

Rigid-Body Dynamics with Friction and Impact*

David E. Stewart[†]

Abstract. Rigid-body dynamics with unilateral contact is a good approximation for a wide range of everyday phenomena, from the operation of car brakes to walking to rock slides. It is also of vital importance for simulating robots, virtual reality, and realistic animation. However, correctly modeling rigid-body dynamics with friction is difficult due to a number of discontinuities in the behavior of rigid bodies and the discontinuities inherent in the Coulomb friction law. This is particularly crucial for handling situations with large coefficients of friction, which can result in paradoxical results known at least since Painlevé [*C. R. Acad. Sci. Paris*, 121 (1895), pp. 112–115]. This single example has been a counterexample and cause of controversy ever since, and only recently have there been rigorous mathematical results that show the existence of solutions to his example.

The new mathematical developments in rigid-body dynamics have come from several sources: “sweeping processes” and the measure differential inclusions of Moreau in the 1970s and 1980s, the variational inequality approaches of Duvaut and J.-L. Lions in the 1970s, and the use of complementarity problems to formulate frictional contact problems by Lötstedt in the early 1980s. However, it wasn’t until much more recently that these tools were finally able to produce rigorous results about rigid-body dynamics with Coulomb friction and impulses.

Key words. rigid-body dynamics, Coulomb friction, contact mechanics, measure-differential inclusions, complementarity problems

AMS subject classifications. Primary, 70E55; Secondary, 70F40, 74M

PII. S0036144599360110

I. Rigid Bodies and Friction. Rigid bodies are bodies that cannot deform. They can translate and rotate, but they cannot change their shape. From the outset this must be understood as an approximation to reality, since no bodies are perfectly rigid. However, for a vast number of applications in robotics, manufacturing, biomechanics (such as studying how people walk), and granular materials, this is an excellent approximation. It is also convenient, since it does not require solving large, complex systems of partial differential equations, which is generally difficult to do both analytically and computationally. To see the difference, consider the problem of a bouncing ball. The rigid-body model will assume that the ball does not deform while in flight and that contacts with the ground are instantaneous, at least while the ball is not rolling. On the other hand, a full elastic model will model not only the contacts and the resulting deformation of the entire ball while in contact, but also the elastic oscillations of the ball while it is in flight. Apart from the computational complexity of all this, the analysis of even linearly elastic bodies in contact with a

*Received by the editors July 26, 1999; accepted for publication (in revised form) August 5, 1999; published electronically January 24, 2000. This work was supported by NSF grant DMS-9804316.

<http://www.siam.org/journals/sirev/42-1/36011.html>

[†]Department of Mathematics, University of Iowa, Iowa City, IA 52242 (dstewart@math.uiowa.edu).

rigid surface subject to Coulomb friction using a Signorini contact condition is not completely developed even now [22, 23, 56, 30, 57, 62, 65]. Even if all this can be done, most of the details of the motion for the fully elastic body are not significant on the time- or length-scales of interest in many of the applications described above. For more information about applications of rigid-body dynamics, see, for example, [20, 21, 92] regarding granular flow and [11, 108] regarding virtual reality and computer animation.

There are some disadvantages with a rigid-body model of mechanical systems. The main one is that the velocities must be discontinuous. Consider again a bouncing ball. While the ball is in flight, there are no contact forces acting on it. But when the ball hits the ground, the negative vertical velocity must become a nonnegative vertical velocity instantaneously. The forces must be impulsive; they are no longer ordinary functions of time but rather *distributions* or *measures*. While there has been considerable work on differential equations with impulsive right-hand sides, these are usually concerned with situations where the impulsive part is known a priori and is not part of the unknown solution. (The work of Bainov et al., for example, has this character [8, 7, 71]. In these works Bainov et al. can allow for some dependence of the *time* of the discontinuity on the solution, but the way the solution changes at the discontinuity is assumed to be known, and problems like bouncing balls, where the ball comes to rest in finite time after infinitely many bounces, are beyond the scope of their approach.)

The rigid body model with Coulomb friction has been subject to a great deal of controversy, mostly due to a simple model problem of Painlevé which appears not to have solutions. The list of papers on this problem is quite extensive and includes [11, 12, 27, 28, 36, 49, 68, 76, 77, 78, 84, 82, 88, 89, 91, 109, 131, 128]. The modern resolution of Painlevé's problem involves impulsive forces and still generates controversy in some circles.

In this article, an approach is described that combines impulsive forces (measures) with convex analysis. It develops a line of work begun by Schatzman [117] and J. J. Moreau [87, 88, 90, 91] and continued by Monteiro Marques, who produced the first rigorous results in this area [83, 84]. Related work has been done by Brogliato, which is directed at the control of mechanical systems with friction and impact, and is based on the approach of Moreau and Monteiro Marques; Brogliato's book [14] gives an accessible account of many of these ideas. The intellectual heritage used in this work is extensive: convex analysis, measure theory, complementarity problems, weak* compactness, and convergence are all used in the theory, along with energy dissipation principles and other more traditional tools of applied mathematics.

A number of aspects of rigid-body dynamics nonetheless remain controversial and unresolved. These include the proper formulation of impact laws and how to correctly handle multiple simultaneous contacts. These are discussed below in sections 1.2 and 4.4. Neither of these issues affects the internal consistency of rigid-body models; rather, they deal with how accurately they correspond to experimentally observed behavior.

The structure of this article is as follows. This introduction continues with subsections dealing with Coulomb friction and discontinuous ODEs (section 1.1); impact models (section 1.2); the famous problem of Painlevé (section 1.3); complementarity problems, which are useful tools for formulating problems with discontinuities (section 1.4); and measure differential inclusions (section 1.5). Section 2 discusses how to formulate rigid-body dynamics, first as a continuous problem (section 2.1) and then as a

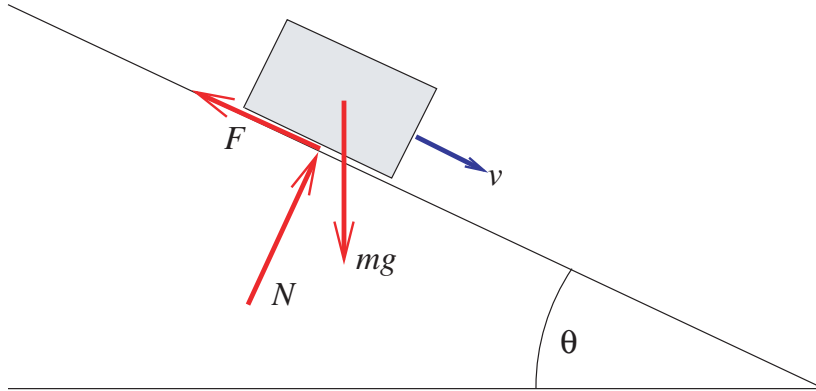


Fig. 1.1 *Brick on a frictional ramp.*

numerical problem (section 2.2), and concludes with a discussion of practicalities and numerical results (section 2.3). Section 3 is about convergence and existence theory for the solutions of rigid-body dynamics problems with impact and friction. Section 4 discusses variants on the ideas presented in the preceding sections. In particular, there are discussions of how to treat rigid-body dynamics as a singular perturbation problem (section 4.1), how to apply symplectic integration methods and difficulties in using them (section 4.2), and how to handle velocity-dependent friction coefficients (section 4.3), multiple contact problems (section 4.4), and the treatment of extended elastic bodies (section 4.5).

1.1. Coulomb Friction and Discontinuous Differential Equations. The Coulomb law is the most common and practical model of friction available. It is, however, a discontinuous law. Coulomb's famous law was derived from a great deal of experimental work that was published in 1785 [26] in his *Théorie des machines simples* (*Theory of simple machines*). While Coulomb's law still arouses controversy, and there are many variants on his basic law, it is a suitable starting point and has been successfully used in practice. In its simplest form, Coulomb's law says that the friction force is bounded in magnitude by the normal contact force (N) times the coefficient of friction (μ); if the contact is sliding, then the magnitude of the friction force is exactly μN in the opposite direction to the relative velocity at the contact. As an example, consider a brick sliding on a ramp, as illustrated in Figure 1.1.

If the brick is sliding down the ramp ($v > 0$), then since $N = mg \cos \theta$, the friction force is $F = +\mu mg \cos \theta$. If the brick is sliding up the ramp ($v < 0$), then $F = -\mu mg \cos \theta$. The differential equation for the velocity v is

$$(1.1) \quad m \frac{dv}{dt} = mg \sin \theta - \mu mg \cos \theta \operatorname{sgn} v,$$

where $\operatorname{sgn} v$ is $+1$ if $v > 0$, -1 if $v < 0$, and 0 if $v = 0$. The right-hand side is clearly a discontinuous function of the state variable v . If it were only discontinuous in t , then we could apply Carathéodory's theorem [24, section 2.1] to establish existence of a solution. But Carathéodory's theorem is not applicable. What is worse is that the typical behavior of bricks in this situation is that they stop and stay stopped; that is, we will have $v = 0$ in finite time, and v will stay at zero for at least an interval of positive length. Understanding the discontinuity is essential for understanding the solution. In fact, solutions do not exist for this differential equation as it is stated,

since if $v = 0$ and $dv/dt = 0$, then we get $0 = mg \sin \theta - \mu mg \operatorname{sgn} 0$, which can only be true if $\sin \theta = 0$. To solve a discontinuous differential equation like this, we need to extend the concept of differential equations to *differential inclusions* [38, 39, 40], which were first considered by A. F. Filippov around 1960. A differential inclusion has the form

$$\frac{dx}{dt} \in F(t, x),$$

where F is a set-valued function. There are some properties that F should have. Its graph $\{(x, y) \mid y \in F(t, x)\}$ should be a closed set. The values $F(t, x)$ should all be closed, bounded, convex sets. And $F(t, x)$ should satisfy a condition to prevent “blow-up” in finite time, such as $x^T z \leq C(1 + \|x\|^2)$ for all $z \in F(t, x)$. Numerical methods for discontinuous ODEs need to use this differential inclusion formulation if high accuracy is desired. If this is not done, then the methods typically have first order convergence due to rapid “chattering” of the numerical trajectories around the discontinuities for simple discontinuous ODEs. The first published results on numerical methods for discontinuous ODEs and differential inclusions were those by Taubert [137] in 1976. Further work on numerical methods for differential inclusions and discontinuous ODEs includes [31, 35, 60, 61, 74, 95, 96, 126, 127, 136, 137, 138]. Of these, only [60, 61, 126, 127] give methods with order higher than one.

Excellent overviews of numerical methods for differential inclusions can be found in Dontchev and Lempio [31] or Lempio and Veliov [75].

Since the rigidity of objects is only an approximation, it is reasonable to consider approximating the Coulomb law for the friction force by a continuous or smooth law. The discontinuity in the Coulomb friction law has an important physical consequence: a block on an inclined ramp will not move down the ramp as long as the applied tangential forces do not exceed μN . If the Coulomb law were replaced by a smooth law, then the block would creep down the ramp at a velocity probably proportional to the tangential force divided by μN . Experimentally, very little if any creep is observed in typical situations with dry friction, which demands a friction force function that is discontinuous or very close to being discontinuous. On the other hand, using a continuous approximation for numerical purposes leads to a stiff ODE. Applying implicit time-stepping procedures then results in solutions that are very close to the solution obtained by applying the implicit method to the corresponding differential inclusion. In summary, the physics points to real discontinuities, and there is little advantage numerically in smoothing the discontinuity. The discontinuity is here to stay.

So far we have considered only one-dimensional friction laws where the set of possible friction forces is one-dimensional. For ordinary three-dimensional objects in contact, the plane of relative motion is two-dimensional, and so the set of possible friction forces is two-dimensional. In this case, to allow for complications such as anisotropic friction, we need a better approach. A better basis for formulating physically correct friction models is the *maximum dissipation principle*. This says that given the normal contact force c_n , the friction force c_f is the one that maximizes the rate of energy dissipation $-c_f^T v_{rel}$, where v_{rel} is the relative velocity at the contact, out of all possible friction forces allowed by the given normal contact force c_n . To be more formal, there is a set FC_0 which is the set of possible friction forces c_f for $c_n = 1$. The set FC_0 is assumed to be closed, convex, and *balanced* ($FC_0 = -FC_0$). So the maximum dissipation principle says that

$$(1.2) \quad c_f \text{ maximizes } -c_f^T v_{rel} \text{ over } c_f \in c_n FC_0.$$

In the case of two-dimensional isotropic friction acting on a particle, FC_0 is a disk of radius μ (the coefficient of friction) in the plane of contact. If n is the normal direction to the contact surface, then the total contact force is $n c_n + c_f$. The set of possible contact forces is the friction cone, which is given by

$$(1.3) \quad FC = \{ n c_n + c_f \mid c_n \geq 0 \text{ and } c_f \in c_n FC_0 \}.$$

Of course, as the contact point changes, so does the plane of the possible friction forces. So we must allow FC_0 and FC to depend on the configuration of the system: $FC_0 = FC_0(q)$ and $FC = FC(q)$. Some pathologies should be prevented, such as having the normal direction n lying in the vector subspace generated by FC_0 . Note that FC_0 and FC are again set-valued functions. However, they are generally continuous ones on the boundary of the admissible region. In the interior of the admissible region, the normal and frictional contact forces must both be zero, so in that case, $FC(q) = \{0\}$.

1.2. Impact Models. The behavior of impacting bodies is a topic in rigid-body dynamics that does not arise in formulating other problems in mechanics. It would normally be considered to be the *result* of the model, rather than an ingredient in building the model. However, for rigid-body dynamics, an impact is regarded as an atomic (i.e., indivisible) event and must be modeled as such. The use of measure differential inclusions below does not save us from having to decide. And it is a fact of nature that some materials behave more elastically than others on impact. Some seem to be inelastic with little or no “bounce,” and some are very elastic and seem to lose very little energy in an impact. As a general rule, modelers use a *coefficient of restitution*, usually denoted here by ε between zero and one to describe the impact behavior of a pair of bodies or materials. For $\varepsilon = 0$ we have purely inelastic impacts, and for $\varepsilon = 1$ the impacts are purely elastic.

There are two generally pervasive approaches to modeling impact behavior. One (the *Newtonian* approach) relates pre- and postimpact velocities’ normal components (typically $n^T v(t^+) = -\varepsilon n^T v(t^-)$) [94]; the other (the *Poisson* approach) divides the impact into compression and decompression phases and relates the impulse in the decompression phase to the impulse in the compression phase: $N_{decompr} = -\varepsilon N_{compr}$ [110, 114]. The value of the contact impulse for the compression phase of the contact should be determined by the impulse needed for *inelastic impact*. Each approach has been found to produce an *increase* in the total mechanical energy in certain situations! For difficulties with the Newtonian approach in its naive form (even with only one contact in two dimensions), see Stronge’s article [135] and the references therein and also Keller’s short article [63].

Whichever approach is used, the case of inelastic impacts is an important reference case that both approaches must handle: $\varepsilon = 0$. For one contact, this means that $n^T v(t^+) = 0$ for any time t when there is contact. For a fuller discussion of how the Newton and Poisson approaches can be used and modeled, the reader is again referred to the excellent book of Brogliato [14]. Another discussion of the Newton and Poisson formulations is given in Chatterjee and Ruina [18], who also discuss alternative collision laws. Note that both the Newton and Poisson impact laws have serious defects, which are discussed in [18, 135], for example.

