\mathsf{f}} \int_0^T \dot{\mathsf{q}}^T \mathsf{f} + I_{\mathsf{F}(\alpha,\mathsf{q})}(\mathsf{f})\, dt \ni 0, \tag{3.20}$$

---

[1] Extensions to stiction laws can be obtained by reformulating $\boldsymbol{F}_k$ to also be parameterized by $\dot{\mathsf{q}}$.

[2] The direct sum of two arbitrary subsets, U and V, is given by $U \bigoplus V = \{x + y : x \in U, y \in V\}$.

we now obtain a second multi-valued optimality condition in DI form,

$$-\dot{\mathsf{q}} \in \partial I_{\mathsf{F}(\alpha,\mathsf{q})}(\mathsf{f}), \tag{3.21}$$

which corresponds to the optimality condition for the Maximal Dissipation principle.

3.2.6 A Lagrangian Dual Formulation

Equivalently, to expose the friction force magnitudes, β, we can also reformulate the Maximal Dissipation Principle using Lagrangian duality. We first concatenate the set of all possible local friction inequality constraints into a single global constraint

$$\breve{\boldsymbol{F}}(\alpha,\beta,\mathsf{q}) \stackrel{\text{def}}{=} \left(\breve{\boldsymbol{F}}_1(\alpha_1,\beta_1,\mathsf{q}),...,\breve{\boldsymbol{F}}_{|\mathscr{C}|}(\alpha_{|\mathscr{C}|},\beta_{|\mathscr{C}|},\mathsf{q})\right)^T. \tag{3.22}$$

Similarly we concatenate the friction magnitudes

$$\boldsymbol{\beta} \stackrel{\text{def}}{=} \left(\beta_1,...\beta_{|\mathscr{C}|}\right)^T \tag{3.23}$$

and the local friction bases

$$\boldsymbol{T} \stackrel{\text{def}}{=} \left(\boldsymbol{T}_1^T,...\boldsymbol{T}_{|\mathscr{C}|}^T\right)^T, \tag{3.24}$$

so that the friction equality constraints can be given by

$$\left(\boldsymbol{f}_1^T,...\boldsymbol{f}_{|\mathscr{C}|}^T\right)^T = \boldsymbol{T}\boldsymbol{\beta}. \tag{3.25}$$

We now can reformulate the Frictional Functional as

$$\int_0^T (\Gamma\dot{\mathsf{q}})^T(\boldsymbol{T}\boldsymbol{\beta}) - \lambda^T\breve{\mathsf{F}}(\alpha,\beta,\mathsf{q}) - I_{\mathbb{R}^m_+}(\lambda)\,dt. \tag{3.26}$$

Extremizing this action with respect to $\boldsymbol{\beta}$ and λ then gives the equivalent stationarity condition in the form of the pair

$$\boldsymbol{T}^T\Gamma\dot{\mathsf{q}} \in \lambda\,\partial_{\boldsymbol{\beta}}\breve{\mathsf{F}}(\alpha,\beta,\mathsf{q}), \tag{3.27}$$

$$\breve{\mathsf{F}}(\alpha,\beta,\mathsf{q}) \in -\partial I_{\mathbb{R}^m_+}(\lambda). \tag{3.28}$$

3.3 Joint Optimality

We have now independently derived the two Variational Principles for frictional contact. We have additionally shown how, despite the original nonconvexity of the Nonsmooth Hamilton's Principle, it can be further modified, using a theorem from Moreau, to expose the underlying convex structure of the contact problem. In each principle, however, essential unknowns from the other principle are required to correctly formulate the problem. In this section we will discuss this coupling and its implications for numerically solving frictional contact problems. First, however, we will discuss and clarify the superpotential interpretation of the obtained optimality conditions using Fenchel-Legendre conjugates.

3.3.1 Conjugate Optimality Forms

In Equation (3.21) the inclusion shows that the generalized velocity, $\dot{\mathsf{q}}$, is effectively constrained by the generalized friction response, f. In this inclusion $I_{\mathsf{F}(\alpha,\mathsf{q})}(\mathsf{f})$ functions as a conjugate form of the friction potential since, via an application of the Legendre-Fenchel transform, the equivalent conjugate DI optimality form of Equation (3.21) is

$$\mathsf{f} \in \partial I^{*}_{\mathsf{F}(\alpha,\mathsf{q})}(-\dot{\mathsf{q}}). \tag{3.29}$$

Similarly, applying the Legendre-Fenchel transform to Equation (3.11) we obtain

$$\dot{\mathsf{q}} \in \partial I^{*}_{\mathsf{T}(\mathsf{q})}(-\mathsf{N}\alpha), \tag{3.30}$$

and see that the generalized velocity is similarly constrained by the generalized contact response.

3.3.2 The Coupled Optimality System

With the addition of these last two conjugate terms we can now fully characterize the coupled frictional contact response system. The following diagram summarizes the overall relationship, including the dependency of the friction optimality terms on α (we use "$L-F$" to indicate the application of the Legendre-Fenchel Transform),

$$\begin{array}{ccc} -\mathsf{N}\alpha \in \partial I_{\mathsf{T}(\mathsf{q})}(\dot{\mathsf{q}}) & \xrightarrow{L-F} & \dot{\mathsf{q}} \in \partial I^*_{\mathsf{T}(\mathsf{q})}(-\mathsf{N}\alpha) \\ \alpha \big\downarrow & & \\ -\dot{\mathsf{q}} \in \partial I_{\mathsf{F}(\alpha,\mathsf{q})}(\mathsf{f}) & \xrightarrow{L-F} & \mathsf{f} \in \partial I^*_{\mathsf{F}(\alpha,\mathsf{q})}(-\dot{\mathsf{q}}). \end{array} \tag{3.31}$$

Recalling the defintion of the subdifferential, we can cast the diagram explicitly as a set of optimizations. Then we can equivalently view the relationship as

$$\begin{array}{ccc} \dot{\mathsf{q}} = \operatorname{argmin}_{\mathsf{y}} \left\{ \mathsf{y}^T \mathsf{N}\alpha + I_{\mathsf{T}(\mathsf{q})}(\mathsf{y}) \right\} & \xrightarrow{L-F} & \mathsf{N}\alpha = \operatorname{argmin}_{\mathsf{y}} \left\{ \mathsf{y}^T \dot{\mathsf{q}} + I^*_{\mathsf{T}(\mathsf{q})}(-\mathsf{y}) \right\} \\ \alpha \big\downarrow & & \\ \mathsf{f} = \operatorname{argmin}_{\mathsf{y}} \left\{ \mathsf{y}^T \dot{\mathsf{q}} + I_{\mathsf{F}(\alpha,\mathsf{q})}(\mathsf{y}) \right\} & \xrightarrow{L-F} & \dot{\mathsf{q}} = \operatorname{argmin}_{\mathsf{y}} \left\{ \mathsf{y}^T \mathsf{f} + I^*_{\mathsf{F}(\alpha,\mathsf{q})}(-\mathsf{y}) \right\}. \end{array} \tag{3.32}$$

3.3.3 Coupled Potentials

Based on the above coupled formulation, $I_{\mathsf{T}(\mathsf{q})}(\dot{\mathsf{q}})$ and $I^*_{\mathsf{F}(\alpha,\mathsf{q})}(-\dot{\mathsf{q}})$ function as the *convex* potentials for contact and friction response, respectively. Joint optimality for the system's instantaneous frictional contact response, $\mathsf{r} = \mathsf{c} + \mathsf{f}$, is thus given by the coupled system of the two subdifferentials

$$\begin{aligned} \mathsf{c} = \mathsf{N}\alpha \in -\partial I_{\mathsf{T}(\mathsf{q})}(\dot{\mathsf{q}}), \\ \mathsf{f} \in \partial I^*_{\mathsf{F}(\alpha,\mathsf{q})}(-\dot{\mathsf{q}}). \end{aligned} \tag{3.33}$$

The Euler-Lagrange equation,

$$\mathsf{M}\ddot{\mathsf{q}} = \mathsf{g} + \mathsf{f} + \mathsf{c}, \tag{3.34}$$

then completes the description of the contacting system's dynamics by describing the correct relation between the frictional-contact response and the non-contacting portion of the system. Note that classical dissipative forces, such as damping and external forcing, can also be included in the Euler-Lagrange equation, i.e., Equation (3.8), in the usual manner, and from the perspective of the above derivation can equivalently be added into the original variational formulation via the Lagrange-d'Alembert principle.

3.3.4 Discussion: Two-Way Interdependence Revisited

Since the friction potential is parameterized by α, the correct formulation of the friction potential is directly dependent on the subdifferential of the contact potential. The correct formulation of the contact potential, conversely, is dependent on the subdifferential of the friction potential, since the system's contact constraint status is directly effected by changes imposed by friction terms via the Euler-Lagrange equation. These interdependencies pose the primary difficulties encountered in frictionally contacting systems. In the later chapters we will clarify exactly how these difficulties manifest in the discrete time setting and just what aspects of this two way coupling are the chief culprits.

Chapter 4

Variational Methods for the Numerical Integration of Frictional Contact

4.1 Overview

Recently attention has been focused on numerical integration methods that preserve geometric invariants of continuous systems [Hairer, Lubich, & Wanner 2002]. The chief observation is that while numerical accuracy is desirable, there will always be error, due to discretization, for all numerical schemes. Thus the preservation of invariants, which leads to *qualitatively* good long term behavior, is essential for simulating the flow of differential equations.

One such class of methods are discrete variational integrators [Marsden & West 2001]. Instead of discretizing the equations of motion in the usual manner, discrete analogs of original minimization/stationarity principles that lead to a system's descriptive differential equations are derived. By paralleling in the discrete domain the classical development of the Lagrangian/Hamiltonian formulation, numerical integration schemes are obtained. They can be formulated to any order of accuracy desired and are symplectic in the discrete sense [West 2004].

For frictionally contacting systems, where, as we have already shown, minimization is a fundamental and unavoidable feature, the variational framework potentially offers several advantages. As we will show in this chapter, applying variational methods to contacting systems we can directly expose the underlying discrete constrained minimization systems that are necessary to accurately model contact and friction, while avoiding measure-based issues relating to Painlevé's paradox. Variational discretizations, once properly formulated, will also offer a wide range of possible integration approaches that directly couple the treatment of nonsmooth contact and friction terms with the classical numerical treatment of continuum and rigid-body mechanics. In order to properly formulate frictional contact in the discrete variational setting,

however, several technical hurdles will need to be overcome in this chapter. One bar to applying discrete variational techniques – formulating appropriate Lagrangians for frictional contact has already, however, been cleared in the last chapter. As we will see later in this chapter, the coupled Principles Formulation, introduced in the last chapter, when properly handled, provides a suitable basis for formulating discrete variational integrators for frictional contact.

Kane *et al.* [1999] first proposed applying nonsmooth extensions of variational integration methods in the frictionless contact setting. Although variational integration schemes have also been proposed for both conservative systems and dissipative systems with forcing and/or Rayleigh-type damping [Kane *et al.* 2000], extending these approaches to frictional contact has proved difficult. An extension of this latter approach, employing Newmark integration for frictional contact, has also been formulated [Pandolfi *et al.* 2002]. However, because it treats friction as effectively an uncoupled damping force, it requires that the contacting system treated in this model have a closed form solution for the evaluation of contact Lagrange multipliers, α. This is an unlikely property that generally only holds for very simple systems that do not exhibit the coupling difficulties that are characteristic of many interesting frictional contact problems. In part to address this issue, the model in Pandolfi *et al.* [2002] was later reformulated using a penalty based approach [Pandolfi & Ortiz 2007].

Even without friction, appropriately integrating contact into the discrete variational setting still remains challenging. Stewart [2000] notes that, despite good conservative properties in the smooth setting, symplectic methods do not necessarily translate well for nonsmooth contact problems. In particular, he observes that symplectic methods like the midpoint rule and even the variational approach of Kane *et al.* [1999] loose their good (discrete) energy conserving properties when subject to contact.

In this chapter, we will show how a combined application of the coupled principles formulation (developed in the last chapter) and variational discretization techniques allows for the natural construction of variational frictional contact integrators that avoid many of the issues discussed above. We will first introduce useful discrete analogues of the friction sets, normal cones, and other elements introduced in earlier chapters. Next we will motivate our derivation by presenting a brief tutorial on variational integration methods (without contact). We will introduce our general construction strategy for deriving variational frictional contact integrators

from the coupled principles formulation and then, finally, focus on the particular example of deriving a symmetric variational integrator for frictional contact. Examples follow in Section 4.8 to demonstrate the performance, effectiveness, and good energy preservation properties of the derived integrators.

4.2 Discrete Contacting Systems

In the discrete setting, individual constraints describe non-penetration conditions, pairwise, between geometric primitives, such as boundary elements (e.g., edges, faces, algebraic surface representations). For each such constraint, g_k, with an active value at time t, i.e., $g_k(\mathsf{q}^t) \leq 0$, we can associatea discrete spatial contact location, $\boldsymbol{x}_k$, and tangent plane, $\boldsymbol{T}_k$. The corresponding discrete contacting system, at time t, is then described by the contact set, $\mathscr{C}_t$, which contains all such potentially active constraints.

4.2.1 Discrete Constraint Gradients, Cones and Sets

In the remaining part of this section we will describe some further, time-discrete analogues of useful geometric structures first introduced in Chapter 2.

Letting

$$\mathsf{n}_k \stackrel{\text{def}}{=} \nabla g_k, \tag{4.1}$$

for all $k \in \mathscr{C}_t$, the gradients of the contact set are concatenated to form a subspace of normal directions

$$\mathsf{N}(\mathsf{q}^t) = (\mathsf{n}_1, ..., \mathsf{n}_{|\mathscr{C}_t|}), \tag{4.2}$$

the corresponding discrete normal cone is then

$$\mathsf{C}(\mathsf{q}^t) = \{\mathsf{N}\mathsf{y} : \mathsf{y} \geq 0\}, \tag{4.3}$$

while the discrete tangent cone is

$$\mathsf{T}(\mathsf{q}^t) = \{\mathsf{x} : \mathsf{N}^T\mathsf{x} \geq 0\}. \tag{4.4}$$

Finally, the discrete friction subset is given by

$$\mathsf{F}(\alpha^t, \mathsf{q}^t) \stackrel{\text{def}}{=} \bigoplus_{k \in \mathscr{C}_t} \mathsf{F}_k(\alpha_k^t, \mathsf{q}^t). \tag{4.5}$$

4.2.2 Discrete Frictional-Contact Terms

In the following sections the contact terms, r, f, and c, *with superscripts* will be *exclusively* reserved to indicate discrete *impulse* terms. For example, time will be super-scripted using t, $t+1$, etc..., while intermediary predictor values will be indicated using the superscript, p.

4.3 Introduction to Variational Integrators

An alternative approach to creating numerical integration methods with good properties (symplectic, invariant preserving, etc...) is to begin with derivations that respect discrete versions of variational principles [Cadzow 1970; Logan 1973; Maeda 1981; Marsden & West 2001]. Mimicking the continuous derivation of the Lagrangian equations of motion we wish to extremize a discrete Lagrangian on the time interval $t \in [0,T]$ over all possible discrete paths of the form $\{\mathsf{q}^0,...,\mathsf{q}^k,...,\mathsf{q}^N\}$, where $N = T/h$. We can define a discrete Lagrangian,

$$L_d(\mathsf{q}^k,\mathsf{q}^{k+1},h) \approx \int_{kh}^{(k+1)h} L(\mathsf{q},\dot{\mathsf{q}})\ dt \tag{4.6}$$

as the quadrature on a finite interval of the continuous Lagrangian. The quadrature of the Lagrangian over the entire discrete path, called the *discrete action*, is then the sum of the discrete Lagrangians over all intervals:

$$\sum_{k=0}^{N-1} L_d(\mathsf{q}^k,\mathsf{q}^{k+1},h) \approx \int_0^{Nh} L(\mathsf{q},\dot{\mathsf{q}})\ dt. \tag{4.7}$$

Then the minimization of the discrete Lagrangian, over all discrete paths with fixed endpoints, is obtained by finding the variations of the discrete action,

$$\delta \sum_{k=0}^{N-1} L_d(\mathsf{q}^k,\mathsf{q}^{k+1},h).$$

This is computed as

$$\sum_{k=0}^{N-1} [\ D_1L_d(\mathsf{q}^k,\mathsf{q}^{k+1},h)\cdot\delta_k + D_2L_d(\mathsf{q}^k,\mathsf{q}^{k+1},h)\cdot\delta_{k+1}\ ],$$

where D_n is the partial derivative with respect to the n^{th} variable and δ_k is the variation of the k^{th} point on the discrete path. Noting that our endpoints are fixed, we have $\delta_0 = \delta_N = 0$. Then, rearranging our summation we have

$$\sum_{k=1}^{N-1} [\ D_2L_d(\mathsf{q}^{k-1},\mathsf{q}^k,h) + D_1L_d(\mathsf{q}^k,\mathsf{q}^{k+1},h)\ ]\cdot\delta_k.$$

Finally, by imposing zero action for all possible δ_k, we obtain our Discrete Euler-Lagrange equations [Marsden & West 2001]:

$$D_2L_d(\mathsf{q}^{k-1},\mathsf{q}^k,h)+D_1L_d(\mathsf{q}^k,\mathsf{q}^{k+1},h)=0. \tag{4.8}$$

Discrete Euler-Lagrange equations obtained in this way are called variational integrators. Given q^k, and q^{k-1} they advance our system to a new configuration, q^{k+1}. Thus velocity phase space, in the discrete domain, has been replaced by two copies of the configuration space.

In terms of accuracy, the order of the quadrature scheme we use to generate our discrete Lagrangian entirely specifies the order of our variational integrator [West 2004]. This means that we can derive variational integrators of any accuracy simply by selecting appropriate discrete Lagrangians.

4.3.1 An Example

As an example consider a physical system for which we have a generalized inertia tensor, M. Kinetic energy is then $K=\frac{1}{2}\,\dot{\mathsf{q}}^T\,\mathsf{M}\,\dot{\mathsf{q}}$, while the system's Hamiltonian will simply represent the system's total energy. A simple quadrature for a discrete Lagrangian is the generalized midpoint rule,

$$L_d(\mathsf{q}^k,\mathsf{q}^{k+1},h)=h\,L\!\left((1-\alpha)\mathsf{q}^k+\alpha\mathsf{q}^{k+1},\frac{\mathsf{q}^{k+1}-\mathsf{q}^k}{h}\right), \tag{4.9}$$

with $\alpha\in[0,1]$. A choice of $\alpha=\frac{1}{2}$ leads to a second order scheme while all other choices lead to first order schemes. With $\alpha=0$ we have

$$L_d=\frac{h}{2}\left(\frac{\mathsf{q}^{k+1}-\mathsf{q}^k}{h}\right)\,\mathsf{M}\,\left(\frac{\mathsf{q}^{k+1}-\mathsf{q}^k}{h}\right)-h\,V(\mathsf{q}^k), \tag{4.10}$$

which gives us

$$D_1L_d(\mathsf{q}^k,\mathsf{q}^{k+1},h)=-\mathsf{M}\left(\frac{\mathsf{q}^{k+1}-\mathsf{q}^k}{h}\right)-h\nabla V(\mathsf{q}^k) \tag{4.11}$$

and

$$D_2L_d(\mathsf{q}^k,\mathsf{q}^{k+1},h)=\mathsf{M}\left(\frac{\mathsf{q}^{k+1}-\mathsf{q}^k}{h}\right). \tag{4.12}$$

Substituting into Equation (4.8) our discrete variational integrator is

$$\mathsf{q}^{k+1}=\mathsf{q}^k+h\left(\frac{\mathsf{q}^k-\mathsf{q}^{k-1}}{h}\right)-h^2\,\mathsf{M}^{-1}\,\nabla V(\mathsf{q}^k),$$

which is just a first order symplectic Euler scheme [Hairer, Lubich, & Wanner 2002].

4.4 Variational Integration for Frictional Contact

Kane *et al.* [1999] first proposed applying nonsmooth discrete variational methods in the frictionless contact setting; however, this model formulates contact constraints in configuration space. This requires a nonconvex minimization solve just for resolving contact at each time-step and may be difficult to formulate for systems with nonlinear configurations such as are commonly found in multi-body and rigid-body mechanics. In this section we will show how to overcome these issues via an application of Moreau's Transform (introduced in Section 3.1.7). In particular, we will show that, in the discrete setting, this is simply equivalent to a local constraint linearization, but has important consequences for the resulting problem. We will also discuss the consistency problem for discrete variational friction and show how this too can be appropriately handled in the variational setting. We will then show how the application of these developments together with the coupled principles formulation, introduced in the last chapter, allows for the natural construction of variational frictional contact integrators. In this section we will first introduce the general construction strategy that this approach leads to. This method allows for the potential construction of a wide range of possible variational integrators suitable for frictionally contacting systems. Finally, to ground this section, we will focus on a particular (and potentially useful) example by constructing a symmetric variational integrator for frictional-contact.

4.4.1 Discretization Strategy

At a high level, our strategy for formulating discrete variational integrators for frictional contact begins with the coupled formulation from the last chapter. We extract two functions: a Lagragian for contact and dissipation rate function for friction, from Equations (3.2) and (3.19) respectively. We then generate a discretization of each. To do this, unlike in classical Discrete Variational methods, we also need to make the discrete friction impulses, discrete contact impulses *and* velocity (in the dissipation rate function) variables of the quadratures.

Once we have the quadratures defined, we then generate discrete functionals for each, as in the non-contacting case. Next we extremize each action independently to get a pair of time-discrete differential inclusions: a contact inclusion and a Friction inclusion. At the end of this

process our goal is to obtain a pair of minimizations that correspond to these optimality terms. To formulate the minimizations, however, there are several steps that need to be taken to ensure that the two inclusions are both consistent with one another and lead to convex minimizations.

First, we need to address the discrete velocity variable in the friction inclusion. The contact inclusion we obtain from the Discrete Action will always implicitly define a discrete velocity term. Our next step then is to enforce the discrete definition of velocity in the friction inclusion, so that it is kinematically consistent with the contact inclusion.

Second, the discrete contact inclusion will, as in the time-continuous case, have a nonconvex term, obtained from the indicator function on the nonconvex feasible set A. Here we will apply a discrete analogue of Moreau's Transform (Section 3.1.7) to regain the stable and consistent energetic behavior of classic variational integrators as well as to remove nonconvexity from the inclusion.

Finally, using the definition of the subdifferential we are able to reformulate the resulting inclusions as a pair of coupled, convex optimization problems.

In the next few sections we will cover these steps in more detail, before we apply them to an example discretization in Section 4.5.

4.4.2 Discrete Lagrangian and Action

Extrapolating from Equation (3.2) in Section 3.1, we can define the frictionally-contacting system's extended value Lagrangian using

$$L(\mathsf{q},\dot{\mathsf{q}}) = \frac{1}{2}\,\dot{\mathsf{q}}^T \mathsf{M}\dot{\mathsf{q}} - V(\mathsf{q}) - I_{\mathsf{A}}(\mathsf{q}) + \mathsf{f}^T\mathsf{q}. \tag{4.13}$$

Mirroring standard variational integration techniques, we then define a discrete Lagrangian map, L_d, using quadrature, so that

$$L_d(\mathsf{q}^k,\mathsf{q}^{k+1},\mathsf{f}^{k+1}) \cong \int_{kh}^{(k+1)h} L(\mathsf{q},\dot{\mathsf{q}})dt. \tag{4.14}$$

Likewise, we can construct a corresponding discrete analogue of the Hamiltonian Action using the sum

$$A_d = \sum_{k=0}^{N-1} L_d(\mathsf{q}^k,\mathsf{q}^{k+1},\mathsf{f}^{k+1}). \tag{4.15}$$

4.4.3 Extremizing the Discrete Action

Stationarity for A_d with respect to the discrete path, $(\mathsf{q}^0, ..., \mathsf{q}^N)$, is then given by the inclusion

$$D_1 L_d(\mathsf{q}^{t+1}, \mathsf{q}^{t+2}, \mathsf{f}^{t+2}) + D_2 L_d(\mathsf{q}^t, \mathsf{q}^{t+1}, \mathsf{f}^{t+1}) \ni 0, \tag{4.16}$$

Equation (4.16) in turn implies the discrete position update inclusion,

$$\mathsf{p}^t \in -D_1 L_d(\mathsf{q}^t, \mathsf{q}^{t+1}, \mathsf{f}^{t+1}), \tag{4.17}$$

and the discrete momentum update inclusion,

$$\mathsf{p}^{t+1} \in D_2 L_d(\mathsf{q}^t, \mathsf{q}^{t+1}, \mathsf{f}^{t+1}). \tag{4.18}$$

Here p^k is used to denote the discrete system's momentum at time k.

4.4.4 Discrete Dissipation Rate Function and Friction Functional

Similarly, extrapolating from Equation (3.19) in Section 3.2, the system's frictional dissipation rate function is then

$$FL(\mathsf{q}, \dot{\mathsf{q}}, \alpha, \mathsf{f}) = \dot{\mathsf{q}}^T \mathsf{f} - I_{\mathsf{F}(\alpha, \mathsf{q})}(\mathsf{f}). \tag{4.19}$$

Correspondingly, we can also define a discrete dissipation rate map, FL_d, using quadrature, so that

$$FL_d(\mathsf{q}^k, \dot{\mathsf{q}}^{k+1}, \alpha^{k+1}, \mathsf{f}^{k+1}) \cong \int_{kh}^{(k+1)h} FL(\mathsf{q}, \dot{\mathsf{q}}, \alpha, \mathsf{f}) dt. \tag{4.20}$$

Because the time step size, h, does not explicitly appear in this equation, it may not, at first, appear to be a proper quadrature; however, note that we have substituted the force, f, with an impulse, f^{t+1}, that integrates the friction force response over the time interval defined by the step size. Finally, a discrete analogue of the Friction Functional can be constructed using the sum

$$FA_d = \sum_{k=0}^{N-1} FL_d(\mathsf{q}^k, \dot{\mathsf{q}}^{k+1}, \alpha^{k+1}, \mathsf{f}^{k+1}). \tag{4.21}$$

4.4.5 Extremizing the Discrete Friction Functional

Stationarity for FA_d with respect to the discrete impulse arc, $(\mathsf{f}^0, ..., \mathsf{f}^N)$, is then given by the inclusion

$$D_4 FL_d(\mathsf{q}^t, \dot{\mathsf{q}}^{t+1}, \alpha^{t+1}, \mathsf{f}^{t+1}) \ni 0. \tag{4.22}$$

4.4.6 The General Recipe

With these preliminaries in place we can now describe, at a high level, our general strategy for generating coupled discrete variational integrators for frictionally contacting systems. Using the above derivations, we compose our numerical integrators using the set of discrete optimality conditions given by Equations (4.17), (4.18), and (4.22). We then use the first two inclusions to enforce consistency across variables by reducing $\dot{\mathsf{q}}^{t+1}$, p^{t+1} and q^{t+1} to a single unknown. Next, we substitute this unknown in place of $\dot{\mathsf{q}}^{t+1}$ in the friction inclusion to ensure kinematic consistency between inclusions. Finally, motivated by Section 3.1.7, we apply a discrete analogue of Moreau's Transform that allows us to convexify the discrete problem. In the next section we will go through this process in detail.

4.5 A Symmetric, Half-Step Integrator

As an important example, and to highlight many novel and non-trivial details required to derive these integrators, we discuss in detail the construction of a symmetric, half-step Integrator. The derivation of coupled, Newmark-type integrators, and other useful schemes for frictional contact integration follow from the application of similar steps.

4.5.1 Discrete Lagrangian, Dissipation Rate Function, and Inclusions

Discrete Lagrangian

We consider a discretization of the Lagrangian of the form

$$L_d(\mathsf{q}^k,\mathsf{q}^{k+1},\mathsf{f}^{k+1}) = hL\Big(\frac{\mathsf{q}^k+\mathsf{q}^{k+1}}{2},\frac{\mathsf{q}^{k+1}-\mathsf{q}^k}{h},\frac{\mathsf{f}^{k+1}}{h}\Big). \tag{4.23}$$

Evaluating we then obtain

$$\begin{aligned} L_d(\mathsf{q}^k,\mathsf{q}^{k+1},\mathsf{f}^{k+1}) = \frac{h}{2}\Big(\frac{\mathsf{q}^{k+1}-\mathsf{q}^k}{h}\Big)^T\mathsf{M}\Big(\frac{\mathsf{q}^{k+1}-\mathsf{q}^k}{h}\Big) - hV\Big(\frac{\mathsf{q}^k+\mathsf{q}^{k+1}}{2}\Big) \\ - I_{\mathsf{A}}\Big(\frac{\mathsf{q}^k+\mathsf{q}^{k+1}}{2}\Big) + \Big(\frac{\mathsf{q}^k+\mathsf{q}^{k+1}}{2}\Big)^T\mathsf{f}^{k+1}. \end{aligned} \tag{4.24}$$

Stationary Terms

Taking the generalized gradients of the discrete Lagrangian we obtain the first inclusion,

$$D_1 L_d(\mathsf{q}^k, \mathsf{q}^{k+1}\mathsf{f}^{k+1}) \in \mathsf{M}\Big(\frac{-\mathsf{q}^{k+1}+\mathsf{q}^k}{h}\Big) - \frac{h}{2}\nabla V\Big(\frac{\mathsf{q}^k+\mathsf{q}^{k+1}}{2}\Big) - \partial I_{\mathsf{A}}\Big(\frac{\mathsf{q}^k+\mathsf{q}^{k+1}}{2}\Big) + \frac{1}{2}\mathsf{f}^{k+1}, \tag{4.25}$$

and second inclusion,

$$D_2 L_d(\mathsf{q}^k, \mathsf{q}^{k+1}, \mathsf{f}^{k+1}) \in \mathsf{M}\Big(\frac{\mathsf{q}^{k+1}-\mathsf{q}^k}{h}\Big) - \frac{h}{2}\nabla V\Big(\frac{\mathsf{q}^k+\mathsf{q}^{k+1}}{2}\Big) - \partial I_{\mathsf{A}}\Big(\frac{\mathsf{q}^k+\mathsf{q}^{k+1}}{2}\Big) + \frac{1}{2}\mathsf{f}^{k+1}. \tag{4.26}$$

Momentum and Position Update Inclusions

Noting Equations (4.17) and (4.18), Equations (4.25) and (4.26) generate the position map,

$$\mathsf{p}^t \in \mathsf{M}\Big(\frac{\mathsf{q}^{t+1}-\mathsf{q}^t}{h}\Big) + \frac{h}{2}\nabla V\Big(\frac{\mathsf{q}^{t+1}+\mathsf{q}^t}{2}\Big) - \frac{1}{2}\mathsf{f}^{t+1} + \partial I_{\mathsf{A}}\Big(\frac{\mathsf{q}^{t+1}+\mathsf{q}^t}{2}\Big), \tag{4.27}$$

and the momentum map,

$$\mathsf{p}^{t+1} \in \mathsf{M}\Big(\frac{\mathsf{q}^{t+1}-\mathsf{q}^t}{h}\Big) - \frac{h}{2}\nabla V\Big(\frac{\mathsf{q}^{t+1}+\mathsf{q}^t}{2}\Big) + \frac{1}{2}\mathsf{f}^{t+1} - \partial I_{\mathsf{A}}\Big(\frac{\mathsf{q}^{t+1}+\mathsf{q}^t}{2}\Big). \tag{4.28}$$

Variational Contact Integrator

Subtracting Equation (4.27) from (4.28) then gives us the discrete variational contact integrator,

$$\mathsf{p}^{t+1} - \mathsf{p}^t + h\nabla V\Big(\frac{\mathsf{q}^{t+1}+\mathsf{q}^t}{2}\Big) - \mathsf{f}^{t+1} \in -\partial I_{\mathsf{A}}\Big(\frac{\mathsf{q}^{t+1}+\mathsf{q}^t}{2}\Big). \tag{4.29}$$

Discrete Dissipation Rate Function

We analogously discretize the frictional dissipation rate function using

$$FL_d(\mathsf{q}^k, \dot{\mathsf{q}}^{k+1}, \alpha^{k+1}, \mathsf{f}^{k+1}) = FL(\mathsf{q}^k, \dot{\mathsf{q}}^{k+1}, \alpha^{k+1}, \mathsf{f}^{k+1}), \tag{4.30}$$

so that we have

$$FL_d(\mathsf{q}^k, \dot{\mathsf{q}}^{k+1}, \alpha^{k+1}, \mathsf{f}^{k+1}) = (\dot{\mathsf{q}}^{k+1})^T \mathsf{f}^{k+1} + I_{\mathsf{F}(\alpha^{k+1}, \mathsf{q}^k)}(\mathsf{f}^{k+1}). \tag{4.31}$$

Variational Friction Integrator

Applying Equation (4.22) to Equation (4.31) then gives us the discrete variational friction integrator,

$$\dot{\mathsf{q}}^{t+1} + \partial I_{\mathsf{F}(\alpha^{t+1},\mathsf{q}^t)}(\mathsf{f}^{t+1}) \ni 0. \tag{4.32}$$

4.5.2 Coupled Variational Integrator System

The system of coupled inclusions given by Equations (4.29) and (4.32) now form the base for our variational integrator (given the above discrete Lagrangian and dissipation rate function). At the moment, however, there are too many uncorrelated unknowns. In particular, between the two inclusions, we currently have the unknowns: $\mathsf{q}^{t+1}, \mathsf{p}^{t+1}, \dot{\mathsf{q}}^{t+1}, \alpha^{t+1}$, and f^{t+1}. Given that we only have two inclusions, we need to reduce the number of unknowns as well.

Update Equation

Our next step is to add Equation (4.27) to (4.28) obtain the update equation,

$$\mathsf{p}^{t+1} = 2\mathsf{M}\left(\frac{\mathsf{q}^{t+1} - \mathsf{q}^t}{h}\right) - \mathsf{p}^t, \tag{4.33}$$

This equation describes the relationship between new positions and momenta and makes explicit how we perform a position update.

Primary Unknown

Motivated by the above equation, we then define a time-step scaled change in configuration,

$$\delta^{t+1} \stackrel{\text{def}}{=} \frac{\mathsf{q}^{t+1} - \mathsf{q}^t}{h}. \tag{4.34}$$

Note that now, using the update equation, we can define all unknowns, excepting friction, with respect to δ^{t+1},

$$\mathsf{q}^{t+1} = \mathsf{q}^t + h\delta^{t+1}, \tag{4.35}$$

$$\mathsf{p}^{t+1} = 2M\delta^{t+1} - \mathsf{p}^t, \tag{4.36}$$

and

$$\dot{\mathsf{q}}^{t+1} = 2\delta^{t+1} - \dot{\mathsf{q}}^t. \tag{4.37}$$

Coupled Integration System

Substituting these relationships into Equations (4.29) and (4.32) we then obtain the coupled inclusion system that forms the basis of the discrete variational integration method:

$$\mathsf{M}\delta^{t+1} - \mathsf{p}^t + \frac{h}{2}\nabla V\left(\mathsf{q}^t + \frac{h}{2}\delta^{t+1}\right) - \frac{1}{2}\mathsf{f}^{t+1} \in -\partial I_{\mathsf{A}}\left(\mathsf{q}^t + \frac{h}{2}\delta^{t+1}\right), \tag{4.38}$$

$$(2\delta^{t+1} - \dot{\mathsf{q}}^t) + \partial I_{\mathsf{F}_{(\alpha^{t+1},\mathsf{q}^t)}}(\mathsf{f}^{t+1}) \ni 0. \tag{4.39}$$

In its current form, the above system comprises a complete numerical integration scheme for frictionally contacting systems. There are several factors that remain to be addressed, however, to fully enforce good energetic behavior in the dissipative setting, as well as to reduce the computational burden of solving such systems. In the next section we will discuss some of these potential difficulties and then show how to formulate extensions of the above scheme that avoid them.

4.5.3 Configuration and Kinematic Constraints

Applying the definition of the generalized gradient [Clarke 1983] (see also Section 2.5.2) to Equation (4.38) we find that it is equivalent to the local, nonlinear minimization

$$\delta^{t+1} = \underset{\delta}{\operatorname{argmin}}\left\{\frac{1}{2}\delta^T\mathsf{M}\delta - \delta^T\left(\mathsf{p}^t + \frac{1}{2}\mathsf{f}^{t+1}\right) + V\left(\mathsf{q}^t + \frac{h}{2}\delta\right) : (\mathsf{q}^t + \frac{h}{2}\delta^{t+1}) \in \mathsf{A}\right\} \tag{4.40}$$

Leaving aside the interpretation of the objective for a moment, we see that the constraint imposed requires the midpoint position to lie in the feasible set. Ideally, of course, we'd like to constrain the final position, at the end of the time-step, to lie in the feasible set; this can be generated using, alternate, non-symmetric discrete Lagrangians (for instance a Newmark-type discretization).

In either case, however, constraining the final configuration to the feasible set can be computationally expensive for two reasons. First, to enforce position constraints, we need to precompute and then maintain, throughout the course of a simulation, constraint functions for all pairs of boundary simplices. While this is feasible for small meshings [Kane *et al.* 1999], this can very quickly become impractical for systems with large and/or complex boundary geometries. Efficient pruning strategies, however, could be devised to accelerate this process. We

expect that this could potentially be an exciting and profitable direction of future research to pursue. Second, presuming we do have an efficient strategy to formulate and update configuration constraints, the resulting constraint set will generally be nonconvex. Kane *et al.* [1999] give a simple motivational example of a particle in a nonconvex box. Even this trivial example results in a nonconvex constraint set. Thus, even with efficient, well formulated configuration constraints, a nonconvex optimization solver needs to be employed.

Moreau Transform

To redress energy behavior in the nonsmooth setting and to remove nonconvexity we invoke a discrete interpretation of Moreau's transform. Following our discussion in Section 3.1.7, we manipulate Equation (4.38) to form the new DI

$$\mathsf{M}\delta^{t+1} - \mathsf{p}^t + \frac{h}{2}\nabla V\left(\mathsf{q}^t + \frac{h}{2}\delta^{t+1}\right) - \frac{1}{2}\mathsf{f}^{t+1} \in -\partial I_{\mathsf{T}(\mathsf{q}^t)}\left(\delta^{t+1}\right). \tag{4.41}$$

The corresponding minimization (see Equation (4.50)) that we obtain from Equation (4.41), now simply enforces the convex constraint $\delta^{t+1} \in \mathsf{T}(\mathsf{q}^t)$. This requires a time scaled change in configuration to lie in the discrete tangent cone, and thus will be a *convex* problem, amenable to robust convex programming algorithms.

To understand the implications of applying Moreau's transform in the discrete context, consider that, in order for Equation 4.38 to hold, we must have $(\mathsf{q}^t + h\delta^{t+1}) \in \mathsf{A}$ (otherwise $\partial I_{\mathsf{A}}\left(\mathsf{q}^t + h\delta^{t+1}\right)$ is an empty set). Given this feasibility condition, we simply apply a first order taylor expansion for each constraint $k \in \mathscr{C}_t$, to get

$$\mathsf{g}_k(\mathsf{q}^t + h\delta^{t+1}) \approx \mathsf{g}_k(\mathsf{q}^t) + h\nabla\mathsf{g}_k(\mathsf{q}^t)^T\delta^{t+1} = h\nabla\mathsf{g}_k(\mathsf{q}^t)^T\delta^{t+1}. \tag{4.42}$$

Then, to satisfy these linearized constraints we need

$$\nabla\mathsf{g}_k(\mathsf{q}^t)^T\delta^{t+1} \geq 0; \tag{4.43}$$

or, concatenating all constraints in $\mathscr{C}_t$ together and recalling our earlier notation from Section 4.2.1,

$$\mathsf{N}(\mathsf{q}^t)^T\delta^{t+1} \geq 0. \tag{4.44}$$

In turn, we then must have

$$\delta^{t+1} \in \mathsf{T}(\mathsf{q}^t). \tag{4.45}$$

Applying Moreau's transform in this context is thus effectively a linearization of the local contact constraint set around q^t.

4.5.4 Fully Implicit Friction Inclusion

To solve the coupled system it will generally be useful to formulate the friction inclusion in a fully implicit form. We find that this generally helps to stabilize potential solution methods.

We first introduce an evaluated acceleration notation [Kane *et al.* 2000] where we let

$$\mathsf{a}^{t+1} \stackrel{\text{def}}{=} \mathsf{M}^{-1}\nabla V\left(\mathsf{q}^t + \frac{h}{2}\delta^{t+1}\right). \tag{4.46}$$

Then, substituting into Equation (4.29), we obtain

$$\mathsf{p}^{t+1} = \mathsf{p}^t + h\mathsf{M}\mathsf{a}^{t+1} + \mathsf{f}^{t+1} + \mathsf{c}^{t+1}. \tag{4.47}$$

This allows us to equivalently describe the discrete velocity as

$$\dot{\mathsf{q}}^{t+1} = \dot{\mathsf{q}}^t + h\mathsf{a}^{t+1} + \mathsf{M}^{-1}\mathsf{f}^{t+1} + \mathsf{M}^{-1}\mathsf{c}^{t+1}. \tag{4.48}$$

Substituting this discrete velocity equation into the discrete friction optimality terms (Equation (4.32)), then gives us an alternate form of the friction inclusion

$$(\dot{\mathsf{q}}^t + h\mathsf{a}^{t+1} + \mathsf{M}^{-1}\mathsf{f}^{t+1} + \mathsf{M}^{-1}\mathsf{c}^{t+1}) + \partial I_{\mathsf{F}(\alpha^{t+1},\mathsf{q}^t)}(\mathsf{f}^{t+1}) \ni 0. \tag{4.49}$$

4.5.5 Coupled Minimization Solution

We now can formulate a single step of the integration using the coupled minimizations

$$\delta^{t+1} = \underset{\delta}{\operatorname{argmin}}\left\{\frac{1}{2}\delta^T\mathsf{M}\delta - \delta^T\left(\mathsf{p}^t + \frac{1}{2}\mathsf{f}^{t+1}\right) + V\left(\mathsf{q}^t + \frac{h}{2}\delta\right) : \mathsf{N}(\mathsf{q}^t)^T\delta \geq 0\right\} \tag{4.50}$$

and

$$\mathsf{f}^{t+1} = \underset{\mathsf{f}}{\operatorname{argmin}}\left\{\mathsf{f}^T\left(2\delta^{t+1} - \dot{\mathsf{q}}^t\right) : \mathsf{f} \in \mathsf{F}(\alpha^{t+1}, \mathsf{q}^t)\right\}. \tag{4.51}$$

Or, alternately, we can replace Equation (4.51), with an equivalent, implicit minimization form,

$$\mathsf{f}^{t+1} = \underset{\mathsf{f}}{\operatorname{argmin}} \Big\{ \frac{1}{2} \mathsf{f}^T \mathsf{M}^{-1} \mathsf{f} + \mathsf{f}^T \Big(\dot{\mathsf{q}}^t + h \mathsf{a}^{t+1} + \mathsf{M}^{-1} \mathsf{c}^{t+1} \Big) : \mathsf{f} \in \mathsf{F}(\alpha^{t+1}, \mathsf{q}^t) \Big\}, \tag{4.52}$$

obtained from Equation (4.49). While equivalent, this latter form exposes properties that allow us to obtain implicit stationary point relations, useful for the solution of these systems. Either way, once a solution to the coupled minimization system is obtained, the new state, at time $t+1$, is given by

$$\mathsf{q}^{t+1} = \mathsf{q}^t + h\delta^{t+1}, \tag{4.53}$$

and

$$\mathsf{p}^{t+1} = 2\mathsf{M}\delta^{t+1} - \mathsf{p}^t. \tag{4.54}$$

4.6 A Velocity-Level Interpretation of the Contact Integrator

In the following it will be useful to consider an equivalent, velocity-level interpretation of the contact differential inclusion.

Starting again with Equation (4.29) we see that this is, in turn, equivalent to

$$\mathsf{p}^{t+1} - \mathsf{p}^t - \mathsf{f}^{t+1} + h\nabla V \Big(\frac{1}{2} \big(\mathsf{q}^t + \frac{h}{2} \dot{\mathsf{q}}^t \big) + \frac{h}{4} \dot{\mathsf{q}}^{t+1} \Big) + \partial I_{\mathsf{A}} \Big(\frac{1}{2} \big(\mathsf{q}^t + \frac{h}{2} \dot{\mathsf{q}}^t \big) + \frac{h}{4} \dot{\mathsf{q}}^{t+1} \Big) \ni 0. \tag{4.55}$$

Then applying the discrete Moreau Transform (Section 3.1.7) to replace the generalized gradient,

$$\partial I_{\mathsf{A}_t} \Big(\frac{1}{2} \big(\mathsf{q}^t + \frac{h}{2} \dot{\mathsf{q}}^t \big) + \frac{h}{4} \dot{\mathsf{q}}^{t+1} \Big), \tag{4.56}$$

by

$$\partial I_{\mathsf{T}\left(\frac{1}{2} \left(\mathsf{q}^t + \frac{h}{2} \dot{\mathsf{q}}^t \right) \right)} \big(\dot{\mathsf{q}}^{t+1} \big), \tag{4.57}$$

and defining a predictor position as

$$\mathsf{q}^p = \frac{1}{2} \big(\mathsf{q}^t + \frac{h}{2} \dot{\mathsf{q}}^t \big), \tag{4.58}$$

we obtain

$$\mathsf{p}^{t+1} - \mathsf{p}^t - \mathsf{f}^{t+1} + h\nabla V \Big(\mathsf{q}^p + \frac{h}{4} \dot{\mathsf{q}}^{t+1} \Big) + \partial I_{\mathsf{T}(\mathsf{q}^p)} \big(\dot{\mathsf{q}}^{t+1} \big) \ni 0. \tag{4.59}$$

Noting that $\dot{\mathsf{q}}^{t+1}$ (or equivalently p^{t+1}) is the unknown in the above DI, we then obtain the equivalent convex, velocity-level minimization problem

$$\dot{\mathsf{q}}^{t+1} = \underset{\mathsf{v}}{\operatorname{argmin}} \left\{ \mathsf{v}^T \mathsf{M} \mathsf{v} - \mathsf{v}^T \left(\mathsf{f}^{t+1} + \mathsf{p}^t \right) + 4V \left(\mathsf{q}^p + \frac{h}{4} \mathsf{v} \right) : \mathsf{N}(\mathsf{q}^p)^T \mathsf{v} \geq 0 \right\}. \tag{4.60}$$

4.7 A Two-step Variant

Kane *et al.* [2000] suggest an efficient approach for approximating the solution of complex dissipative systems. The general strategy is to first obtain a predictor state by integrating forward only the conservative portion of the dissipative system. Then one applies a corrector step that solves the dissipative portion of the system based on the predictor estimates. In this spirit we briefly investigate below how this strategy can be applied to approximate the symmetric integrator from Section 4.5.5. The resulting predictor-corrector minimization system will be discussed in detail in the following two chapters where it will be used as a practical integration method (see Section 5.4.5).

We first note that if we presume that the evaluated acceleration, a^{t+1}, is known a priori then, following the same steps as in Section 4.6, we obtain an implicit form of the velocity-level minimization

$$\dot{\mathsf{q}}^{t+1} = \underset{\mathsf{v}}{\operatorname{argmin}} \left\{ \mathsf{v}^T \mathsf{M} \mathsf{v} - \mathsf{v}^T \left(\mathsf{f}^{t+1} + h\mathsf{M}\mathsf{a}^{t+1} + \mathsf{p}^t \right) : \mathsf{N}(\mathsf{q}^p)^T \mathsf{v} \geq 0 \right\}. \tag{4.61}$$

A semi-implicit-type discretization then suggests itself in which we replace a^{t+1} with an evaluated acceleration, a^p, computed from the predictor configuration estimate. This in turn generates an associated predictor velocity, $\dot{\mathsf{q}}^p$. A predictor step is the given by

$$\mathsf{q}^p = \frac{1}{2} \left(\mathsf{q}^t + \frac{h}{2} \dot{\mathsf{q}}^t \right), \tag{4.62}$$

$$\mathsf{a}^p = -\mathsf{M}^{-1} \nabla V \left(\mathsf{q}^p \right), \tag{4.63}$$

$$\dot{\mathsf{q}}^p = h\mathsf{a}^p + \dot{\mathsf{q}}^t. \tag{4.64}$$

The above implicit velocity-level contact optimization, (4.61), and the friction optimization, (4.52), then reduce to a pair of convex Quadratic Programs (QP), that form the corrector step

system

$$\dot{\mathsf{q}}^{t+1} = \underset{\mathsf{v}}{\operatorname{argmin}} \Big\{ \mathsf{v}^T \mathsf{M} \mathsf{v} - \mathsf{v}^T \left(\mathsf{f}^{t+1} + \mathsf{M} \dot{\mathsf{q}}^p \right) : \mathsf{N}(\mathsf{q}^p)^T \mathsf{v} \geq 0 \Big\}, \tag{4.65}$$

$$\mathsf{f}^{t+1} = \underset{\mathsf{f}}{\operatorname{argmin}} \Big\{ \frac{1}{2} \mathsf{f}^T \mathsf{M}^{-1} \mathsf{f} + \mathsf{f}^T \left(\mathsf{M}^{-1} \mathsf{c}^{t+1} + \dot{\mathsf{q}}^p \right) : \mathsf{f} \in \mathsf{F}(\alpha^{t+1}, \mathsf{q}^t) \Big\}. \tag{4.66}$$

The final configuration is then updated by applying

$$\mathsf{q}^{t+1} = \mathsf{q}^t + \frac{h}{2} (\dot{\mathsf{q}}^{t+1} + \dot{\mathsf{q}}^t). \tag{4.67}$$

4.7.1 Dual Two-Step Variational Formulation

In the following Chapters it will often be useful to invoke the Lagrangian dual form of the above corrector system. Because Equation (4.65) is convex, we can also solve the problem using its Lagrangian dual formulation [Boyd & Vandenberghe 2004]. Additionally, recalling from Section 2.4.2, that the friction impulse is constrained by $\mathsf{f}^{t+1} = \mathsf{D}\beta$, we can also reformulate Equation (4.66) by substitution. Applying these two observations together we then obtain an equivalent dual Predictor-Corrector optimization system, in which we directly minimize over discrete friction and contact impulse magnitudes,

$$\alpha^{t+1} = \underset{\alpha}{\operatorname{argmin}} \Big\{ \frac{1}{2} \alpha^T \mathsf{N}^T \mathsf{M}^{-1} \mathsf{N} \alpha + \alpha^T \mathsf{N}^T (\mathsf{M}^{-1} \mathsf{D} \beta^{t+1} + \dot{\mathsf{q}}^p) : \ \alpha \geq 0 \Big\}, \tag{4.68}$$

$$\beta^{t+1} = \underset{\beta}{\operatorname{argmin}} \Big\{ \frac{1}{2} \beta^T \mathsf{D}^T \mathsf{M}^{-1} \mathsf{D} \beta + \beta^T \mathsf{D}^T (\mathsf{M}^{-1} \mathsf{N} \alpha^{t+1} + \dot{\mathsf{q}}^p) : \breve{\mathsf{F}}(\alpha^{t+1}, \beta) \geq 0 \Big\}. \tag{4.69}$$

The final velocity is then obtained using

$$\dot{\mathsf{q}}^{t+1} = \dot{\mathsf{q}}^p + \mathsf{M}^{-1} \mathsf{D} \beta^{t+1} + \mathsf{M}^{-1} \mathsf{N} \alpha^{t+1}, \tag{4.70}$$

before being fed to the position update, (4.67).

4.8 Examples

4.8.1 One Degree of Freedom Particle System

We begin by examining a simple frictionless example of a one dimensional particle with mass, m, configuration, $\mathsf{q} \in \mathbb{R}$, and a simple contact constraint $\mathsf{g}(q) = \mathsf{q} \geq 0$. Letting g be the magnitude of gravity, the potential for this example is just $V(\mathsf{q}) = g\mathsf{q}$. Stewart [2000] uses this

simple example to illustrate potential issues involved in applying symplectic integration methods to contacting systems. Poor energy conservation behaviors are observed for this case by the author for both the (symplectic) implicit midpoint rule [Hairer, Lubich, & Wanner 2002] and the variational, contact Newmark scheme from Kane *et al.* [1999]. See Figures 4.1(a) and 4.1(b) for these results. In particular, Stewart [2000] notes that the energy of these systems is not conserved and that the effective coefficient of restitution varies with state for both of these integrators.

In Figures 4.1(c) and 4.1(d) we show the results of applying the variational contact integrator, developed above and given in Equation (4.50), to the same particle system. We show the trajectories obtained in Figure 4.1(d), for step sizes of $h = 10^{-1}$ (red circle), 10^{-2} (green circle), and 10^{-3} (black line). For all three of these trajectories, energy is closely conserved. We compare this with the trajectories given by the implicit midpoint rule in Figure 4.1(a) (at $h = 10^{-3}$) and contact Newmark in Figure 4.1(b) (at $h = 10^{-2}$). We also show, in Figure 4.1(c), that the effective coefficient of restitution remains stable over perturbations of initial conditions and also compare this with the test in Figure 4.1(b).

4.8.2 Two Degree of Freedom Nonlinear Oscillator

To examine both nontrivial potential energies and friction, we next consider the example of a two-degree of freedom, non-linear oscillator. We start with a particle in the plane, with mass, m, and configuration, $\mathsf{q} \in \mathbb{R}^2$. The non-linear potential for this system is given by

$$V(\mathsf{q}) = \| \mathsf{q} \|^2 (\| \mathsf{q} \|^2 - 1)^2. \tag{4.71}$$

Contact Free Integration

To get a sense of the non-contacting behavior of our proposed integrator, in this example, we first apply the variational contact integrator, from Equation (4.50) to a contact-free system and observe that we continue to obtain consistently good (approximate) energy conserving behavior (see Figure 4.2 (a)). Note that, in particular, we obtain the characteristic fluctuating behavior of energy, typical of many variational (and symplectic) methods in general; while energy varies between time-steps, the integrator maintains the initial energy of the system, at any point, up

to an additive constant [Hairer, Lubich, & Wanner 2002]. For reference, consider Kane *et al.* [2000] where this same example system is simulated in a contact free setting using a variety of high-order and symplectic methods.

Frictionless Contact

We then add a simple inequality constraint, $\mathsf{g}(\mathsf{q}) = \mathsf{q}_2 \geq 0$, restricting the particle to the upper half-plane, and again apply our variational contact integrator to this system. Similar energy conserving properties, as noted above for the contact-free system, are observed here as well; however, in this example, the energy fluctuations of the plot are no longer symmetric, due to the ordering of collision events. See Figure 4.2 (b).

Frictional Contact

Finally, we apply our coupled variational integrator, given in Equations (4.50) and (4.52), to simulate the frictional contact between the particle and constraint surface. In the context of dissipation, the desired attributes of a variational integrator incorporating dissipation, are a bit less clear [Kane *et al.* 2000]; however, as with the conserving case, we expect that the (dissipative) energy behavior at contacts should be accurately mirrored at all reasonable step-sizes, just as energy conservation should be maintained in free flight. We examine the behavior of our contacting system by starting the mass particle at $\mathsf{q} = (4,4)^T$, and then simulating the contacting system for three time-step sizes: $h = 10^{-2}$ (red line), $h = 5 \times 10^{-3}$ (green line), and $h = 10^{-3}$ (black line) and two coefficients of friction: $\mu = 0.5$ and $\mu = 5$.

In each example we observe a period of strong tangential dissipation, followed by a leveling off of the angle of energy decay, as the motion of the oscillator moves into a mode that is mostly normal to the contact surface. See Figures 4.2 (c) and (d). We also note that the behavior is dissipative during tangential contact and conservative for all motion sequences corresponding to free flight or purely normal contact.

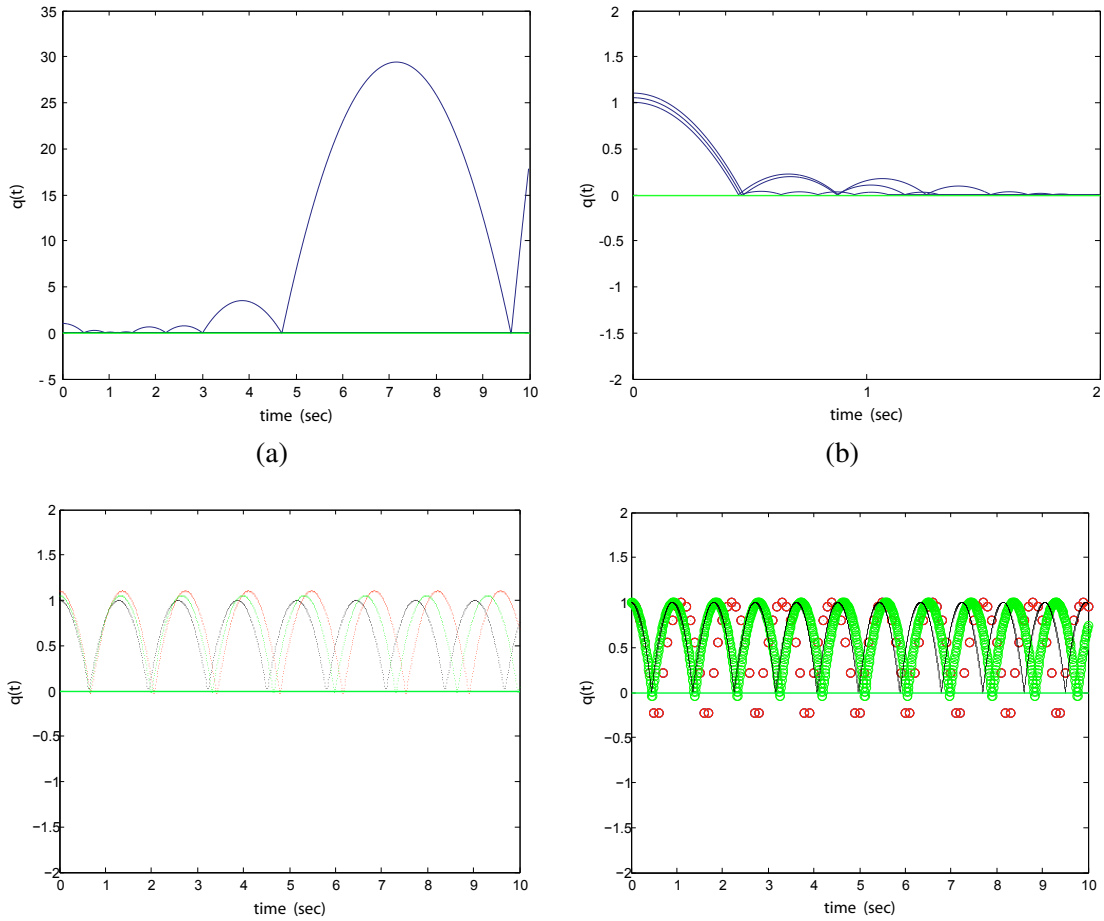


Figure 4.1: **Contacting Particle Example:** In these figures we examine a single point mass example in a one dimensional system subject to gravity and a single inequality (contact) constraint. In **(a)** the trajectory of the (symplectic) implicit midpoint rule is plotted (time-step: $h = 10^{-2}$). In **(b)** the trajectory of the contact Newmark scheme [Kane *et al.* 1999] is plotted (time-step: $h = 10^{-3}$) for three close initial conditions. In **(c)** we show the results of the same test of three close initial conditions using our proposed variational contact integrator (EQ (4.50)) and contrast its consistent behavior with the indeterminant changes of these other two symplectic approaches. In **(d)** we show the results for our variational contact integrator over a range of step sizes: $h = 10^{-1}$ (red circle), $h = 10^{-2}$ (green circle), and $h = 10^{-3}$ (black line). We again note the consistent, energy conserving behavior obtained. Plots (a) and (b) duplicate results from Stewart [2000].

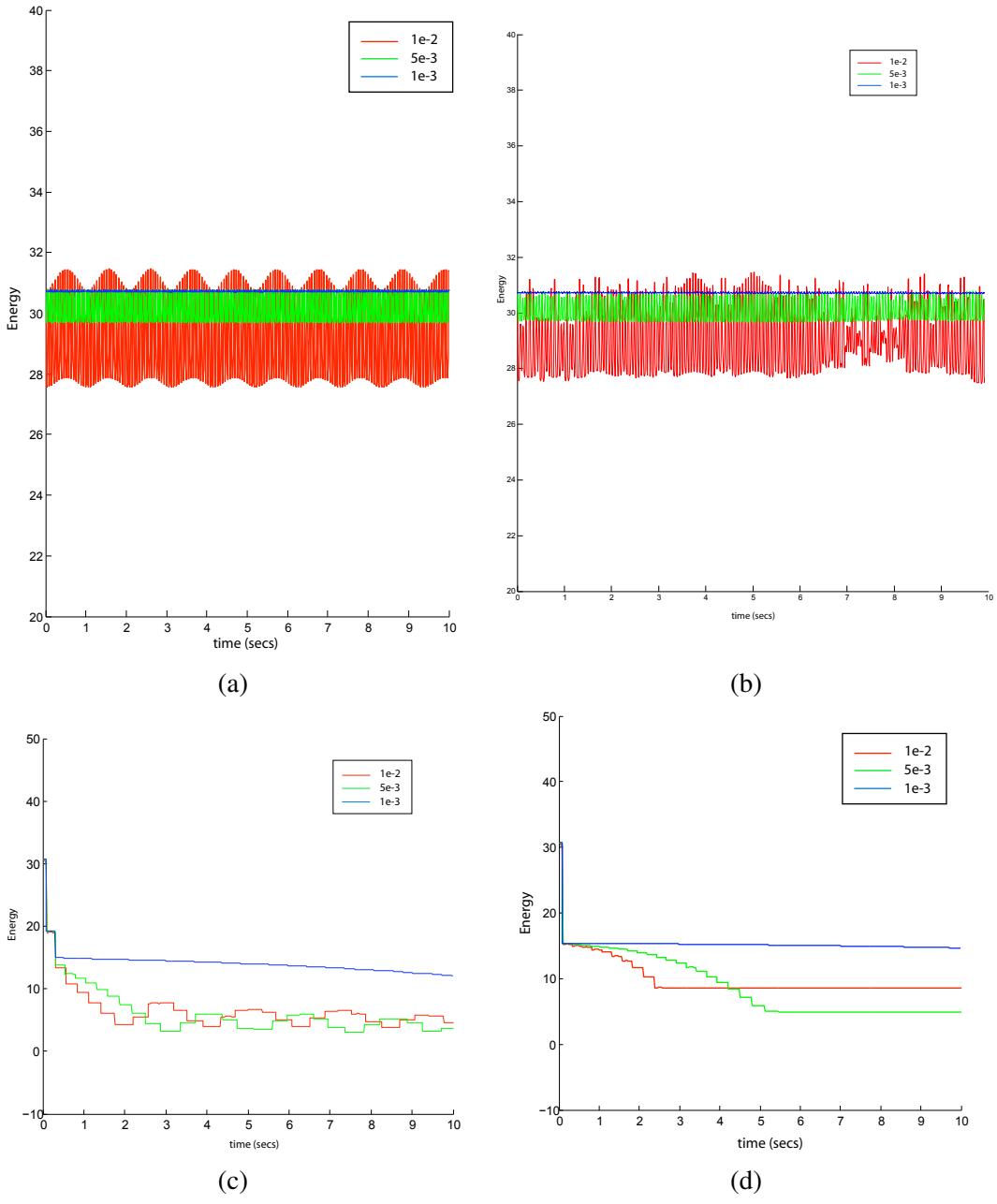


Figure 4.2: **Non-linear Oscillator Example:** In these figures we examine the behavior of our proposed coupled variational integrator on a non-linear oscillator example in two dimensions. In **(a)** we examine the energy behavior for the non-contacting case. In **(b)** the energy behavior of the frictionless contact case is plotted. In **(c)** we look at contact *and* friction ($\mu = 0.5$). In **(d)** we look at a case of higher friction ($\mu = 5$). We again note the consistent, energy conserving behavior for the frictionless examples, as well as accurate dissipative behavior for the examples with friction.

Chapter 5

The Two-Step Variational Method and Staggered Projections

5.1 Overview

While the generalized discrete variational framework, introduced in the last chapter, is a promising and ongoing area of research, the specific case of the Two-Step Variational integrator, introduced in Section 4.7, is of particular interest both because of its efficiency and, as we will show in Section 5.5.4, due to its close relationship to the popular Stewart-Trinkle Linear Complementarity Programming (LCP) formulation for frictional contact [Stewart & Trinkle 1996].

Formulations based on Linear Complementarity Programming are standard for frictional contact problems involving linear systems. Initially introduced for contact simulation by Lötstedt, newer variants, such as the popular Stewart-Trinkle formulation, now guarantee solution existence; however, solving these optimization problems accurately and efficiently has remained difficult. We have found that that, in practice, standard optimization methods such as Lemke's algorithm, Projected Gauss-Seidel, and interior point methods, that have generally been presumed suitable for solving these contact-related optimization problems,fail entirely for many important examples of frictional contact.