One of the abiding difficulties in this area is the lack of understanding of the physical mechanisms behind impact processes. It has long been believed that three mechanisms are responsible for energy dissipation in impact: (1) localized plastic deformation; (2) viscous damping in the material; and (3) energy transfer to elastic

vibrations. Until recently it has not been clear which is the most important. As a result, models have lagged in terms of their physical accuracy and sophistication. Recent work has thrown fresh light onto this issue.

Recent experimental and simulation work, particularly by Stoianovici and Hurmuzlu [133], has highlighted the importance of elastic vibrations, although localized plastic deformation appears to play a role. Viscous damping seems to play very little direct role in the impact behavior. This is mostly because the time-scale of the impact (typically between 10 μs and 10 ms for metal objects of sizes between 1 cm and 1 m) is too short for significant viscous damping, except for very high frequency modes.

What Stoianovici and Hurmuzlu did was to drop steel bars onto a hard, massive block, record the impacts using high-speed video recorders, and identify contacts by measuring the current flowing from the bars to the block underneath. One of the main results of their experimental investigation was the way the kinematic coefficient of restitution $\varepsilon = -(n^T v^+) / (n^T v^-)$ depended on the angle of the bar relative to the upper surface of the block (ϕ). Figure 1.2 is taken from Stoianovici and Hurmuzlu [133] with the kind permission of the authors and the *ASME Journal of Applied Mechanics*. As can be seen in Figure 1.2, if the bar is dropped vertically ($\phi = 90^\circ$), then ε ranges between 0.8 and 0.9. As the angle is decreased, ε decreases until ε is between 0.1 and 0.2, at an angle that appears to approach 90° as the bar becomes more slender. As ϕ is further reduced, ε increases in a sometimes erratic way until it reaches a value of around 0.6 for $\phi = 0^\circ$. The results of Stoianovici and Hurmuzlu [133] also include simulation results, which show excellent agreement with the experimental results. The simulation results do not include plastic deformation, but viscous damping is included at the contact point. The viscous damping parameter was calibrated to fit the experimental results.

If we assume that plastic deformation and viscous damping are insignificant during impact, then an effective coefficient of restitution can be computed for a given body using only linear elasticity and taking the material stiffness (i.e., Young's modulus) to infinity to recover a rigid limit. This gives a coefficient of restitution $\varepsilon = \varepsilon(q)$. However, to make $\varepsilon(q)$ well defined, we need to assume that prior to contact there are no excited modes of elastic vibration. If the body undergoes a rapid series of impacts, or if other damping mechanisms are too slow, then later impacts will have excited modes of vibration, which will invalidate the assumptions. Significant energy could be transferred from the elastic modes back into rigid-body modes. While common experience suggests that these effects are probably not large, they may still be important for accurate simulations. Appropriate modeling of vibrational effects in a rigid-body framework has not yet been attempted. An open question here is, "*In the rigid limit, do the phases (rather than just the amplitudes) of elastic vibrations play a significant role in the impact behavior of the body?*" If the answer is "yes," then practical prediction will be extremely difficult, because the elastic vibrations have frequencies that are typically in the acoustic range of 100 Hz to 1 kHz, and future impact times will have to be computed to within a fraction of the period of these vibrations (i.e., probably within 100 μs) in order to accurately simulate the behavior. In robotics, where speeds often range up to 1 ms^{-1} , computing impact times to this accuracy would require knowledge of distances to within 0.1 mm, which is a very high accuracy requirement. If only the amplitudes of the elastic vibrations are important, then a plausible approach to modeling is to consider the motion of the body as consisting of rigid motions plus small high-frequency elastic vibrations. The elastic vibrations would generally decay while the body is in free flight but could change

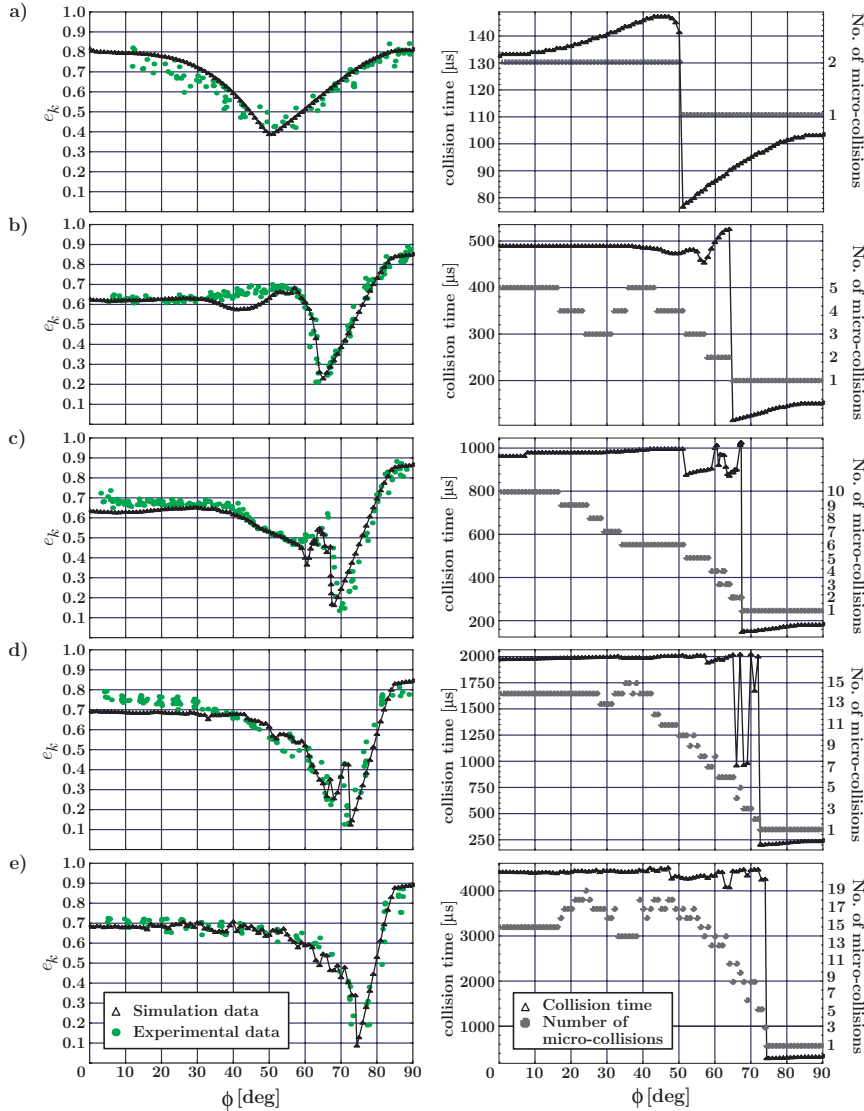


Fig. 1.2 Results of Stoianovici and Hurmuzlu showing the effective kinematic coefficient of restitution (left), collision time and number of “micro-collisions” (right) for bars of lengths (a) 100 mm, (b) 200 mm, (c) 300 mm, (d) 400 mm, and (e) 600 mm. Reprinted from D. Stoianovici and Y. Hurmuzlu, *ASME Journal of Applied Mechanics*, 63 (1996), with permission from the American Society of Mechanical Engineers.

rapidly (instantaneously in the rigid limit!) during impact. This approach may allow the development of new low-order models that can accurately capture the motion of common objects.

In the remainder of the paper we will consider only the Newton and Poisson types of impact models.

1.3. Painlevé’s Problem. In 1895 P. Painlevé published a paper [97] in which he presented a simple rigid-body dynamics problem which appears not to have a solution.

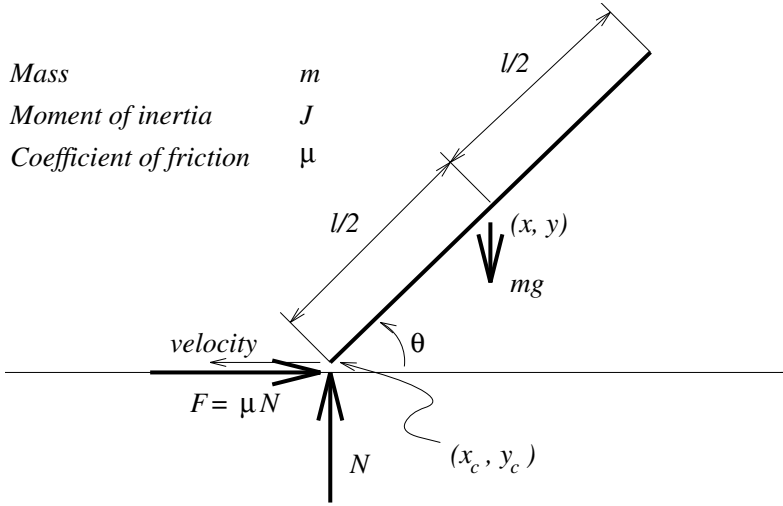


Fig. 1.3 *Painlevé's problem.*

The scenario is much like getting chalk to screech on a blackboard. It is illustrated in Figure 1.3.

The parameter values that we are most interested in are large μ and small J/ml^2 . We let (x, y) be the coordinates of the center of mass; θ is the angle of the rod with respect to the horizontal (counterclockwise); the coordinates of the point where the rod contacts the table are (x_c, y_c) . If $\dot{x} < 0$ and $\dot{\theta} = 0$, then $\dot{x}_c < 0$ and the relative sliding velocity (of the rod with respect to the table) is to the left. Therefore, the friction force F should have magnitude $+\mu N$ to the right. The equations of motion for (x, y, θ) are

$$(1.4) \quad \begin{aligned} m\ddot{x} &= F, \\ m\ddot{y} &= N - mg, \\ J\ddot{\theta} &= (l/2)[+F \sin \theta - N \cos \theta]. \end{aligned}$$

Of particular importance is the equation for y_c , since the rigid-body condition is that $y_c \geq 0$. If $y_c = 0$ and $\dot{y}_c = 0$, then to prevent penetration we need $\ddot{y}_c \geq 0$. This can be computed in terms of F and N since $y_c = y - (l/2) \cos \theta$. Differentiating twice gives

$$\ddot{y}_c = \ddot{y} + (l/2) \sin \theta \ddot{\theta} + (l/2) \cos \theta \dot{\theta}^2.$$

Substituting $F = +\mu N$ into these equations gives

$$(1.5) \quad \ddot{y}_c = \left[\frac{1}{m} - \frac{l^2}{4J} \cos \theta (\mu \sin \theta - \cos \theta) \right] N + (l/2) \cos \theta \dot{\theta}^2 - g.$$

We require that $\ddot{y}_c \geq 0$ and $N \geq 0$ (since there is no adhesion to the surface—that is, there is no glue). If $\ddot{y}_c > 0$, then contact is broken immediately and so (provided there are no impulsive forces) we must take N and also F to be zero. Conversely, if $N > 0$, then contact must be maintained and so $\ddot{y}_c = 0$. Both cases are covered by the simple complementarity condition

$$(1.6) \quad 0 \leq \ddot{y}_c \quad \perp \quad N \geq 0.$$

(Note that “ $a \perp b$ ” indicates that $ab = 0$ if a and b are scalars and that $a^T b = 0$ if they are vectors; for a vector “ $a \geq 0$ ” means that the *components* a_i of a are nonnegative.)

Now consider the situation with large coefficient of friction μ and small J/ml^2 . For μ large, the contact force (the vector sum of F and N) is applied to the contact point at a shallow angle. This means that the torque from the contact forces is *counterclockwise*. If J/ml^2 is sufficiently small, the effect of this torque will overpower the effect of the vertical contact force N in preventing penetration. When this happens, the torque will cause the rod to rotate *into the table*. Mathematically, the coefficient $(1/m) - (l^2/4J) \cos \theta (\mu \sin \theta - \cos \theta) < 0$ in this situation. If $(l/2) \cos \theta \dot{\theta}^2 - g < 0$ as well, then no value of $N \geq 0$ will give $\dot{y}_c \geq 0$, and penetration appears impossible to prevent.

This line of analysis has been used to argue that rigid-body dynamics and the Coulomb law of friction are inconsistent. For a recent example of this line of reasoning, see the otherwise excellent textbook of Pfeiffer and Glocker [109, section 5.3, pp. 61–63]. However, this line of reasoning is flawed. The flaw is the assumption that *all forces are bounded functions of time*. In particular, it rules out the possibility that the horizontal component of the velocity could be brought to zero *instantaneously* by impulsive contact forces. If \dot{x}_c was driven to zero during the impact, then we would no longer require that $F = +\mu N$, the angle of the contact forces does not need to be shallow, and solutions can exist. This example also shows that it is important to allow for impulsive forces in rigid-body dynamics, even when there are no collisions. It also shows that care should be taken in developing formulations of rigid-body dynamics that can directly incorporate impulsive forces and discontinuous solutions.

Note that when Pfeiffer and Glocker [109, section 8.2] discuss impact laws with friction, they implicitly allow for this resolution of Painlevé’s problem. However, this would require explicitly invoking the impact law in a situation without impacts.

The reader wishing to look at the extensive discussion of Painlevé’s problem is invited to peruse the references [11, 12, 27, 28, 36, 49, 68, 76, 84, 82, 89, 91, 109].

On a final note, the Painlevé problem also has a bearing on formulating impact laws since applying the Newton impact law would imply that \dot{y}_c immediately after impact should be zero, even for the “perfectly elastic” case, $\varepsilon = 1$. However, simulations and asymptotic analysis with the table represented by a spring of stiffness k show that as $k \rightarrow \infty$ there is a definite nonzero limit of \dot{y}_c after impact. This corresponds to $\varepsilon = +\infty$ using Newton’s impact law, since \dot{y}_c just before impact is zero. A Poisson impact law would be more realistic, since it would give a positive value for \dot{y}_c just after the impact for any $\varepsilon > 0$, due to the impulsive contact forces.

1.4. Complementarity Problems. Complementarity problems give a uniform way of representing a very large range of conditions that would require modeling using discontinuous functions. For example, the contact conditions (1.6) of the previous section avoid the need for discontinuous or infinite functions that would arise if we tried to represent N as $N(y_c)$, for example.

Complementarity problems typically have the following form: Find $z \in \mathbf{R}^n$ such that $0 \leq z \perp f(z) \geq 0$, where $f : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is a given continuous function. Note that this is equivalent to requiring that $z_i \geq 0$ and $f_i(z) \geq 0$ for all $i = 1, \dots, n$, and that $z_i = 0$ or $f_i(z) = 0$ for all $i = 1, \dots, n$. Linear complementarity problems (LCPs) are complementarity problems where f is affine: $f(z) = Mz + q$, where M is a given matrix and q is a given vector. Not all (not even “almost all”) linear complementarity problems have solutions. It should be understood that LCPs are truly nonlinear problems, in spite of their linear-combinatorial structure. For a comprehensive overview

of LCPs, see Cottle, Pang, and Stone [25]. A more recent overview of applications of linear and nonlinear complementarity problems is the article by Ferris and Pang [37].

Complementarity problems arise in many contexts; one of the most important of these is constrained optimization. If we consider an optimization problem

$$(1.7) \quad \min_x f(x) \quad \text{subject to} \quad g_i(x) \geq 0, \quad i = 1, \dots, m,$$

then provided a suitable constraint qualification [41] holds, there are Lagrange multipliers λ_i such that

$$(1.8) \quad \begin{aligned} \nabla f(x) - \sum_{i=1}^m \lambda_i \nabla g_i(x) &= 0, \\ \lambda_i, g_i(x) &\geq 0, \\ \lambda_i g_i(x) &= 0 \end{aligned}$$

hold at the minimizing x . These are the Kuhn–Tucker (or Karush–Kuhn–Tucker) conditions for optimality [41, 43]. Note that the last two lines of (1.8) are actually complementarity conditions: $0 \leq \lambda \perp g(x) \geq 0$.