In this chapter we will show that these difficulties can be explained, in part, by noting that optimization problems, obtained from frictional contact, are equivalent to nonconvex problems in global optimization (which are generally NP-hard). We believe that this is the first time that a complexity argument has been made via the connection to global minimization for the Stewart-Trinkle model.

We address these limitations with an efficient and effective solution strategy, the Staggered Projections algorithm. Applying a fixed point scheme, derived from the coupled principles formulation, Staggered Projections efficiently obtains accurate solutions to optimization problems

for many frictional contact problems that were previously impractical to solve. Staggered Projections performs well in practice as evidenced by both our simulations of complex multibody frictional contact of rigid and large deformation models (see Chapter 7), and our extensive comparison of a Staggered Projections implementation with other available solvers (Sections 5.7, 7.5, and Chapter 6).

In this chapter, for completeness, we will first present an alternate derivation of the Two-Step method [Kaufman *et al.* 2008] that avoids the inclusion based formulation presented in the last chapter and applies the usual linearization of the Coulomb Constraint discussed in Section 2.4.4. We will then focus on how the Two-Step method naturally suggests the Staggered Projections Algorithm and discuss how it generalizes the Stewart-Trinkle formulation. We will then analyze the difficulties involved in solving the resulting optimization problems posed by these formulations. Finally, we will individually discuss specific existing solution methods and potential issues unique to each. In the following chapter we will focus further on the Staggered Projections algorithm by analyzing stability and convergence behaviors.

5.2 Two-Step, Predictor-Corrector Derivation

The two-step algorithm for frictional contact can be directly obtained using a predictor-corrector discretization. We first define a velocity-level, predictor-corrector form of the discrete Euler-Lagrange equations (DEL) of motion for integrating the contacting system from time t to $t+1$, with a step-size of h,

$$\mathsf{M}\left(\dot{\mathsf{q}}^{t+1}-\dot{\mathsf{q}}^{t}\right)=h\mathsf{g}(\mathsf{q}^{p},\dot{\mathsf{q}}^{p})+h\mathsf{f}_{ext}^{p}+\mathsf{r}^{t+1} \tag{5.1}$$

Here f_{ext} is a specified external force, r^{t+1} is the frictional contact response impulse, and the generalized force is

$$\mathsf{g}(\mathsf{q},\dot{\mathsf{q}}) \;=\; -\nabla V(\mathsf{q})+\left(-\dot{\mathsf{M}}\dot{\mathsf{q}}+\nabla_{\mathsf{q}}K(\mathsf{q},\dot{\mathsf{q}})\right) \;=\; -\nabla V(\mathsf{q})+\mathsf{g}_{qv}, \tag{5.2}$$

where V and K are the potential and kinetic energies of the system respectively, and g_{qv} is the *quadratic velocity vector* that provides Coriolis and centrifugal forces in multibody systems Shabana [2005].

5.2.1 Predictor

Depending on the physical models in use, an appropriate integration method is selected based on stability requirements of the underlying conservative system and modeling needs. We then use the selected method to integrate *only* the terms indexed by p in (5.1). This amounts to solving the system

$$\mathsf{M}\left(\dot{\mathsf{q}}^{p}-\dot{\mathsf{q}}^{t}\right)=h\mathsf{g}(\mathsf{q}^{p},\dot{\mathsf{q}}^{p})+h\mathsf{f}_{ext}^{p} \tag{5.3}$$

and generates a predictor velocity, $\dot{\mathsf{q}}^{p}$.

5.2.2 Corrector

Then, by subtracting (5.3) from (5.1), the remaining unknown, r^{t+1}, is determined by

$$\mathsf{M}\left(\dot{\mathsf{q}}^{t+1}-\dot{\mathsf{q}}^{p}\right)=\mathsf{r}^{t+1}. \tag{5.4}$$

5.2.3 Discrete Signorini-Fichera Condition

We then apply an impulse-level implicit discretization of the *Signorini-Fichera Condition* [Signorini 1933; Fichera 1964]. The Signorini-Fichera Condition directly enforces nonlinear boundary conditions between final velocities at contact points and the corrective contact impulses applied at those points. In this case we have

$$0\leq\mathsf{N}^{T}\dot{\mathsf{q}}^{t+1}\perp\alpha^{t+1}\geq 0. \tag{5.5}$$

The left-hand side of the above complementarity condition is simply the contact constraint inequality [1] that requires velocities to be positive along contact normals, while the right-hand side correspondingly requires all contact impulses to be non-negative along the normals; i.e., no adhesive impulses are allowed. Note that, as discussed above, the full contact impulse is given by $\mathsf{c}^{t+1}=\mathsf{N}\alpha^{t+1}$. Finally, the complementarity condition itself ensures that contact impulses are only applied at points where the body *is not* moving away from the contacting surface.

[1] For rigid body systems we can also include a Newtonian restitution model by modifying the contact constraints to $\mathsf{n}_{k}^{T}\left(\dot{\mathsf{q}}^{t+1}+e\dot{\mathsf{q}}^{t}\right)\geq 0$. Here the coefficient of restitution is given by $e\in[0,1]$.

5.2.4 Maximal Dissipation Principle

Letting $\boldsymbol{f}_k$ denote a frictional force applied at a sliding contact point, $\boldsymbol{x}_k$, we recall that the Maximal Dissipation Principle [Moreau 1973] requires that friction maximize the rate of negative work at the contact, given by $-\boldsymbol{f}_k^T \dot{\boldsymbol{x}}_k$. Combined with a suitable friction constraint, as described in Section 2.4, this simply provides a variational interpretation of the more familiar Coulomb friction.

To extend the Maximal Dissipation Principle to generalized coordinates we enforce it simultaneously at *all* contact points to obtain

$$\max_{\boldsymbol{f}_k} \sum_{k \in \mathscr{C}} \left(-\boldsymbol{f}_k^T \dot{\boldsymbol{x}}_k \right) = \min_{\boldsymbol{f}_k} \Big[\sum_{k \in \mathscr{C}} \boldsymbol{f}_k^T \Gamma_k \Big] \dot{\mathsf{q}} = \min_{\mathsf{f}} \mathsf{f}^T \dot{\mathsf{q}}. \tag{5.6}$$

This, in turn, requires that the total generalized friction impulse is given by $\mathsf{f} = \sum_{k \in \mathscr{C}} \Gamma_k^T \boldsymbol{f}_k$.

Then, using the above identity and remembering our earlier definition of the friction impulse magnitude vector, $\boldsymbol{\beta}$, from Section 2.4.2, the generalized friction impulse at each contact is $\mathsf{f}_k = \Gamma_k^T \ \boldsymbol{T}_k \boldsymbol{\beta}_k$. Note that, by construction, friction now appropriately applies an equal and opposite impulse at each point of contact.

A generalized basis for friction impulses at contact k is then given by

$$\mathsf{D}_k \stackrel{\text{def}}{=} \Gamma_k^T \ \boldsymbol{T}_k, \tag{5.7}$$

and the corresponding subspace of all generalized friction impulses is

$$\mathsf{D} = (\mathsf{D}_1 ... \mathsf{D}_{|\mathscr{C}|}). \tag{5.8}$$

Finally, by applying an implicit discretization to (5.6), and using the linearized Coulomb constraint from Section 2.4.4, we obtain a discrete generalized maximal dissipation principle,

$$\boldsymbol{\beta}^{t+1} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \Big(\boldsymbol{\beta}^T \mathsf{D}^T \dot{\mathsf{q}}^{t+1} \ \ : \ \ E^T \boldsymbol{\beta} \leq \operatorname{diag}(\mu) \boldsymbol{\alpha}^{t+1}, \ \ \boldsymbol{\beta} \geq 0 \Big). \tag{5.9}$$

5.2.5 Discrete Hamilton's Principle

In light of the above discussion, the friction impulse at time $t+1$, is given by $\mathsf{D}\boldsymbol{\beta}^{t+1}$, while the contact impulse is $\mathsf{N}\boldsymbol{\alpha}^{t+1}$, so that

The full frictional contact response impulse is now given by [2]

$$\mathsf{r}^{t+1} = \mathsf{c}^{t+1} + \mathsf{f}^{t+1} = \mathsf{N}\alpha^{t+1} + \mathsf{D}\beta^{t+1}. \tag{5.10}$$

Equation (5.10) together with (5.4) and (5.5) now form the Karush-Kuhn-Tucker (KKT) conditions for a second minimization [Moreau 1966; Boyd & Vandenberghe 2004]:

$$\dot{\mathsf{q}}^{t+1} = \underset{\mathsf{u}}{\operatorname{argmin}}\Big(\frac{1}{2}\mathsf{u}^T\mathsf{M}\mathsf{u} - \mathsf{u}^T\left(\mathsf{M}\dot{\mathsf{q}}^p + \mathsf{D}\beta^{t+1}\right) \;:\; \mathsf{N}^T\mathsf{u} \geq 0\Big). \tag{5.11}$$

This minimization implicitly solves for the contact impulse, c^{t+1}, and, when combined with (5.1), can be interpreted (as we saw in Chapter 4) as a discrete analog of Hamilton's Stationarity Principle.

5.3 The Staggered Projections Algorithm

To solve the corrector step for the unknown frictional-contact response r^{t+1} we employ projections in impulse space, using the kinetic metric. We define these projections, so that, given an arbitrary convex set S, the projection of some v onto S is given by

$$\mathsf{P}_\mathsf{S}(\mathsf{v}) \stackrel{\text{def}}{=} \underset{\mathsf{u}\in\mathsf{S}}{\operatorname{argmin}}\ (\mathsf{u}-\mathsf{v})^T\mathsf{M}^{-1}(\mathsf{u}-\mathsf{v}). \tag{5.12}$$

5.3.1 Projective System

If the generalized velocity, $\dot{\mathsf{q}}$, has n DoFs, the set of possible normal impulse directions, given by N, forms a polyhedral cone in $\mathbb{R}^n$, while the linearized Coulomb constraint, (2.10), similarly defines a more complex scaled, convex subset of $\mathbb{R}^n$.

More rigorously we can define the polyhedral cone of all possible contact impulses as

$$\mathsf{C} \stackrel{\text{def}}{=} \{\mathsf{N}\alpha : \alpha \geq 0\}, \tag{5.13}$$

the polyhedral, scaled, convex set of possible friction impulses as

$$\mathsf{F}(\alpha) \stackrel{\text{def}}{=} \{\mathsf{D}\beta : E^T\beta \leq \operatorname{diag}(\mu)\alpha,\ \beta \geq 0\}, \tag{5.14}$$

[2] Note that because r^{t+1}, by construction, only opposes relative sliding and relative normal velocities, it will conserve the linear and angular momentum of the contacting system it is applied to.

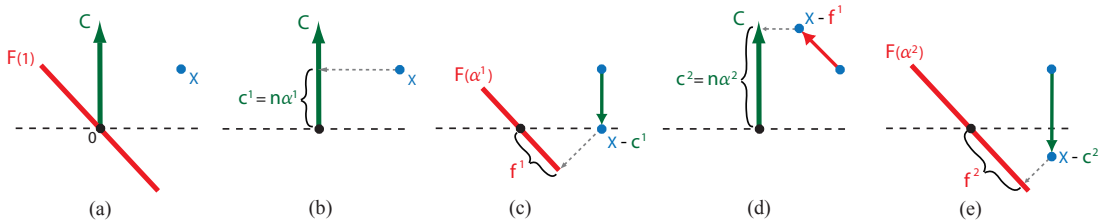


Figure 5.1: **Staggered Projection Sequence:** A simplified example of a few steps in a possible staggered sequence is shown above. **(a)** We start the staggered sequence with the negative momentum predicted by (5.3) and depicted above by the point x. We also start with the set of all possible contact normal directions, in this case composed of only a single vector, and depicted above as the green arrow C. Finally, we also initialize the unscaled friction set, F, depicted above as a red line segment. Note that, in this example, we show the likely case where the friction set and the contact normal directions are not orthogonal. **(b)** Starting with $\mathsf{f}^0 = 0$, for the initial (warm start) estimate of the friction impulse, we project $(\mathsf{x} - \mathsf{f}^0)$ onto C to obtain the first contact impulse estimate, c^1. **(c)** We then scale F by $\alpha^1 = |\mathsf{c}^1|$ and project $(\mathsf{x} - \mathsf{c}^1)$ onto it, to get the first friction impulse estimate f^1. This completes the first subsequence of the staggered projection sequence. Steps **(d)** and **(e)** then repeat the process for the second subsequence, which will converge, in this simple example, after two more subsequences are completed, to the optimal solution shown in Figure 5.2.

and then, to simplify the discussion, we also let $\mathsf{x} = -\mathsf{M}\dot{\mathsf{q}}^P$ denote the negative momentum of the predictor velocity.

Using these definitions, a solution to the corrector step is given by two coupled projections [Kaufman *et al.* 2008],

$$\mathsf{f}^{t+1} = \mathsf{P}_{\mathsf{F}(\alpha^{t+1})}(\mathsf{x} - \mathsf{c}^{t+1}), \tag{5.15}$$

$$\mathsf{c}^{t+1} = \mathsf{P}_{\mathsf{C}}(\mathsf{x} - \mathsf{f}^{t+1}). \tag{5.16}$$

See Figure 5.2 (left) for a simplified geometric example of the coupled relationship.

5.3.2 Equivalence

To see the equivalence between Equations (5.15) and (5.9), we first multiply out Equation (5.12). We then drop the constant $\mathsf{v}^T\mathsf{M}^{-1}\mathsf{v}$ term and the minimizer remains the same. We then substitute in $\mathsf{u} = \mathsf{D}\beta$ and $\mathsf{U} = \mathsf{F}(\alpha^{t+1})$ into Equation 5.12 and notice that the unknown in the minimization is now β.

We then note that the optimality (KKT) conditions for Equation (5.9) are

$$\begin{aligned} 0 \leq \beta \perp \mathsf{D}^T \dot{\mathsf{q}}^{t+1} + E\lambda \geq 0 \\ 0 \leq \lambda \perp \mathrm{diag}(\mu) - E^T \beta \geq 0. \end{aligned} \tag{5.17}$$

If we then substitute into these conditions the final discrete velocity, given by $\dot{\mathsf{q}}^{t+1} = \dot{\mathsf{q}}^p + \mathsf{M}^{-1}\mathsf{N}\alpha^{t+1} + \mathsf{M}^{-1}\mathsf{D}\beta$, the resulting system gives the optimality conditions for the projection in Equation (5.12), transformed to minimize the variable β, as described above.

To see the equivalence between Equations (5.16) and (5.23) we again multiply out Equation (5.12) and drop the constant $\mathsf{v}^T\mathsf{M}^{-1}\mathsf{v}$ term. We then substitute in $\mathsf{u} = \mathsf{N}\alpha$ and $\mathsf{U} = \mathsf{C}$ into Equation (5.12) and notice that the unknown in the minimization is now α. This resulting minimization is then the same as the Lagrangian dual minimization form of Equation (5.23) which is given by

$$\alpha^{t+1} = \underset{\alpha}{\mathrm{argmin}} \left\{ \frac{1}{2}\alpha^T \mathsf{N}^T \mathsf{M}^{-1} \mathsf{N}\alpha + \alpha^T \mathsf{N}^T (\dot{\mathsf{q}}^p + \mathsf{M}^{-1}\mathsf{D}\beta^{t+1}) : \alpha \geq 0 \right\} \tag{5.18}$$

Since Equation (5.23) is convex, solving the dual minimization is equivalent to solving the primal minimization and we get equivalence here too.

5.3.3 A Fixed-Point Projective Property

Then, by substituting (5.15) into (5.16), we obtain the fixed-point property that characterizes all solutions of the coupled contact problem:

$$\mathsf{N}\alpha^{t+1} = \mathsf{P}_{\mathsf{C}}\big(\mathsf{x} - \mathsf{P}_{\mathsf{F}(\alpha^{t+1})}(\mathsf{x} - \mathsf{N}\alpha^{t+1})\big). \tag{5.19}$$

By construction, any fixed-point solution of (5.19), together with the resulting friction response given by (5.15), defines a solution to the frictional contact problem and, as we will show later in Section 5.6.2 effectively defines a global minimum to a nonconvex optimization problem.

5.3.4 Base Algorithm

The Staggered Projections algorithm then applies a staggered sequence of the two projections (see Figure 5.1), to obtain the solution to the corrector step. Each subsequence is defined so

that we obtain the $i+1$ solution estimate by applying

$$\begin{aligned} \mathsf{f}^{i+1} &\leftarrow \mathsf{P}_{\mathsf{F}(\alpha^i)}(\mathsf{x}-\mathsf{N}\alpha^i), \\ \mathsf{N}\alpha^{i+1} &\leftarrow \mathsf{P}_{\mathsf{C}}(\mathsf{x}-\mathsf{f}^{i+1}). \end{aligned} \tag{5.20}$$

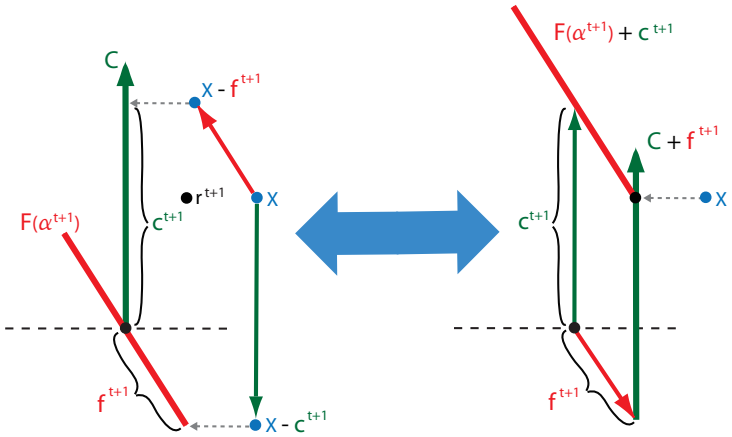


Figure 5.2: **Geometric Optimality:** The optimal solution at time $t+1$ is given by the pair c^{t+1} and f^{t+1}. A simplified example of an optimal geometry characterizing this solution is depicted above. **Left:** A pair, c and f, are optimal if and only if the projection of $(\mathsf{x}-\mathsf{f})$ onto the contact cone, here depicted by the larger green arrow, C, returns c, and if the projection of $(\mathsf{x}-\mathsf{c})$ onto the scaled friction set, here depicted by the red line segment, $\mathsf{F}(\alpha)$, similarly returns f. If this is true, then the sum of c and f gives r^{t+1}, an optimal frictional-contact response. **Right:** Equivalently, c and f are optimal when the projection of x onto both of the sets, $\{\mathsf{F}(\alpha)+\mathsf{c}\}$ (with $|\mathsf{c}| = \alpha$) and $\{\mathsf{C}+\mathsf{f}\}$, return the same point. If this is true, then this projection point is r^{t+1}, an optimal frictional-contact response.

5.3.5 Closest Point Geometry

The proposed approach is described as a staggered projection method to differentiate it from alternating projection methods Bauschke [2000] which superficially resemble our approach, but have very different properties. Effectively, the staggered projection sequence maintains the unchanging point x which is repeatedly projected, in a staggered order, onto a pair of converging sets, $\{\mathsf{F}(\alpha)+\mathsf{N}\alpha\}$ and $\{\mathsf{C}+\mathsf{f}\}$. Geometrically, convergence is satisfied when the closest points to x, in both sets, are the same (see Figure 5.2, right). This type of minimum property can not be satisfied by alternating projection methods which can only apply to finding *arbitrary* feasible points in set intersections.

The staggered sequence of projections is thus essentially a self-correcting process, alternating between two changing sets. The output of each successive projection onto one set corrects

for any change it might impose on the current predicted solution. From a physical standpoint these modifications account for the effect of changes in a friction (or contact) impulse, just obtained from the last projection, may have on the next projection. For instance, after subsequence i, a projection onto C, offset by the current friction impulse estimate, f^i, will generate a new contact impulse that takes into account any possible changes in constraint status that could be caused by applying f^i.

5.3.6 Subproblem Dimensions and Complexity

Both of the projections in (5.20) can be implemented as convex Quadratic Programming (QP) solves. Roughly speaking the cost of each successive QP solve will be polynomial in its dimension, and linear in the number of constraints. In practice, in our implementations, for QPs with dimension n we've found this to be approximately $O(n^3)$.

For systems and complex contacting scenarios where we expect the system's DoFs to be much smaller than the number of contacts, i.e., $|\dot{\mathsf{q}}| \ll |\mathscr{C}|$, P_C can be formulated as a primal QP solves with a dimension proportional to the system's DoFs rather than the number of contacts. Alternately for systems where we expect $|\dot{\mathsf{q}}| \gg |\mathscr{C}|$ we can instead solve P_C using a dual QP formulation with a dimension proportional to the number of contacts.

In either case we note that for any solve of P_C, at most $\min(|\mathscr{C}|, |\dot{\mathsf{q}}|)$ entries of the output, α, will be non-zero. This is because the associated primal QP can have, at most, $|\dot{\mathsf{q}}|$ active constraints for *any* given constraint set (and certainly can't have more active constraints than the number of contacts). This allows us to formulate, for each solve of $\mathsf{P}_{\mathsf{F}(\alpha)}$, a much smaller QP corresponding only to the non-zero entries in α and correspondingly bring down the dimension of each friction solve to at most $\min(|\mathscr{C}|, |\dot{\mathsf{q}}|)$ times the tangent sample size. This corresponds to the physically intuitive restriction that inactive contacts should not generate frictional impulses.

5.3.7 Warm Start

The structure of the staggered projection formulation also encourages the reuse of computations. Because of the fixed-point property in (5.19), initializing a staggered projection algorithm with *either* an optimal contact impulse *or* an optimal friction impulse brings the system

to convergence after a single projection. Similarly, we find that initializing the algorithm with a *close to optimal* impulse dramatically speeds up convergence. Because most time-steps of contacting simulations maintain a reasonable degree of coherence, we can leverage this warm start property to accelerate contact resolution solves. In particular, many of the most difficult frictional-contact configurations, such as those maintaining complex but stable sticking behavior, exhibit very high temporal coherence that allows for the reuse of solutions over many time steps. Overall we typically observe up to two orders of magnitude speed-up when applying warm starts (see §7.2).

5.4 Staggered Projections in Practice

While the essential components to the Staggered Projections Algorithm are described above, in practice we have applied the following strategies to formulate the final version of the algorithm.

5.4.1 Sequence Ordering

Because storing and reusing prior contact correction impulses generally requires more overhead, it is often simpler, and more efficient, to warm start using friction impulses. We thus begin each staggered projection sequence with a contact projection that applies the friction impulse obtained from the prior step, in order to encourage rapid convergence.

5.4.2 Graceful Degradation

Each staggered projection sequence finishes its sequence with a final contact projection. Ordering the sequences in this way guarantees that the final solution obtained will satisfy both the Discrete Euler-Lagrange equation and the Signorini-Fichera boundary conditions and so will generate robust behavior, regardless of whether convergence has been obtained. For time-critical applications, where hard time limits may require setting a fixed upper bound on the maximum number of iterations, this allows for graceful degradation in case of early exits from the projective sequence.

5.4.3 Convergence Criteria

Given the staggered sequence ordering, we determine numerical convergence at the end of subsequence i, to a fixed-point solution of (5.19), using the relative, kinetic metric error,

$$\texttt{rel\_err} = \frac{(\mathsf{f}^i - \mathsf{f}^{i-1})^T \mathsf{M}^{-1} (\mathsf{f}^i - \mathsf{f}^{i-1})}{\mathsf{f}^{i-1^T} \mathsf{M}^{-1} \mathsf{f}^{i-1}}, \tag{5.21}$$

and a user supplied friction error threshold.

5.4.4 Non-Monotonicity

While generally the decrease in residual is monotone, the staggered projection sequences often exhibit local non-monotone residual behavior. Thus, for time-critical applications we can cache and update current best solutions, based on residual values, for applications where early exiting may be necessary. Here a useful residual, at the end of subsequence i, is given by the sum,

$$\texttt{residual} = \sum_{k \in \mathscr{C}} \left| \alpha_k^i \mathsf{n}_k^T (\dot{\mathsf{q}}^p + \mathsf{M}^{-1} \mathsf{N} \boldsymbol{\alpha}^i + \mathsf{M}^{-1} \mathsf{D} \boldsymbol{\beta}^i) \right|. \tag{5.22}$$

5.4.5 Algorithm

Once an acceptable friction error tolerance, ε, is selected, the Staggered Projection algorithm for stepping a contacting system from time t to $t+1$ is given in Algorithm 1. As in the preceding sections, we continue to use the superscripts t and $t+1$ to indicate discrete time increments and reserve the superscripts i and $i+1$ to denote iteration numbers within a single time step.

5.5 Analysis of the Corrector Step

In this section, for convenience, we will first recover the coupled QP formulation of frictional contact and then use it to discuss how friction-contact subspace orthogonality effects the Staggered Projection sequence. As discussed in Chapter 2.8, nonorthogonality between friction and contact vectors is a fundamental feature of Painleve's Paradox, Erdmann's condition and other possible descriptions of frictional contact difficulty. We will show that in this discrete setting the converse also holds: when orthogonality holds it implies that the frictional contact problem is, in some senses, made much easier. Later in this chapter (and in the following chapter) we

Algorithm 1 Staggered Projections.

$\dot{\mathsf{q}}^p \leftarrow$ solve: $\mathsf{M}\ddot{\mathsf{q}}^p = \mathsf{g}(\mathsf{q}^p, \dot{\mathsf{q}}^p) + \mathsf{f}^p_{ext}$
$\mathsf{f}^0 \leftarrow \mathsf{f}^t$ // warm start
$i \leftarrow 1$, `min_res` $\leftarrow \infty$, `rel_err` $\leftarrow \infty$
while `rel_err` $> \varepsilon$ **and** $i <$ `max_iters` **do**
  $\mathsf{N}\alpha^i \leftarrow \mathsf{P}_{\mathsf{C}}(-\mathsf{M}\dot{\mathsf{q}}^p - \mathsf{f}^{i-1})$
  $\mathsf{f}^i \leftarrow \mathsf{P}_{\mathsf{F}(\alpha^i)}(-\mathsf{M}\dot{\mathsf{q}}^p - \mathsf{N}\alpha^i)$
  compute `rel_err` // Equation (5.21)
  compute `residual` // Equation (5.22)
  if `residual` $<$ `min_res` **then**
    `min_res` $\leftarrow$ `residual`
    $\mathsf{f}^* \leftarrow \mathsf{f}^i$ // cache best solution
  end if
  $i \leftarrow i+1$
end while
$\mathsf{f}^{t+1} \leftarrow \mathsf{f}^*$
$\mathsf{c}^{t+1} \leftarrow \mathsf{N}\alpha^{t+1} \leftarrow \mathsf{P}_{\mathsf{C}}(-\mathsf{M}\dot{\mathsf{q}}^p - \mathsf{f}^{t+1})$
$\dot{\mathsf{q}}^{t+1} \leftarrow \dot{\mathsf{q}}^p + \mathsf{M}^{-1}(\mathsf{c}^{t+1} + \mathsf{f}^{t+1})$

will then show that nonorthogonality also implies very specific computational difficulties must be overcome.

First, however, we will use the coupled QP formulation to clarify the relationship between the predictor-corrector formulation and the Stewart-Trinkle LCP formulation. This will allow us to make several general observations about the difficulty of the discrete frictional contact problem that apply to both methods. In particular, we will use this relationship to understand why accurately resolving Stewart-Trinkle LCPs has remained an outstanding problem.

5.5.1 Coupled QP Form

As discussed in Section 5.2, once the predictor velocity, $\dot{\mathsf{q}}^p$, has been generated by the predictor step, the corrector step then resolves contact constraints and friction by solving a coupled system of projections. Presuming the linearized Coulomb Law is enforced we then can equivalently define the solution of the projective system as a pair of coupled QPs

$$\alpha^{t+1} = \underset{\alpha}{\operatorname{argmin}}\Big\{\frac{1}{2}\alpha^T\mathsf{N}^T\mathsf{M}^{-1}\mathsf{N}\alpha + \alpha^T\mathsf{N}^T(\mathsf{M}^{-1}\mathsf{D}\beta^{t+1} + \dot{\mathsf{q}}^p) : \alpha \geq 0\Big\}, \tag{5.23}$$

$$\beta^{t+1} = \underset{\beta}{\operatorname{argmin}}\Big\{\frac{1}{2}\beta^T\mathsf{D}^T\mathsf{M}^{-1}\mathsf{D}\beta + \beta^T\mathsf{D}^T(\mathsf{M}^{-1}\mathsf{N}\alpha^{t+1} + \dot{\mathsf{q}}^p) : \beta \geq 0, \quad \operatorname{diag}(\mu)\alpha^{t+1} - E^T\beta \geq 0\Big\}. \tag{5.24}$$

5.5.2 Subspace Orthogonality and Convergence of Staggered Projections

While the general convergence behavior of Staggered Projections will be discussed in the next chapter, the special case where the friction subspace, given by the span of D, is mass orthogonal to the contact subspace, given by the span of N, is of special interest. In particular, for this case, we can show that Staggered Projections will converge after a single iteration.

First we notice that subspace orthogonality, as we've defined it, implies $\mathsf{N}^T\mathsf{M}^{-1}\mathsf{D} = 0$. Examining the above QP formulation, this implies, in turn, that the friction unknown, β^{t+1}, disappears from the contact solve in Equation (5.23). Thus we can first solve the contact QP in Equation (5.23) independently from friction. We then can use the obtained contact solution, α^{t+1}, to solve the friction QP in Equation (5.24). After this single sequence we are guaranteed a coupled solution that satisfies all frictional contact conditions.

5.5.3 Joint Optimality Conditions

Because each of the above QPs is convex, their individual KKT optimality conditions are both necessary *and* sufficient for optimality [Boyd & Vandenberghe 2004]. First we note that the KKT conditions for Equation (5.23) are (by construction) the discrete Signorini-Fischera Boundary Conditions:

$$0 \leq \alpha \perp \mathsf{N}^T\mathsf{M}^{-1}\mathsf{N}\alpha + \mathsf{N}^T(\mathsf{M}^{-1}\mathsf{D}\beta^{t+1} + \dot{\mathsf{q}}^p) \geq 0 \tag{5.25}$$

while the KKT conditions for Equation (5.24) are

$$\begin{aligned} &0 \leq \beta \perp \mathsf{D}^T\mathsf{M}^{-1}\mathsf{D}\beta + \mathsf{D}^T(\mathsf{M}^{-1}\mathsf{N}\alpha^{t+1} + \dot{\mathsf{q}}^p) + E\lambda \geq 0 \\ &0 \leq \lambda \perp \mathrm{diag}(\mu)\alpha^{t+1} - E^T\beta \geq 0 \end{aligned} \tag{5.26}$$

Then, by combining the KKT conditions for both of the above QPs, we obtain an LCP system that fully characterizes *all* solutions of the frictional contact problem for the predictor-corrector formulation,

$$0 \leq \mathsf{A}\mathsf{y} + \mathsf{b} \perp \mathsf{y} \geq 0 \tag{5.27}$$

where we let

$$\mathsf{A} \stackrel{\text{def}}{=} \begin{pmatrix} \mathsf{N}^T\mathsf{M}^{-1}\mathsf{N} & \mathsf{N}^T\mathsf{M}^{-1}\mathsf{D} & 0 \\ \mathsf{D}^T\mathsf{M}^{-1}\mathsf{N} & \mathsf{D}^T\mathsf{M}^{-1}\mathsf{D} & E \\ \text{diag}(\mu) & -E^T & 0 \end{pmatrix}, \tag{5.28}$$

$$\mathsf{b} \stackrel{\text{def}}{=} \begin{pmatrix} \mathsf{N}^T\dot{\mathsf{q}}^p \\ \mathsf{D}^T\dot{\mathsf{q}}^p \\ 0 \end{pmatrix}, \tag{5.29}$$

and

$$\mathsf{y} \stackrel{\text{def}}{=} \begin{pmatrix} \alpha \\ \beta \\ \lambda \end{pmatrix}. \tag{5.30}$$

5.5.4 Discussion: Stewart-Trinkle LCPs

If we ignore the update method by which the predictor state estimate is obtained, and likewise the method by which the system's state is updated after the corrector solve, then the joint optimality conditions of Equation (5.27) generate the pure form [Anitescu & Potra 1997] of the Stewart-Trinkle LCP formulation for frictional contact [Stewart & Trinkle 1996]. Thus any solution of the corrector step satisfies the Stewart-Trinkle LCP and conversely any analysis of the corrector system applies to the Stewart LCP. This important as we will be able to use these properties in the foregoing discussion both to help understand the difficulties that are generally encountered when attempting to solve the Stewart-Trinkle LCP directly (Section 5.6.2) and how the Staggered Projections algorithm obtains solutions for problems where other practical solution approaches generally fail (Section 6.7).

5.6 Analysis of the Corrector Step

5.6.1 LCP Properties

While it can be shown that the matrix A is Copositive [Anitescu & Potra 1997], it is neither semi-definite nor definite and is additionally non-symmetric[3] so that Equation (5.27) does not correspond to the KKT conditions of any QP, convex or otherwise [Cottle, Pang, & Stone 1992]. By multiplying the complementarity condition through, however, and applying all inequality constraints, we obtain a non-convex, constrained QP,

$$\min_{\mathsf{y}}\{\mathsf{y}^T\left(\mathsf{A}\mathsf{y}+\mathsf{b}\right) : \mathsf{A}\mathsf{y}+\mathsf{b}\geq 0,\ \mathsf{y}\geq 0\}, \tag{5.31}$$

whose global minima, which have the objective value of zero, all solve the above LCP (in Equation 5.27) exactly.

5.6.2 Global Optimization Form

Substituting and simplifying, we then find that obtaining the coupled solution to Equations (5.23) and (5.24) is equivalent to finding a *global* minimizer of

$$\begin{aligned}\min_{\alpha,\beta,\lambda}\Big\{ \| \mathsf{N}\alpha+\mathsf{D}\beta \|^2_{\mathsf{M}^{-1}} + (\mathsf{N}\alpha+\mathsf{D}\beta)^T\dot{\mathsf{q}}^p + \lambda^T\mathrm{diag}(\mu)\alpha :\\ \mathsf{N}^T(\dot{\mathsf{q}}^p+\mathsf{M}^{-1}\mathsf{N}\alpha+\mathsf{M}^{-1}\mathsf{D}\beta)\geq 0,\\ \mathsf{D}^T(\dot{\mathsf{q}}^p+\mathsf{M}^{-1}\mathsf{N}\alpha+\mathsf{M}^{-1}\mathsf{D}\beta)+E\lambda\geq 0,\\ \mathrm{diag}(\mu)\alpha-E^T\beta\geq 0,\\ \alpha\geq 0,\beta\geq 0,\lambda\geq 0\Big\}.\end{aligned} \tag{5.32}$$

This formulation is interesting for several reasons. First, it clearly shows that the complementarity matrix A is indeed Copositive. The second observation is that all constraints in the above minimization are convex and, similarly, with exception of the term, $\lambda^T\mathrm{diag}(\mu)\alpha$, all elements of the objective are also convex. This single non-convex term, however, is sufficient to effectively make this global minimization problem nonconvex.

---

[3]Essentially, from the coupled minimization perspective, this non-symmetry comes from taking the Lagrange multipliers, α, from the contact QP, and incorporating them into the Friction QP's constraint. Otherwise we would have a symmetric linear system. Anitescu & Hart [2004] force symmetry into the LCP matrix by effectively adding the Coulomb constraint to the contact QP as well as the friction QP. As discussed in Section 1.3 this has several possible negative consequences.

5.6.3 Computational Complexity and Feasible Minima

As we will soon see in Section 5.7, existing methods for solving these LCP systems generally have a great amount of difficulty solving frictional contact problems for higher coefficients of friction and/or complex constraint sets. To help motivate why many methods run into trouble and generally why these problems can be hard we will consider here the implications of Section 5.6.2.

From a worst case analysis perspective, the global minimization of a nonconvex QP is an NP-hard problem in optimization [Murty & Kabadi 1987]. This problem, however, is far from being a generic nonconvex QP minimization problem. There's several elements of structure that this problem inherits from the effective convexity of the two parent optimizations. First the quadratic term of the QP will be nonegative in the positive orthant given by $\alpha, \lambda \geq 0$. Second, unlike most nonconvex problems, the coupled optimality criteria in Equation (5.27) offer a set of first order optimality conditions that fully describe all global minimums of the constrained problem. Third, we know, a priori, what the minimum value of this constrained minimization problem is, since Equation 5.32 guarantees that the minimum value at all feasible optimal solutions will be zero.

While this is promising, the global minimization problem does not, unfortunately, inherit the strong duality property [Boyd & Vandenberghe 2004] of convex minimizations. Because of this, the duality gap for this problem will be, in general, non-zero, and thus first order optimality (KKT) conditions for this constrained minimization are, as in the majority of nonconvex problems, only necessary and are *not* sufficient – unlike a convex minimization problem [Boyd & Vandenberghe 2004]. See the following section (Section 5.6.4) for a more detailed analysis of Equation (5.32)'s KKT system. Also on the negative side, solutions may not be unique and, possibly even more worrisome, the solution sets are potentially nonconvex [Anitescu & Hart 2004].

Descent methods for nonconvex problems, such as active set algorithms and, even more generally, Sequential QP algorithms, are designed to generate descent sequences that decrease approximations of the objective function and/or infeasibility measures along the sequence. By construction, however, these algorithms are only guaranteed to descend to first order optimal

points, i.e., KKT points, and do not necessarily even find true local minima. Unfortunately an an analysis of the KKT conditions for Equation (5.32) shows that arbitrary KKT points do not, in general, satisfy crucial aspects of the frictional contact problem (see §5.6.4). A surprising and possibly non-intuitive corollary is that despite the reasonable expectation that local minimization should be a more tractable problem than global minimization, finding local minima for a non-convex problem is, in general, also an NP-hard problem in optimization [Murty & Kabadi 1987; Vavasis 1991].

While it is still tempting to apply local descent methods to Equation (5.32), experiments conducted both with active set algorithms for indefinite QPs and with more general SQP algorithms have consistently returned non-minimal KKT points that do not satisfy the frictional-contact problem (see Section 6.7).

5.6.4 KKT Points

The KKT conditions for Equation (5.32) are

$$\begin{aligned}
0 \leq \mathsf{N}^T\Big(\mathsf{v}_l + \mathsf{M}^{-1}\mathsf{N}(2\alpha - \gamma_\alpha) + \mathsf{M}^{-1}\mathsf{D}(2\beta - \gamma_\beta)\Big) - \mathrm{diag}(\mu)(\lambda - \gamma_\lambda) \perp \alpha \geq 0 \\
0 \leq \mathsf{D}^T\Big(\mathsf{v}_l + \mathsf{M}^{-1}\mathsf{N}(2\alpha - \gamma_\alpha) + \mathsf{M}^{-1}\mathsf{D}(2\beta - \gamma_\beta)\Big) + E\gamma_\lambda \perp \beta \geq 0 \\
0 \leq \mathrm{diag}(\mu)\alpha - E^T\gamma_\beta \perp \lambda \geq 0 \\
0 \leq \mathsf{N}^T\big(\mathsf{v}_l + \mathsf{M}^{-1}\mathsf{N}\alpha + \mathsf{M}^{-1}\mathsf{D}\beta\big) \perp \gamma_\alpha \geq 0 \\
0 \leq \mathsf{D}^T\big(\mathsf{v}_l + \mathsf{M}^{-1}\mathsf{N}\alpha + \mathsf{M}^{-1}\mathsf{D}\beta\big) + E\lambda \perp \gamma_\beta \geq 0 \\
0 \leq \mathrm{diag}(\mu)\alpha - E^T\beta \perp \gamma_\lambda \geq 0.
\end{aligned} \tag{5.33}$$

Here $\gamma_\alpha, \gamma_\beta, \gamma_\lambda$ are the lagrange multipliers that correspond to the first three inequality constraints in Equation (5.32). If a stationary point satisfying both the above KKT conditions in Equation (5.33) *and* the side condition $(\gamma_\alpha, \gamma_\beta, \gamma_\lambda) = (\alpha, \beta, \lambda)$ is obtained, then it is a true solution that satisfies the coupled frictional contact problem. In particular, this point will correspond to a (global) minimum of the constrained problem. This can be seen by noting that, in this special case, Equation (5.33) reduces to the joint optimality conditions given in Equation (5.27). In general, however, these KKT conditions can also be clearly satisfied by points that will not satisfy the original coupled frictional-contact problem.

5.6.5 Row Sufficiency

One way to understand how arbitrary KKT points will not necessarily satisfy global optimality is via row sufficiency. A matrix $\boldsymbol{M}$ is called row sufficient if and only if

$$\boldsymbol{x}_i^T(\boldsymbol{M}^T\boldsymbol{x})_i \leq 0 \;\Rightarrow\; \boldsymbol{x}_i^T(\boldsymbol{M}^T\boldsymbol{x})_i = 0, \tag{5.34}$$

for all vectors $\boldsymbol{x}$. All KKT points of Equation (5.33) will satisfy Equation (5.27) if and only if matrix A is row sufficient [Cottle, Pang, & Stone 1992, Lemma 3.5.4]. One way to show that A is not row sufficient is by noting that, setting $\beta = 0$ and, for all $j \neq i$, $\lambda_j, \alpha_j = 0$, we have

$$\mathsf{y}_i^T(\mathsf{A}^T\mathsf{y})_i = \alpha_i(\mathsf{N}^T\mathsf{M}^{-1}\mathsf{N}\alpha + \mathrm{diag}(\mu)\lambda)_i = \alpha_i\mathsf{n}_i^T\mathsf{M}^{-1}\mathsf{n}_i\alpha_i + \alpha_i\mu_i\lambda_i. \tag{5.35}$$

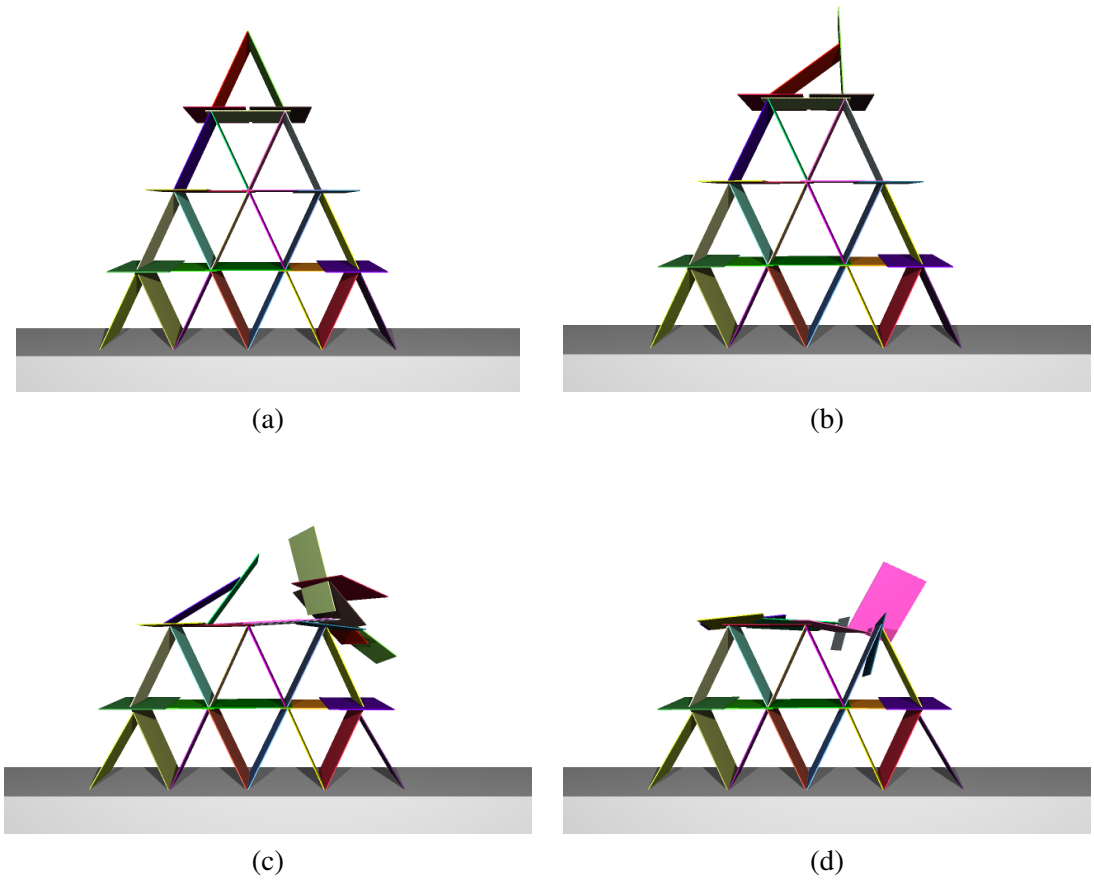
It then follows that

$$\alpha_i\mathsf{n}_i^T\mathsf{M}^{-1}\mathsf{n}_i\alpha_i + \alpha_i\mu_i\lambda_i \leq 0 \;\Rightarrow\; \alpha_i\mathsf{n}_i^T\mathsf{M}^{-1}\mathsf{n}_i\alpha_i + \alpha_i\mu_i\lambda_i = 0 \tag{5.36}$$

does *not* hold *for all* possible values of α_i, λ_i so that A is not a row sufficient matrix. Applying Lemma 3.5.4 in Cottle, Pang, & Stone [1992] we then find that first order optimal KKT points (i.e., points satisfying Equation (5.33)) are not sufficient to guarantee a global minimum for the optimization problem given by Equation (5.32) and so do not necessarily solve the coupled frictional contact problem.

5.7 LCP Algorithms

While the above observations help us to understand the potential issues and difficulties that existing LCP solvers may face, here, in this section, to examine the practical implications for existing LCP algorithms, we discuss problems we observe in our implementations of these algorithms. All algorithms discussed in the following sections start with the same base implementation for collision detection, predictor generation, and contact sampling. We will cover in this section issues pertinent to each method and also note several common problems across algorithms. Later at the end of Chapter 6 we will also compare the behavior of the algorithms with each other and Staggered Projections quantitatively .



(a) (b)

(c) (d)

Figure 5.3: **Pivoting LCP Example:** Here we show snapshots from four early steps in a rigid card house simulation where the frictional contact resolution (corrector) step was solved using the PATH LCP solver [Ferris & Munson 1998]. We start the card house in a stable configuration, see **(a)**, with a high coefficient of friction ($\mu = 0.8$). An exact solution for this simulation problem (when left alone) generates a stable house that stays in this configuration at all steps (see Section 5.8). Initially small, but significant, errors causes cards to incorrectly slip, as in **(b)**. At other steps, however, generally corresponding to reported stationary points, the contacting system blows up. See **(c)** and **(d)**.

5.7.1 Direct Pivoting Methods

Generally, it has been widely assumed that, when fully accurate and robust solutions to frictional contact LCPs were desired, direct, pivoting-based LCP algorithms, such as Lemke's [Cottle, Pang, & Stone 1992], could be applied, provided enough computation time was available. It has been recently noted, however, that direct LCP solvers do not, in practice, scale beyond relatively small-scale examples [Anitescu & Hart 2004; Erleben 2007].

To directly examine the reported behavior of LCP pivoting methods in practice, we implemented the Stewart-Trinkle LCP formulation using the PATH LCP solver [Ferris & Munson

1998]. PATH supports large sparse LCP solves and is the most broadly referenced direct LCP solver in the frictional contact literature. In our experience the PATH based solver, although certainly slow, returned, for the frictionless examples, robust and accurate solutions, even for complex and redundant contact configurations. When friction was added to the LCP the PATH solver still sometimes returned robust and accurate solutions to individual problem instances, in other cases it returned solutions that were robust but with error. Generally these were steps that still maintained a relatively stable simulation, but generated incorrect frictional behavior (e.g, sliding when there should be sticking); however, equally often, PATH instead returned stationary-points. Unfortunately, for these cases, the returned solutions consistently generated highly unstable solutions. See Figure 5.3 for an example.

5.7.2 Iterative LCP Methods

In part, to address the perceived high computational cost of direct LCP algorithms, recent attention has been focused on iterative LCP solution methods [Murty 1988; Cottle, Pang, & Stone 1992] customized for contacting systems [Moreau 1994; Jourdan, Alart, & Jean 1998; Jean 1998; Guendelman, Bridson, & Fedkiw 2003; Duriez *et al.* 2006; Erleben 2007]. These methods, generally based on Gauss-Seidel variants, can potentially be quite fast, and, for LCP problems obtained from frictionless contacting systems, have reasonably nice convergence guarantees [Cottle, Pang, & Stone 1992]. When friction is added, however, these guarantees are lost and behavior is generally inaccurate.

To directly examine the behavior of iterative LCP methods in practice, we implemented the Stewart-Trinkle LCP formulation with two different iterative solvers. For iterative LCP solutions to the frictional contact problem, far and away the most popularly referenced solver is the Projected Gauss-Seidel method [Murty 1988], applied to the Stewart-Trinkle LCP formulation with a box constraint approximation for friction (see Erleben [2007] for example). Box constraints, in these methods, are used to approximate the friction disk with a square, i.e, four samples.

Generally the reason cited for using the simple square approximation to the friction disk

is computational efficiency, since this approach generally adds only four [4] extra unknowns per contact. There is however, an additional, generally unacknowledged, reason why using the friction box constraint is not only efficient but necessary for applying Projected Gauss-Seidel to the frictional contact LCP.

In the simple case of a box constraint, the two friction sample directions are orthogonal so that the magnitudes of the two friction contributions are effectively uncoupled. This allows all friction magnitudes to be constrained independently and thus removes the lower diagonal zero block in the LCP matrix A. As with traditional Gauss-Seidel solvers, Projected Gauss-Seidel requires a positive block diagonal; so this simplification is necessary to implement the method.

To examine the behavior of iterative LCP methods we implemented both the usual box-constrained Projected Gauss-Seidel method and also, for comparison in examples with better friction cone approximations, we also implemented a regularized iterative LCP method [Cottle, Pang, & Stone 1992] for the full Stewart-Trinkle LCP formulation. In our experience both solvers returned, for the frictionless examples, reasonably robust and accurate solutions, provided we allowed for a sufficiently high number of iterations. When friction was added to the LCP however, both solvers failed to return solutions, regardless of the number of iterations we allowed. Generally contact constraints were reasonably well enforced in the friction case as well, but friction behavior was generally incorrect and weak. See Figure 5.4 for an example.

5.7.3 Sequential Quadratic Programming

Finally we also implemented a solver using the NLQLP Sequential Quadratic Programming solver [Schittkowski 2006], applied to the global minimization formulation of EQ (5.32). NLQLP is quite robust and generally accurate, although somewhat slow in comparison to the iterative LCP methods. In all examples NLQLP reported success according to its standards, i.e, it found at least a local first order minimizer. As might be expected, however, the simulations generated by these stationary points and local minimizers obtained unstable behavior, similar to the PATH implementation described above. See Figure 5.5 for an example. This is again consistent with the above results suggesting that stationary points are insufficient for accurate frictional

[4] An added optimization actually reduces this to two.

contact.

5.8 Accurate Solutions Via Staggered Projections and Experimental Validation

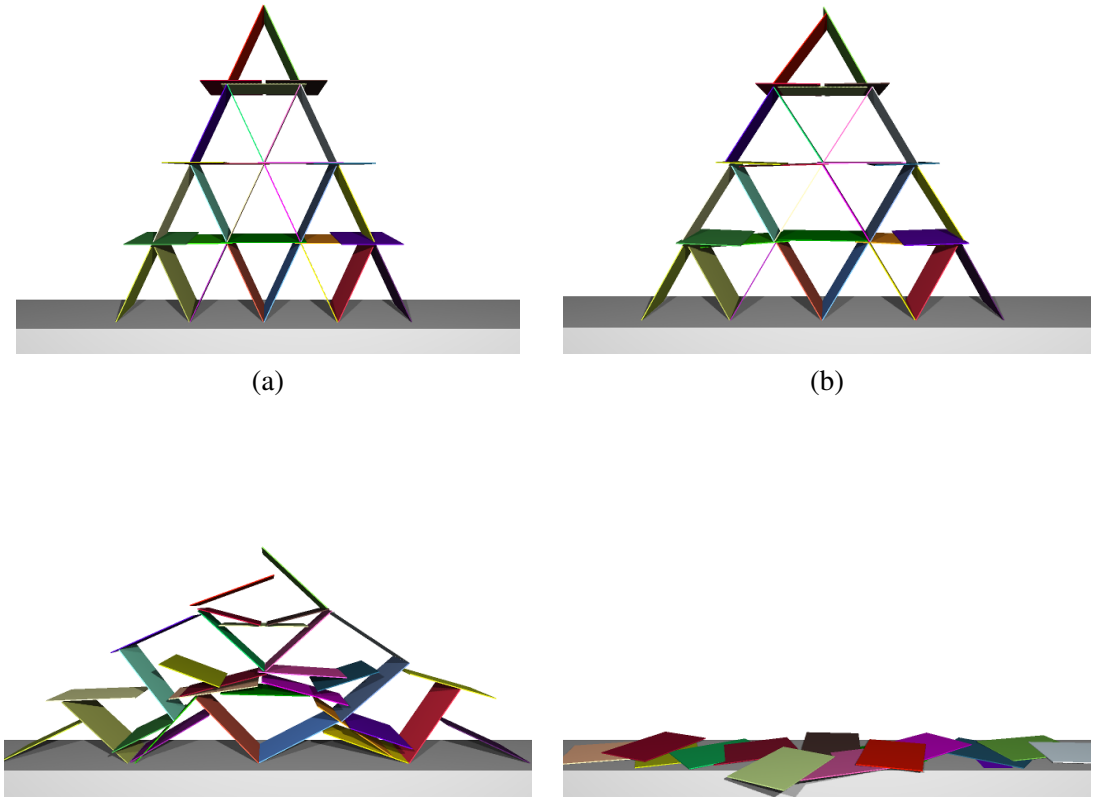
We note that a low error solution ($\texttt{residual} \simeq \texttt{10}^{-6}$) for the same four level card house example is obtained using our Staggered Projections algorithm. We do not show snapshots from the corresponding simulation, as the card house stands stably over all times step in the simulation given by accurate solutions and so corresponds to the initial configuration (as in Figure 5.5(a) for example) throughout. It seems, however, fair to ask whether the accurate solution of these LCPs correspond, in any meaningful way, to expected real-world behaviors.

To begin to answer this question we set up a simple experimental example. Starting with a set of thin metal plates, roughened up with sandpaper, we were easily able to build a rigid card house that maintained a long-term, stable configuration, similar to the the solution obtained by Staggered Projections. See the first frames of Figure 5.6 top, and bottom, respectively.

Then, to examine a more dynamic problem's behavior, we dropped small wooden blocks on the stable structure and filmed the resulting behavior. We observed that under successive impacts from small blocks, the house repeatedly has sections fall down and then regains equilibrium in fairly characteristic way. See the film snapshot sequence in Figure 5.6, top. We then attempted to duplicate the process by dropping blocks on the rigid card house example, simulated using Staggered Projections. The resulting simulation qualitatively captures the stability and characteristic falling behavior observed in the experimental example of the card house. See simulation snapshot sequence in Figure 5.6, bottom.

5.9 Conclusion and Foreshadowing

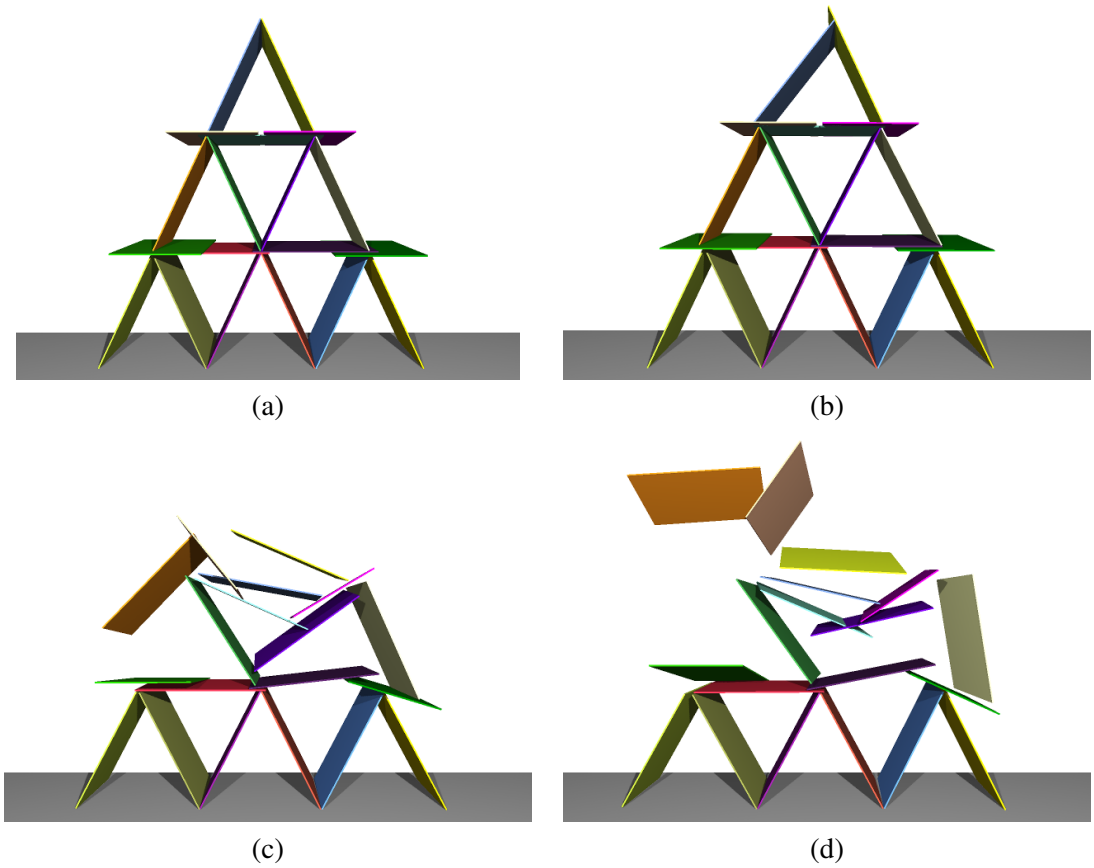
In this chapter we've derived the Staggered Projections algorithm and discussed how it can be applied to solve frictional contact resolution problems were many existing approaches fail. We've also presented new results that help to clarify what the difficulties involved in solving these problems are. In the next chapter we will analyze the Staggered Projections method. We will attempt to clarify the stability and convergence behavior of Staggered Projections, as well as its potential limitations and suitability in applications.



(a) (b)

(c) (d)

Figure 5.4: **Iterative LCP Example:** Here we show snapshots from four early steps in a rigid card house simulation where the frictional contact resolution (corrector) step was solved using a Projected-Gauss-Seidel LCP solver [Murty 1988]. We start the card house in a stable configuration, see **(a)**, with a high coefficient of friction ($\mu = 0.8$). An exact solution for this simulation problem (when left alone) generates a stable house that stays in this configuration at all steps (see Section 5.8). Nonconvergence (even for very high iteration counts, e.g., $\gg 1000$) causes cards to quickly begin to incorrectly slip, as in **(b)** eventually causing the full collapse of the house **(c)** and **(d)**. While constraint stabilization techniques can be applied to augment these approaches [Erleben 2007], they generally make the overall behavior in difficult examples, like this, worse. See Section 7.5.



(a) (b)

(c) (d)

Figure 5.5: **Sequential Quadratic Programming Example:** Here we show snapshots from four early steps in a rigid card house simulation where the frictional contact resolution (corrector) step was solved using a Sequential Quadratic Programming solver [Schittkowski 2006], applied to the global minimization formulation in Equation (5.32). We start the card house in a stable configuration, see **(a)**, with a high coefficient of friction ($\mu = 0.8$). An exact solution for this simulation problem (when left alone) generates a stable house that stays in this configuration at all steps (see Section 5.8). While in all examples the SQP solver reported success (i.e, it found at least a local first order minimizer) and, in some steps, only generated small slipping errors, as in **(b)**, solutions obtained also often generated very unstable, blow-up behaviors similar to those obtained from the pivoting LCP implementation, shown above in Figure 5.3. See **(c)** and **(d)**.

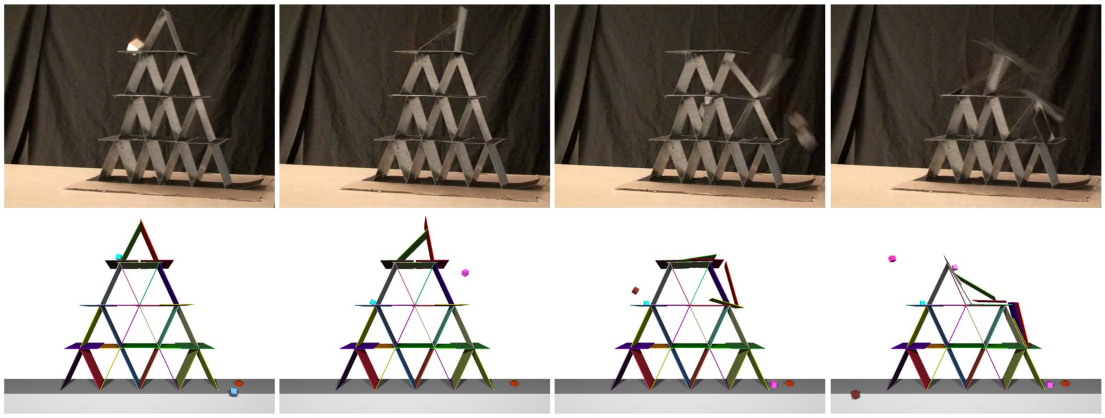


Figure 5.6: **Top:** A stable house of rigid cards is made out of thin metal plates with high friction. In an experiment, under successive impacts from small blocks, the house repeatedly has sections fall down and then regains equilibrium. **Bottom:** An instrumented example simulation of a card house composed of rigid body models using a high coefficient of friction ($\mu = 0.8$). We use the Staggered Projections algorithm to obtain an accurate solution to the underlying frictional contact resolution problem. Dropping blocks in this simulation we obtain qualitatively similar behavior to the experiment.

Chapter 6

Staggered Projections: Stability, Convergence Behavior, and Performance

6.1 Overview

In the last chapter we derived the Staggered Projections algorithm and discussed its application to difficult frictional contact resolution problems. In this chapter we will attempt to clarify the stability and convergence behavior of Staggered Projections, as well as its potential limitations and suitability for use in applications. In the following sections we will first redefine the Staggered Projections as Quadratic Programming (QP) minimizations over unknown contact impulse magnitudes, α, and friction impulse magnitudes, β. We will then use this reformulation to derive stability results for the Staggered Projections algorithm (Section 6.3). We will then further reformulate the Staggered Projections method in Section 6.4 to explicitly include the extra global variable λ (originally introduced in Section 5.5.3) to facilitate a convergence analysis (Section 6.5) performed in the full global space of the contact problem. This analysis will lead to a pair of useful convergence bounds (Section 6.5.6) that will help to clarify convergence behavior. We will then investigate the numerical convergence behavior of Staggered Projections and perform a comparison with alternate solution approaches (Section 6.7). Finally, we will end this chapter with some conclusions and a discussion.

6.2 Staggered Projections Reprise

To facilitate the stability and convergence analysis that will be discussed in the following sections it will be helpful to formulate the Staggered Projections algorithm using a pair of coupled QPs once again,

$$\boldsymbol{a}(\beta) \stackrel{\text{def}}{=} \underset{\alpha}{\text{argmin}}\Big\{\frac{1}{2}\alpha^T \mathsf{N}^T \mathsf{M}^{-1}\mathsf{N}\alpha + \alpha^T \mathsf{N}^T(\mathsf{M}^{-1}\mathsf{D}\beta + \dot{\mathsf{q}}^p) : \ \alpha \geq 0\Big\}, \tag{6.1}$$

and

$$\begin{aligned}\boldsymbol{b}(\alpha) \stackrel{\text{def}}{=} \underset{\beta}{\text{argmin}}\Big\{\frac{1}{2}\beta^T \mathsf{D}^T \mathsf{M}^{-1}\mathsf{D}\beta + \beta^T \mathsf{D}^T(\mathsf{M}^{-1}\mathsf{N}\alpha + \dot{\mathsf{q}}^p) : \beta \geq 0,\\ \text{diag}(\mu)\alpha - E^T\beta \geq 0\Big\}.\end{aligned} \tag{6.2}$$

The contact and friction projections defined in Section 5.4.5 can then be solved using

$$\mathsf{P}_{\mathsf{C}}(-\mathsf{D}\beta - \mathsf{M}\dot{\mathsf{q}}) = \mathsf{N}\boldsymbol{a}(\beta), \tag{6.3}$$

and

$$\mathsf{P}_{\mathsf{F}(\alpha)}(-\mathsf{N}\alpha - \mathsf{M}\dot{\mathsf{q}}) = \mathsf{D}\boldsymbol{b}(\alpha). \tag{6.4}$$

respectively.