The existence of solutions and algorithms to compute them is known for a wide class of LCPs. The best-known algorithms for LCPs are pivoting methods, which are closely related to the simplex method for linear programming. The best known of these is Lemke’s method [25, section 4.4]. The most useful result for our purpose shows that Lemke’s method computes a solution to the LCP $0 \leq z \perp Mz + q \geq 0$, where M is a *copositive* matrix (defined below) and $q^T y \geq 0$ whenever y solves the homogeneous LCP $0 \leq y \perp My \geq 0$ [25, Thms. 3.8.6 and 4.4.12]. A matrix M is *copositive* if $y \geq 0$ implies that $y^T My \geq 0$. The application of complementarity problems to general unilateral contact problems seems to have begun with the work of Lötstedt [76, 77, 78, 79]. In this work, Lötstedt was able to show the existence of solutions of certain LCPs provided the coefficient of friction was sufficiently small. Since then, there has been much work applying complementarity problems to contact problems [4, 5, 66, 67, 103, 102, 131, 139, 101], and the question of the existence of solutions to the complementarity problems has attracted a great deal of attention.

Complementarity problems can be extended to a more general setting which can include infinite-dimensional problems [55, 99]: if X is a Banach space and K is a closed convex cone in X , then the dual cone is $K^* = \{y \in X^* \mid \langle y, x \rangle \geq 0 \text{ for all } x \in K\}$. For a function $f : X \rightarrow X^*$ the complementarity problem $CP(f)$ is the problem of finding $z \in K$ such that $f(z) \in K^*$ and $\langle f(z), z \rangle = 0$. One way in which we will use this idea is to take X to be the set of continuous functions on $[0, T]$ so that X^* is the set of finite-valued Borel measures on $[0, T]$. The cone $K \subset C[0, T]$ is the cone of nonnegative continuous functions, so K^* is the cone of finite-valued nonnegative Borel measures. We understand complementarity between a measure μ and a function ϕ on $[0, T]$ to mean that $\mu(E) \geq 0$ for all Borel E , $\phi(t) \geq 0$ for all t , and $\langle \mu, \phi \rangle = \int \phi(t) \mu(dt) = 0$. This we denote by the shorthand $0 \leq \mu \perp \phi \geq 0$.

1.5. Measure Differential Inclusions. General rigid-body dynamics requires a new approach which can combine impulsive forces with differential inclusions. A framework for this approach can be built using *measure differential inclusions* [87, 90, 91]. For Moreau, this work developed out of a study of *sweeping processes* [16, 69, 70, 85, 86]: in a sweeping process, there is a set-valued function $C(t)$ which is convex for all t and a “particle” at $x(t)$ which is “swept” along by the $C(t)$ so that $x(t) \in C(t)$ for all t . If $x(t)$ is in the interior of $C(t)$, then $x'(t)$ will be zero, and if $x(t)$ is on the boundary of $C(t)$, then $x'(t)$ will be in the normal cone of $C(t)$ at

$x(t)$. If $C(t)$ is allowed to be discontinuous as a set-valued function, then $x(t)$ may also have to “jump”: $x(t^+) - x(t^-)$ must then belong to the normal cone of $C(t)$ at $x(t^+)$. Measure differential inclusions (although not always called that) can be found in other contexts in the work of Schatzman [117, 118, 119], for example.

In a measure differential inclusion,

$$(1.9) \quad \frac{dv}{dt} \in F(t, x), \quad \frac{dx}{dt} = g(t, x, v),$$

$F(t, x)$ does not need to be bounded, and $v(\cdot)$ is only required to have bounded variation. The other properties assumed about F for differential inclusions should hold: $F(t, x)$ should be a closed convex set, and the graph of F should be closed. The difficulty is in interpreting “ $dv/dt \in F(t, x)$ ” when $v(\cdot)$ is not absolutely continuous or is discontinuous. Then “ dv/dt ” is not an integrable function, or perhaps not even a function at all, but a distribution or measure. We do, however, suppose that v has bounded variation on a finite interval $[0, T]$. That is, the supremum of

$$\sum_{i=0}^{N-1} \|v(t_{i+1}) - v(t_i)\|$$

over all N and choice of $0 \leq t_0 \leq t_1 \leq \dots \leq t_N \leq T$ is finite. This supremum is denoted $\bigvee_0^T v$ and is called the variation of v on $[0, T]$. Then we can use Riemann–Stieltjes integrals to define $\int \phi(t) dv(t)$ for continuous functions ϕ ; see, for example, [134, pp. 281–284] or [72, Chap. X]. By the Riesz representation theorem, the linear functional $\phi \mapsto \int \phi(t) dv(t)$ is a continuous linear functional and is therefore equivalent to integration against a measure: $\int \phi(t) dv(t) = \int \phi(t) \mu(dt)$, where μ is a Borel measure [72, Thm. 2.7, Chap. IX, pp. 264–265]. We write $\mu = Dv$ to denote the distributional derivative of $v(\cdot)$. The expression “ dv/dt ” can be thought of as a *Radon–Nikodym* derivative of Dv with respect to the ordinary Lebesgue measure Dt . However, this will work only if Dv is an absolutely continuous measure with respect to Dt , which amounts to requiring that $v(\cdot)$ is an absolutely continuous function. If $v(\cdot)$ is not an absolutely continuous function we need to extend the notion of the Radon–Nikodym derivative.

Fortunately it is possible to split measures into absolutely continuous and singular parts: the Lebesgue decomposition of Dv is $Dv = \mu_s + a Dt$, where $a(\cdot)$ is a Lebesgue integrable function and μ_s is a Borel measure that is singular with respect to the Lebesgue measure [64, pp. 111–113]. The singular part μ_s contains all the forces and impulses “at infinity”; this singular part is supported on a set that has Lebesgue measure zero. So we should require that $a(t) \in F(t, x(t))$ for Lebesgue almost all t . For μ_s we want to look at the part of $F(t, x)$ “at infinity.” For closed convex $F(t, x)$ we can use the *asymptotic* or *regression cone* [53]. This asymptotic cone of a closed convex set K is the set of directions in K “at infinity”:

$$(1.10) \quad K_\infty = \left\{ x \mid x = \lim_{k \rightarrow \infty} t_k x_k, x_k \in K \right\}.$$

While we can’t take the Radon–Nikodym derivative of μ_s with respect to the Lebesgue measure, we can “normalize” μ_s by using its variation $|\mu_s|$ which is a nonnegative Borel measure: any finite-valued Borel measure μ has variation $|\mu|$, which is another finite-valued measure defined by

$$|\mu|(E) = \sup_{\{E_i\}} \sum_i \|\mu(E_i)\|,$$

where $\{E_i\}$ ranges over all countable Borel partitions of a Borel set E . Then clearly $\|\mu(E)\| \leq |\mu|(E)$ for all Borel sets E , and μ is an absolutely continuous measure with respect to $|\mu|$. So we define “ $dv/dt \in F(t, x)$ ” to mean that

$$(1.11) \quad a(t) \in F(t, x(t)), \quad Dt \text{ a.e.},$$

$$(1.12) \quad \frac{d\mu_s}{d|\mu_s|}(t) \in F(t, x(t))_\infty, \quad |\mu_s| \text{ a.e.}$$

An alternative definition that is useful for handling problems of convergence is the following: for each continuous function $\phi \geq 0$, not everywhere zero,

$$(1.13) \quad \frac{\int \phi(t) dv(t)}{\int \phi(t) dt} \in \overline{\text{co}} \bigcup_{\tau: \phi(\tau) \neq 0} F(\tau, x(\tau)).$$

The definition based on (1.11), (1.12) I call the *strong* definition, and the definition based on (1.13) I call the *weak* definition. Under conditions (C1)–(C3) below, the two definitions are equivalent [130].

- (C1) The graph of F is closed, and $F(t, x)$ is a closed convex set for all (t, x) .
- (C2) The asymptotic cone $F(t, x)_\infty$ is always *pointed*. (A pointed cone K is a cone where $K \cap (-K) = \{0\}$.)
- (C3) $\min\{\|z\| \mid z \in F(t, x)\}$ is a bounded function of (t, x) .

The strong definition is useful for proving properties about the solutions, while the weak definition is useful for obtaining existence results. Consider, for example, a sequence of functions $\{v^h\}_{h>0}$ generated by some numerical procedure. If it can be proved that $\bigvee v^h \leq M$ for some constant M and $v^h(0) = v_0$ for all $h > 0$, then by the Helly selection theorem [93], there is a pointwise convergent subsequence $v^{h_k}(\cdot)$ which converges to a function $\hat{v}(\cdot)$ with $\bigvee \hat{v} \leq M$. In this subsequence, the measures $Dv^{h_k} \rightharpoonup D\hat{v}$ weak*. If v^h satisfies a measure differential inclusion $dv^h/dt \in F_h(t, x(t))$, and $0 < h' < h$ implies that $F_{h'}(t, x) \subset F_h(t, x)$, then using the weak formulation it is easy to show that $d\hat{v}/dt \in F_h(t, x(t))$ for all $h > 0$, and consequently $d\hat{v}/dt \in \bigcap_{h>0} F_h(t, x(t))$.

While measure differential inclusions cannot satisfactorily treat all aspects of rigid-body dynamics, they give structure to a large part of it. The additional conditions can be specified in terms of complementarity conditions, usually between measures and functions.

2. Formulation and Simulation. The paradoxes uncovered by Painlevé show the need for care in formulating the equations of rigid-body dynamics with contacts. When it comes to formulating *numerical* methods for simulating rigid-body systems, we must be even more careful. Current methods for handling rigid-body dynamics fall into several categories:

1. For each possible configuration of contacts, solve for the forces that could be generated and feed the result into an ODE or differential algebraic equation (DAE) solver. This method is vulnerable to Painlevé’s problem as well as being very cumbersome.
2. Formulate the problem as in case 1 but use a complementarity problem to decide at each step which contacts are active. (This is basically the approach of Pfeiffer and Glocker [109].) This is still vulnerable to Painlevé’s problem.
3. Use a penalty formulation of the no-interpenetration condition. This corresponds to approximating the rigid bodies by very stiff bodies. This avoids the theoretical existence questions and avoids impulses but raises new questions

about singular perturbations and the accuracy of the computational methods. A penalty approach can also be used to approximate the Coulomb friction law. All of these modifications give very stiff differential equations.

4. Use a time-stepping formulation based on integrals of the contact forces rather than their instantaneous values. Complementarity conditions or optimization conditions are used to resolve whether contact is maintained or broken.

The approach I will present here is approach 4: use a time-stepping formulation. This approach directly incorporates impulsive forces. In fact, in a single simulation with fixed step-size, there is no way of determining if there actually is an impulsive force, since only integrals are represented or computed.

2.1. The Continuous Problem. In many ways it is easier to write down a numerical method for rigid-body dynamics than it is to say exactly what the method is trying to compute. Nevertheless, with the tools of measure differential inclusions, functions of bounded variation, and complementarity problems, this can be fairly straightforward for many situations. To keep matters simple at this stage, we consider a one contact problem.

To obtain the equations of motion we first need the equations of motion for a system without contacts. To do this we use *generalized coordinates* q , which can contain rectilinear coordinates of centers of mass as well as angles and orientation parameters and other types of coordinates. Associated with this are the *generalized velocities* $v = dq/dt$ and the Lagrangian $L(q, v) = \frac{1}{2}v^T M(q)v - V(q)$, where $\frac{1}{2}v^T M(q)v$ is the kinetic energy ($M(q)$ is the mass matrix) and $V(q)$ is the potential energy. If there are only internal forces, the equations of motion are given by Lagrange's equations of motion:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial v} \right) - \frac{\partial L}{\partial q} = 0 \quad \text{or} \quad M(q) \frac{dv}{dt} = k(q, v) - \nabla V(q),$$

where $k_l(q, v) = -\frac{1}{2} \sum_{r,s} [\partial m_{ir} / \partial q_s + \partial m_{is} / \partial q_r - \partial m_{rs} / \partial q_i]$. With external and contact forces we can write this as

$$(2.1) \quad M(q) \frac{dv}{dt} = n(q) c_n + D(q) \beta + k(q, v) - \nabla V(q) + F_{ext}(t).$$

The admissible region of configuration space is given by $\{x \mid f(q) \geq 0\}$ for a suitable function f . The normal direction vector at q is $n(q) = \nabla f(q)$. The contact conditions that we need are essentially the Signorini contact conditions: $0 \leq f(q) \perp c_n \geq 0$. We represent the set of possible friction forces through $FC_0(q) = \{D(q)\beta \mid \beta \in \mathbf{R}^d, \psi(\beta) \leq \mu\}$. The function ψ used for defining $FC_0(q)$ should be convex, positively homogeneous ($\psi(\alpha\beta) = |\alpha|\psi(\beta)$ for all $\alpha \in \mathbf{R}$), and coercive ($\psi(\beta) \rightarrow \infty$ as $\|\beta\| \rightarrow \infty$). Positive homogeneity implies that $D(q)\beta \in c_n FC_0(q)$ is equivalent to $\psi(\beta) \leq \mu c_n$.

The maximal dissipation principle states that we choose β so as to maximize the dissipation rate $v^T D(x)\beta$ over all $\beta \in c_n FC_0(x)$. Because of discontinuities, we need to be careful to make some distinctions between $v^+(t) = v(t^+) = \lim_{s \downarrow t} v(s)$, $v^-(t) = v(t^-) = \lim_{s \uparrow t} v(s)$, and $v(t)$. For inelastic impacts, for example, we require that $n^T v^+(t) = 0$ for all t where there is contact: $f(q(t)) = 0$. Also, in the maximum dissipation principle we should minimize $(v^+)^T D(q)\beta$ over $\beta \in c_n FC_0(q)$, that is,

$$(2.2) \quad \min_{\beta} (v^+)^T D(q)\beta \quad \text{subject to} \quad \psi(\beta) \leq \mu c_n.$$

This is a convex program (with linear objective function), although it is nonsmooth. If $c_n > 0$, then using a Slater condition, there exists a Lagrange multiplier $\lambda \geq 0$ such that $\partial_\beta h(\beta, \lambda) = 0$ where $h(\beta, \lambda) = v^T D(q)\beta - \lambda(\mu c_n - \psi(\beta))$ and $\partial f(z)$ is the generalized gradient of f at z [19, Thms. 6.1.1 and 6.3.1]. Furthermore, $0 \leq \lambda \perp \mu c_n - \psi(\beta) \geq 0$. That is,

$$(2.3) \quad \begin{aligned} 0 &\in \mu D(q)^T v^+ + \lambda \partial\psi(\beta), \\ 0 &\leq \lambda \perp \mu c_n - \psi(\beta) \geq 0. \end{aligned}$$

If $c_n = 0$, then $\beta = 0$ by the condition $\mu c_n - \psi(\beta) \geq 0$; since $\partial\psi(0)$ contains a neighborhood of the origin, there is a value of $\lambda \geq 0$ such that $0 \in D(q)^T v^+ + \lambda \partial\psi(0)$, no matter what v^+ and $D(q)$ are.