6.3 Lipschitz Continuity

In this section we will obtain a stability result for staggered projections by concatenating the non-expansion property of projections with a stability result for QPs over varying parameters.

As we discussed in Section 5.4.5, Equation (6.3) effectively projects $(-\mathsf{D}\beta - \mathsf{M}\dot{\mathsf{q}}^p)$ onto the convex contact cone. Since projections are non-expansive maps [Clarke 1983], we obtain, for an arbitrary x,

$$\| \mathsf{P}_{\mathsf{C}}(-\mathsf{D}\beta^i + \mathsf{x}) - \mathsf{P}_{\mathsf{C}}(-\mathsf{D}\beta^j + \mathsf{x}) \|_{\mathsf{M}^{-1}} \leq \| \mathsf{D}\beta^i - \mathsf{D}\beta^j \|_{\mathsf{M}^{-1}} . \tag{6.5}$$

For a fixed value of α, $\mathsf{P}_{\mathsf{F}(\alpha)}(\cdot)$, can similarly be viewed as a projection onto the friction subset. When, however, α is an input parameter to the optimization, as in the case of Staggered Projections, non-expansion no longer necessarily holds. This is because the friction subset, onto which each friction minimization projects, varies with each input α. We can instead, however, apply Robinson's [1982] sensitivity analysis that, when applied to Equation (6.2), guarantees Lipschitz continuity for the QP over varying parameters. Thus we have a Lipschitz constant $L_{\mathsf{F}} > 0$ such that

$$\| \boldsymbol{b}(\alpha^i) - \boldsymbol{b}(\alpha^j) \| \leq L_{\mathsf{F}} \| \alpha^i - \alpha^j \|, \tag{6.6}$$

or, alternatively, in order to keep our notation consistent with Equation (6.5), we equivalently have

$$\| \mathsf{P}_{\mathsf{F}(\alpha^i)}(-\mathsf{N}\alpha^i + \mathsf{x}) - \mathsf{P}_{\mathsf{F}(\alpha^j)}(-\mathsf{N}\alpha^j + \mathsf{x}) \|_{\mathsf{M}^{-1}} \leq L_{\mathsf{F}} \| \mathsf{N}\alpha^i - \mathsf{N}\alpha^j \|_{\mathsf{M}^{-1}} . \tag{6.7}$$

Recalling that the staggered projections algorithm is given by the mapping

$$\chi(\mathsf{N}\alpha) \stackrel{\text{def}}{=} \mathsf{P}_{\mathsf{C}}\Big(-\mathsf{P}_{\mathsf{F}(\alpha)}(-\mathsf{N}\alpha - \mathsf{M}\dot{\mathsf{q}}^P) - \mathsf{M}\dot{\mathsf{q}}^P \Big),$$

and then applying Equations (6.5) and (6.7), we obtain, for arbitrary x, the bound

$$\begin{aligned} \| \chi(\mathsf{N}\alpha) - \mathsf{N}\alpha^* \|_{\mathsf{M}^{-1}} &= \| \mathsf{P}_{\mathsf{C}}\big(-\mathsf{P}_{\mathsf{F}(\alpha)}(-\mathsf{N}\alpha + \mathsf{x}) + \mathsf{x}\big) - \mathsf{P}_{\mathsf{C}}\big(-\mathsf{P}_{\mathsf{F}(\alpha^*)}(-\mathsf{N}\alpha^* + \mathsf{x}) + \mathsf{x}\big) \|_{\mathsf{M}^{-1}} \\ &\leq \| \mathsf{P}_{\mathsf{F}(\alpha)}(-\mathsf{N}\alpha + \mathsf{x}) - \mathsf{P}_{\mathsf{F}(\alpha^*)}(-\mathsf{N}\alpha^* + \mathsf{x}) \|_{\mathsf{M}^{-1}} \\ &\leq L_{\mathsf{F}} \| \mathsf{N}\alpha - \mathsf{N}\alpha^* \|_{\mathsf{M}^{-1}} . \end{aligned} \tag{6.8}$$

The Staggered Projections algorithm is thus Lipschitz continuous (with respect to the Kinetic metric) with a Lipschitz constant that is dependent only on the properties of the friction projection.

6.4 Parallel Iterations

In order to analyze the convergence behavior of Staggered Projections we will need to recall the global minimization formulation in Equation (5.32) from Section 5.6.2. We note that Staggered Projections effectively ignores the third, and so far neglected variable of this global minimization, λ. This neglected variable is, from the perspective of the Staggered Projections formulation, a dependent (or Lagrangian dual) variable, given by a set of optimal Lagrange multipliers associated with the Coulomb constraints in the friction minimization (Equation (6.2)). To discuss the convergence of Staggered Projections to a global solution, however, we need to explicitly include λ in the following. To do this, for each Staggered Projections solve, we use the function $\boldsymbol{l}(\alpha)$ to generate the (possibly non-unique) Lagrange multipliers, λ, consistent with an optimal solution obtained by the QP solution $\boldsymbol{b}(\alpha)$ generated by Equation (6.2).

Then, to examine global convergence properties, we reformulate the Staggered Projections algorithm so that it redundantly iterates on y (i.e., doing twice the necessary work at each iteration). For each iteration i of Staggered Projections, we solve for *all* of the global variables,

simultaneously, using the update step

$$\mathsf{y}^{i+1} = \begin{pmatrix} \alpha^{i+1} \\ \beta^{i+1} \\ \lambda^{i+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a}(\beta^i) \\ \boldsymbol{b}(\alpha^i) \\ \boldsymbol{l}(\alpha^i) \end{pmatrix}. \tag{6.9}$$

Note that iterating in this way simply reformulates the Staggered Projections sequence as two sets of redundant and interlacing solves.

Since this formulation unnecessarily doubles the workload of the algorithm, this is not a computationally efficient approach; however, this is an equivalent formulation of the Staggered Projections algorithm and will be useful for discussing convergence to optimal solutions given by

$$\mathsf{y}^* = \begin{pmatrix} \alpha^* \\ \beta^* \\ \lambda^* \end{pmatrix}. \tag{6.10}$$

Using this parallel iteration formulation we are now in a position to analyze a single iteration of Staggered Projections going from iteration i to $i+1$. After solving this iteration using the two convex QPs given by Equations (6.1) and (6.2), each of these two QPs must satisfy its first order optimality (KKT) conditions. Then if the parallel formulation (Equation (6.9)) obtained the iterates

$$\mathsf{y}^i = \begin{pmatrix} \alpha^i \\ \beta^i \\ \lambda^i \end{pmatrix} \tag{6.11}$$

and

$$\mathsf{y}^{i+1} = \begin{pmatrix} \alpha^{i+1} \\ \beta^{i+1} \\ \lambda^{i+1} \end{pmatrix}, \tag{6.12}$$

these KKT conditions, of the two solved QPs, can be combined and jointly written out as the

complementarity conditions

$$0 \leq \begin{pmatrix} \alpha^{i+1} \\ \beta^{i+1} \\ \lambda^{i+1} \end{pmatrix} \perp \begin{pmatrix} \mathsf{N}^T\mathsf{M}^{-1}\mathsf{N} & 0 & 0 \\ 0 & \mathsf{D}^T\mathsf{M}^{-1}\mathsf{D} & E \\ 0 & -E^T & 0 \end{pmatrix} \begin{pmatrix} \alpha^{i+1} \\ \beta^{i+1} \\ \lambda^{i+1} \end{pmatrix} + \begin{pmatrix} 0 & \mathsf{N}^T\mathsf{M}^{-1}\mathsf{D} & 0 \\ \mathsf{D}^T\mathsf{M}^{-1}\mathsf{N} & 0 & 0 \\ \mathrm{diag}(\mu) & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha^{i} \\ \beta^{i} \\ \lambda^{i} \end{pmatrix} + \begin{pmatrix} \mathsf{N}^T\dot{\mathsf{q}}^p \\ \mathsf{D}^T\dot{\mathsf{q}}^p \\ 0 \end{pmatrix} \geq 0. \tag{6.13}$$

Then, to simplify this complementarity system, we set

$$\mathsf{B} \stackrel{\text{def}}{=} \begin{pmatrix} \mathsf{N}^T\mathsf{M}^{-1}\mathsf{N} & 0 & 0 \\ 0 & \mathsf{D}^T\mathsf{M}^{-1}\mathsf{D} & E \\ 0 & -E^T & 0 \end{pmatrix} \tag{6.14}$$

and

$$\mathsf{C} \stackrel{\text{def}}{=} \begin{pmatrix} 0 & \mathsf{N}^T\mathsf{M}^{-1}\mathsf{D} & 0 \\ \mathsf{D}^T\mathsf{M}^{-1}\mathsf{N} & 0 & 0 \\ \mathrm{diag}(\mu) & 0 & 0 \end{pmatrix}. \tag{6.15}$$

Equation (6.13) can now be compactly given as

$$0 \leq \mathsf{y}^{i+1} \perp \mathsf{B}\mathsf{y}^{i+1} + \mathsf{C}\mathsf{y}^{i} + \mathsf{b} \geq 0. \tag{6.16}$$

6.4.1 Global Optimality of Fixed-Points

The above complementarity conditions of the parallel iteration formulation also provide another way of seeing that a fixed-point obtained by Staggered Projections must satisfy global optimality of the frictional contact problem. We note that if a fixed-point y^* is obtained, the fixed-point joint optimality conditions are given by

$$0 \leq \mathsf{y}^* \perp \mathsf{B}\mathsf{y}^* + \mathsf{C}\mathsf{y}^* + \mathsf{b} \geq 0. \tag{6.17}$$

Since $\mathsf{B}\mathsf{y}^* + \mathsf{C}\mathsf{y}^* + \mathsf{b} = \mathsf{A}\mathsf{y}^* + \mathsf{b}$, each such fixed-point must clearly also satisfy the global optimality conditions given by Equation (5.27).

6.5 Convergence Analysis

Despite all of the above effort to include λ in our discussion, we will now attempt to ignore λ as much as we can. Intuitively, we can consider the λ variable as one of the chief sources of difficulty in the frictional contact problem. One way to note this, is by observing that λ only appears in our global objective in the nonconvex term and, in fact, without λ nonconvex terms drop out of the optimization problem entirely. In the following, we will first examine convergence behavior for a series of cases where we can, without loss of generality, ignore λ. Then, we will be forced to consider cases where λ can not be excluded. As might be expected, these will correspond to potentially difficult and even possibly expansive steps. As we will see, however, it will still be better to consider λ in these cases than to ignore it altogether. To help us partition off the λ terms and to better generate norm bounds in the following, we will reformulate the joint optimality conditions of the Staggered Projections sequence in Equation (6.16).

We first define the symmetric matrices

$$\hat{\mathsf{B}} \stackrel{\text{def}}{=} \begin{pmatrix} \mathsf{N}^T\mathsf{M}^{-1}\mathsf{N} & 0 \\ 0 & \mathsf{D}^T\mathsf{M}^{-1}\mathsf{D} \end{pmatrix} \tag{6.18}$$

and

$$\hat{\mathsf{C}} \stackrel{\text{def}}{=} \begin{pmatrix} 0 & \mathsf{N}^T\mathsf{M}^{-1}\mathsf{D} \\ \mathsf{D}^T\mathsf{M}^{-1}\mathsf{N} & 0 \end{pmatrix}, \tag{6.19}$$

along with variable

$$\mathsf{z} \stackrel{\text{def}}{=} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \tag{6.20}$$

6.5.1 Inequality

We now can use the known optimality conditions of the Staggered Projections iterate to generate an inequality that will be the basis for all of our upcoming norm bounds. Multiplying through the complementarity conditions given in Equation (6.16) and applying the inequalities from the same we obtain

$$0 \geq \left(\mathsf{y}^{i+1} - \mathsf{y}^{i}\right)^T \left(\mathsf{B}\mathsf{y}^{i+1} + \mathsf{C}\mathsf{y}^{i} + \mathsf{b} - (\mathsf{B}\mathsf{y}^{i} + \mathsf{C}\mathsf{y}^{i-1} + \mathsf{b})\right). \tag{6.21}$$

Rearranging we then have

$$0 \geq (\mathsf{y}^{i+1} - \mathsf{y}^{i})^{T}\mathsf{B}(\mathsf{y}^{i+1} - \mathsf{y}^{i}) + (\mathsf{y}^{i+1} - \mathsf{y}^{i})^{T}\mathsf{C}(\mathsf{y}^{i} - \mathsf{y}^{i-1}) \tag{6.22}$$

and, finally, substituting in the matrix and vector definitions given above we obtain the inequality

$$0 \geq (\mathsf{z}^{i+1} - \mathsf{z}^{i})^{T}\hat{\mathsf{B}}(\mathsf{z}^{i+1} - \mathsf{z}^{i}) + (\mathsf{z}^{i+1} - \mathsf{z}^{i})^{T}\hat{\mathsf{C}}(\mathsf{z}^{i} - \mathsf{z}^{i-1}) + (\lambda^{i+1} - \lambda^{i})^{T}\mathrm{diag}(\mu)(\alpha^{i} - \alpha^{i-1}). \tag{6.23}$$

6.5.2 Preliminaries

In the following, we will ignore the case where $\mathsf{N}^{T}\mathsf{M}^{-1}\mathsf{D} = 0$ since, as discussed in Section 5.5.2, in this case convergence is guaranteed by applying a single contact solve followed by a single friction solve. We will also presume, in the following, without loss of generality, that the matrices N and D are both full rank. If not (which will certainly be the case for D), we can redefine the maximally independent set of columns of N and D (with respect to the inverse mass matrix, M^{-1}) as being the new matrices N and D, and, correspondingly, redefine all considered values of α and β so that evaluations of $\mathsf{N}\alpha$ and $\mathsf{D}\beta$ remain consistent.

We then define $\tilde{\mathsf{B}}$ to be the Cholseky factorization of $\hat{\mathsf{B}}$, such that

$$\tilde{\mathsf{B}}^{T}\tilde{\mathsf{B}} = \hat{\mathsf{B}}. \tag{6.24}$$

Finally, for convenience we also define the individual terms of the above inequality

$$\begin{aligned} \nu &\stackrel{\text{def}}{=} (\lambda^{i+1} - \lambda^{i})^{T}\mathrm{diag}(\mu)(\alpha^{i} - \alpha^{i-1}), \\ \gamma &\stackrel{\text{def}}{=} (\mathsf{z}^{i+1} - \mathsf{z}^{i})^{T}\hat{\mathsf{B}}(\mathsf{z}^{i+1} - \mathsf{z}^{i}), \\ \delta &\stackrel{\text{def}}{=} (\mathsf{z}^{i+1} - \mathsf{z}^{i})^{T}\hat{\mathsf{C}}(\mathsf{z}^{i} - \mathsf{z}^{i-1}). \end{aligned} \tag{6.25}$$

6.5.3 Case by Case

After a few further observations we will see that, given the above preliminaries, in order to obtain convergence bounds on the iterates of the Staggered Projections algorithm there are effectively only two main cases to consider along with three subcases for the latter.

The Cases:

We first note that, for all cases, we have $\gamma \geq 0$. Together with Equation (6.23) this implies that we must also always have $\nu + \delta \leq 0$. Additionally, if $\gamma = 0$ convergence has been obtained. Thus, in the following, we will only consider cases for which $\nu + \delta \leq 0$ and $\gamma > 0$.

Given these additional observations, there are now two remaining cases to consider:

Case 1: $\nu \geq 0$. This implies that $\gamma + \delta \leq 0$ and, given that $\gamma > 0$, we must also have $|\gamma| \leq |\delta|$.

Case 2: $\nu < 0$. Here we must consider three subcases:

(a) $\gamma \leq \delta$. Given that $\gamma > 0$ this also implies $|\gamma| \leq |\delta|$.

(b) $\gamma + \delta \leq 0$. As in Case 1 this implies $|\gamma| \leq |\delta|$.

(c) $\gamma > \delta$ and $\gamma + \delta > 0$. This final case is considered below.

We then obtain the following bounds for each case:

Cases 1, 2(a), and 2(b):

All of these cases, as discussed above, imply $|\gamma| \leq |\delta|$. In turn this inequalty implies that the inequality

$$\| \tilde{\mathsf{B}}\left(\mathsf{z}^{i+1} - \mathsf{z}^{i}\right)\|_2^2 \leq \| \tilde{\mathsf{B}}\left(\mathsf{z}^{i+1} - \mathsf{z}^{i}\right)\|_2 \| \tilde{\mathsf{B}}^{-T} \hat{\mathsf{C}} \tilde{\mathsf{B}}^{-1} \tilde{\mathsf{B}}\left(\mathsf{z}^{i} - \mathsf{z}^{i-1}\right) \|_2, \tag{6.26}$$

also holds. Dividing both sides by $\| \tilde{\mathsf{B}}\left(\mathsf{z}^{i+1} - \mathsf{z}^{i}\right)\|_2$ and pulling out a matrix norm from the right hand side we then obtain

$$\| \tilde{\mathsf{B}}\left(\mathsf{z}^{i+1} - \mathsf{z}^{i}\right)\|_2 \leq \| \tilde{\mathsf{B}}^{-T} \hat{\mathsf{C}} \tilde{\mathsf{B}}^{-1} \|_2 \| \tilde{\mathsf{B}}\left(\mathsf{z}^{i} - \mathsf{z}^{i-1}\right) \|_2 \, . \tag{6.27}$$

Case 2(c):

On first examination, if we, as above, only consider the α and β variables in Case 2(c), we apparently have $|\gamma| < |\delta|$. This would seem to potentially imply that iterations covered by this case are always expansive; however, this analysis ignores the full set of global variables.

Here, unlike the other cases examined above, we can no longer remove the λ variable from our inequality. Instead we actually want to find a way to nontrivially reinsert an entry into the

γ term that includes λ. We can do this by creating new matrix $\hat{\mathsf{B}}_\varepsilon$ that augments the matrix $\hat{\mathsf{B}}$ with an additional row and column, to correspond to entries for λ.

Next we note that we can rule out the possibility of $(\lambda^{i+1} - \lambda^i) = 0$ since this would imply that $\nu = 0$ and thus would be covered by Case 1.

Then, with the assumption that $(\lambda^{i+1} - \lambda^i) \neq 0$, we define a new non-negative variable,

$$\varepsilon \in \Big(0, \frac{-(\gamma + \delta + \nu)}{\| \lambda^{i+1} - \lambda^i \|_2^2} \Big], \tag{6.28}$$

and the augmented matrix

$$\hat{\mathsf{B}}_\varepsilon \stackrel{\text{def}}{=} \begin{pmatrix} \mathsf{N}^T\mathsf{M}^{-1}\mathsf{N} & 0 & 0 \\ 0 & \mathsf{D}^T\mathsf{M}^{-1}\mathsf{D} & 0 \\ 0 & 0 & \varepsilon I \end{pmatrix}. \tag{6.29}$$

Then the inequality in Equation (6.23) implies that the inequality

$$0 \geq (\mathsf{y}^{i+1} - \mathsf{y}^i)^T \hat{\mathsf{B}}_\varepsilon (\mathsf{y}^{i+1} - \mathsf{y}^i) + (\mathsf{y}^{i+1} - \mathsf{y}^i)^T \mathsf{C} (\mathsf{y}^i - \mathsf{y}^{i-1}), \tag{6.30}$$

must also hold. This, in turn, allows us (in a similar fashion as above) to compose the bound

$$\| \tilde{\mathsf{B}}_\varepsilon \big(\mathsf{y}^{i+1} - \mathsf{y}^i\big) \|_2 \leq \| \tilde{\mathsf{B}}_\varepsilon^{-T} \mathsf{C} \tilde{\mathsf{B}}_\varepsilon^{-1} \|_2 \| \tilde{\mathsf{B}}_\varepsilon \big(\mathsf{y}^i - \mathsf{y}^{i-1}\big) \|_2, \tag{6.31}$$

using the Cholesky decomposition

$$\tilde{\mathsf{B}}_\varepsilon^T \tilde{\mathsf{B}}_\varepsilon = \hat{\mathsf{B}}_\varepsilon. \tag{6.32}$$

6.5.4 Evaluating the Matrix Norms

To understand potential convergence behavior implied by the above bounds, in both cases we need to evaluate the matrix norms over all iterate cases. In this section we will formulate the norms in a useful way so that, in the next section, we can determine under what conditions these matrices indicate that each case is contractive, non-expansive, or possibly expansive.

Cases 1, 2(a), and 2(b):

To examine the behavior of Cases 1, 2(a), and 2(b), we need to evaluate the matrix norm $\| \tilde{\mathsf{B}}^{-T} \hat{\mathsf{C}} \tilde{\mathsf{B}}^{-1} \|_2$. Recalling the definition of the matrix two-norm and the Cholesky factorization we have

$$\| \tilde{\mathsf{B}}^{-1} \|_2^2 = \| \hat{\mathsf{B}}^{-1} \|_2 = 1, \tag{6.33}$$

so that the matrix norm can be bounded as

$$\| \tilde{\mathsf{B}}^{-T} \hat{\mathsf{C}} \tilde{\mathsf{B}}^{-1} \|_2 \leq \| \tilde{\mathsf{B}}^{-1} \|_2 \| \hat{\mathsf{C}} \|_2 \| \tilde{\mathsf{B}}^{-1} \|_2 = \| \hat{\mathsf{C}} \|_2 . \tag{6.34}$$

In turn, examining the structure of $\hat{\mathsf{C}}$ (Equation (6.19)) we have, for Cases 1, 2(a), and 2(b), the bound

$$K_A \stackrel{\text{def}}{=} \| \tilde{\mathsf{B}}^{-T} \hat{\mathsf{C}} \tilde{\mathsf{B}}^{-1} \|_2 \leq \| \hat{\mathsf{C}} \|_2 = \| \mathsf{N}^T \mathsf{M}^{-1} \mathsf{D} \|_2 . \tag{6.35}$$

Case 2(c):

Similarly, to understand the bound in Case 2(c) we examine the matrix norm $\| \tilde{\mathsf{B}}_\varepsilon^{-T} \mathsf{C} \tilde{\mathsf{B}}_\varepsilon^{-1} \|_2$. Again recalling the definition of the matrix two-norm and the Cholesky factorization we have

$$\| \tilde{\mathsf{B}}_\varepsilon^{-1} \|_2^2 = \| \hat{\mathsf{B}}_\varepsilon^{-1} \|_2 = \varepsilon^{-1}, \tag{6.36}$$

so that the matrix norm can be bounded as

$$\| \tilde{\mathsf{B}}_\varepsilon^{-T} \mathsf{C} \tilde{\mathsf{B}}_\varepsilon^{-1} \|_2 \leq \| \tilde{\mathsf{B}}_\varepsilon^{-1} \|_2 \| \mathsf{C} \|_2 \| \tilde{\mathsf{B}}_\varepsilon^{-1} \|_2 = \varepsilon^{-1} \| \mathsf{C} \|_2 . \tag{6.37}$$

In turn, examining the structure of C (Equation (6.15)) we have

$$\mathsf{C}^T \mathsf{C} = \begin{pmatrix} (\mathsf{N}^T \mathsf{M}^{-1} \mathsf{D})^2 + \operatorname{diag}(\mu)^2 & 0 & 0 \\ 0 & (\mathsf{N}^T \mathsf{M}^{-1} \mathsf{D})^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{6.38}$$

which implies that

$$\| \mathsf{C} \|_2^2 = \| (\mathsf{N}^T \mathsf{M}^{-1} \mathsf{D})^2 + \operatorname{diag}(\mu)^2 \|_2 \leq \| (\mathsf{N}^T \mathsf{M}^{-1} \mathsf{D}) \|_2^2 + \mu_{max}^2, \tag{6.39}$$

where μ_{max} indicates the largest entry in the vector μ (i.e., the largest coefficient of friction in the contacting system). Then, for case 2(c), we have the bound

$$K_B \stackrel{\text{def}}{=} \| \tilde{\mathsf{B}}_\varepsilon^{-T} \mathsf{C} \tilde{\mathsf{B}}_\varepsilon^{-1} \|_2 \leq \frac{1}{\varepsilon} \Big(\| (\mathsf{N}^T \mathsf{M}^{-1} \mathsf{D}) \|_2^2 + \mu_{max}^2 \Big)^{\frac{1}{2}}. \tag{6.40}$$

6.5.5 Minimum Subspace Angle

The minimum principle angle between two subspaces $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{R}^n$ (see for instance Golub & VanLoan [1996]) is defined as

$$\theta_{U,V} = \cos^{-1} \Big(\max_{\boldsymbol{x} \in \boldsymbol{U}, \|\boldsymbol{x}\|=1} \; \max_{\boldsymbol{y} \in \boldsymbol{V}, \|\boldsymbol{y}\|=1} \boldsymbol{x}^T \boldsymbol{y} \Big). \tag{6.41}$$

This gives a measure of how close these spaces are to being linearly dependent – effectively the smaller the angle, the closer the subspaces are.

For the kinetic metric we can generalize the principle angle formulation so that, presuming the columns of N and D are mass normalized, the minimum principle angle between the span of N and the span of D is defined as

$$\theta_{\mathsf{N},\mathsf{D}} = \cos^{-1}\Big(\max_{\mathsf{x}\in\mathsf{N},\|\mathsf{x}\|_{\mathsf{M}^{-1}}=1} \; \max_{\mathsf{y}\in\mathsf{D},\|\mathsf{y}\|_{\mathsf{M}^{-1}}=1} \mathsf{x}^T\mathsf{M}^{-1}\mathsf{y}\Big). \tag{6.42}$$

In our discussion this is useful, as it generalizes our notion of the degree of the possible coupling between friction and contact subspaces and correspondingly, gives a measure for multipoint contact configurations, of the angle from the Painlevé example for single point contacts. In particular, notice that the recurring bound term $\| \mathsf{N}^T\mathsf{M}^{-1}\mathsf{D} \|_2$ effectively measures this coupling since it is simply the cosine of this minimum spanning angle:

$$\begin{aligned} \| \mathsf{N}^T\mathsf{M}^{-1}\mathsf{D} \| &= \max_{\boldsymbol{x}\in dim(\mathsf{D}),\|\boldsymbol{x}\|=1} \| \mathsf{N}^T\mathsf{M}^{-1}\mathsf{D}\mathsf{x} \| \\ &= \max_{\mathsf{x}\in\mathsf{N},\|\mathsf{x}\|_{\mathsf{M}^{-1}}=1} \; \max_{\mathsf{y}\in\mathsf{D},\|\mathsf{y}\|_{\mathsf{M}^{-1}}=1} \mathsf{x}^T\mathsf{M}^{-1}\mathsf{y}. \end{aligned} \tag{6.43}$$

In other words, this norm captures the degree to which the friction and contact subspaces may overlap and thus interact. Note that, at worst, this measure is one; when the two subspaces are linearly dependent. Otherwise, it measures the potential amount of coupling between possible friction and contact impulses.

6.5.6 Contraction and Non-Expansion Criteria

In this section we examine the conditions under which the bounds obtained in Section 6.5.4 imply contractive or nonexpansive behavior, while in the next section we will cover the implications for expected convergence rates and physical interpretation of these quantities. In this and the following sections we will presume, without loss of generality, that the columns of N and D are also mass normalized with respect to the Kinetic metric.

Recalling, for Cases 1, 2(a) and 2(b), that the bound is given by

$$K_A = \| \tilde{\mathsf{B}}^{-T}\hat{\mathsf{C}}\tilde{\mathsf{B}}^{-1} \|_2 \leq \| \hat{\mathsf{C}} \|_2 = \| \mathsf{N}^T\mathsf{M}^{-1}\mathsf{D} \|_2 \, . \tag{6.44}$$

we find that $K_B \leq 1$ so that we obtain either non-expansion or contraction for all iterates. If

the iterate case corresponds to a strict inequality of $|\gamma| < |\delta|$ then this corresponds to strict inequality in Equation (6.27) and thus the corresponding iterate is a contraction.

Otherwise, the difference between a contractive and non-expansive step effectively depends on the maximum principle angle, as discussed above, between the active constraint directions (i.e., the span of vectors in N corresponding to non-negative entries of α) and the active friction directions (i.e., the span of vectors in D corresponding to non-negative entries of β). See the next section for a more detailed discussion.

For case 2(c), based on the analysis in Section 6.5.4, we see that in order for the iterate to contract we require

$$K_B = \| \tilde{\mathsf{B}}_\varepsilon^{-T} \mathsf{C} \tilde{\mathsf{B}}_\varepsilon^{-1} \|_2 \leq \frac{1}{\varepsilon} \Big(\| (\mathsf{N}^T \mathsf{M}^{-1} \mathsf{D}) \|_2^2 + \mu_{max}^2 \Big)^{\frac{1}{2}} < 1. \tag{6.45}$$

or equivalently

$$\varepsilon > \Big(\cos(\theta_{\mathsf{N},\mathsf{D}})^2 + \mu_{max}^2 \Big)^{\frac{1}{2}}. \tag{6.46}$$

Then, recalling the bounds on ε (Equation (6.28)) we see that if

$$\frac{-(\gamma + \delta + \nu)}{\| \lambda^{i+1} - \lambda^i \|_2^2} > \Big(\cos(\theta_{\mathsf{N},\mathsf{D}})^2 + \mu_{max}^2 \Big)^{\frac{1}{2}}, \tag{6.47}$$

we will obtain a contraction for any iterate corresponding to Case 2(c).

While this criterion may not be particularly intuitive, the main point to take away is that, in these potentially difficult iterates, the likelihood of a contractive step increases both as the size of the coefficients of friction decrease and the minimum subspace angle increases. This coincides with the reasonable expectation that problems with higher coefficients of friction are somehow harder and that likewise problems with less coupling are easier. We will formalize these issues some more and discuss their impact on convergence rates as well as their relation to the underlying global minimization problem in the next section.

6.6 Convergence Rates and Their Relationship to Contacting System Properties

The above analysis implies that the difference between a contractive and nonexpansive step depends both on the minimum principle angle (see above) between the active constraint directions (i.e., the span of vectors in N corresponding to non-negative entries of α) and the active friction

directions (i.e., the span of vectors in D corresponding to non-negative entries of β) as well as the size of the coefficients of friction at active contacts in each iterate.

As shown above, the matrix two-norm of $\mathsf{N}^T\mathsf{M}^{-1}\mathsf{D}$ is equivalent to the minimum principle angle between the two subspaces spanned by N and D correspondingly. It follows for Cases 1, 2(a) and 2(b) then that if this angle is greater than zero we have a contraction and otherwise non-expansion. More generally, we again see that as the angle between the friction and contact subspaces increases the Lipschitz constants in Equation (6.27) and Equation (6.31) decrease correspondingly.

At a high level, the bounds from the last section suggest that, globally, for each problem instance, more iterations are required as the maximum μ value increases and, likewise, more iterations are required as the minimum subspace angle decreases.

Note, however, that these global conclusions do not tell the whole story, since they do not consider the individual, local convergence behavior at each iterate. The above bounds specifically imply that the local contractive behavior at each iterate is based entirely on the minimum principle angle between subspaces spanned *only* by the the constraint and friction directions that have been active in recent iterates *and* the μ values associated with those active constraints. That is, the active constraint directions that correspond to nonzero values of α^{i-1}, α^i,or α^{i+1}, the active friction directions that correspond to nonzero values of β^{i-1}, β^i, or β^{i+1}, and the maximum among μ values that correspond to nonzero values of α^i or α^{i+1}.

This helps to explain the overall observed good behavior of the algorithm despite the regular occurence of strong linear dependence between N and D in large and/or complex contacting configurations. It also suggests that local convergence behavior will be strongly influenced by the properties of the active friction and constraint sets at solutions.

6.7 Numerical Behavior

In this section we will examine the change in both residual and error, with respect to iteration count, for the Staggered Projections algorithm. First we investigate the variation in convergence behavior both when varying the minimum principle angle and the coefficient of friction.

Next we will investigate the change in residual and error over the course of individual Staggered Projections solves. Finally, we will also, in the following section, compare these results with the final residual obtained by Lemke's algorithm and a projected Gauss-Seidel variant, applied to the LCP formulation of EQ (5.27) as well as SQP, applied to the global minimization formulation of EQ (5.32).

Note that in our analysis we use the difference between the current evaluation of the global objective given in EQ (5.32), and the known global minimum value, 0, as our residual. In this sense the residual under discussion here does not have the usual linear relationship to error that we would expect for linear systems. Instead, this residual gives the height of the nonconvex energy surface at the current iterate. Error, however, is defined in the usual way as the distance between the current iterate and a given minimizer.

6.7.1 Convergence Domains for Painlevé-Type Configurations

To examine the domain of convergence of Staggered Projections we implemented a simple two-dimensional case where the normal cone is a single vector and the friction set is a line segment at an angle θ to the normal direction (as in Figure 5.1). For example, this covers the case of the Painleve rod in SE(2) for one point of contact (see below) and includes potentially problematic configurations such as high friction and low minimum angle examples. In each of the following graphs we will either show the number of iterations to converge to a relative error of 10^{-7}. We will plot these datum for each possible initial starting point for a range of predictor values.

In each plot we also draw the generalized normal direction as a white arrow, the feasible half-space boundary, as a white dotted line, the generalized friction set, as a red dotted line segment , and the two boundaries of the generalized friction cone [Erdmann 1994] (which will be useful in the following analysis) as two yellow dotted lines.

Note that in the plotted domain, each point represents the location of the starting negative predicted velocity. Thus all points corresponding to locations below the feasible boundary actually correspond to problems that began with predicted velocities that were feasible and so converged after a single iteration in all cases. Points corresponding to locations above the feasible boundary begin with predicted velocities that are infeasible and so are the interesting (and potentially difficult) cases to examine in each plot.

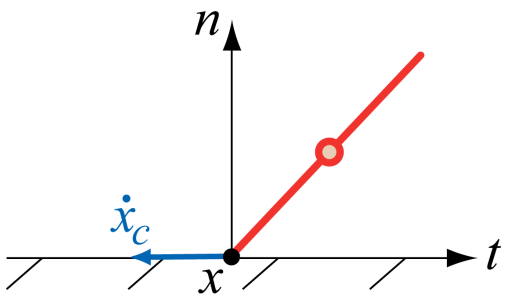


Figure 6.1: **Painlevé Rod Example**

A Single Point Painlevé Rod Example

To help build physical intuition for this example we revisit the Painlevé rod example discussed in Section 2.8.3. The Painlevé rod is one possible physical configuration that corresponds to the above problem. Consider the simple $\mathbb{R}^2$ Painlevé Rod example illustrated in Figure 6.1. Recall that we have a one dimensional rod with a single point of contact at $\boldsymbol{x}$. to a fixed surface with the normal $\boldsymbol{n}$ and tangent $\boldsymbol{t}$. Here the degrees of freedom of the system are given by

$$\mathsf{q} = (a, \boldsymbol{p})^T \in SE(2), \tag{6.48}$$

where a gives the angle of rotation about the center of mass, and $\boldsymbol{p}$ indicates the $\mathbb{R}^2$ coordinates of the center of mass. The generalized velocity is then

$$\dot{\mathsf{q}} = (\dot{a}, \dot{\boldsymbol{p}})^T \in se(2). \tag{6.49}$$

The body coordinates of the contact point could then be given by

$$\hat{\boldsymbol{x}} = (0, -1)^T. \tag{6.50}$$

The world coordinates of the contacting point are the given by

$$\boldsymbol{x}(\mathsf{q}) = R(a)\hat{\boldsymbol{x}} + \boldsymbol{p}, \tag{6.51}$$

Where $R(a)$ is the rotation specified by a; i.e.,

$$R(a) = \begin{pmatrix} \cos(a) & -\sin(a) \\ \sin(a) & \cos(a) \end{pmatrix}. \tag{6.52}$$

Then, we have

$$\dot{\boldsymbol{x}} = \begin{pmatrix} \cos(a) & 1 & 0 \\ \sin(a) & 0 & 1 \end{pmatrix} \dot{\mathsf{q}}, \tag{6.53}$$

so that the jacobian for contact point $\boldsymbol{x}$ is

$$\Gamma_{\boldsymbol{x}} = \begin{pmatrix} \cos(a) & 1 & 0 \\ \sin(a) & 0 & 1 \end{pmatrix}, \tag{6.54}$$

with $\dot{\boldsymbol{x}} = \Gamma_{\boldsymbol{x}} \dot{\mathsf{q}}$.

Because there is only a single point of contact, we can formulate the Staggered Projection minimizations in the two-dimensional contact space of the contact location rather than in the effectively three-dimensional space of planar rigid-body impulses (i.e., $se^*(2)$). In particular we can define the magnitude of the friction impulses along $\boldsymbol{t}_1 \stackrel{\text{def}}{=} \boldsymbol{t}$ and $\boldsymbol{t}_2 \stackrel{\text{def}}{=} -\boldsymbol{t}$, as β_1 and β_2 respectively so that the friction unknowns are $\beta = (\beta_1, \beta_2)^T$, while, as usual, the magnitude of the contact impulse, along $\boldsymbol{n}$ is given by α. Finally we let $\dot{\boldsymbol{x}}^p$ denote the initially predicted velocity.

Then, by rescaling, the resulting contact QP minimization can be given as

$$\boldsymbol{a}(\beta) = \underset{\alpha}{\operatorname{argmin}} \Big\{ \frac{1}{2}\alpha^2 + \alpha \boldsymbol{n}^T \Gamma_{\boldsymbol{x}} \boldsymbol{M}^{-1} \Gamma_{\boldsymbol{x}}^T (\boldsymbol{t}_1, \boldsymbol{t}_2)\beta + \alpha \boldsymbol{n}^T \dot{\boldsymbol{x}}^p) : \ \alpha \geq 0 \Big\}, \tag{6.55}$$

and similarly the friction QP is

$$\boldsymbol{b}(\alpha) = \underset{\beta}{\operatorname{argmin}} \Big\{ \frac{1}{2}\beta^2 + \beta^T (\boldsymbol{t}_1, \boldsymbol{t}_2)^T \Gamma_{\boldsymbol{x}} \boldsymbol{M}^{-1} \Gamma_{\boldsymbol{x}}^T \boldsymbol{n}\alpha + \beta^T (\boldsymbol{t}_1, \boldsymbol{t}_2)^T \dot{\boldsymbol{x}}^p) : \ \beta \geq 0, \ \mu\alpha - \beta_1 - \beta_2 \geq 0 \Big\}. \tag{6.56}$$

Next, to simplify even a bit more, presume $\boldsymbol{M} = \boldsymbol{I}_3$. Then notice that, in the above, the matrix

$$\boldsymbol{U} \stackrel{\text{def}}{=} \Gamma_{\boldsymbol{x}} \boldsymbol{M}^{-1} \Gamma_{\boldsymbol{x}}^T \tag{6.57}$$

effectively rotates and scales the tangent vector $\boldsymbol{t}$, so that the tangent and normal are no longer orthogonal. I.e., $\boldsymbol{n}\boldsymbol{U}\boldsymbol{t} \neq 0$. In particular, for this simple case, the minimum principle angle is then given by

$$\theta = \Big| \boldsymbol{n}^T \frac{\boldsymbol{U}\boldsymbol{t}}{\| \boldsymbol{U}\boldsymbol{t} \|} \Big|. \tag{6.58}$$

Staggered Projections can then be solved (in this particular case) using the very simple contact and friction projections

$$\mathsf{P}_{\boldsymbol{n}}(-\boldsymbol{U}(\boldsymbol{t}_1, \boldsymbol{t}_2)\beta - \dot{\boldsymbol{x}}^p) = \boldsymbol{n}\boldsymbol{a}(\beta), \tag{6.59}$$

and

$$\mathsf{P}_{\alpha \boldsymbol{U}(\boldsymbol{t}_1, \boldsymbol{t}_2)}(-\boldsymbol{n}\alpha - \dot{\boldsymbol{x}}^p) = (\boldsymbol{t}_1, \boldsymbol{t}_2)\boldsymbol{b}(\alpha). \tag{6.60}$$

Variations of θ and μ for Solutions Over the Predictor Domain

Varying θ (the angle between the friction subspace and the normal direction), and μ (the coefficient of friction), for a set of solves each of which ranges over the domain of possible predictor velocities we observe that globally (see Figures 6.2 - 6.5) , for each Staggered Projections solve:

1. More iterations are required as the value of μ increases.

2. Likewise, more iterations are required as the minimum angle between the friction subspace and the normal direction decreases.

Both of these observations correspond well with the bounds developed in the last three sections. We also, however, based on the above bounds expect to see an interaction between how changes in the minimum angle and μ effect convergence behavior.

In these examples we notice that, holding θ fixed, as we vary μ, a distinct border between better and worse convergence behavior clearly follows the two edges of the configuration space friction cone:

- For large values of μ the better convergence behavior is clearly inside the cone, which corresponds to the sticking case.

- As we decrease μ, there is always comes a transition point where there's a flip. At this point the interior of the cone becomes harder to converge and the exterior easier.

Having the interior of the friction cone converge faster for large values of μ is consistent with the above bounds. Points starting in the interior of the cone will all correspond to sticking cases (i.e., zero final velocity) where $\lambda = 0$ at the solution. Then near these solutions we can expect iterates to have $v = 0$ so that the first case, where the contraction bound is given only by $\cos(\theta)$ can be expected to dominate. An alternate and complementary viewpoint is that since $\lambda = 0$ near these solutions we can also expect non-convexity to correspondingly also not be an issue near these solutions.

Decreasing the value of μ, the observed switch to where convergence behavior inside the generalized friction cone becomes harder, generally always occurs when the farther edge of the

friction cone becomes orthogonal to the friction segment. See Figure 6.2 examples. Interestingly, this corresponds to the moment when the cosine term in the bound begins to dominate the friction term. I.e., at the orthogonal point we have $\cos(\theta) = \mu$ and then as we decrease μ further, $\cos(\theta) > \mu$. At this point the lower values of μ dominate the convergence behavior outside the friction cone, while the larger cosine term dominates the interior of the cone.

Finally, for the extreme cases of $\theta = 0$ (equivalently π) and $\theta = \pi/2$ (equivalently $3\pi/2$) convergence occurrs after 1 and 2 iterations respectively, *for all* values of μ. These are simply the cases where friction is either exactly colinear or exactly orthogoal to the normal, so the issues with the principle angle are not at play here.

Note that overall, however, despite this observed switching behavior, the global convergence behavior still follows observations (1) and (2) above. The local switching behavior simply differentiates which regions of the predictor domain have worse convergence behavior as we vary μ and/or θ.

Finally, in all of the above examples, we tested physically plausible coefficients of friction. In most engineering texts, experimentally obtained values of dry friction generally have a ceiling of around $\mu \simeq 1.2$ (e.g., for dry aluminum on aluminum sliding). Of course, we may wish to examine behavior at higher coefficients and here, in this simplified example, Staggered Projections starts hitting some potential limitations for larger values. For larger coefficients of friction the generalized friction cone widens and thus correspondingly increases the area of the stick region were we continue to experience good convergence behavior; however, at the same time, the sliding region becomes correspondingly more difficult to converge.

As μ continues to increase we eventually transition (see Figures 6.6 - 6.8) into the case where the sliding region changes from a high, but converging, iteration count, to non-convergence. Past this point the sliding region does not generally converge, although the residual generally remains low in most of this region. As as we increase μ further, the number of iterations required to converge, inside the generalized cone remains essentially static, while the absolute residual, in a relatively narrow subset of the non-converging region hits a maximum value of ~ 3 (with respect to the usual residual of $\sim 10^{-6}$ in the converging region). This final configuration corresponds to cases where a combined low angle and high μ value, in the above bounds, no longer necessarily guarantee convergence. Interestingly, we note that we still obtain good

convergence to solutions for sticking behavior, as suggested above, while, for sliding behavior, we only note large residual error for some cases. Finally, overall we observe the interesting relationship that while, effectively, convergence generally becomes harder as we increase μ, the region of greater difficulty correspondingly becomes smaller.

6.7.2 Convergence Behavior: Residual and Error

While the above examples may be useful to understand the general behavior of the Staggered Projections algorithm, they also miss the additional complexities imposed in the more general case when there are a large number of possibly redundant constraint and friction directions. In this section we instrumented an example of a larger scale difficult frictional contact configuration to examine the individual behavior of Staggered Projections iterates over the course of the algorithm.

Card House Example

We instrumented a four level rigid card house example with a high coefficient of friction ($\mu = 0.8$). See Figure 5.6 (bottom) for snapshots from this example. For each time-step's Staggered Projections solve we compute the error and global objective value at each iteration. In Figure 6.9 we show the results for the first four time-steps.

In each case we see a (characteristic) close to monotonic decrease in error, with a few small regions of expansion. This is consistent with the analysis in Section 6.5. Similarly, as suggested by the analysis in Section 6.3, we also see that the error appears to be well bounded. Interestingly, however, the change in the objective value (residual) is strongly non-monotonic. Effectively we see that both hill climbing and gradient descent occur regularly throughout the iterations of the Staggered Projections sequence.

Finally, combining these observations, it is suggestive that smooth paths of mostly decreasing error follow such a non-monotonic energy trajectory. This too may help us understand why gradient descent approaches might be expected to have difficulty in these problems.

6.7.3 Solution Comparison

Using the same base code we also attempted to solve the corrector step by applying Lemke's algorithm to the LCP formulation of Equation (5.27), using the PATH solver [Ferris & Munson 1998], and applying Sequential Quadratic Programming to the global minimization formulation of Equation (5.32), using the NLQLP solver [Schittkowski 2006]. Instrumenting the same four level card house example, as above, we compare the obtained residuals for all three algorithms over a simulation sequence stepped at $h = 10^{-3}$ with $\mu = 0.8$. In general, we observe close to zero residual for Staggered Projections throughout. PATH generates low residual in many instances but also produces large positive and negative residuals at many steps. In general these steps physically correspond in the generated simulation to incorrect sliding and blowup behaviors as discussed above. Finally, in many steps NLQLP generates large, but always positive residuals, which again correspond to incorrect sliding and blowup behaviors in the simulation. Note also that the positivity in NLQLP is obtained because SQP always finds a feasible, but not necessarily optimal point. See Figure 6.10 for the results. The average time for a call to PATH, NLQLP, and Staggered Projections over these simulations, was 57 seconds, 123 seconds, and 0.8 seconds respectively.

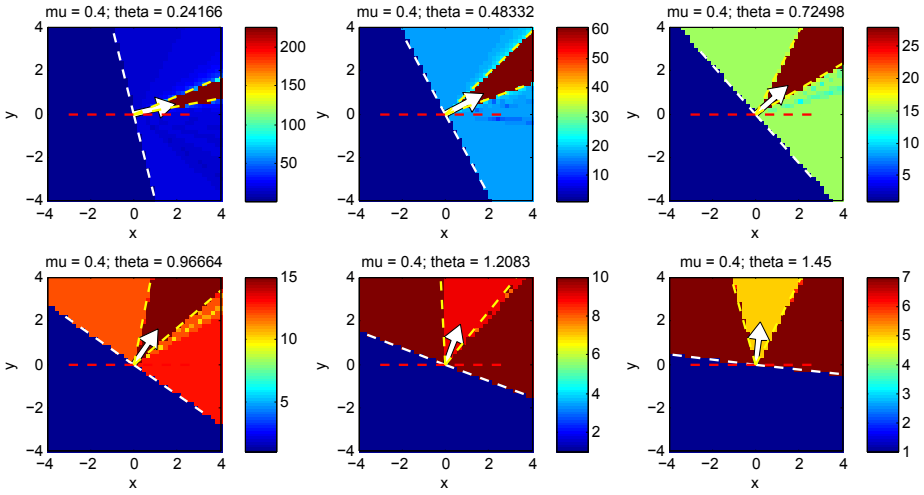
6.8 Conclusions and Discussion

In this chapter we analyzed the stability and convergence behavior of the Staggered Projections Algorithm, as well as its potential limitations and suitability for use in applications. This analysis lead to a pair of useful convergence bounds (Section 6.5.6). We then discussed the potential implications of these bounds for the practical application of the Staggered Projections algorithm and confirmed that this behavior was reflected in our numerical results.

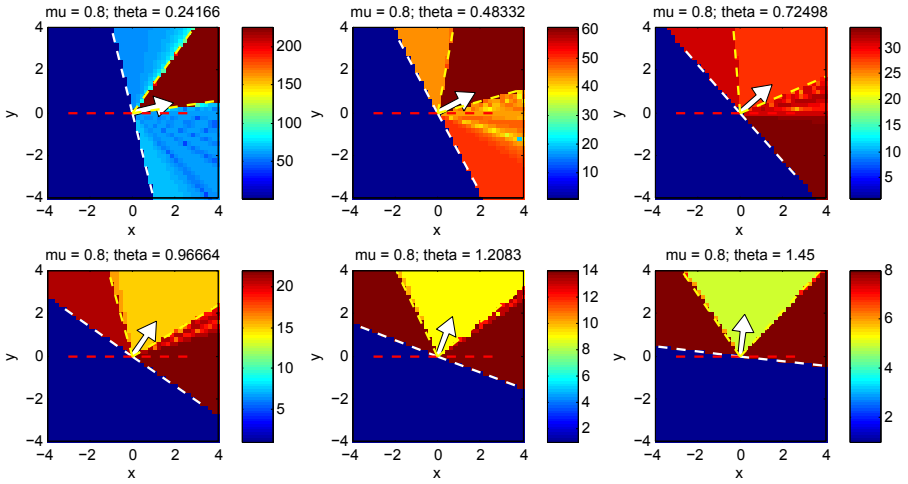
In particular, we found that, as might be expected, there is a strong correlation between reducing the coefficient of friction and the number of iterations needed. Unlike prior algorithms, however, we found that Staggered Projections converged over a large range of plausible larger coefficients.

Perhaps less obvious, we also found that the subspace angle between the friction and contact subspaces equally dominates the convergence behavior of Staggered Projections (in conjunction with μ). In particular, we observed that for a large number of cases the convergence bound for individual iterates was actually *independent* of μ, and only dependent on this angle. This might help explain the strong performance of Staggered Projections for reasonably high coefficients of friction.

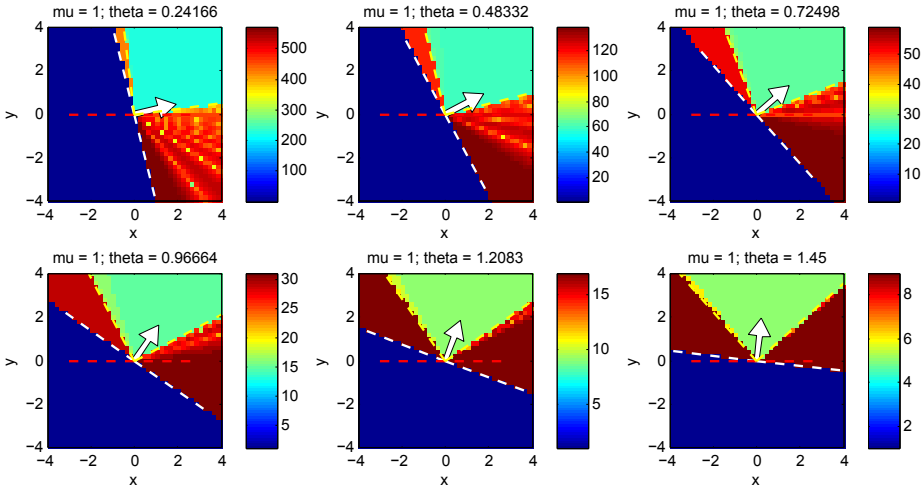
We have shown that, in the discrete setting, this subspace angle analagously encodes the degree to which Painlevé-Type Conditions may manifest for individual multi-point contact problems and, specifically, via the bounds in this chapter, have shown that this angle, together with μ, is an equally necessary parameter for understanding the computational difficulties of frictional contact problems.



Example 1: $\mu = 0.4$



Example 2: $\mu = 0.8$



Example 3: $\mu = 1$

Figure 6.2: **Painlevé Convergence Examples:** keeping μ fixed at three successively higher values in these examples, we varying θ and examine the number of iterations to converge over the domain of possible predictor points. Note that convergence behavior is symmetric on all three other quadrants. The residual was $\sim 10^{-6}$ over the domain.

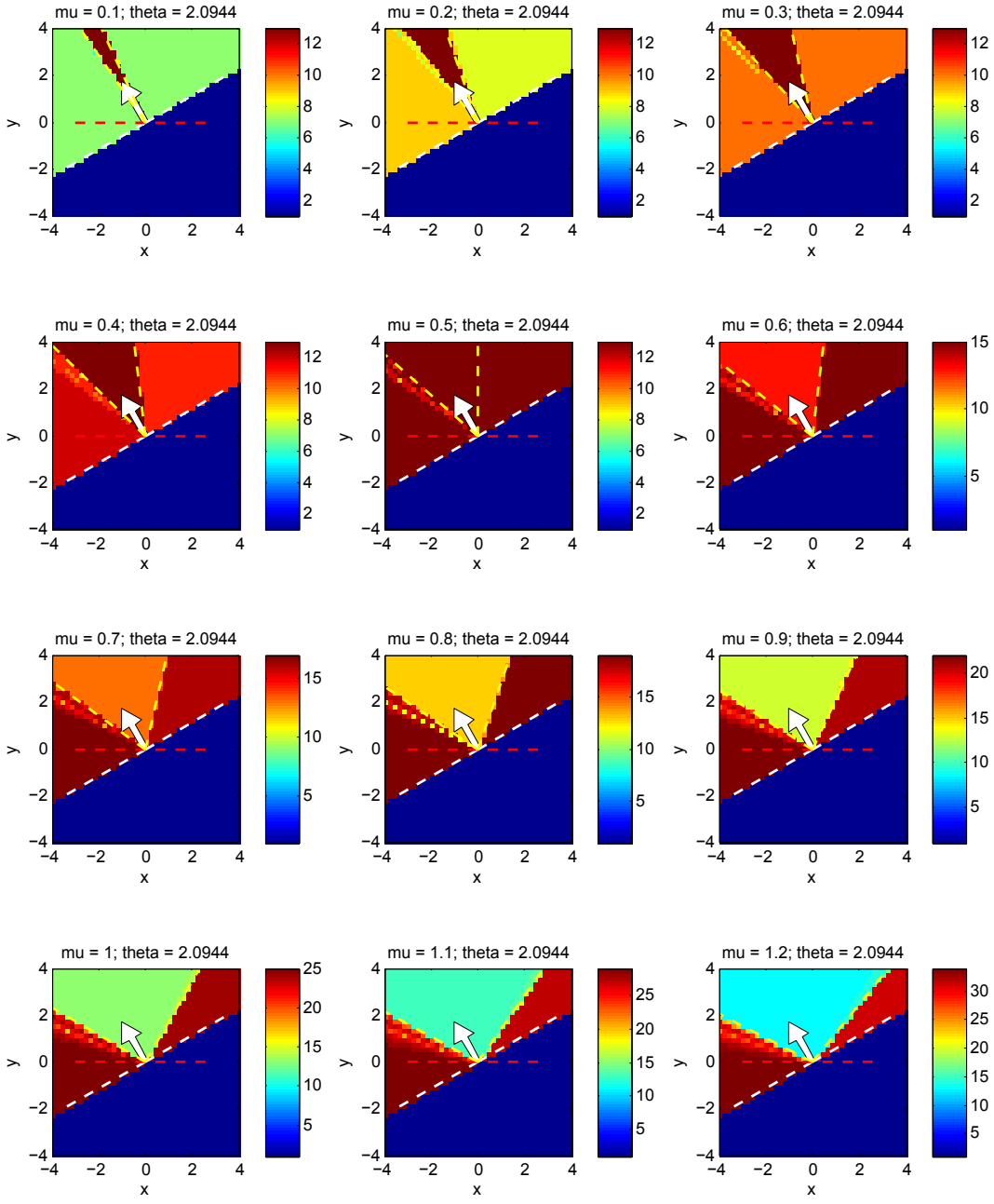


Figure 6.3: **Painlevé Convergence Example:** keeping $\theta = \frac{2\pi}{3}$ fixed, we varying μ over a range of realistic values and examine the number of iterations to converge over the domain of possible predictor points. The residual was $\sim 10^{-6}$ over the domain.

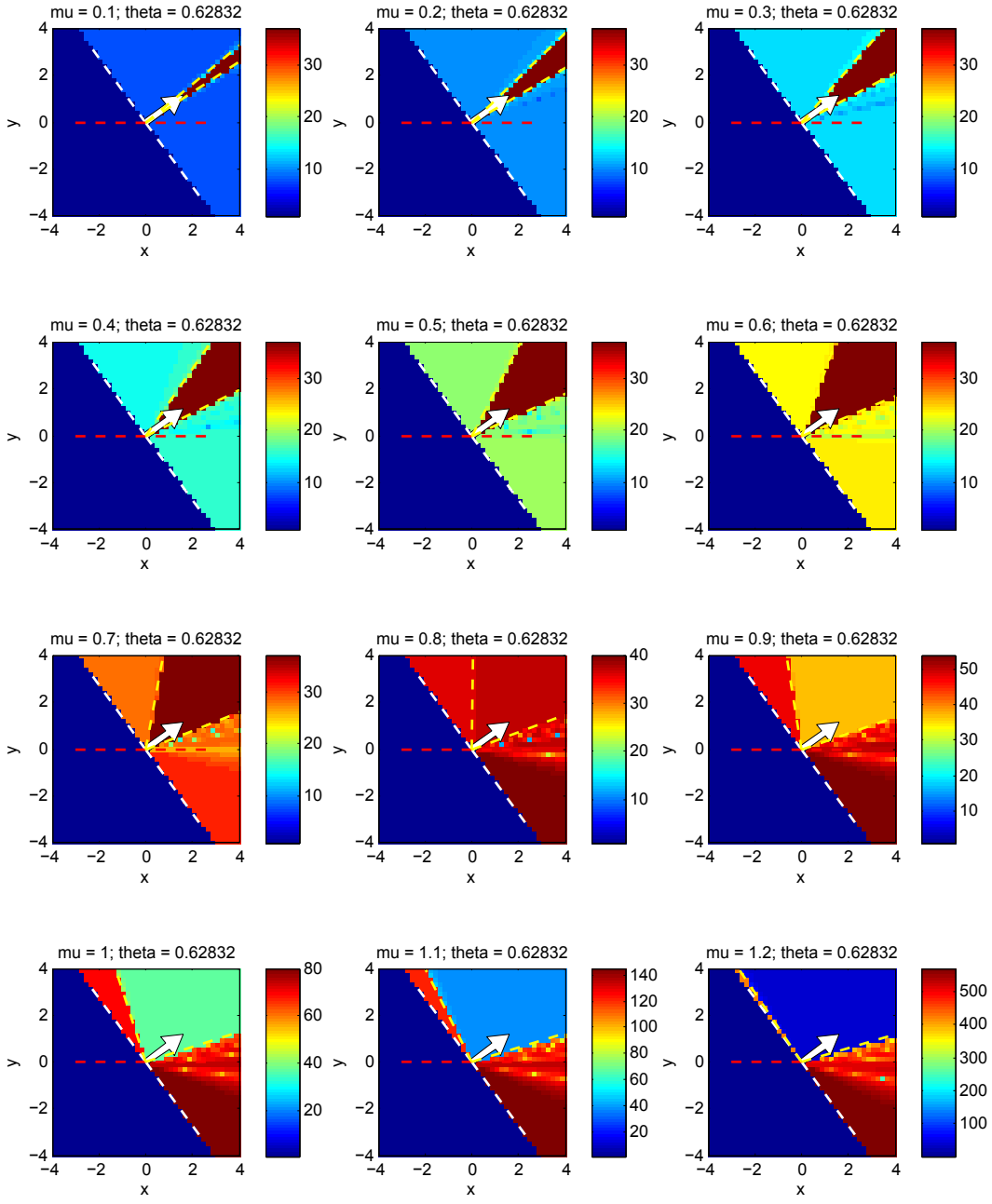


Figure 6.4: **Painlevé Convergence Example:** keeping $\theta = \frac{\pi}{5}$ fixed, we varying μ over a range of realistic values and examine the number of iterations to converge over the domain of possible predictor points. The residual was $\sim 10^{-6}$ over the domain.

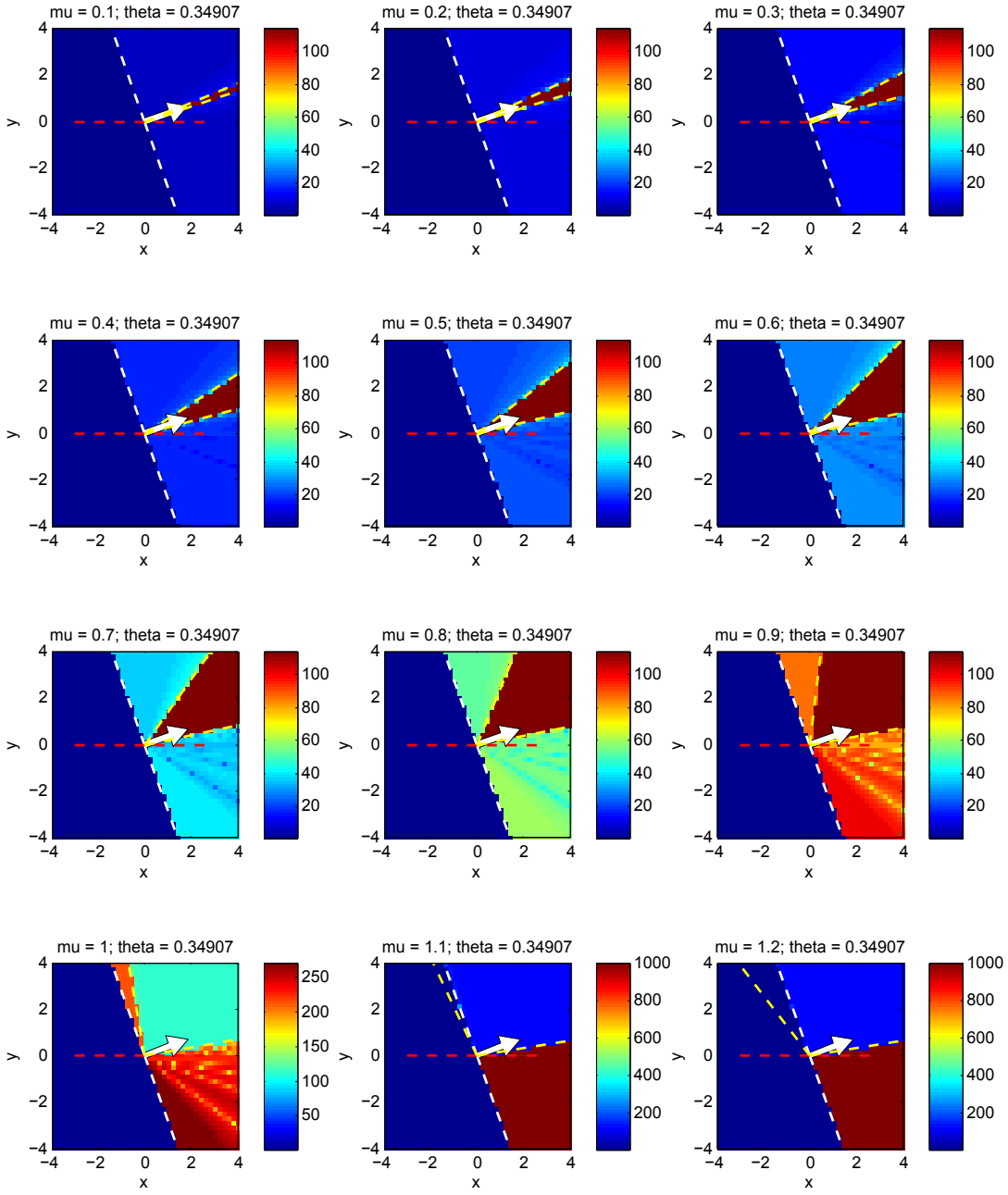


Figure 6.5: **Painlevé Convergence Example:** keeping $\theta = \frac{\pi}{9}$ fixed, we varying μ over a range of realistic values and examine the number of iterations to converge over the domain of possible predictor points. The residual was $\sim 10^{-6}$ over the domain.