For a simple particle, we can take $\psi(\beta) = \|\beta\|_2 = \sqrt{\beta^T \beta}$. Then as long as the columns of $D(q)$ span the friction plane, we have a representation of the isotropic Coulomb law of friction.

2.1.1. Formulations and Function Spaces. The Signorini contact condition that $0 \leq f(q(t)) \perp c_n \geq 0$ involves complementarity between a continuous function $t \mapsto f(q(t))$ and a measure c_n , so the complementarity condition can be represented by $\int f(q(t)) c_n(dt) = 0$ as well as the usual inequality conditions: $f(q(t)) \geq 0$ for all t and $c_n(E) \geq 0$ for all Borel sets $E \subset [0, T]$. The integrated maximum dissipation principle $\min_\beta \int (v^+)^T D(q)\beta$ subject to $\psi(\beta) \leq \mu c_n$ can be formulated where β is a measure. The main difficulty is that $\psi(\beta)$ must be interpreted as a measure. Convex functions of measures were studied as measures in [29]. Since $\psi(\cdot)$ is a positively homogeneous function, $\psi(\beta) = \psi(d\beta/d|\beta|) |\beta|$, and the condition $\psi(\beta) \leq \mu c_n$ is equivalent to $\psi(d\beta/d|\beta|) |\beta| \leq \mu c_n$ or even $\psi(d\beta/dc_n) \leq \mu$ a.e. Note that the derivatives $d\beta/d|\beta|$ and $d\beta/dc_n$ are all Radon–Nikodym derivatives. The function λ is a bounded Borel function; we can bound $\|\lambda\|_\infty$ by $\max_t \mu \|D(q(t))^T v^+(t)\|_\infty / C_\psi$, where $C_\psi = \min\{\|z\|_\infty \mid z \in \partial\psi(\beta), \beta \neq 0\}$.

2.1.2. Complete Formulations. The continuous problem with one contact and inelastic impacts can be formulated as follows. The data of the problem consists of the mass matrix $M(q)$, the contact constraint $f(q)$, its gradient $n(q) = \nabla f(q)$, the matrix of direction vectors defining plane of the friction cone $D(q)$, the coefficient of friction μ , and the function $\psi(\beta)$ so that the friction cone is $FC(q) = \{n(q)c_n + D(q)\beta \mid \psi(\beta) \leq c_n\}$. Then we wish to find the trajectory $q(\cdot)$ which should be absolutely continuous, the velocity function $v(\cdot)$ which should be of bounded variation, along with the measures c_n for the normal contact forces and β to describe the frictional forces, and a bounded Borel-measurable function λ where

$$(2.4) \quad M(q) \frac{dv}{dt} = n(q) c_n + D(q) \beta - \nabla V(q) + k(q, v) + F_{ext}(t),$$

$$(2.5) \quad \frac{dq}{dt} = v,$$

$$(2.6) \quad 0 \leq c_n \perp f(q) \geq 0,$$

$$(2.7) \quad 0 \in \mu D(q)^T v^+ + \lambda \partial\psi(\beta),$$

$$(2.8) \quad 0 \leq \lambda \perp \mu c_n - \psi(\beta) \geq 0,$$

$$(2.9) \quad 0 = n(q(t))^T v^+(t) \quad \text{if} \quad f(q(t)) = 0.$$

For partly or fully elastic impacts with coefficient of restitution $0 \leq \varepsilon \leq 1$, we can replace (2.9) with

$$(2.10) \quad n(q(t))^T (v^+ + \varepsilon v^-) = 0 \quad \text{if} \quad f(q(t)) = 0.$$

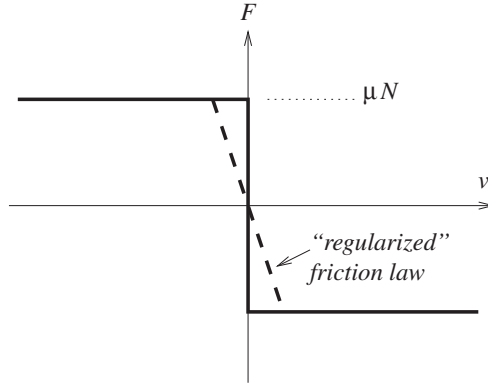


Fig. 2.1 Regularized Coulomb law to avoid discontinuity.

Related models and theory for partly elastic impacts, at least for the frictionless case, can be found in the work of Mabrouk [80, 81]. Note that this is a Newton-style model of partly elastic impact. Interestingly, this formulation of the Newton approach always dissipates energy, unlike the formulation of the Newton approach discussed by Stronge in [135]. A Poisson approach for partly elastic impacts is developed in Anitescu and Potra [5]. A new Poisson-type formulation of impact recently developed by Pang and Tzitzouris [104] is based on complementarity problems and is guaranteed to be dissipative.

It has been argued that complementarity conditions between the normal contact force and the normal velocity of the form of (2.6) are not always valid [17]. The phenomena discussed in Chatterjee [17] are associated with multiple simultaneous contacts and are discussed in section 4.4 of this article.

2.2. Numerical Methods. Numerical methods for rigid-body dynamics without friction or impact have been studied extensively in the engineering and also mathematics literature. See, for example, [34, 143, 124, 122, 52, 2, 123, 111, 33, 107, 46, 42, 142, 13], in reverse chronological order. One of the reasons for this is the need to simulate mechanical systems, especially the complex mechanical systems that arise in manufacturing processes and robotics. Even simple grasping problems involve problems of contact, impact, and friction. So far, most of the numerical methods are based on ODEs or DAEs or both. To avoid the difficulties with the discontinuity in Coulomb’s law, for instance, a regularized version is used, as shown in Figure 2.1. In this paper, a different strategy is recommended.

2.2.1. Time-Stepping Methods for Problems with Impulses. In order to handle problems like those arising with the Painlevé problem, a time-stepping approach which uses the integrals of the force functions (or measures) c_n and β over each time-step interval $[t_l, t_{l+1}]$ is used. Two different numerical formulations are presented here. The first is based on linear complementarity problems and uses a polyhedral approximation $\widehat{FC}(q)$ to the friction cone $FC(q)$. The second is a nonlinear complementarity formulation which uses $\psi(\beta)$ directly.

The polyhedral approximation to the friction cone is the cone generated by $\{n(q) + \mu d_i(q) \mid i = 1, 2, \dots, m\}$, where $\mu d_i(q)$ is a collection of direction vectors in $FC_0(q)$. Write $\widetilde{D}(q) = [d_1(q), d_2(q), \dots, d_m(q)]$. The friction forces $D(q)\beta$ are

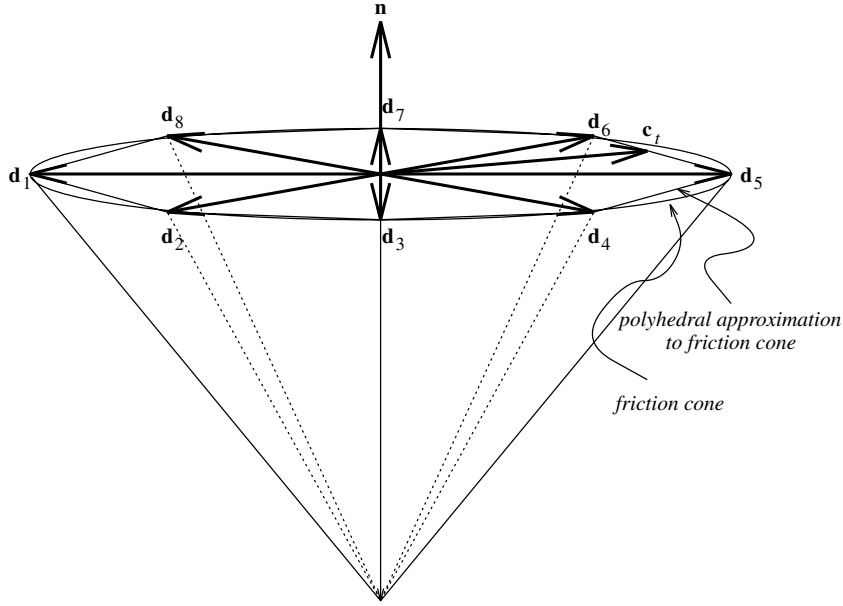


Fig. 2.2 Polyhedral approximation to the friction cone.

approximated by $\tilde{D}(q)\tilde{\beta}$, where $\tilde{\beta}_i \geq 0$ and $\sum_i \tilde{\beta}_i \leq \mu c_n$. The relationship between $\widehat{FC}(q)$ and $FC(q)$ is illustrated in Figure 2.2. It is assumed that for each i there is a j , where $d_i(q) = -d_j(q)$. This is related to the assumption that $FC_0(q)$ is a *balanced* set: $FC_0(q) = -FC_0(q)$.

The discretization of (2.4)–(2.9) is the problem of finding q^{l+1} and v^{l+1} (and the force integrals c_n^{l+1} , $\tilde{\beta}^{l+1}$ and Lagrange multiplier λ^{l+1}) given q^l and v^l for a time-step of size $h > 0$ that satisfy the following conditions:

$$(2.11) \quad M(q^{l+1})(v^{l+1} - v^l) = n(q^l)c_n^{l+1} + \tilde{D}(q^l)\tilde{\beta}^{l+1} \\ + h[-\nabla V(q^l) + k(q^l, v^l) + F_{ext}(t)],$$

$$(2.12) \quad q^{l+1} - q^l = h v^{l+1},$$

$$(2.13) \quad 0 \leq c_n^{l+1} \perp n(q^l)^T(v^{l+1} + \varepsilon v^l) \geq 0,$$

$$(2.14) \quad 0 \leq \tilde{\beta}^{l+1} \perp \lambda^{l+1} e + \tilde{D}(q^l)^T v^{l+1} \geq 0,$$

$$(2.15) \quad 0 \leq \lambda^{l+1} \perp \mu c_n^{l+1} - e^T \tilde{\beta}^{l+1} \geq 0,$$

where $f(q^l + h v^l) < 0$; if $f(q^l + h v^l) \geq 0$, then we set $c_n^{l+1} = 0$ and $\tilde{\beta}^{l+1} = 0$ and solve the first two equations. Note that e is a vector of ones of the appropriate size.

This discretization is a partly implicit Euler method. Therefore it can give only $O(h)$ accuracy at best. However, unlike conventional discretizations, it can handle impulsive forces, in particular Painlevé's problem. Note that the complementarity condition $0 \leq f(q) \perp c_n \geq 0$ does not appear explicitly in (2.11)–(2.15); (2.13) is essentially the differentiated form of this condition. Using the differentiated constraint only can result in the true constraint “drifting” into the inadmissible region, which is an effect that has been noticed in relation to DAE formulations of rigid-body dynamics with bilateral (i.e., equality) constraints [15]. It is tempting to replace (2.13) with

$0 \leq f(q^{l+1}) \perp c_n^{l+1} \geq 0$. This does not work: the resulting discretization behaves as if it had a “random” coefficient of restitution when impacts occur. (The effective coefficient of restitution depends on the time within the time-step $[t_l, t_{l+1}]$ that contact occurs.) Since (2.13) uses differentiated constraints, it may be occasionally advisable to project q^{l+1} back to the feasible region. This can be done without disturbing the time-stepping, since the time-stepping method is a one-step method.

If we ignore the dependence of M on q and substitute for v^{l+1} and q^{l+1} in terms of c_n^{l+1} , $\tilde{\beta}^{l+1}$, and λ^{l+1} , then (2.11)–(2.15) with $\varepsilon = 0$ can be reduced to a pure LCP which has the form (superscripts suppressed)

$$(2.16) \quad 0 \leq \begin{bmatrix} n^T M^{-1} n & n^T M^{-1} \tilde{D} & 0 \\ \tilde{D}^T M^{-1} n & \tilde{D}^T M^{-1} \tilde{D} & e \\ \mu & -e^T & \end{bmatrix} \begin{bmatrix} c_n \\ \tilde{\beta} \\ \lambda \end{bmatrix} + \begin{bmatrix} n^T b_1 \\ \tilde{D}^T b_2 \\ 0 \end{bmatrix} \perp \begin{bmatrix} c_n \\ \tilde{\beta} \\ \lambda \end{bmatrix} \geq 0.$$

The 3×3 block matrix in this LCP is a *copositive* matrix; a solution to this LCP exists and can be constructed by Lemke’s algorithm [25, Cor. 4.4.12].

To avoid approximations of the friction cone, we can use the integrated maximal dissipation principle: β^{l+1} maximizes $-(v^{l+1})^T D(q^l) \beta^{l+1}$ over all β^{l+1} satisfying $\psi(\beta^{l+1}) \leq \mu c_n^{l+1}$. The Kuhn–Tucker conditions for this problem replace (2.14), (2.15) above, giving

$$(2.17) \quad 0 \in D(q^l)^T v^{l+1} + \lambda^{l+1} \partial \psi(\beta^{l+1}),$$

$$(2.18) \quad 0 \leq \lambda^{l+1} \perp \mu c_n^{l+1} - \psi(\beta^{l+1}) \geq 0.$$

Of course, (2.11) should be replaced by

$$(2.19) \quad M(q^{l+1})(v^{l+1} - v^l) = n(q^l) c_n^{l+1} + D(q^l) \beta^{l+1} \\ + h[-\nabla V(q^l) + k(q^l, v^l) + F_{ext}(t_l)].$$

Solving these systems requires more sophisticated methods since we have to solve for an inclusion $0 \in D(q^l)^T v^{l+1} + \lambda^{l+1} \partial \psi(\beta^{l+1})$ as well as for nonlinear complementarity conditions $0 \leq \lambda^{l+1} \perp \mu c_n^{l+1} - \psi(\beta^{l+1}) \geq 0$.

The existence of solutions to the one-time-step conditions above can be shown under mild conditions via the results in [129] for the formulation using $\tilde{\beta}$ and extended via [101] to the formulation(s) in β , using the nonlinear condition $\mu c_n - \psi(\beta) \geq 0$.

2.3. Practicalities. If we use the discretization (2.11)–(2.15), we can use Lemke’s algorithm in an iteration: given an estimate $q^{l+1,k}$ of q^{l+1} , we can obtain an estimate $q^{l+1,k+1}$ by solving (2.11)–(2.15) with q^{l+1} replaced by $q^{l+1,k}$. This iteration will usually converge (and converge quickly), giving a solution of the mixed nonlinear complementarity problem (2.11)–(2.15) [132].