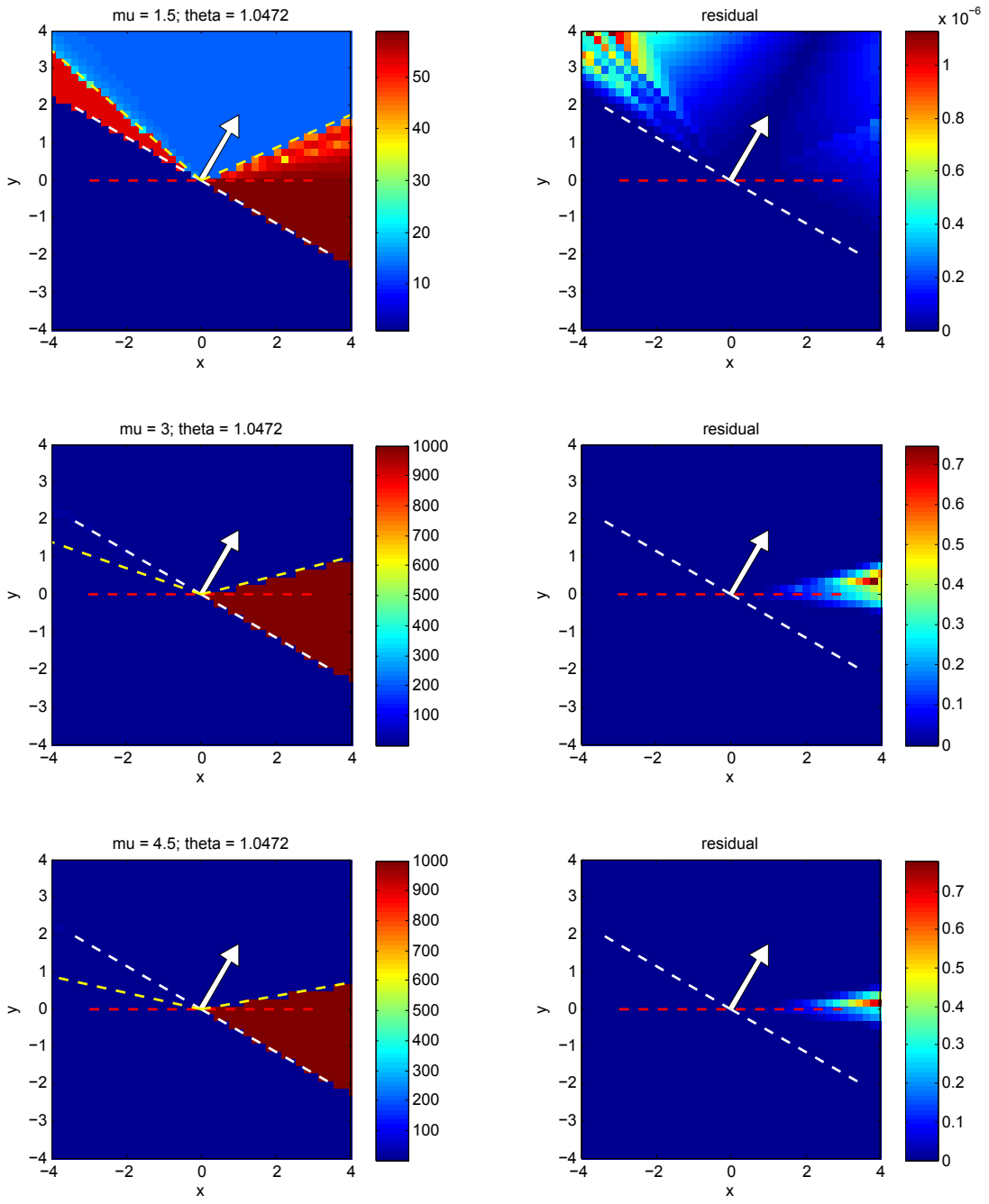


Figure 6.6: **Painlevé Convergence Example:** keeping $\theta = \frac{2\pi}{3}$ fixed, we varying μ over a range of larger friction values. In the left column we show the number of iterations to converge, while in the right column we show the corresponding absolute value residual.

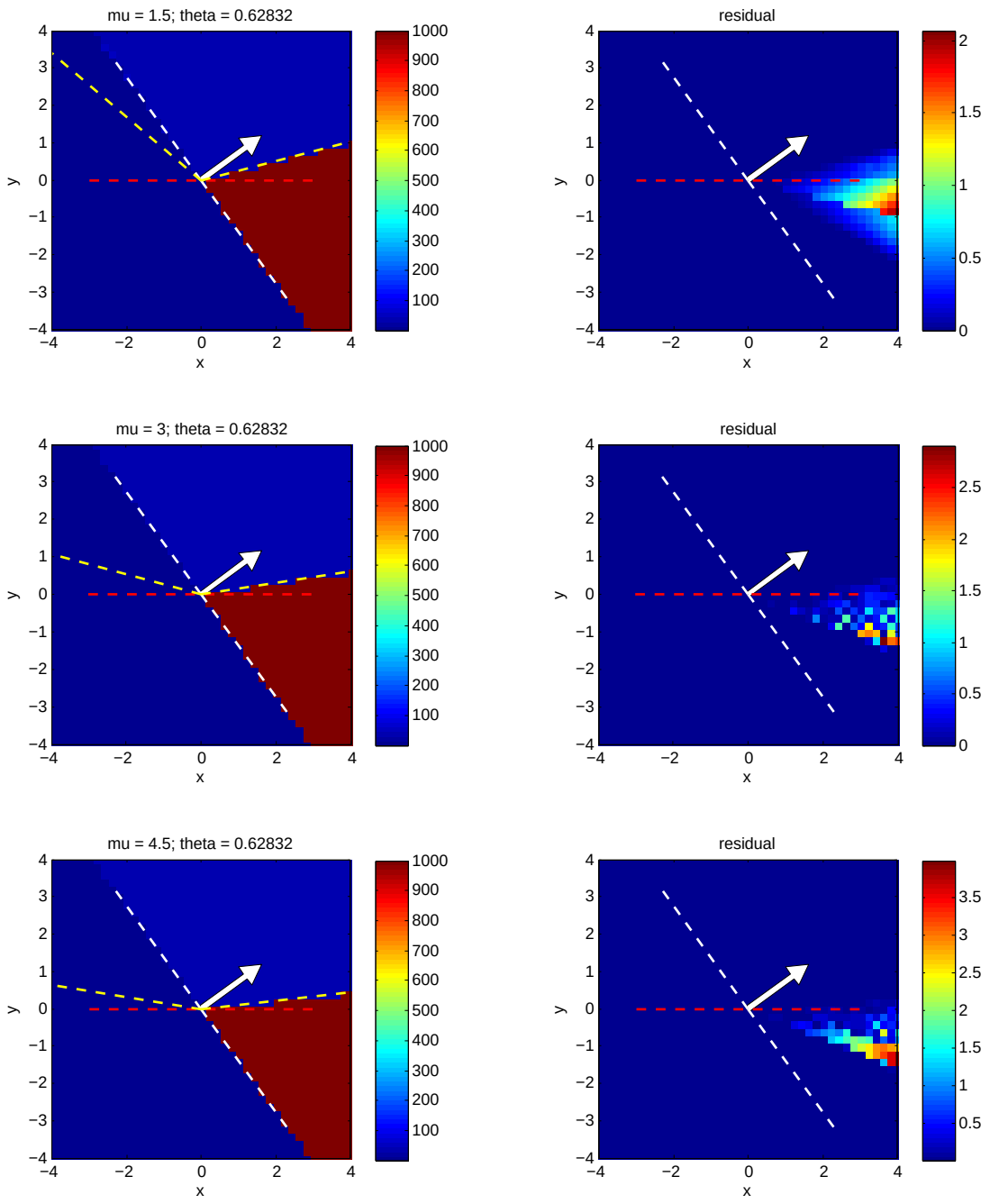


Figure 6.7: **Painlevé Convergence Example:** keeping $\theta = \frac{\pi}{5}$ fixed, we varying μ over a range of larger friction values. In the left column we show the number of iterations to converge, while in the right column we show the corresponding absolute value residual.

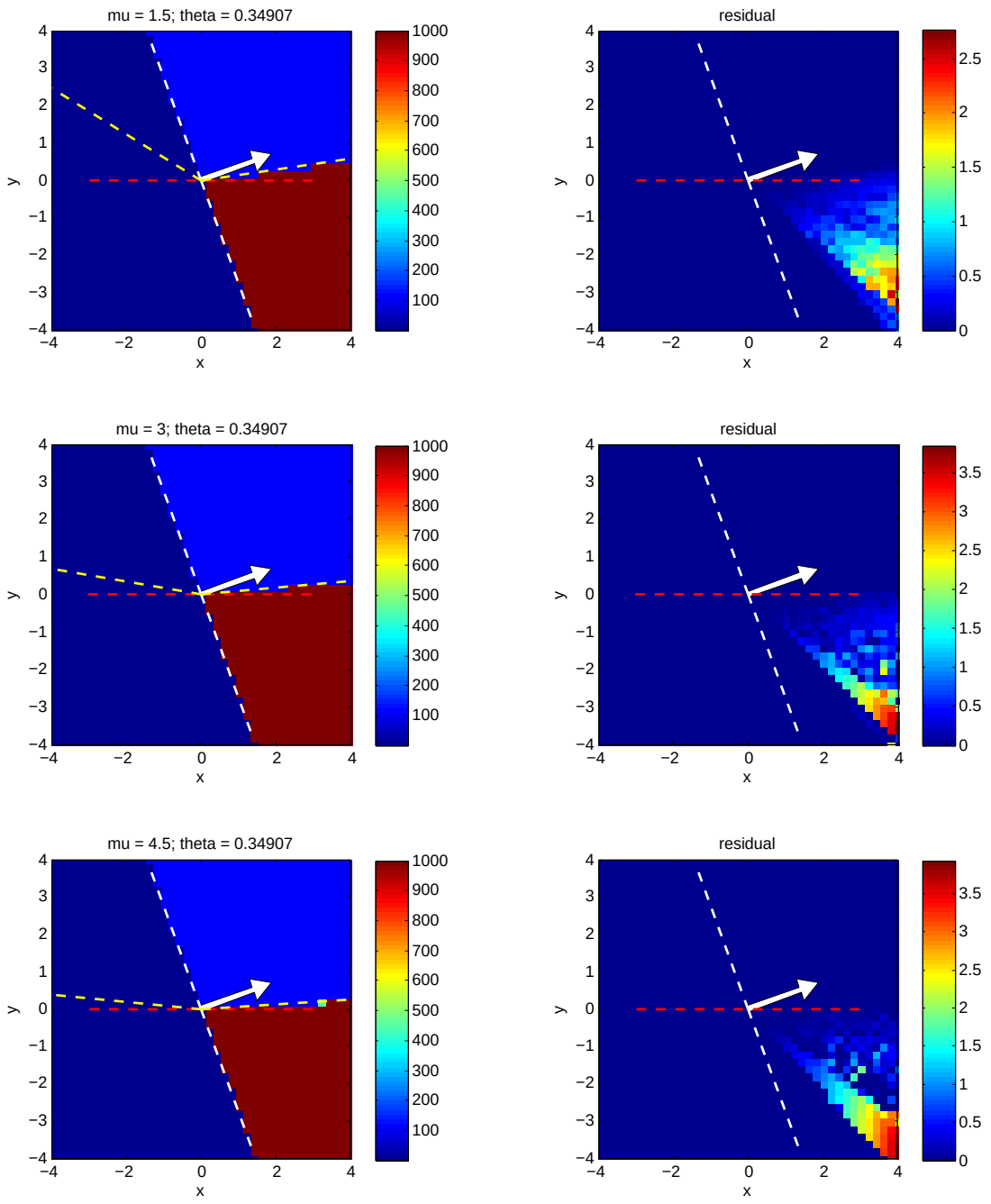


Figure 6.8: **Painlevé Convergence Example:** keeping $\theta = \frac{\pi}{9}$ fixed, we varying μ over a range of larger friction values. In the left column we show the number of iterations to converge, while in the right column we show the corresponding absolute value residual.

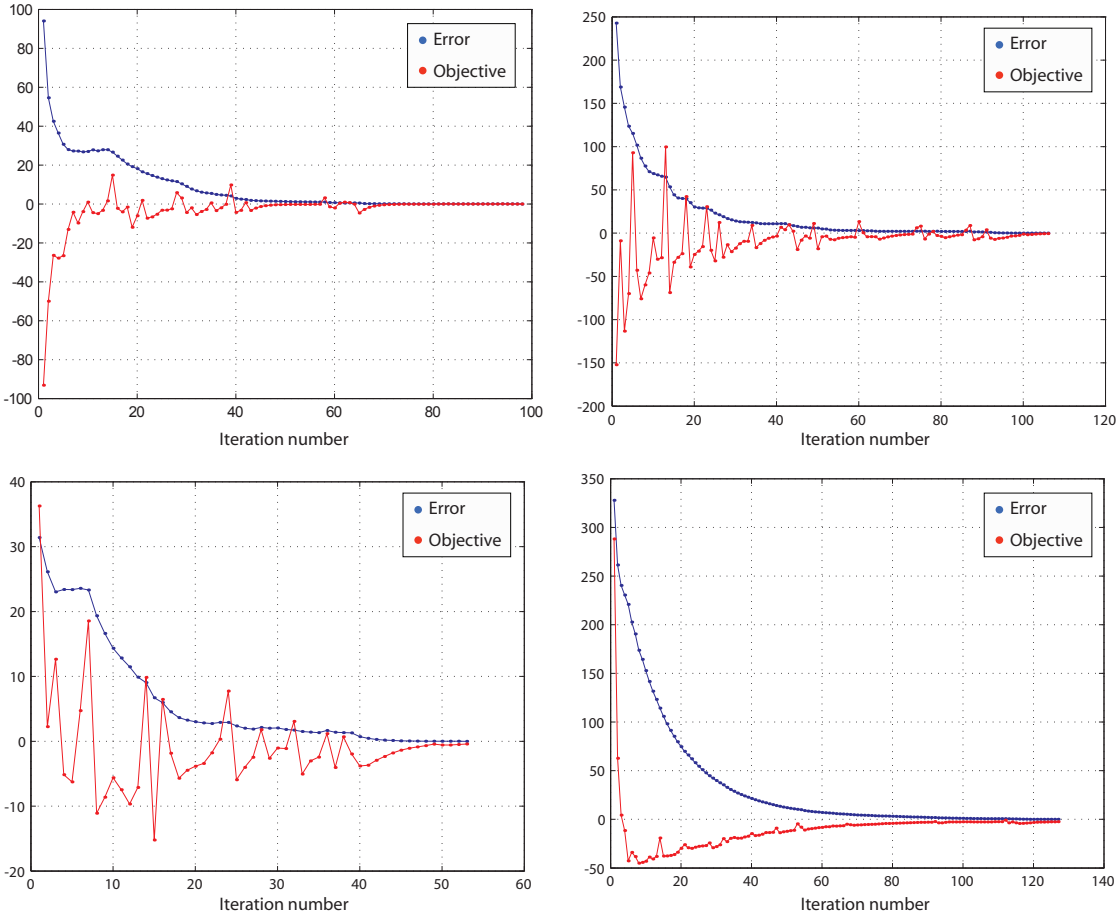


Figure 6.9: **Staggered Projections solves** from the first four time steps of the four level rigid card house example. Red plots the objective value of the corresponding nonconvex minimization which will be zero at an exact solution. Blue plots indicate reanalysis error to the global minimizer.

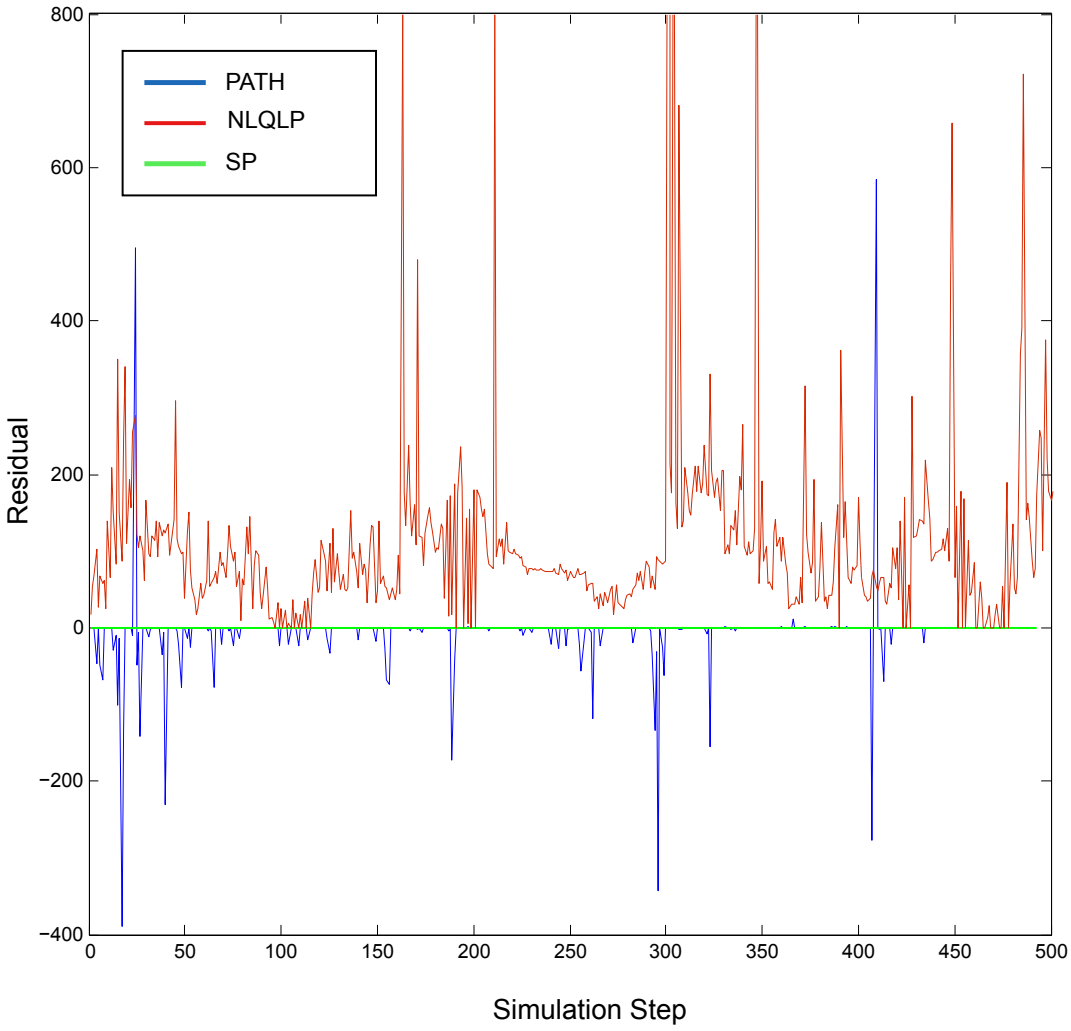


Figure 6.10: **Algorithm Comparison:** This plot shows a comparison of the residual values, for each of 500 simulation steps of the standing card house example, obtained by the Staggered Projections (green line), Lemke's (blue line) and Sequential Quadratic Programming (red line) algorithms. The average time for a call to Staggered Projections, Lemke's (implemented using PATH), and Sequential Quadratic Programming (implemented using NLQLP) over these simulations, was 0.8 seconds, 57 seconds, and 123 seconds respectively.

Chapter 7

Applications and Examples

7.1 Overview

In this Chapter we investigate the use of Staggered Projections in several applications. We also investigate the plausibility and validity of the behavior of the obtained simulations and compare with several popular frictional contact simulation packages.

7.2 Implementation

All timings in the following sections were measured on a PC with a 3 GHz Intel Pentium D processor, 2G RAM. We used Java (JDK 1.6) for our implementation. Note that while faster timings could be obtained via parallelization (Section 7.4) all simulations were run single threaded to get accurate timings.

In all of the following applications we simulate contacting systems composed of large-deformation reduced St.Venant-Kirchhoff (StVK) models [Barbič & James 2005] and/or linear modal models [Shabana 1991], both augmented with a floating frame [Shabana 2005]. We also include rigid body models. In our implementation, narrow-phase collision detection is performed using a sphere-based Bounded Deformation Tree (BD-Tree) [James & Pai 2004] for reduced deformable bodies, a customized bounding-sphere hierarchy for rigid bodies [Kaufman, Sueda, & Pai 2007], and a uniform subdivision for broad-phase culling. No stabilization was employed. Contact samples are generated by sampling the base geometries of deformable bodies, and the zero level set of pre-computed distance fields of rigid bodies [Kaufman, Sueda, & Pai 2007]. To facilitate comparison with existing packages, we wrote a Java wrapper for ODE's [Smith 2006] box/box collision detection and contact sampling implementation. We use this to compute narrow-phase collision detection and contact sampling between box-based

rigid body geometries.

A variety of numerical integrators can be used to both advance the unconstrained state and to generate predictor velocities during contact. In this example implementation, we use a hybrid implicit-explicit (IMEX) integration scheme that is explicit in rigid motion and the quadratic velocity forces g_{qv}, but implicit in the internal deformation forces (and damping). For the latter, we use the semi-implicit Newmark subspace integrator of Barbič & James [2005].

All convex QP solves were implemented using QL [Schittkowski 2005], a robust implementation of the dual active set algorithm of Goldfarb & Idnani [1983]. QL is a dense solver so speedups could potentially be obtained with sparse solvers.

7.3 Validation Tests

We tested our algorithm on a wide range of difficult frictional contact scenarios to verify that we obtain plausible and expected frictional contact behavior.

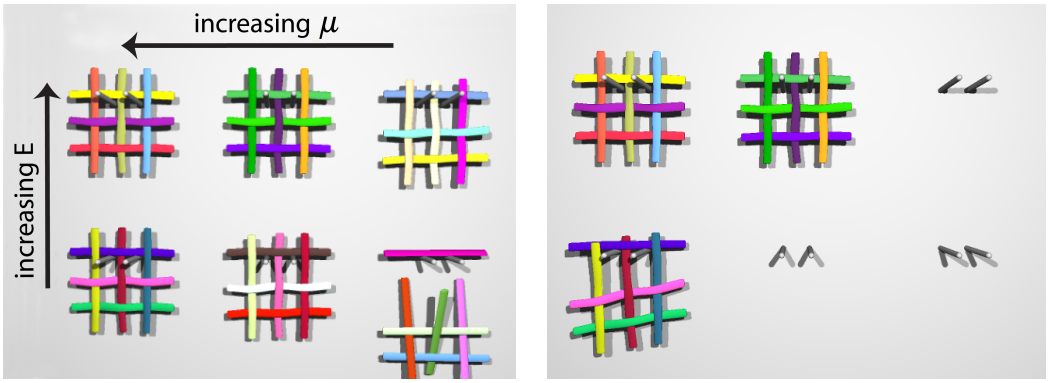


Figure 7.1: **Hanging woven elastic-frictional composites:** We increase stiffness (bottom to top) and the coefficient of friction (right to left). The woven composites come apart for all cases where the combined friction and stiffness are insufficient to resist gravity.

7.3.1 Elastic-Frictional Composites

We wove flattened strands of a stiff rubbery material into frictional composites. These structures demonstrate the importance of tight coupling between friction and deformation. In our first experiment we hang a grid of composites on pegs. We vary μ, the coefficient of friction, right to left, with the increasing values $0.1, 0.3$, and 0.5, and increase stiffness slightly for the upper row. On the top row the two leftmost composites stay cohered, due to a combined high friction and stiffness, while the right hand composite comes apart due to low friction. On the

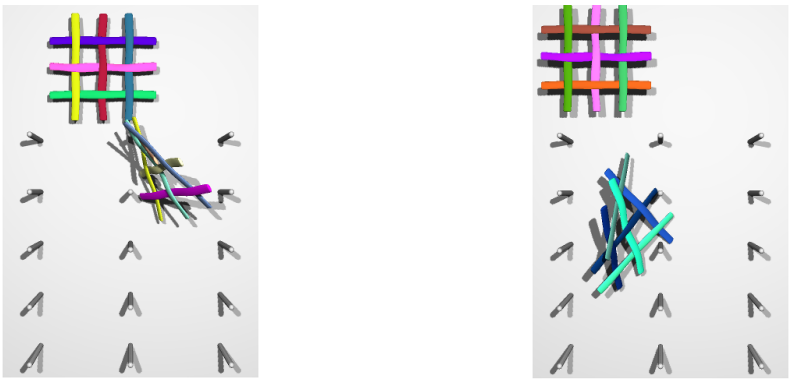


Figure 7.2: **Thrown woven elastic-frictional composites:** We throw stiff composites, with $\mu = 0.5$, at a peg board. The composite's cohesive behavior varies with the type of the impact.

bottom row the two rightmost composites come apart quickly due to a combined low friction and stiffness, while the left composite slowly creeps apart due to low stiffness but high friction. See Figure 7.1. In our second example we throw stiff composites, with $\mu = 0.5$, at a peg board. The composite's cohesive behavior varies with the type of the impact, in some cases deforming as a whole, while for other impacts coming apart quickly.

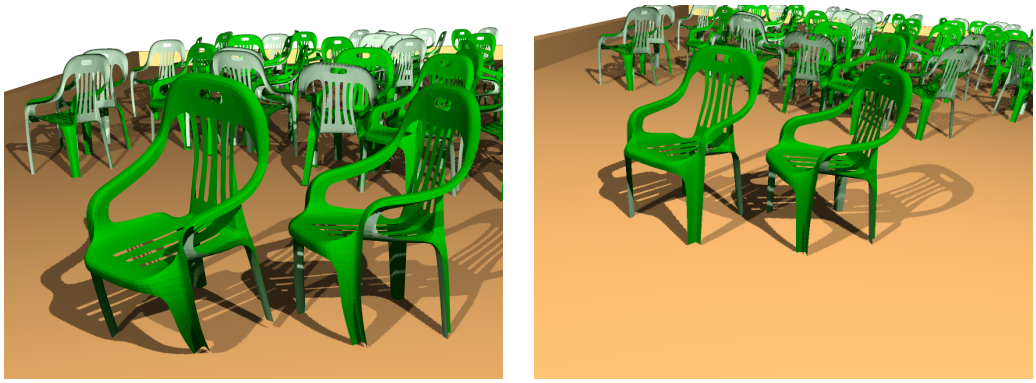


Figure 7.3: **Stick-slip instability:** Plastic chair legs chatter while sliding on an inclined surface due to fictionally induced oscillations. This behavior is obtained from accurate solutions of the coupled friction, contact and deformation modes.

7.3.2 Stick-Slip Instability

Stick-slip oscillation is an important frictionally induced instability in deformation dynamics. High friction during sliding creates a buildup of elastic energy in contacting systems. This energy is partially stored in the global deformation of the system, but also builds up at contacting interfaces. Elastic energy is then released suddenly when the magnitude of the friction

force, opposing sliding contact, is exceeded by the tangential stiffness at the contact. This interaction between the friction forces and the sliding velocities is generally periodic, and so can induce self-excited oscillations that involve the buildup and dissipation of significant amounts of energy. In this example we show that accurate solutions of the coupled friction, contact and deformation modes in our algorithm capture the stick slip induced instability behavior in the chattering legs of plastic chairs sliding on an inclining surface. See Figure 7.3.

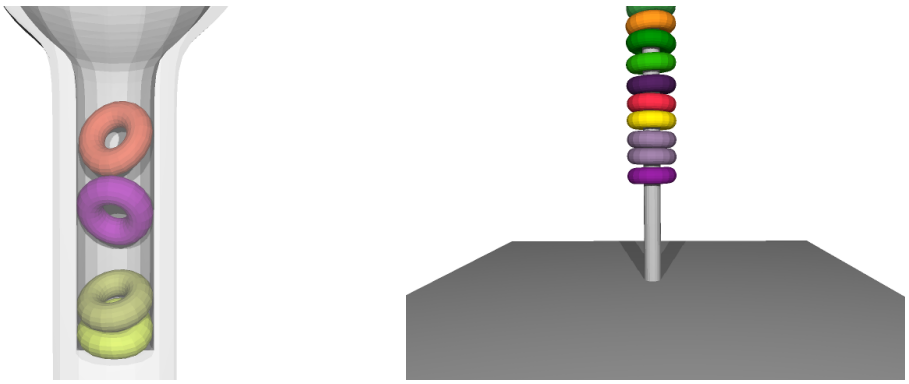


Figure 7.4: **Jamming:** Stable deformable jamming is obtained for a range of tight fitting configurations.

7.3.3 Jamming

Another behavior that arises due to deformation and friction is *jamming*. Stable deformable jamming is another difficult frictional phenomenon to simulate. We dropped deformable and rigid objects into tight configurations with varying coefficients of friction, obtaining stable and plausible changes in jamming behavior as μ decreases. See Figure 7.4 for examples. We also demonstrate robust and stable deformable frictional jamming at real-time rates in our haptics examples. See Section 7.8, below.

7.3.4 Friction-Dependent Masonry

For suitably high coefficients of friction, complex masonry structures, such as arches, can be constructed *without* adhesives. Difficulties in accurately simulating such structures are discussed in Section 2.8.1. We stably simulate a catenary arch using rigid blocks with $\mu = 0.6$. We first drop blocks onto the arch's keystone (the most stable point in the arch) and note that the structure deforms globally but does not fall apart; instead it finds a new stable equilibrium.

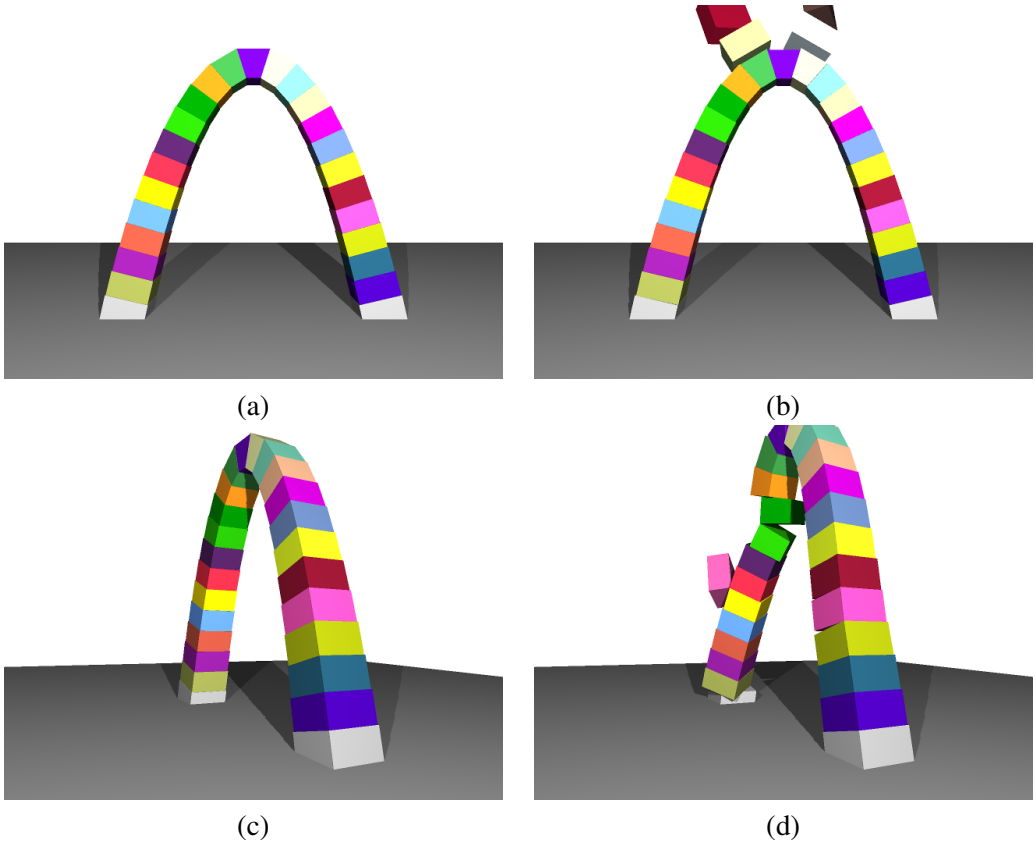


Figure 7.5: **Friction-Dependent Masonry:** A stable Catenary arch is constructed in this simulation with high friction ($\mu = 0.6$), and without adhesives. **(a), (b):** We drop blocks onto the arch's keystone and it remains standing. **(c), (d):** We throw blocks at one of the arch's legs (a less stable point) and the arch falls apart.

We then throw blocks at one of the arch's legs (a less stable point) and the arch falls apart. See Figure 7.5.

7.3.5 Rigid Card House

In this example, a stable house of cards is made out of rigid cards with high friction. We use $e = 0.1$ and $\mu = 0.8$ for all cards. The card house stands stably long term and then, under successive impacts from small blocks, it repeatedly has sections fall down and then regains equilibrium (see Figure 7.6). Difficulties in accurately simulating these behaviors are also discussed in Section 2.8.1.