Replacing (2.14)–(2.15) with (2.17)–(2.18), using the nonlinear condition $\mu c_n \geq \psi(\beta)$ instead of the conditions $\mu c_n \geq e^T \tilde{\beta}$ and $\tilde{\beta} \geq 0$, leads to highly nonlinear complementarity problems. Since ψ is nonsmooth, it may be desirable to replace it with smooth functions. For example, for isotropic friction, where $\psi(\beta) = \|\beta\|_2$, we can replace the condition $\mu c_n \geq \psi(\beta)$ with $(\mu c_n)^2 \geq \|\beta\|_2^2 = \beta^T \beta$. The difficulty in using this condition is that if $c_n = 0$, the constraint qualification fails, and Lagrange multipliers may not exist for the maximal dissipation principle in this form. However, it can be restored by using a version of the Fritz John conditions (see [43, section 2.2, pp. 190–203], for example): replace (2.17) with

$$0 = c_n^{l+1} D(q^l)^T v^{l+1} + 2\lambda^{l+1} \beta^{l+1}.$$

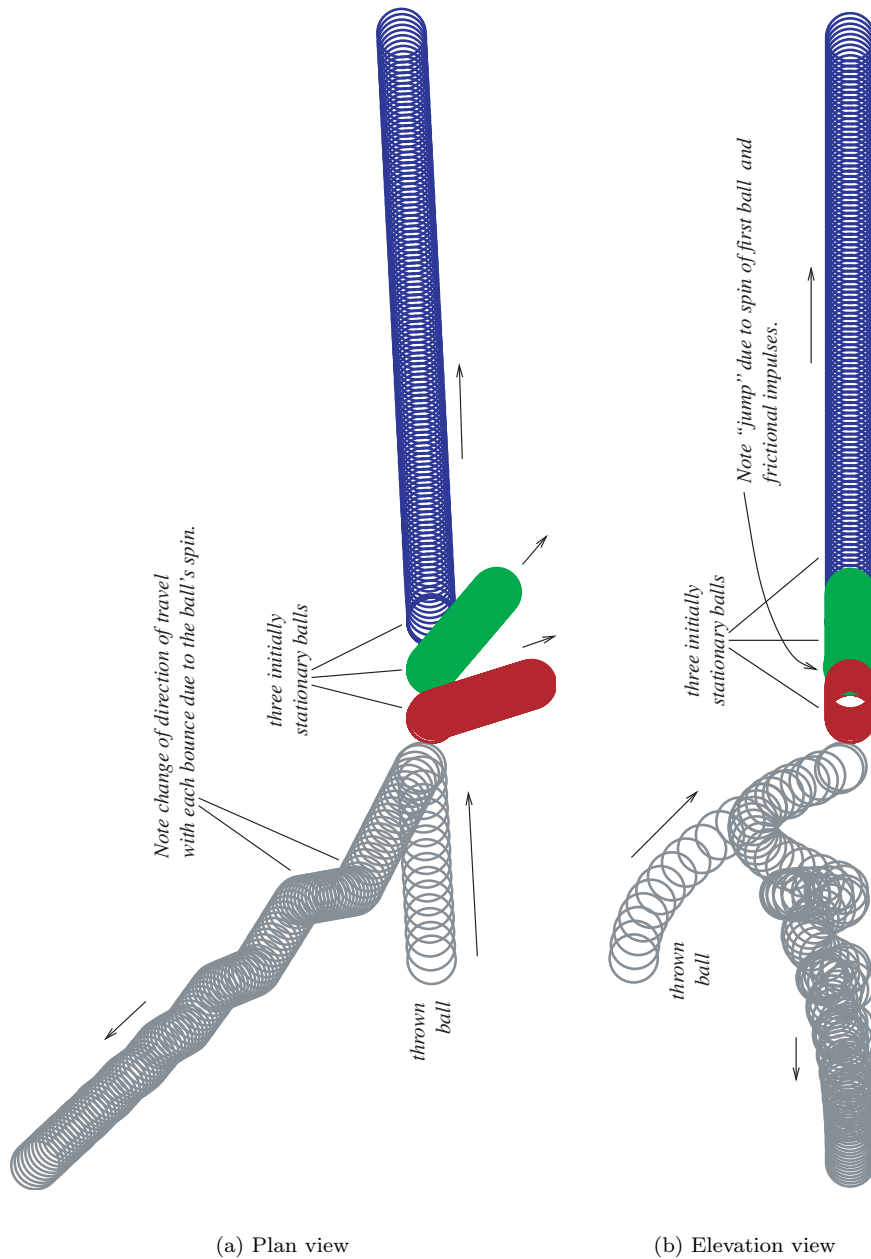


Fig. 2.3 Elevation and plan views of “billiards” problem with partly elastic impacts.

These nonlinear complementarity problems can be solved using nonsmooth Newton methods, which converge at least locally (see, for example, [51, 98, 100, 112]). These methods can be made global by using continuation methods (see, for example, [1, 140] for overviews of practical homotopy/continuation methods and [101] for a theoretical analysis of the basis for using continuation methods for frictional contact problems).

Some numerical results are presented graphically in Figure 2.3. This shows one ball being thrown and colliding with the first of three balls on a table. The coefficient

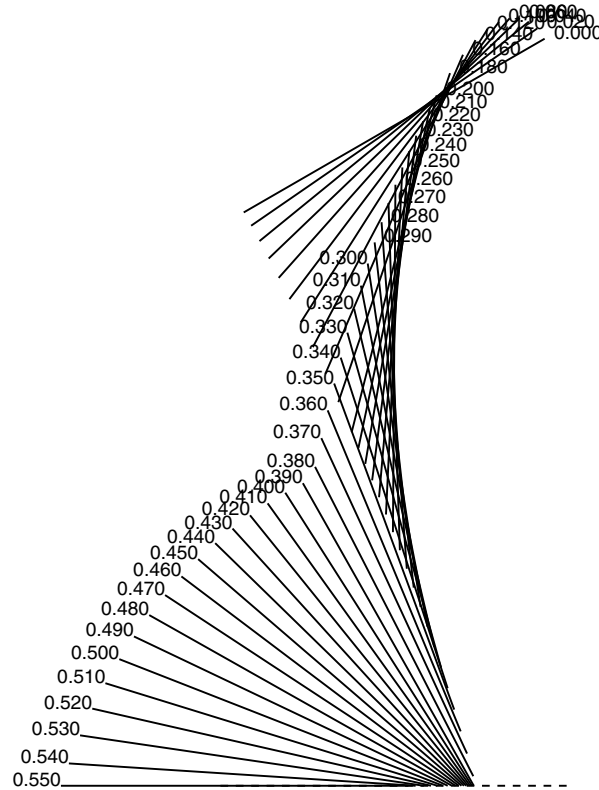


Fig. 2.4 Painlevé's problem: impact version.

of restitution is 0.9, which is nearly perfectly elastic. Note that most of the momentum is transferred to the third ball on the table. Since the impact is slightly oblique, the balls do not remain in a line but move off in different directions. Also note some other effects: the thrown ball rebounds with a strong spin, and when it bounces, the spin reverses direction, resulting in an alternating pattern in terms of direction and speed of its bounces. This effect can be clearly seen when “superballs” (small solid rubber balls designed to bounce well, sold in toy stores) are made to spin rapidly. More subtly, the second of the balls on the table bounces up slightly on impact. An analysis of the frictional and normal impulses will reveal that in a similar situation with n colinear balls struck by a rolling ball, all of the even-numbered originally stationary balls will receive an upward frictional impulse.

Another numerical example is the Painlevé example in an impact situation which is illustrated in Figure 2.4. Any method which assumes that the horizontal component of the contact velocity does not change during the moment of impact cannot compute the solution of this problem correctly. The numerical results shown are for a rod that is initially spinning counterclockwise with stationary center of mass. It then falls while spinning until it hits the table. The solution shown is for inelastic impacts. The numbers shown are the times for each configuration in seconds.

3. Convergence and Existence. The fundamental question is, “Do the numerically computed trajectories satisfying (2.11)–(2.15) converge to exact solutions of (2.4)–(2.9)?” If this question can be answered affirmatively, then it also gives an *existence* theorem for solutions to rigid-body dynamics.

Certain situations are harder than others to deal with; we have already seen how Painlevé’s famous problem causes difficulties for analysis. The difficulties that arise with Painlevé’s problem can be related to the fact that $M(q)^{-1}FC(q)$ points outside the set of admissible velocities. This can be avoided if we assume Erdmann’s condition that for any $q \in \partial C$ and $z \in FC(q)$, $n(q)^T M(q)^{-1}z \geq 0$ [36]. Note that Erdmann’s condition holds automatically for frictionless problems, because $z \in FC(q) = \text{cone}(n(q))$ implies that $z = \alpha n(q)$ ($\alpha \geq 0$) and $n(q)^T M(q)^{-1}z = \alpha n(q)^T M(q)^{-1}n(q) \geq 0$ since $M(q)$ is positive definite. Also, Erdmann’s condition implies that there cannot be any impulses without collisions; that is, $n(q(t))^T v(t^-) \geq 0$ implies that $v(t^+) = v(t^-)$ under Erdmann’s condition.

The question of convergence is answered affirmatively in [129] for one inelastic contact ($\varepsilon = 0$) in the case of one-dimensional friction (that is, $FC_0(q)$ is a one-dimensional set), or if Erdmann’s condition holds [36]. This result includes Painlevé’s problem. It also extends the fundamental results of Monteiro Marques [84] for inelastic frictional dynamics of a particle.

In this section we will review how to prove convergence (and thus existence) of solutions to rigid-body dynamics problems. This is based on the results in [129], which should be consulted for more details. A more complete summary of the proof is in [128].

3.1. The Easy Part. The main part of the theorem can be handled by standard techniques once a few basic facts are established. The first is the existence of solutions to the mixed complementarity conditions (2.11)–(2.15) at each time-step. This is done in [129] by a combination of an approximation argument and the Brouwer fixed-point theorem. The numerical solutions are approximately dissipative. This is based on the result that for constant M , n , D , and linear $V(q)$, the numerical solutions are exactly dissipative, which is proved using $(v^{l+1})^T M(v^{l+1} - v^l) = \frac{1}{2}(v^{l+1})^T M v^{l+1} - \frac{1}{2}(v^l)^T M v^l + \frac{1}{2}(v^{l+1} - v^l)^T M(v^{l+1} - v^l)$ and substituting for $M(v^{l+1} - v^l)$ in terms of c_n^{l+1} and β^{l+1} . Using a generalization of the discrete Gronwall lemma to nonlinear ODEs, the energy of the computed trajectory is shown to be bounded on some sufficiently small interval $[t_0, t_1]$, $t_1 > t_0$. This means that the computed velocities $v^h(\cdot)$ are bounded on $[t_0, t_1]$. Since $dq/dt = v$, the numerically computed functions $q^h(\cdot)$ are uniformly Lipschitz and thus equicontinuous. By the Arzelá–Ascoli theorem, there is a uniformly convergent subsequence. We restrict attention to the subsequence.

Assuming that the friction cones $FC(q)$ are pointed and that $q \mapsto FC(q)$ has a closed graph, it can be concluded from the uniform boundedness of $v^h(\cdot)$ that the variations $\bigvee v^h$ are also uniformly bounded. By the Helly selection theorem, there is a convergent subsequence where $v^h(\cdot) \rightarrow v(\cdot)$ pointwise, and the differential measures $dv^h \rightharpoonup dv$ weak*. By the weak* closedness results in [130], the limit $v(\cdot)$ satisfies the measure differential inclusion in the continuous formulation.

3.2. The Hard Part. The hard part involves nonstandard arguments and complementarity conditions between functions that converge pointwise and measures that converge weak*.

First, the limits $q(\cdot)$ and $v(\cdot)$ can be proved to be exactly dissipative in the sense that the total energy of the solution $\frac{1}{2}v(t)^T M(q(t))v(t) + V(q(t))$ is a nonincreasing function of t . This could not be proved before, because the argument used at this stage needs a uniform bound on $\bigvee_0^T v^h$. From the exact dissipativity result, we can go on to show that the solution grows at most exponentially on any interval where the limit exists. By “bootstrapping” this argument with the local boundedness result, it can be shown that limits $q(\cdot)$ and $v(\cdot)$ exist on $[0, \infty)$.

The next step is to show that $f(q(t)) \geq 0$ for all t . This is needed because the time-stepping formulation only ensures that the differentiated constraint $n(q^l)^T v^{l+1} \geq 0$ at each step where $f(q^l + hv^l) < 0$. The other step that is needed is to show that the support of the limiting measure c_n is contained in the set $\{t \mid f(q(t)) = 0\}$. This is equivalent to showing that $c_n(\{t \mid f(q(t)) > 0\}) = 0$. This can be done by showing that $c_n(\{t \mid f(q(t)) \geq \epsilon\}) = 0$ for any $\epsilon > 0$. Thus we have the complementarity condition that $0 \leq f(q) \perp c_n \geq 0$ between a (continuous) function and a Borel measure. This is the correct limiting contact condition.

The next condition to consider is the inelastic impact rule. The proof that the limits $q(\cdot)$ and $v(\cdot)$ satisfy the condition $f(q(t)) = 0 \Rightarrow n(q(t))^T v^+(t) = 0$ in [129] is restricted to the one-contact case. Simple examples show that this condition cannot be directly generalized to multiple contacts. The proof is based on the result that for the one-contact case, $n(q^{l+1})^T v^{l+1} \leq \max(0, (n(q^l)^T v^l)) + O(h)$, which is obtained from the complementarity condition $0 \leq n(q^l)^T v^{l+1} \perp c_n^{l+1} \geq 0$ for $f(q^l + hv^l) < 0$. (If $f(q^l + hv^l) \geq 0$, then there are no contact forces, and the result holds trivially.)

Finally, the Coulomb law in its maximal dissipation version must be satisfied. That is, we need to show that

$$\beta^T D(q)^T v^+ = -\mu c_n \|D(q)^T v^+\|_\infty$$

as measures. Since the numerical velocity functions v^h converge pointwise, and the measures β^h and c_n^h only converge weak*, we cannot directly take limits, even though this property holds for the discrete formulation. First, it is shown that $n^T v - \mu \|D(q)^T v\|_\infty$ is a right lower semicontinuous function of time. This is obtained from the discrete formulation, where $(v^{l+1} - v^l)^T (n(q^l)^T c_n^{l+1} + D(q^l)^T \beta^{l+1})$ is expanded two ways: one by expanding $(v^{l+1} - v^l)$ using the discrete equations of motion, and the other by multiplying the expression out and using the complementarity and optimization conditions. The next step is to show that any weak* limit ν of any subsequence of $c_n^h [n(q^h)^T v^h - \mu \|D(q^h)^T v^h\|_\infty]$ is bounded below by the measure $c_n [n(q)^T v^+ - \mu \|D(q)^T v^+\|_\infty]$. The next step is to use this inequality as part of a detailed accounting of the energy in the limit. Part of these computations involve forming the differential measure of the energy:

$$d\left(\frac{1}{2} v^T M(q) v + V(q)\right) = \frac{1}{2} (v^+ + v^-) M(q) dv + \frac{1}{2} v^T \left(\frac{d}{dt} M(q)\right) v dt + \nabla V(q)^T v dt.$$

The first term comes from a result of Moreau for differential measures of quadratic functions [85]. Corresponding formulas for the numerical approximations are developed. One of the difficulties in this step is attempting to take limits of the measures $((v^h)^+ - (v^h)^-)^T M(q) dv^h$. Unfortunately, because the numerical acceleration measures dv^h only converge weak* and not weakly, the usual weak lower semicontinuity arguments (based on Mazur's lemma) do not hold, and other means are needed to prove the desired inequalities.

At this point, the argument breaks up into three cases. The first is where the limit $v(t)$ is continuous; note that the measure $(v^+ - v^-) M(q) dv$ has its support on the discontinuities in v . This enables the proof of the validity of the Coulomb law for the limit at every point of continuity of v . Of course, the real interest is in the discontinuities of v . The second case is where the condition of Erdmann [36] holds: the friction cone transformed by $M(q)^{-1}$ is strictly inside the tangent cone of the feasible region. That is, $n(q)^T M(q)^{-1} z > 0$ for any $z \in FC(q)$ and $z \neq 0$. The violation of this condition seems to be essential for Painlevé's and related examples. The argument used will not be described here, except that it uses the inelastic impact

result. The final case includes the Painlevé case: it is the case where the friction plane is one-dimensional ($\dim \text{span } D(q) = 1$). A general argument shows that the only way the Coulomb law can fail for the limit is if the numerical friction impulses have oscillations that are not $O(h)$ in magnitude within a small time interval. In the one-dimensional case, the only way the numerical friction impulses can oscillate in this way is if the sign of $(D(q)^T v)_1$ oscillates while c_n^{l+1} is “large”; this is shown to be impossible.

While these results do not cover all cases of interest (especially two-dimensional friction planes), they do provide a satisfying resolution of some well-known “paradoxes.” Further, to extend these results to two-dimensional friction planes, all that is needed is to rule out some pathological behavior by the numerical methods.

3.3. Other Aspects.

3.3.1. Uniqueness. It is well known that solutions to rigid-body dynamics problems can have multiple solutions. For example, consider Painlevé’s problem with slightly different starting values: $\dot{\theta} < -\sqrt{2g/l} \cos \theta$ ensures that $\dot{x}_c < 0$, but the complementarity problem obtained using the analysis of section 1.3 is

$$0 \leq \ddot{y}_c = \left[\frac{1}{m} - \frac{l^2}{4J} \cos \theta (\mu \sin \theta - \cos \theta) \right] N + (l/2) \cos \theta \dot{\theta}^2 - g \perp N \geq 0.$$

Since the quantity in brackets is negative and $(l/2) \cos \theta \dot{\theta}^2 - g > 0$, there are two solutions to this complementarity problem, one of which corresponds to continued sliding, the other of which corresponds to the contact breaking ($\ddot{y}_c > 0$). There is, in fact, a third impulsive solution, which cannot be obtained from this analysis, just as the impulsive resolution of Painlevé’s original problem cannot be obtained through his analysis. It turns out that the “continued sliding” solution is extremely unstable. To see this, a singular perturbation analysis of the rigid limit of stiff contact must be carried out; as the stiffness k of the contact is increased, the time for a perturbation to double in magnitude decreases as $O(1/\sqrt{k})$.

In general, nonuniqueness of solutions can be seen as the result of extreme instability in stiff approximations. To predict which of the possible solutions actually occurs requires knowledge of the microscopic details of the contact, which is not available at this level of modeling. In practice, there is little point in trying to predict which solution occurs in reality, because other unmodeled aspects of a real system will outweigh the effects of imprecise modeling of the contact phenomena.

The lack of uniqueness also signals another difficulty: the solutions are not necessarily continuous in the data. The best that can be done in these circumstances is to repeat the simulations, but with random disturbances, so that a set of possible outcomes can be determined. Determining *all* possible outcomes computationally is not feasible at present. Even if all of the time-stepping problems (2.11)–(2.15) had unique solutions, this does not imply that the continuous problem has unique solutions. (An example related to this is Ballard’s existence and nonuniqueness proof for quasi-static contact problems; the solutions are nonunique for arbitrarily small friction coefficients $\mu > 0$ [9]!)

3.3.2. Control and Optimization. Much of the work done on rigid-body dynamics has a view to controlling mechanical systems; this is a clear motivation in robotics, for example. See [14]. Classical optimal control methods such as the Pontryagin principle and the calculus of variations require a certain amount of smoothness, even

for the nonsmooth versions in Clarke [19], which do not hold in general for these problems. Even if the normal contact force N is known, at least as a function of the configuration, and the map from control functions to solution trajectories is well defined and Lipschitz, the variational approach to optimal control is beyond current tools. Partial results in this direction have been achieved by Frankowska [44] based on the differential inclusion formulation, for example; however, these results assume that the right-hand side $F(x)$ in $dx/dt \in F(x)$ is a Lipschitz set-valued function in the Hausdorff metric. Note that the Hausdorff metric for closed bounded sets in \mathbf{R}^n is given by

$$d_H(A, B) = \max(\max\{d(a, B) \mid a \in A\}, \max\{d(b, A) \mid b \in B\}),$$

where $d(x, C)$ is the distance from x to C . By contrast, the right-hand side for the Coulomb friction law is not even continuous in the Hausdorff metric on sets.

Since this area is of practical importance, it will see a great deal of interest. Several avenues are open: One is to regularize the Coulomb law and the contact conditions, and apply Pontryagin to the regularized system. This leads to stiff equations and a singular perturbation approach. Another is to attempt to handle discontinuous differential equations directly; since the computation of the adjoint or dual variables in the Pontryagin approach amounts to a differentiation, the adjoint functions can be expected to be discontinuous for discontinuous ODEs, even though the trajectories are continuous. Another approach is to apply a pattern search to these problems in order to compute optimal trajectories. None of these approaches is perfect, and much needs to be done.

4. Open Questions, Other Ideas. Many people have worked on rigid-body dynamics from different points of view, and there are many aspects of these problems that deserve attention. Some of this related work is briefly discussed in this section. This is not an exhaustive discussion of these issues, but rather an introduction to some of the interesting unresolved issues in this area.

From an applications point of view, rigid-body dynamics is just the limit of dynamics of elastic bodies as the coefficients of elasticity go to infinity. Some work has been done on justifying impact laws from this approach. The “holy grail” for this point of view would be complete justification of models of rigid-body dynamics by singular perturbation analysis. This still seems a long way off.

Integrators for mechanical systems with bilateral constraints have become more refined with the development of DAE solvers and symplectic integrators. Rigid bodies can be approximated by very stiff elastic bodies and symplectic integrators can be applied to these problems. By using adaptive step-sizes (which requires some care to keep symplecticness!) solutions can be generated, at least for perfectly elastic impacts. However, I believe that for impact and contact problems, a tighter coupling between the contact conditions and the integrators should be developed.

The standard Coulomb model is inadequate in itself to describe a number of experimentally observed phenomena, such as velocity-dependent coefficients of friction. This brings new difficulties to the theory. The simplest model is a two-coefficients model with a static and a dynamic coefficient of friction. The extra discontinuity makes the theoretical analysis particularly difficult. Another model with a better experimental basis is to have $\mu = \mu(\|v\|)$, which is also better behaved theoretically, although there are a number of open questions.

Multiple contact problems predominate in applications; handling them well requires attention to theory, too.

Finally, dynamic problems with elastic bodies in contact with Coulomb friction are still beyond the reach of our theoretical tools, in spite of the considerable progress and partial or approximate solutions that have been found.

4.1. Singular Perturbations and the Rigid Limit. Rigid-body dynamics is really the study of the limiting case, where the elasticity constants, such as Young's modulus, go to infinity. Ultimately, the justification of rigid-body dynamics should be via a singular perturbation theory for stiff elastic bodies. Such a theory has not yet been developed, although there are some partial results in this direction.

Paoli and Schatzman [105, 106] have proven some singular perturbation results for problems involving particles without friction, but with partly elastic impacts. In [105, 106] the limiting problem is assumed to have a convex admissible set K with nonempty interior and a C^2 boundary. The equations of motion in the interior of K are given by $u'' = f(t, u, u')$, while on the boundary the impact law is $u'(t^+) = -\varepsilon u'_n(t^-) + u'_t(t^-)$, where $u'_n(t) = n(n^T u'(t))$ is the normal component of the velocity $u'(t)$ and $u'_t(t) = u'(t) - u'_n(t)$ is the tangential component of the velocity. To approximate this, Paoli and Schatzman use a penalty law

$$u''_\lambda + \frac{2\eta}{\sqrt{\lambda}}G(u_\lambda - P_K(u_\lambda)) + \frac{1}{\lambda}(u_\lambda - P_K(u_\lambda)) = f(t, u_\lambda, u'_\lambda),$$

where $P_K(z)$ is the projection of z onto K , $G(u, v) = (u^T v)u/\|u\|^2$ if $u \neq 0$ and zero if $u = 0$, and η is related to the coefficient of restitution ε . Note that the quantity $u_\lambda - P_K(u_\lambda)$ acts as a measure of the amount of interpenetration occurring between the particle and the boundary of the admissible region. Paoli and Schatzman show that as $\lambda \downarrow 0$, $u_\lambda \rightarrow u$ with u a solution of the continuous problem with convergence strong in $W^{1,p}$ for $1 \leq p < \infty$ and weak in $W^{1,\infty}$.

Clearly there is a great deal to be done in this area: friction needs to be included, the convexity assumption on the admissible region should be weakened, and the whole theory should be extended to infinite-dimensional problems involving extended elastic bodies. Some relevant issues are discussed below in sections 4.4 (on multiple contacts) and 4.5 (on elastic bodies).

4.2. Symplectic and Variational Methods for Impact. Symplectic integrators have become very popular for integrating the equations of motion of conservative mechanical systems. (See, for example, [45, 48, 73, 113, 116].) Symplectic methods are methods for Hamiltonian systems: given a Hamiltonian $H(q, p)$ defined in terms of the coordinates q and momentum variables p , the Hamiltonian system is

$$\begin{aligned} \frac{dq}{dt} &= + \frac{\partial H}{\partial p}(q, p), \\ \frac{dp}{dt} &= - \frac{\partial H}{\partial q}(q, p). \end{aligned}$$

Symplectic integrators preserve the differential forms $dq_i \wedge dp_i$. They do not exactly conserve the Hamiltonian (which is the total energy), but they nearly conserve a "numerical" Hamiltonian $H_h(q, p)$, where $h > 0$ is the step-size used for the symplectic method [116]. The numerical Hamiltonian $H_h(q, p)$, is usually obtained from H via Taylor series.

Adaptive versions of these methods can deal with certain kinds of singularities, such as those arising from terms of the form $f(q)^{-\alpha}$ for large α which models “soft” contacts, and bilaterally constrained mechanical systems. Because of the possibility of unbounded right-hand sides, implicit methods are necessary. Take, for example, the simplest implicit (nonpartitioned) symplectic method, the implicit midpoint rule for $dx/dt = f(t, x)$:

$$x_{n+1} = x_n + h f((t_n + t_{n+1})/2, (x_n + x_{n+1})/2).$$

Consider the case of a single particle in one dimension with only contact forces and gravitation, and a simple inequality constraint: $q(t) \geq 0$. The Hamiltonian of the system is $H(q, p) = \frac{1}{2m}p^2 + \psi(q) + gq$, where $\psi(q) = 0$ for $q \geq 0$ and $\psi(q) = +\infty$ for $q < 0$. This gives the differential equations and inclusions

$$\begin{aligned} \frac{dq}{dt} &= \frac{1}{m}p, \\ \frac{dp}{dt} &\in -\partial\psi(q) - g. \end{aligned}$$

Note that $\partial\psi(q) = \{0\}$ if $q > 0$, $\partial\psi(q) = \mathbf{R}_+$ if $q = 0$, and $\partial\psi(q) = \emptyset$ if $q < 0$. Applying the implicit midpoint rule to this system gives the numerical scheme

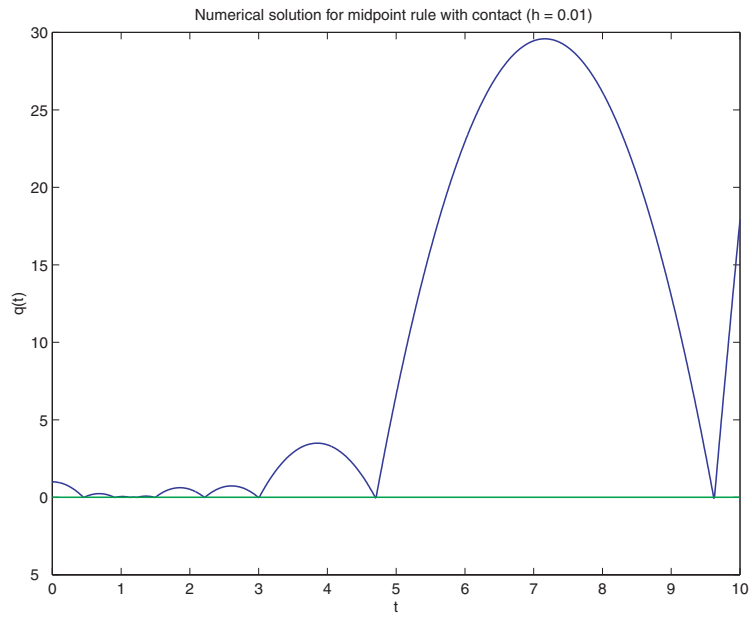
$$\begin{aligned} q_{n+1} &= q_n + \frac{h}{2m}(p_n + p_{n+1}), \\ p_{n+1} &\in p_n - \frac{h}{2}\partial\psi((q_{n+1} + q_n)/2) - hg. \end{aligned}$$

Writing N_{n+1} for the element we choose from $-(h/2)\partial\psi((q_n + q_{n+1})/2)$ (that is, N_{n+1} is the normal contact impulse), we get the mixed complementarity problem

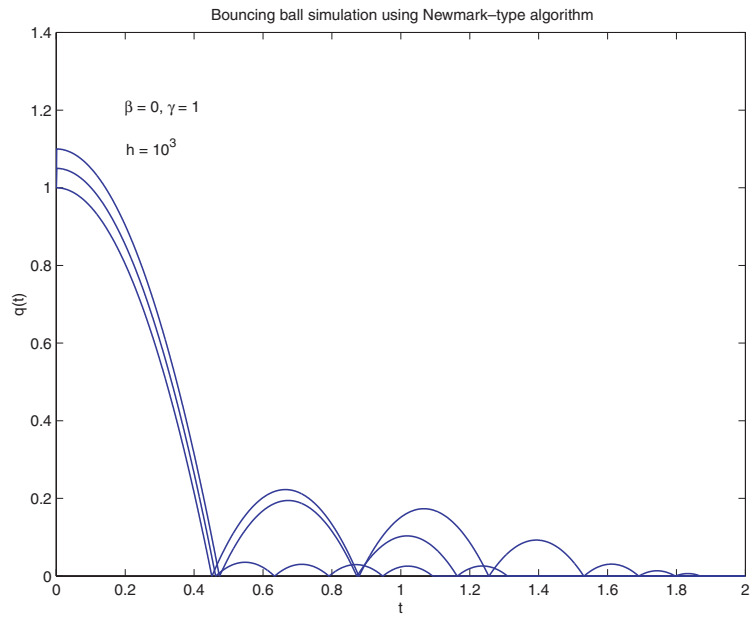
$$\begin{aligned} p_{n+1} &= p_n + N_{n+1} - hg, \\ q_{n+1} &= q_n + \frac{h}{2m}(p_n + p_{n+1}), \\ 0 &\leq N_{n+1} \perp (q_n + q_{n+1})/2 \geq 0. \end{aligned}$$

Since symplectic methods do not conserve $H(q, p)$ and the numerical Hamiltonian $H_h(q, p)$ is obtained assuming a smooth $H(q, p)$, it does not follow that this method is conservative even in the limit as $h \downarrow 0$. In fact, numerical results indicate that it is not conservative: if $q_n < 0$ and $p_n < 0$, then $N_{n+1} > 0$ is needed to ensure that $q_n + q_{n+1} \geq 0$; by complementarity, $q_{n+1} = -q_n > 0$; therefore $p_n + p_{n+1} = 2(2m/h)q_{n+1}$ and $p_{n+1} = -p_n + (4mq_{n+1}/h) > |p_n|$. The effective coefficient of restitution will depend on $q_n/(hp_n)$. A typical numerical trajectory is shown in Figure 4.1(a) for $q(0) = 1$, $p(0) = 0$ with $h = 0.01$, $g = 9.8213$, and $m = 1$. Clearly energy is not conserved, even approximately. This should not be surprising: the Hamiltonian cannot in general be conserved by any fixed step-size unpartitioned method [47], and conservation of momentum is of little use when we have a priori unknown impulsive contact forces!