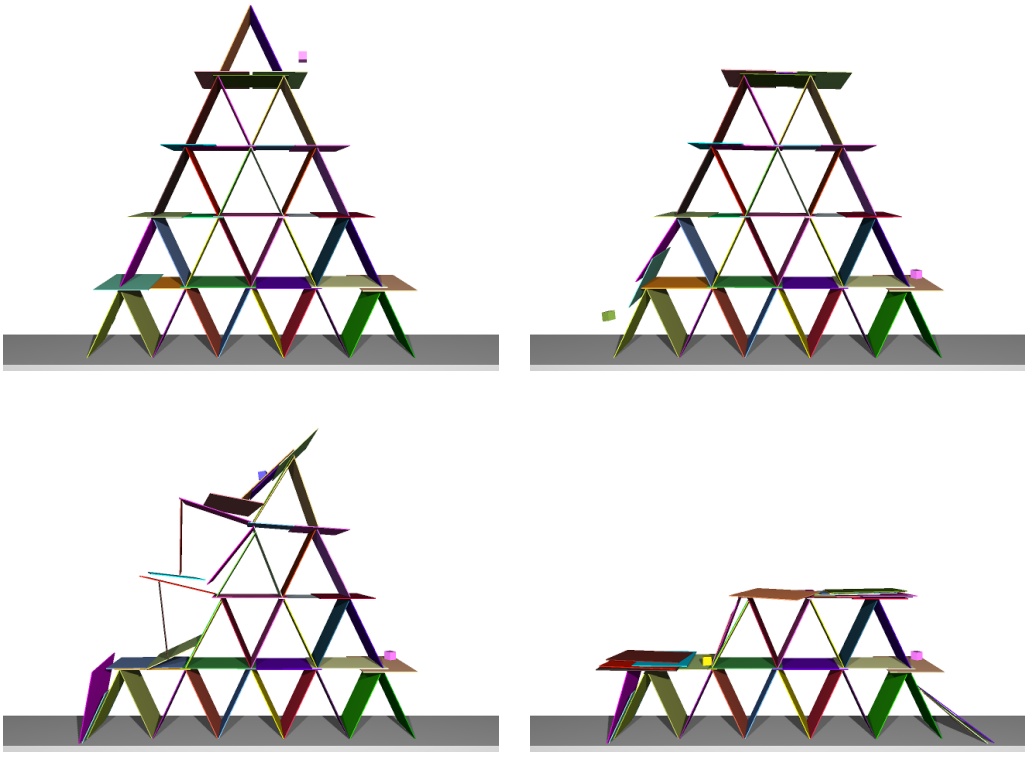


Figure 7.6: **Rigid Card House Example:** A rigid card house ($\mu = 0.8$) is initially stable and then partially knocked down by dropping blocks.

7.3.6 Large Deformation Frictional Contact

In these examples we simulate large deformation frictional contact using reduced StVK models. In the first example dinosaurs are dropped together undergoing fast impacts with $\mu = 0.5$ (see Figure 7.7). In a second example we drop two groups of bunnies, with differing coefficients of friction, onto a ramp and then into collision with semi-circular arches composed of rigid blocks. The red bunnies use $\mu = 0.4$, blue bunnies, $\mu = 0.1$ and the rigid arch blocks, $\mu = 0.5$ (see Figure 7.8). The bunnies exhibit differing sticking, and stick-slip sliding behaviors depending on the incidence of impact and μ. In both examples the models end the simulation in stable piles.

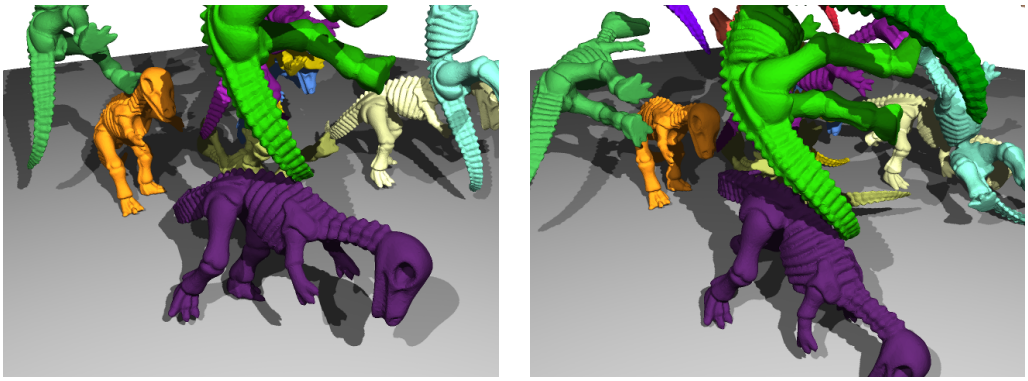


Figure 7.7: **Nonlinear Deformation Example:** A nonlinear deformation example using reduced, deformable StVK dinosaur models.

scene	solve (s)	contacts	iters.	DoFs	time-step (s)	models
composites	0.56	921	6.4	192	3e-3	modal
chairs	0.72	9,025	2.9	1,766	1e-2	modal
catenary arch	1.2	2,042	5.6	162	1e-2	rigid
card house	4.7	528	3.2	348	1e-3	rigid
dinosaurs	0.61	4,690	2.7	315	1e-2	StVK
haptic	0.0032	117	2.4	72	1e-2	modal & rigid

Table 7.1: **Performance Evaluation:** This table summarizes the average solve time, number of contacts and iterations, per simulation step, for the validation examples. The tangent sample size was 8 for all friction solves.

7.4 Algorithm Behavior

Stability and Long Term Integration

To test long term integration and stability properties, all of the above validation examples were stepped for 10 minutes of simulation time. All simulations ran without blow-up, or constraint drift other than possibly initial (small) penetrations sometimes caused by the collision of sharp asperities (bunny ears, cow horns, etc.) during the fast phases of impacts.

7.5 Algorithm Comparison

While in the last chapter we compared and discussed the results of side by side tests of Staggered Projections with industrial grade optimization solvers, all of which use the same base

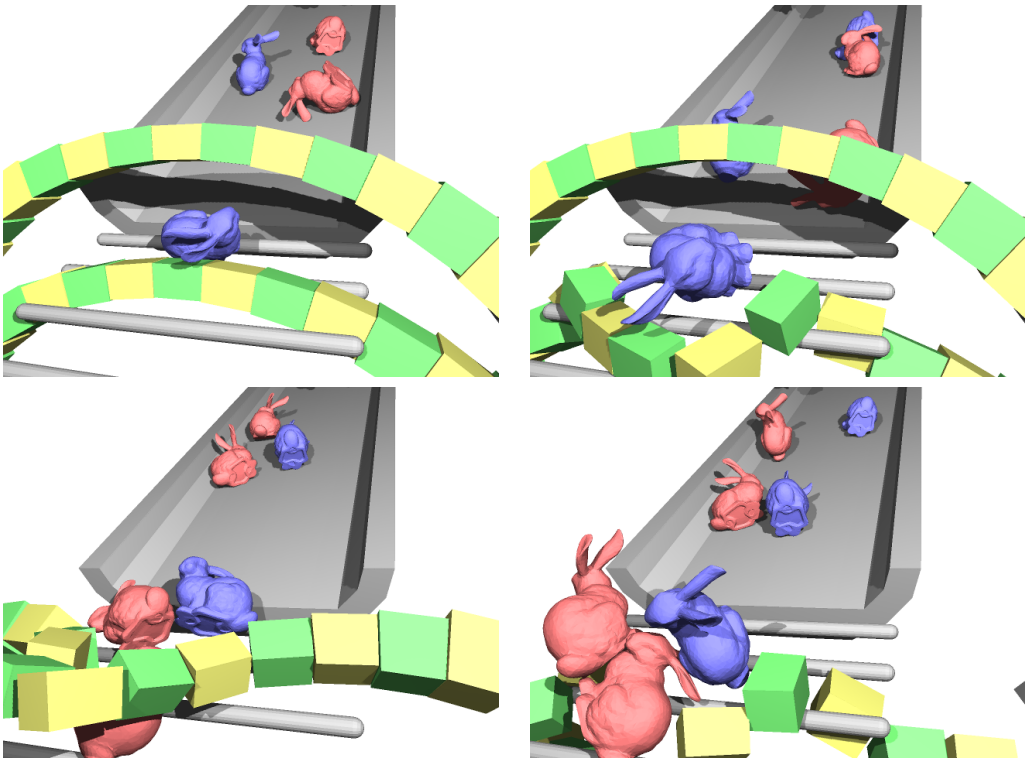


Figure 7.8: **Hybrid Scene Example:** Reduced deformable StVK bunny models with varying coefficients of friction collide with each other and with initially stable semi-circular arches composed of rigid blocks.

code, here, in this section, we will compare the performance of the Staggered Projections implementation with two popular contact simulation packages, ODE and Open-Tissue. ODE Smith [2006] is an open source project and is perhaps the most popular contact resolution package in general use. We will also include our the direct (Lemke's) LCP solver implementation for a second base-line comparison here as well.

7.5.1 Comparison with LCP-based Frictional Contact Simulation Packages

For the following examples we implemented our base collision and contact sampling code with the Stewart-Trinkle [Stewart & Trinkle 1996] velocity-level LCP using the PATH solver [Ferris & Munson 1998]. We also tested examples with the Projected Gauss-Seidel based iterative LCP solvers in the ODE [Smith 2006] and the Open-Tissue [Erleben 2007] packages. In these

examples, all methods employ semi-implicit Euler for non-contacting integration. For collision detection and contact sampling direct LCP, Staggered Projections, and ODE apply ODE's box/box implementation; Open-Tissue applies a box/box variant.

Stacking: Both the ODE and Open-Tissue iterative solvers generate sidewise sliding errors for simple rigid stacking block examples. While these errors build more slowly in Open-Tissue, for both systems this results in stacks falling over after a few seconds. We find that this occurs even when large iteration counts (100) and very small step sizes (10^{-4}) are applied. If we make the influence of dynamics negligible in Open-Tissue, however, by setting Open-Tissue's fraction parameter to ~ 0.002, and thus make Open-Tissue's stabilization method, shock-propagation, almost entirely dominant, we can obtain stable stacks of blocks. The direct LCP slowly generates stable stacking for small numbers of blocks (≤ 15) but can not scale, and begins failing for larger problems. This is consistent with the results reported in Anitescu & Hart [2004]. Finally, the Staggered Projections implementation generates long term stable stacking [1] for both rigid and deformable bodies.

Rigid Card House: We also attempted to duplicate the card house example from the validation tests using both direct and iterative LCP methods. Direct LCP generates, very slowly, a stable solution for a small, two-level card house, but again, does not scale and fails on card house examples of larger sizes. ODE's iterative solver fails for card houses at all sizes. Again this is caused by a sidewise sliding error, no matter how high we set friction. Open-Tissue's iterative solver fails similarly. If we, in addition, also apply Open-Tissue's shock-propagation based stabilization, some of the sliding error is reduced. However, vibratory artifacts induced by the stabilization method still cause the Open-Tissue solver to fail on card houses of all sizes.

7.6 Convergence

While, as discussed in the last chapter, we do not guarantee convergence, all validation examples obtained convergent solutions. All of the above validation examples were run to convergence with a tolerance of $\varepsilon = 10^{-4}$. A tolerance of $\varepsilon = 10^{-1}$, however, was generally sufficient to generate convincing frictional behavior, while convergence was also tested for tolerances

[1] These S.P. simulations were also run out to 10 minutes of virtual time.

down to $\varepsilon = 10^{-8}$. In general, setting ε in this range allows the user detailed control over the accuracy of the frictional response obtained by the algorithm.

scene	warm	w/o warm
composites	6.4	25.8
chairs	2.9	30.9
catenary arch	5.6	133.7
card house	3.2	57.1
dinosaurs	2.7	18.3

scene	solve(s)	DoFs
log house	4.9	414
log house dec.	0.061	414
bunny drop	37.5	12,276

Table 7.2: **Convergence Behavior (left):** The average number of iterations for the validation examples. **Decomposition Performance (right):** The average solve time for the decomposition examples.

7.7 Warm Starts

We note an appreciable acceleration in convergence when warm starting is applied in these examples. In Table 7.2 (left) we summarize the number of iterations to convergence for the validation examples *with* and then *without* warm starts. In many cases a speed-up of up to two orders of magnitude was obtained. Warm starting also generated faster simulations in examples were a hard maximum iteration limit was enforced. In these simulations convergence may not be obtained for individual steps. When temporal coherence is sufficiently high, however, warm starting appears to help convergent behavior emerge over multiple steps.

(a) (b)

Figure 7.9: **(a):** A $\sim$ 100 Hz, force-feedback, haptic simulation of deformable and rigid frictional contact using Staggered Projections. **(b):** A preliminary decomposition approach as validated on a high-dimensional, high-contact count scene, composed of two-thousand rigid-body models.

7.8 Haptic Rate Interaction

We also implemented a highly unoptimized prototype force-feedback haptic rendering of deformable and rigid frictional contact interactions by plugging a Phantom Premium 1.5 haptic device directly into our existing integration cycle. A timing and convergence summary for the interaction session depicted in Figure 7.9 (a) is given in Table 7.1, bottom. Note that, here, hard time limits required us to implement graceful degradation. Even so we found that most ($\sim 80\%$) of the integration steps were still convergent and that complex frictional behaviors such as jamming, sticking and stick-slip were obtained. See Figure 7.9 (a).

7.8.1 Future Work

While the timings for the haptic demo ($\sim$ 100 Hz.) are sufficient for rendering soft contact, a separate code base specifically optimized for haptic rendering should be able to scale the timings down to enable haptic update rates of 1000 Hz.

7.9 Decomposition

Without constraints, each body in a multibody system can be integrated independently. When contacts are imposed, however, the DoFs of individual bodies are effectively "glued" together. Generally, though, multiple *independent* connected sets of bodies are formed by contact constraints, each of which can be solved separately, and, if parallel processing is available, each such set can be solved in parallel. This is a standard feature in multibody packages (e.g., "islands" in ODE Smith [2006]).

Additional decompositions can be applied to Staggered Projections steps to extract further sparsity. This allows us to effectively decompose the contacting system down to even smaller independent sets. Initial investigations reveal the potential for large speed-ups. While *all* of the above examples and timings provided were obtained *without* decomposition, we implemented two examples with a preliminary decomposition approach. We constructed a "Lincoln Log" scene composed of a rigid log house frictionally sliding, jumping a ramp and then colliding. The scene was simulated twice, once with decomposition and the second time without. We also simulated a larger drop of rigid bunnies using decomposition (see Figure 7.9 (b)). The average

solve time, per step, and the number of DoFs per scene for these three simulations is given in Table 7.2 (right).

Chapter 8

Conclusion

In this thesis we have reformulated time-continuous, frictional contact dynamics as a pair of coupled variational principles to explicitly partition the frictional contact problem into two naturally convex subproblems. Using this formulation we exposed and clarified the two mechanisms by which these principles are coupled:

1. Coulomb-type friction constraints,

2. Non-orthogonality of contact and friction subspaces.

We then suggested that both of these coupling mechanisms are fundamental sources of difficulty in accurate frictional contact simulation and showed that the removal of either one makes the frictional contact resolution problem in many senses much easier.

As a useful measure of the degree of this coupling we then introduced the minimum subspace angle as a generalization of the conditions that allow for the so-called Painlevé Paradox. We suggest that these generalized "Painlevé Conditions", along with the coefficients of friction, are important measures of the potential degree of difficulty of related frictional contact simulation problems.

We then motivated the desirability of obtaining discrete variational integrators for frictional contact simulation and discussed known problems and issues with existing variational integration methods for this problem domain. Combining our Coupled Principles Formulation with discrete variational integration methods we then derived new variational integrators that pose the discrete frictional contact problem as a system of coupled minimizations. We showed, by discussion and example, that solutions to these systems avoid known issues with existing variational integration approaches for friction and contact, and capture accurate conservative and dissipative behavior.

We then focused on a specific two-step variant of these variational schemes. We showed how this variant generates a simple generalization of the popular Stewart-Trinkle model [Stewart & Trinkle 1996] and thus essentially requires the solution of the same underlying Linear Complementarity Problem (LCP). Applying the coupled minimization perspective, we suggested that difficulties that have been encountered in solving these LCP systems can be explained, in part, by noting that optimization problems, that are obtained from these frictional contact formulations, are equivalent to nonconvex problems in *global* optimization (which are generally NP-hard).

We then illustrated, by discussion and example, that existing solution methods, that have generally been presumed suitable for solving these contact-related optimization problems, fail entirely for many important examples of frictional contact and addressed these limitations with our Staggered Projections algorithm. Deriving a fixed-point scheme from our Coupled Principles formulation we obtained the Staggered Projections algorithm and showed that it effectively generates accurate solutions to optimization problems for many frictional contact problems that were previously impractical to solve.

To better understand the suitability of Staggered Projections for practical applications and its potential limitations, we performed a detailed convergence analysis of the algorithm. We observed that the generalized Painlevé Condition, introduced earlier, plays a fundamental role in the convergence behavior of Staggered Projections and helps to clarify its good behavior in many cases.

Finally, by offering simulations and instrumented examples, we validated that Staggered Projections performs well in practice; capturing convincing and accurate frictional contact behaviors for both rigid and large deformation models. We also compared our Staggered Projections implementation with other available popular frictional contact solvers and presented sample applications of Staggered Projections for both simulation and haptics.

8.1 Future Work

The variational methods we have discussed and validated in this thesis are still in their early stages. While we have shown that the proposed methods already improve on existing approaches for discrete variational integration of frictional contact, many practical numerical issues for efficiently scaling these methods to large scale systems still remain to be addressed. Other interesting directions to investigate would be the inclusion of viscoelastic and plastic models into the formulation. In some senses the addition of inelastic models into discrete variational integration formulations might be somewhat direct, as suggested by Pandolfi *et al.* [2002]; however, our sense is that many new and challenging problems might need to be addressed.

As we discuss in our convergence analysis in Chapter 6, for combined high coefficients of friction and small subspace angles, we can not necessarily expect Staggered Projections to converge in all cases. While to some degree the fail-safe option of graceful-degradation potentially mitigates this possibility, understanding such cases better, and formulating efficient strategies for handling them, is also an important ongoing project. A related issue is finding additional methods to improve convergence speeds for slower cases and general improved scaling to larger problems for increased efficiency.

While we have shown that Stewart-Trinkle-type problems require the solution of nonconvex problems in *global* optimization, an interesting question is whether all accurate numerical methods for discrete frictional contact do as well. In other words, can we expect both nonconvexity *and* global optimization to always be features of accurate numerical methods for frictional contact, or are they just artifacts of certain discretization choices? This is an important question as we attempt to construct new methods for accurate frictional contact simulation. Our current research, based on the Coupled Principles Formulation, suggests that nonconvex, global optimization, is in fact an intrinsic, unavoidable feature of discrete frictional contact simulation; this is still, however, a subtle and difficult question to fully address and an area of ongoing investigation.

Bibliography

Amontons, G. 1699. De la resistance caus'ee dans les machines.

Anitescu, M., and Hart, G. D. 2004. A fixed-point iteration approach for multibody dynamics with contact and small friction. *Mathematical Programming* 101(1):3–32.

Anitescu, M., and Potra, F. R. 1997. Formulating dynamic multirigid-body contact problems with friction as solvable linear complementarity problems. *ASME Nonlinear Dynamics* 14:231–247.

Baraff, D., and Witkin, A. P. 1992. Dynamics simulation of non-penetrating flexible bodies. In *Computer Graphics (SIGGRAPH 92)*, 303–308.

Baraff, D. 1989. Analytical methods for dynamic simulation of non-penetrating rigid bodies. In *Computer Graphics (SIGGRAPH 89)*, 223–232.

Baraff, D. 1991. Coping with friction for non-penetrating rigid body simulation. In *Computer Graphics (SIGGRAPH 91)*, 31–41.

Baraff, D. 1993. Issues in computing contact forces for non-penetrating rigid bodies. *Algorithmica* 10(2–4):292–352.

Baraff, D. 1994. Fast contact force computation for nonpenetrating rigid bodies. In *Proc. ACM SIGGRAPH 94*, 23–34.

Barbič, J., and James, D. L. 2005. Real-time subspace integration for St. Venant-Kirchhoff deformable models. *ACM Trans. Graph. (SIGGRAPH 05)* 24(3):982–990.

Bauschke, H. 2000. Projection algorithms: results and open problems. In *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, 11–22.

Belytschko, T.; Liu, W. K.; and Moran, B. 2000. *Nonlinear Finite Elements for Continua and Structures*. Wiley.

Boyd, S., and Vandenberghe, L. 2004. *Convex Optimization*. Cambridge.

Bridson, R.; Fedkiw, R. P.; and Anderson, J. 2002. Robust Treatment of Collisions, Contact, and Friction for Cloth Animation. *ACM Trans. Graph. (SIGGRAPH 02)* 21(3):594–603.

Brogliato, B. 1999. *Nonsmooth Mechanics*. Springer-Verlag.

Cadzow, J. A. 1970. Discrete calculus of variations. *Internat. J. Control.* 11:393–407.

Cirak, F., and West, M. 2005. Decomposition contact response (DCR) for explicit finite element dynamics. *International Journal for Numerical Methods in Engineering* 64(8):1078–1110.

Clarke, F. H. 1983. *Optimization and Nonsmooth Analysis*. John Wiley and Sons.

Cottle, R. W.; Pang, J. S.; and Stone, R. E. 1992. *The Linear Complementarity Problem*. Academic Press.

Coulomb, C. A. 1785. Théorie des machines simples, en ayant égard au frottement de leurs parties, et a la roideur des cordages. *Mém. Math. Phys.* 10:161–342.

Dawson, D. 1998. *History of Tribology*. Professional Engineering Publishing.

Duriez, C.; Dubois, F.; Andriot, C.; and Kheddar, A. 2006. Realistic haptic rendering of interacting deformable objects in virtual environments. *IEEE Transactions on Visualization and Computer Graphics* 12:36–47.

Erdmann, M. E. 1994. On a representation of friction in configuration space. *The International Journal of Robotics Research* 13(3):240–271.

Erleben, K. 2007. Velocity-based shock propagation for multibody dynamics animation. *ACM Trans. Graph.* 26(2).

Ferris, M. C., and Munson, T. S. 1998. Complementarity problems in gams and the path solver. *Mathematical Programming Technical Report 98-12*.

Fichera, G. 1964. Problemi elastostatici con vincoli unilaterali: il problems di signorini con ambique condition a contorno. *Mem. Ace. Naz. Lincei* 8(7):91–14.

Galileo, G. 1638. Discorsi e dimostrazioni matematiche, intorno a due nuove scienze.

Goldfarb, D., and Idnani, G. 1983. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming* 27:1–33.

Golub, G. H., and VanLoan, C. F. 1996. *Matrix computations*. Johns Hopkins University Press.

Goyal, S.; Ruina, A.; and Papadopoulos, J. 1991. Planar sliding with dry friction, Part 1. Limit surface and moment function. *Wear* 143:307–330.

Guendelman, E.; Bridson, R.; and Fedkiw, R. 2003. Nonconvex rigid bodies with stacking. *ACM Trans. Graph. (SIGGRAPH 03)* 22(3):871–878.

Hahn, J. K. 1988. Realistic animation of rigid bodies. In *Computer Graphics (SIGGRAPH 88)*, 299–308.

Hairer, E.; Lubich, C.; and Wanner, G. 2002. *Geometric numerical integration: Structure-Preserving Algorithms for Odinary Differential Equations*. Springer.

Harmon, D.; Vouga, E.; Tamstorf, R.; and Grinspun, E. 2008. Robust treatment of simultaneous collisions. *SIGGRAPH 08, ACM TOG*.

Hauser, K. K.; Shen, C.; and O'Brien, J. F. 2003. Interactive deformation using modal analysis with constraints. In *Graphics Interface*, 247–255.

Hertz, H. 1882. On the contact of elastic solids. *J. Reine und. Angewandte Mathmatik* 92(156):156–171.

Irving, G.; Schroeder, C.; and Fedkiw, R. 2007. Volume conserving finite element simulations of deformable models. In *SIGGRAPH 07, ACM TOG*, 13.

James, D. L., and Pai, D. K. 2004. BD-Tree: Output-sensitive Collision Detection for Reduced Deformable Models. *ACM Trans. Graph. (SIGGRAPH 04)* 23(3):393–398.

Jean, M. 1998. The non-smooth contact dynamics method. *Comput. Methods Appl. Mech. Eng.* 155:31–47.

Johnson, K. L. 1985. *Contact Mechanics.* Cambridge.

Jourdan, F.; Alart, P.; and Jean, M. 1998. A gauss-seidel like algorithm to solve frictional contact problems. *Comput. Methods Appl. Mech. Eng.* 155:31–47.

Kane, C.; Repetto, E.; Ortiz, M.; and Marsden, J. 1999. Finite element analysis of nonsmooth contact. *Comp. Meth. Appl. Mech. Engrg. 180 (1999).*

Kane, C.; Marsden, J. E.; Ortiz, M.; and West, M. 2000. Variational integrators and the Newmark algorithm for conservative and dissipative mechanical systems. *International Journal for Numerical Methods in Engineering* 49(10):1295–1325.

Kaufman, D. M.; Sueda, S.; James, D. L.; and Pai, D. K. 2008. Staggered projections for frictional contact in multibody systems. *ACM Transactions on Graphics (SIGGRAPH Asia 08)* 27(5):1–11.

Kaufman, D. M.; Edmunds, T.; and Pai, D. K. 2005. Fast frictional dynamics for rigid bodies. *ACM Transactions on Graphics (SIGGRAPH 05)* 24(3):946–956.

Kaufman, D. M.; Sueda, S.; and Pai, D. K. 2007. Contact trees: Adaptive contact sampling for robust dynamics. In *Technical Sketches, SIGGRAPH 07*.

Kikuchi, N., and Oden, J. 1988. *Contact Problems in Elasticity: A study of variational inequalities and finite element methods.* SIAM Studies in Applied and Numerical Mathematics.

Klarbring, A. 1986. A mathematical programming approach to three-dimensional contact problems with friction. *Comput. Methods Appl. Mech. Eng.* 58:175–200.

Klarbring, A. 1990. Examples of non-uniqueness and non-existence of solutions to quasistatic contact problems with friction. *Ingenieur-Archiv* 56:529–541.

Klarbring, A. 1993. *Computational Methods in Contact Mechanics*. International Series on Computational Engineering. chapter Mathematical Programming in Contact Problems, 233–263.

Logan, J. D. 1973. First integrals in the discrete calculus of variations. *Aequationes Mathematicae* 9:210–220.

Lötstedt, P. 1981. Coulomb friction in two-dimensional rigid-body systems. *Zeitschrift fr Angewandte Mathematik und Mechanik* 605–615.

Lötstedt, P. 1982a. Numerical simulation of time-dependent contact friction problems in rigid body mechanics. *SIAM Journal of Scientific Statistical Computing* 5(2):370–393.

Lötstedt, P. 1982b. Time-dependent contact problems in rigid-body mechanics. *Mathematical Programming Study* 17:103–110.

Maeda, S. 1981. Lagrangian formulation of discrete systems and concept of difference space. *Mathematica Japonica* 27:345–356.

Marsden, J. E., and Ratiu, T. S. 2003. *Introduction to Mechanics and Symmetry*. Springer.

Marsden, J. E., and West, M. 2001. Discrete mechanics and variational integrators. *Acta Numerica* 357–514.

Milenkovic, V. J., and Schmidl, H. 2001. Optimization-based animation. In *Proc. ACM SIGGRAPH 01*, 37–46.

Mirtich, B., and Canny, J. F. 1995. Impulse-based dynamic simulation of rigid bodies. In *Symposium on Interactive 3D Graphics*.

Moore, M., and Wilhelms, J. 1988. Collision detection and response for computer animation. In *Computer Graphics (SIGGRAPH 88)*, 289–298.

Moreau, J. J. 1963. Les liasons unilatérales et le principe de gauss. *Comptes Rendus Academy Science Paris* 256:871–873.

Moreau, J. J. 1966. Quadratic programming in mechanics: One-sided constraints. *Journal SIAM Control* 4(1):153–158.

Moreau, J. J. 1973. On unilateral constraints, friction and plasticity. *New Variational Techniques in Mathematical Physics* 172–322.

Moreau, J. J. 1988. Unilateral contact and dry friction in finite freedom dynamics. *Nonsmooth Mechanics and Applications, CISM Courses and Lectures* (302):1–82.

Moreau, J. J. 1994. Some numerical methods in multibody dynamics: application to granular materials. *European Journal of Mechanics A/Solids* 13(4):93–114.

Murray, R. M.; Li, Z.; and Sastry, S. S. 1994. *A Mathematical Introduction to Robotic Manipulation*. CRC Press.

Murty, K., and Kabadi, S. 1987. Some np-complete problems in quadratic and nonlinear programmings. *Mathematical Programing* 39:117–129.

Murty, K. G. 1988. *Linear Complementarity, Linear and Nonlinear Programming*. Helderman Verlag.

Newton, I. 1868. Philosophi ae naturalis principia mathematica.

Otaduy, M. A.; Germann, D.; Redon, S.; and Gross, M. 2007. Adaptive deformations with fast tight bounds. In *Proc. of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*.

Painlevé, P. 1895. Sur le lois du frottement de glissemment. *C.R. Acad. Sci. Paris* 121:112–115.

Pandolfi, A., and Ortiz, M. 2007. Finite element analysis of nonsmooth frictional contact. In *IUTAM Symposium on Computational Methods in Contact Mechanics*, volume 3, 57–76.

Pandolfi, A.; Kane, C.; Marsden, J. E.; and Ortiz, M. 2002. Time-discretized variational formulation of non-smooth frictional contact. *International Journal for Numerical Methods in Engineering* 53(8):1801–1829.

Pauly, M.; Pai, D. K.; and Guibas, L. 2004. Quasi-rigid objects in contact. In *ACM SIGGRAPH Symposium on Computer Animation*. ACM/Eurographics.

Raghupathi, L., and Faure, F. 2006. Qp-collide: A new approach to collision treatment. In *journées du groupe de travail Animation et Simulation (GTAS)*, 91–101.

Redon, S.; Kheddar, A.; and Coquillart, S. 2002. Gauss' least constraints principle and rigid body simulations. In *IEEE International Conference on Robotics and Automation*.

Robinson, S. M. 1982. Generalized equations and their solutions, part ii: Applications to nonlinear programming. *Math. Progam. Study* 19:200–221.

Rockafellar, R. T., and Wets, R. J. 1998. *Variational Analysis*. Springer.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press.

Schittkowski, K. 2005. QL: A fortran code for convex quadratic programming - user's guide, version 2.11. *Report, Department of Mathematics, University of Bayreuth*.

Schittkowski, K. 2006. NLPQLP: A fortran implementation of sequential quadratic programming algorithm with distributed and non-monotone line search - user's guide, version 2.2. *Report, Department of Computer Science, University of Bayreuth*.

Shabana, A. 1991. *Theory of Vibration, Volume II: Discrete and Continuous Systems*. Springer-Verlag.

Shabana, A. 2005. *Dynamics of Multibody Systems*. Cambridge, 3rd edition.

Signorini, S. 1933. Sopra akune questioni di elastostatica. *Atti della Societa Italiana per il Progresso della Scienze*.

Smith, R. 2006. Open dynamics engine, v0.5, user guide.

Song, P., and Kumar, V. 2003. Distributed compliant model for efficient dynamic simulation of systems with frictional contacts. In *The 2003 ASME Design Engineering Technical Conferences*.

Spillmann, J.; Becker, M.; and Teschner, M. 2007. Non-iterative computation of contact forces for deformable objects. *Journal of WSCG* 15(13).

Stewart, D. E., and Trinkle, J. C. 1996. An implicit time-stepping scheme for rigid body dynamics with inelastic collisions and coulomb friction. *International Journal Numerical Methods Engineering* 39:2673–2691.

Stewart, D. E. 2000. Rigid-body dynamics with friction and impact. *SIAM Rev.* 42(1):3–39.

Stewart, D. E. 2001. Finite-dimensional contact mechanics. *Phil. Trans. R. Soc. Lond. A* 359:2467–2482.

Trinkle, J. C.; Pang, J. S.; Sudarsky, S.; and Lo, G. 1995. On dynamic multi-rigid-body contact problems with coulomb friction. Technical report, Texas A+M University, Department of Computer Science.

Vavasis, S. A. 1991. *Nonlinear Optimization: Complexity Issues*. Oxford.

Wasfy, T. M., and Noor, A. K. 2003. Computational strategies for flexible multibody systems. *Applied Mechanics Reviews* 56(6):553–613.

West, M. 2004. *Variational Integrators*. Ph.D. Dissertation, California Institute of Technology.

Wriggers, P. 2002. *Computational Contact Mechanics*. John Wiley and Sons, Ltd.