If we replace the complementarity condition $0 \leq N_{n+1} \perp q_n + q_{n+1} \geq 0$ with one involving the momenta $0 \leq N_{n+1} \perp p_n + p_{n+1} \geq 0$, the resulting scheme does give perfectly elastic impacts. However, this new scheme cannot be interpreted as a standard method for ODEs. This shows how much care must be taken with numerical simulation of impact phenomena.



(a) Midpoint rule



(b) Newmark scheme

Fig. 4.1 Numerical trajectories using the midpoint rule and a Newmark scheme.

Other schemes, such as the Newmark schemes, can also be applied to contact problems [59]. However, these schemes also suffer from indeterminism in the effective coefficient of restitution. In Figure 4.1(b), results are shown for a Newmark scheme [59] with $\beta = 0$, $\gamma = 1$, and step-size $h = 10^{-3}$ for the same bouncing ball problem with several initial conditions to illustrate the nondeterminism. Even variational methods based on discrete Lagrangian principles [141] do not seem to behave correctly.

4.3. Static vs. Dynamic Friction. It is an experimentally observed phenomenon that the coefficient of friction often depends on the sliding velocity as well as the nature of the materials in contact. This has a number of practical implications. For example, if you need to brake suddenly while driving, the common advice is to “pump” the brakes. (This advice does not apply to vehicles equipped with ABS, which “pumps” automatically.) This is because when the car begins to slide the coefficient of friction appears to decrease, and the car does not decelerate as quickly. (Many people perceive this effect as the car actually accelerating; in fact it is still decelerating, but not as quickly.)

The standard Coulomb model of dry friction does not take this velocity dependence into account. There are several ways of incorporating velocity dependence into the Coulomb model. The simplest is to have two coefficients of friction: the static coefficient μ_{static} and the dynamic coefficient $\mu_{dynamic}$. From the differential inclusion point of view, it is clear that $\mu_{dynamic} \leq \mu_{static}$, for if $\mu_{dynamic} > \mu_{static}$, there could not be any motion until the friction limit $\mu_{dynamic}N$ is overcome. Thus there is no sense having $\mu_{dynamic} > \mu_{static}$. This discontinuous model of friction leads to a number of theoretical difficulties. For example, consider the “brick on a ramp” problem with this model of friction. Treating the problem naively, the equations of motion become

$$m \frac{dv}{dt} \in mg \sin \theta - mg \cos \theta \operatorname{Sgn} v \begin{cases} \mu_{dynamic} & \text{if } v \neq 0, \\ \mu_{static} & \text{if } v = 0, \end{cases}$$

and the right-hand side contains the right-hand side for the original friction model (1.1) with $\mu = \mu_{dynamic}$. This means that if we interpret this simply as a differential inclusion, the brick can begin sliding whenever $\mu_{dynamic}mg \cos \theta < mg \sin \theta < \mu_{static}mg \cos \theta$. Clearly this is missing something crucial about the two-friction model: there is a hysteresis effect. Once $v = 0$, the coefficient of friction is “stuck” at μ_{static} until such time as the external force is sufficient to overcome the static friction force limit $\mu_{static}N$.

This model is vulnerable to some strong criticisms, such as those of Ruina [115, section 9.3.1], who discusses the model from a continuum point of view.

Even when the hysteresis effect is understood and is used to remove most of the nonuniqueness associated with this model, it is still quite possible to get multiple solutions when the velocity function $v(t)$ “grazes” $v = 0$; that is, $v'(t) < 0$ for $t < t^*$, $v(t^*) = 0$, $v'(t^*) = 0$, but $v''(t^*) > 0$. Then the friction coefficient may jump from $\mu_{dynamic}$ to μ_{static} and “grab” the solution, forcing it to stay at zero for an interval. Or perhaps the coefficient of friction might stay at $\mu_{dynamic}$ and allow the velocity to move away from zero again. How to treat this nonuniqueness is an unresolved theoretical issue with this model.

Another model which has a stronger experimental basis is to write $\mu = \mu(\|v\|_2)$, where $\mu'(s) < 0$ and $\lim_{s \rightarrow \infty} \mu(s) = \mu_\infty > 0$. An example of this model is shown in Figure 4.2(c). The behavior of systems with this friction model is very similar to the common understanding of the two-friction-coefficients model. If $v \neq 0$, the fact that

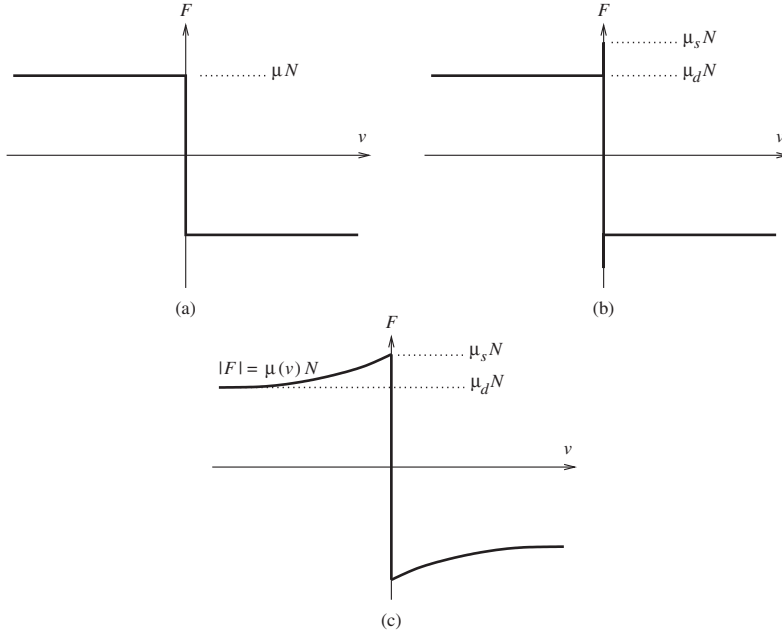


Fig. 4.2 One-dimensional friction models: (a) standard, (b) two-coefficient, (c) continuously dependent coefficient of friction.

$\mu'(\|v\|_2) < 0$ actually makes the system unstable, in spite of the energy dissipation due to $\mu(\|v\|_2) > 0$. Consider again the “brick on a ramp” problem of Figure 1.1 with a small external force $F_{ext}(t)$:

$$(4.1) \quad m \frac{dv}{dt} \in mg \sin \theta - mg \cos \theta \mu(|v|) \operatorname{Sgn} v + F_{ext}(t).$$

Suppose that at time $t = 0$, $v = 0$. Then if we make $F_{ext}(t)$ sufficiently large, so that $mg \sin \theta + F_{ext}(t) > mg \cos \theta \mu(0)$, the brick will begin sliding. But then $m(dv/dt) = mg \sin \theta - F_{ext}(t) - mg \cos \theta \mu(|v|)$. As v increases, $\mu(|v|)$ decreases, so dv/dt will increase, raising the acceleration still higher. This is an example of *friction-induced instability*. If $\mu'(|v|)$ approaches μ_∞ quickly, then $\mu'(|v|)$ will be large and negative, giving a rapid transition to the regime where $\mu(|v|) \approx \mu_\infty$. If we identify μ_{static} with $\mu(0)$ and $\mu_{dynamic}$ with μ_∞ , the two models produce very similar behavior, although the theoretical difficulties of the $\mu = \mu(\|v\|_2)$ model are much less. From the point of view of experiments, the frictional instability means that measurements of these effects require strong viscous damping in order to obtain a proper relationship between μ and v . Nevertheless, there are a number of open questions: For example, in the case of Painlevé problems with velocity-dependent friction, can the sliding velocity be brought to zero instantaneously even if $\mu(\|v^-\|_2)$ is below the threshold normally needed for the Painlevé effect?

4.4. Multiple Contacts and Multiple Impacts. Multiple impact and multiple contact problems can be formulated in the same style as the one-contact formulations here. These can be found in, for example, [5, 6, 132, 129]. The essence of these formulations is that the friction cones in generalized coordinates are summed to give

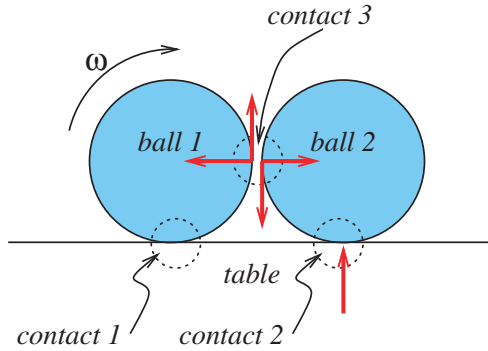


Fig. 4.3 Two balls colliding inelastically.

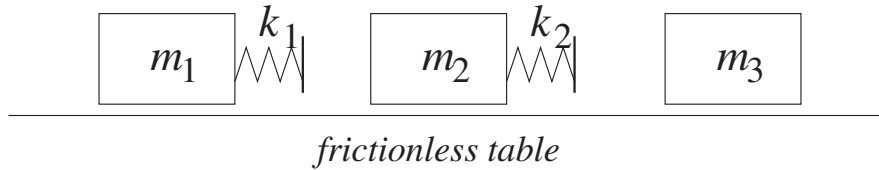


Fig. 4.4 Three masses problem.

the complete friction cone for the entire system: $FC(q) = FC^{(1)}(q) + FC^{(2)}(q) + \dots + FC^{(p)}(q)$ for a system with p contacts. Each contact has its own constraint $f^{(j)}(q) \geq 0$, coefficient of friction $\mu^{(j)}$, coefficient of restitution $\varepsilon^{(j)}$, normal contact force $n^{(j)}(q)c_n^{(j)}$, friction force $D^{(j)}(q)\beta^{(j)}$, and $\lambda^{(j)}$. The complementarity conditions for a single contact are replicated across all of the contacts.

The difficulty with multiple contacts is the correct way of dealing with the continuous formulations of the impact laws. For example, for inelastic impacts, $n^{(j)}(q)^T v^+ = 0$ if $f^{(j)}(q) = 0$ for all j does not hold. Consider, for example, a pair of balls colliding inelastically on a table as illustrated in Figure 4.3. If the coefficients of friction are all strictly positive, with ball 1 initially rolling and ball 2 initially at rest on the table, then on impact ball 1 will jump up due to a vertical frictional impulse at contact 3 on ball 1. This means that while contact 2 and contact 3 both behave inelastically, contact 1 appears not to. Correct and complete impact laws for multiple contacts and impacts are not known even for inelastic impacts. Other difficulties arise with formulating multiple contacts. One is to obtain even numerical formulations which are dissipative where the coefficients of restitution $\varepsilon^{(j)}$ are different for the different contacts. (One formulation put forward by Pfeiffer and Glocker in [50] was later found not to be dissipative [3].)

As well as the difficulties in formulating laws for multiple contacts, there are a number of physical effects that should be considered as well. One of the most important of the properties that are lost in current rigid-body models is the concept of “relative stiffness.” Consider, for example, three masses that can affect each other through springs.

The three masses in Figure 4.4 interact through the two springs with spring constants k_1 and k_2 . The rigid limit is obtained by taking $k_1, k_2 \rightarrow \infty$. However,

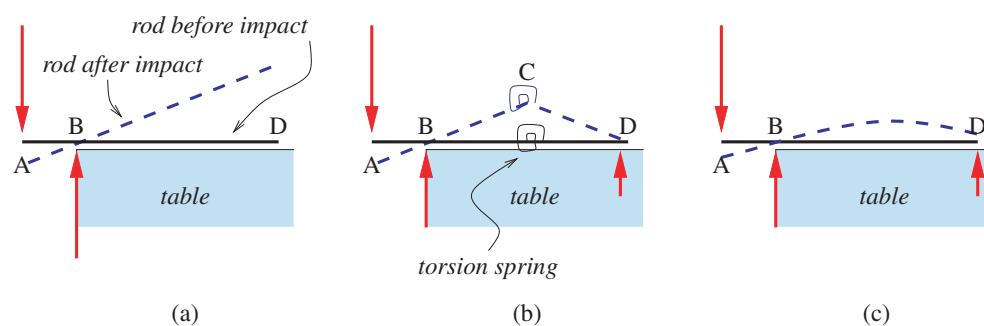


Fig. 4.5 Rod and table example of Chatterjee: (a) rigid bodies; (b) torsion spring approximation; (c) elastic beam.

there are actually many rigid limits depending on the ratio k_1/k_2 . If mass m_2 is in contact with mass m_3 , and m_1 collides with m_2 from the left, then several outcomes are possible. If $k_1/k_2 \ll 1$, then since the time constant for the m_2 - m_3 interaction is much shorter than the time constant for the m_1 - m_2 interaction, m_2 and m_3 will appear to be a single unit, and will leave together. On the other hand, if $k_1/k_2 \gg 1$, then the impacts will appear to be consecutive: momentum is transferred to mass m_2 , which is then transferred to mass m_3 , and only m_3 will leave with a significant velocity. Clearly, models of impact with multiple bodies and contacts should take into account these stiffness-ratio effects.

High-speed photographs of rows of initially touching clear plastic disks under polarized light on impact clearly show that throughout the impact there are usually three or four disks that are simultaneously in contact. This seems to indicate that collisions with multiple contacts cannot be replaced by sequences of two-body collisions. When “stiffness ratio” effects are included, problems with multiple contacts become quite complex. The development of extended rigid-body models that can accurately model these situations is beyond the scope of this paper.

A further issue that is related to multiple simultaneous contacts and vibrations of stiff elastic bodies is raised by Chatterjee [17]. Consider a uniform slender rod overhanging the edge of a frictionless table and then striking the overhanging edge of the rod, as illustrated in Figure 4.5. Treating the bodies simply as rigid bodies would give the picture (a), where a reaction impulse occurs at point B, while at D, the rod moves up off the table so that there is no impulse there. However, the rod is not perfectly rigid, and this can be modeled using a torsion spring at a point C (say, midway between B and D). Then since AC is considered rigid, immediately after A is struck, the vertical velocity at C is positive. There is an upward impulse on CD at C. Since CD is rigid, some calculations show that there must be an upward impulse at D to prevent D penetrating the table. The exact value of this impulse at D depends on the location of C, *but does not depend on the stiffness of the torsion spring*. Since the forces at D can only be upwards, later motion could only increase the total impulse applied at D. Thus the limit as the stiffness of the torsion spring goes to infinity does not recover the rigid-body solution.

Of course, the torsion spring model is only an approximation; a more realistic approximation is to model the rod as an elastic beam. Similar effects occur there, although the analysis has not been carried out yet, in part because of the difficulty of dealing with infinite-dimensional problems. (See the following section for more

on these issues.) Simple experiments with a rod (like a knitting needle) and carbon paper verify that this effect is real. Thus we have both $n_D^T v_D^+ > 0$ and $N_D > 0$ in the limit as $k_{spring} \rightarrow \infty$, contradicting the usual assumption about complementarity: $0 \leq n_D^T v_D^+ \perp N_D \geq 0$. Nevertheless, it should be noted that the usual complementarity conditions hold for all times and spring stiffnesses in (b) and (c). In addition, complementarity between the vertical displacement and the normal contact force remains fundamental, as noted in [17].

Can the rigid-body solution (a) be obtained as some sort of limit of (b) or (c)? The answer is yes, and again is related to stiffness ratios: if k_{table}/k_{spring} goes to zero as both stiffnesses go to infinity, the rigid body solution (a) is recovered. The reason that this is not seen experimentally in [17] is that it is much easier to build a stiff table than it is to find a stiff yet slender rod. Again we come to a question of how to deal with stiffness ratios and elastic vibrations. The problem is that as the material stiffness goes to infinity, the amplitudes of the elastic vibrations go to zero, but the amplitudes of the surface velocities do not. This means that complementarity conditions for rigid motions may not be correct even though complementarity conditions for the true surface velocities may be. This issue is more fundamental than the issue of which impact law to use, since all current impact laws must handle the inelastic case $\varepsilon = 0$ at least as a model of the compression phase of the contact. In short, there are a number of issues regarding multiple contact and multiple body impacts that need resolution, but new low-order models may be able to incorporate the effects of multiple contacts and elastic vibrations.

4.5. Elastica in Impact. The problems in handling the infinite-dimensional problems associated with elastic (and elastoviscous, elastoplastic) materials in frictional contact are substantial. Currently, the best results are either for quasi-static problems where the forces are assumed to always be in equilibrium, or for special dynamic problems with “small” coefficients of friction. Good starting points for understanding these problems are the book by Kikuchi and Oden [65] and the book by Duvaut and Lions [32]. However, the early formulations in terms of variational inequalities had to assume frictionless contact, that the normal contact force is given (Tresca friction), or some other simplifying assumption. The feedback between the friction force and the normal contact force was usually ignored in order to obtain a well-posed variational inequality. If the feedback is used to construct a fixed-point problem, then results can be obtained for coefficients of friction “small enough” by means of a contraction mapping argument. For large coefficients of friction, this approach fails. An alternative that uses a Leray–Schauder argument would be more attractive in this case, although there are a number of technical difficulties with compactness, even for quasi-static problems. Often a mollifier is used to modify the friction law to satisfy the compactness requirements, but this also makes the friction law nonlocal, which does not seem physically appropriate. The following references survey the state of the art in this area [22, 23, 30, 56, 57, 58, 62, 65].

Work on fully dynamic infinite-dimensional contact problems includes that of P. Shi [125], Bamberger and Schatzman [10], Schatzman [120], and Schatzman and Bercovier [121]. The work of P. Shi mainly concerns a one-dimensional wave equation with contact only at one end; the work of Schatzman et al. is mainly about the wave equation with a rigid obstacle, where contact can be made over the whole spatial domain. The problems in this area are characterized by a high level of technical difficulty. A full discussion of the issues in this area is beyond the scope of this article.

REFERENCES

- [1] E. ALLGOWER AND K. GEORG, *Numerical Continuation Methods: An Introduction*, Springer-Verlag, Berlin, Heidelberg, New York, 1990.
- [2] J. ANGELES AND A. KECSKEMÉTHY, EDs., *Kinematics and Dynamics of Multi-Body Systems*, Springer-Verlag, Vienna, 1995.
- [3] M. ANITESCU, *private communication*, Argonne National Laboratories, Argonne, IL, 1997.
- [4] M. ANITESCU, J. F. CREMER, AND F. A. POTRA, *On the existence of solutions to complementarity formulations of contact problems with friction*, in *Complementarity and Variational Problems: State of the Art*, SIAM, Philadelphia, 1997, pp. 12–21.
- [5] M. ANITESCU AND F. A. POTRA, *Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems*, *ASME J. Nonlinear Dynamics*, 14 (1997), pp. 231–247.
- [6] M. ANITESCU, F. A. POTRA, AND D. E. STEWART, *Time-stepping for three-dimensional rigid body dynamics*, *Comput. Meth. Appl. Mech. Engrg.*, 177 (1999), pp. 183–197.
- [7] D. BAINOV AND P. SIMEONOV, *Systems with Impulse Effect: Stability, Theory, and Applications*, Math. Appl., Chichester, New York, 1989.
- [8] D. BAINOV AND P. SIMEONOV, *Impulsive Differential Equations: Periodic Solutions and Applications*, Pitman Monogr. Surveys Pure Appl. Math. 66, Longman, Harlow, 1993.
- [9] P. BALLARD, *A counter-example to uniqueness in quasi-static elastic contact problems with small friction*, *Internat. J. Engrg. Sci.*, 37 (1999), pp. 163–178.
- [10] A. BAMBERGER AND M. SCHATZMAN, *New results on the vibrating string with a continuous obstacle*, *SIAM J. Math. Anal.*, 14 (1983), pp. 560–595.
- [11] D. BARAFF, *Fast contact force computation for non-penetrating rigid bodies*, in *Proceedings, SIGGRAPH '94*, ACM, New York, 1994, pp. 23–34.
- [12] H. BÉGHIN, *Sur certain problemes de frottement*, *Nouv. Ann. Math. Sér. 2*, 5 (1923/1924), pp. 305–312.
- [13] M. BENATI AND A. MORRO, *Equations for multibody dynamics*, *Atti Accad. Peloritana Pericolanti Cl. Sci. Fis. Mat. Natur.*, 68 (1990), pp. 291–308.
- [14] B. BROGLIATO, *Nonsmooth Impact Mechanics: Models, Dynamics and Control*, Lecture Notes in Control and Inform. Sci. 220, Springer, New York, 1996.
- [15] S. L. CAMPBELL AND B. LEIMKUEHLER, *Differentiation of constraints in differential-algebraic equations*, *Mech. Structures Mech.*, 19 (1991), pp. 19–39.
- [16] C. CASTAING AND M. D. P. MONTEIRO MARQUES, *BV periodic solutions of an evolution problem associated with continuous moving convex sets*, *Set-Valued Anal.*, 3 (1995), pp. 381–399.
- [17] A. CHATTERJEE, *On the realism of complementarity conditions in rigid-body collisions*, *Nonlinear Dynamics*, 20 (1999), pp. 159–168.
- [18] A. CHATTERJEE AND A. L. RUINA, *A new algebraic rigid-body collision law based on impulse space considerations*, *ASME J. Appl. Mech.*, 65 (1998), pp. 939–950.
- [19] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, Classics Appl. Math. 5, SIAM, Philadelphia, 1990.
- [20] P. W. CLEARY, *Modelling granular flows with complex boundary geometries*, in *Computational Techniques and Applications: CTAC93*, D. Stewart, D. Singleton, and H. Gardner, eds., World Scientific, Singapore, 1994, pp. 148–156.
- [21] P. W. CLEARY AND N. G. BARTON, *Use of particle methods to design mineral processing equipment*, E. Kreuzer and O. Mahrenholtz, eds., *Z. Angew. Math. Mech.*, 76 (1996), pp. 253–256.
- [22] M. COCU, E. PRATT, AND M. RAOUS, *Existence d'une solution du problème quasi-statique de contact unilatéral avec frottement non local*, *C. R. Acad. Sci. Paris Sér. I Math.*, 320 (1995), pp. 1413–1417.
- [23] M. COCU, E. PRATT, AND M. RAOUS, *Formulation and approximation of quasistatic frictional contact*, *Internat. J. Engrg. Sci.*, 34 (1996), pp. 783–798.
- [24] E. A. CODDINGTON AND N. LEVINSON, *Theory of Ordinary Differential Equations*, Tata-McGraw Hill, New Delhi, 1972.
- [25] R. W. COTTLE, J.-S. PANG, AND R. E. STONE, *The Linear Complementarity Problem*, Comput. Sci. Comput., Academic Press, Boston, San Diego, New York, 1992.
- [26] C. A. DE COULOMB, *Théorie des machines simples, en ayant égard au frottement de leurs parties, et la roideur des cordages. Pièce qui a reporté le Prix double de l'Académie des Sciences pour l'année 1781*, *Mémoires des Savans Etrangers*, X (1785), pp. 163–332. Reprinted by Bachelier, Paris, 1809.

- [27] E. DELASSUS, *Considérations sur le frottement de glissement*, Nouv. Ann. Math. Sér. 4, 20 (1920), pp. 485–496.
- [28] E. DELASSUS, *Sur les lois du frottement de glissement*, Bull. Soc. Math. France, 51 (1923), pp. 22–33.
- [29] F. DEMENGEL AND R. TEMAM, *Convex functions of a measure and applications*, Indiana Univ. Math. J., 33 (1984), pp. 673–709.
- [30] L. DEMKOWICZ AND J. T. ODEN, *On some existence and uniqueness results in contact problems with nonlocal friction*, Nonlinear Anal., 6 (1982), pp. 1075–1093.
- [31] A. DONTCHEV AND F. LEMPIO, *Difference methods for differential inclusions: A survey*, SIAM Rev., 34 (1992), pp. 263–294.
- [32] G. DUVAUT AND J.-L. LIONS, *Inequalities in Mechanics and Physics*, Grundlehren Math. Wiss. 219, Springer-Verlag, Berlin, Heidelberg, New York, 1976.
- [33] E. EICH, C. FÜHRER, AND J. YEN, *On the error control for multistep methods applied to ODEs with invariants and DAEs in multibody dynamics*, Mech. Structures Mach., 23 (1995), pp. 159–179.
- [34] E. EICH-SOELLNER AND C. FÜHRER, *Numerical Methods in Multibody Dynamics*, B. G. Teubner, Stuttgart, 1998.
- [35] C. M. ELLIOTT, *On the convergence of a one-step method for the numerical solution of ordinary differential inclusions*, IMA J. Numer. Anal., 5 (1985), pp. 3–21.
- [36] M. ERDMANN, *On a representation of friction in configuration space*, Internat. J. Robotics Research, 13 (1994), pp. 240–271.
- [37] M. C. FERRIS AND J.-S. PANG, *Engineering and economic applications of complementarity problems*, SIAM Rev., 39 (1997), pp. 669–713.
- [38] A. F. FILIPPOV, *Differential equations with discontinuous right-hand side*, Amer. Math. Soc. Trans., 42 (1964), pp. 199–231; original published in Mat. Sb., 5 (1960), pp. 99–127 (in Russian).
- [39] A. F. FILIPPOV, *Classical solutions of differential equations with multivalued right-hand side*, SIAM J. Control Optim., 5 (1967), pp. 609–621.
- [40] A. F. FILIPPOV, *Differential Equations with Discontinuous Right-Hand Side*, Kluwer Academic, Norwell, MA, 1988.
- [41] R. FLETCHER, *Practical Methods of Optimization*, 2nd ed., John Wiley, Chichester, New York, 1987.
- [42] C. G. FRANCHI, *A general formulation in rigid multibody dynamics*, Internat. J. Numer. Meth. Engrg., 38 (1995), pp. 1985–2016.
- [43] J. FRANKLIN, *Methods of Mathematical Economics*, Springer-Verlag, New York, Heidelberg, Berlin, 1980.
- [44] H. FRANKOWSKA, *The maximum principle for an optimal solution to a differential inclusion with end points constraints*, SIAM J. Control Optim., 25 (1987), pp. 145–157.
- [45] C. FÜHRER AND B. LEIMKÜHLER, *Numerical methods for differential-algebraic equations describing constrained mechanical motion*, in Computational Ordinary Differential Equations (London, 1989), Oxford University Press, New York, 1992, pp. 275–284.
- [46] J. GARCÍA DE JALÓN AND E. BAYO, *Kinematic and dynamic simulation of multibody systems*, Springer-Verlag, New York, 1994.
- [47] Z. GE AND J. E. MARSDEN, *Lie-Poisson Hamilton-Jacobi theory and Lie-Poisson integrators*, Phys. Lett. A, 133 (1988), pp. 134–139.
- [48] C. W. GEAR, B. LEIMKÜHLER, AND G. K. GUPTA, *Automatic integration of Euler-Lagrange equations with constraints*, J. Comput. Appl. Math., 12/13 (1985) pp. 77–90.
- [49] F. GÉNOT AND B. BROGLIATO, *New results on Painlevé paradoxes*, European J. Mech. A Solids, 18 (1999), pp. 653–677. Also available as INRIA Research Report 3366 via <ftp://ftp.inria.fr/INRIA/tech-reports/RR/RR-3366.ps.gz>.
- [50] C. GLOCKER AND F. PFEIFFER, *Multiple impacts with friction in rigid multibody systems*, ASME J. Nonlinear Dynamics, 7 (1995), pp. 471–497.
- [51] S.-P. HAN, J.-S. PANG, AND N. RANGARAJ, *Globally convergent Newton methods for non-smooth equations*, Math. Oper. Res., 17 (1992), pp. 586–607.
- [52] E. J. HAUG, D. NEGRUT, AND M. IANCU, *A state-space-based implicit integration algorithm for differential-algebraic equations of multibody dynamics*, Mech. Structures Mach., 25 (1997), pp. 311–334.
- [53] J.-B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex Analysis and Minimization Algorithms I and II*, Grundlehren Math. Wiss., Springer-Verlag, Berlin, Heidelberg, New York, 1993.
- [54] R. HORST AND P. PARDALOS, EDS., *Handbook of Global Optimization*, Nonconvex Optim. Appl., Kluwer Academic Publishers, Dordrecht, the Netherlands, 1995.

- [55] G. ISAC, *Complementarity problems*, Lecture Notes in Math. 1528, Springer-Verlag, Berlin, 1992.
- [56] J. JARUŠEK, *Contact problems with bounded friction: Coercive case*, Czechoslovak Math. J., 33 (1983), pp. 237–261.
- [57] J. JARUŠEK AND C. ECK, *Dynamic contact problems with friction in linear viscoelasticity*, C. R. Acad. Sci. Paris Sér. I Math., 322 (1996), pp. 497–502.
- [58] J. JARUŠEK AND C. ECK, *Dynamic contact problems with small Coulomb friction for viscoelastic bodies. Existence of solutions*, Math. Models Methods Appl. Sci., 9 (1999), pp. 11–34.
- [59] C. KANE AND M. ORTIZ, *Finite element analysis of non-smooth contact*, Comput. Meth. Appl. Mech. Engrg., 180 (1999), pp. 1–26.
- [60] A. E. KASTNER-MARESCH, *Implicit Runge–Kutta methods for differential inclusions*, Numer. Funct. Anal. Optim., 11 (1990), pp. 937–958.
- [61] A. E. KASTNER-MARESCH, *The implicit midpoint rule applied to discontinuous differential equations*, Computing, 49 (1992), pp. 45–62.
- [62] Y. KATO, *Signorini's problem with friction in linear elasticity*, Japan J. Appl. Math., 4 (1987), pp. 237–268.
- [63] J. B. KELLER, *Impact with friction*, ASME J. Appl. Mech., 53 (1986), pp. 1–4.
- [64] J. L. KELLEY AND T. P. SRINIVASAN, *Measure and Integral, Volume 1*, Grad. Texts in Math. 116, Springer-Verlag, New York, 1988.
- [65] N. KIKUCHI AND J. T. ODEN, *Contact Problems in Elasticity: A Study of Variational Inequalities and Finite Element Methods*, SIAM Stud. Appl. Math. 8, SIAM, Philadelphia, 1988.
- [66] A. KLARBRING, *Mathematical programming in contact problems*, in Computational Methods in Contact Mechanics, M. H. Aliabadi and C. A. Brebbia, eds., Internat. Ser. Comput. Engrg., Computational Mechanics Publications, London, 1993, pp. 233–263.
- [67] A. KLARBRING, *Steady sliding and linear complementarity*, in Complementarity and Variational Problems, M. C. Ferris and J.-S. Pang, eds., SIAM, Philadelphia, 1997, pp. 132–147.
- [68] F. KLEIN, *Zu Painlevés Kritik der Coulombschen Reibungsgesetze*, Zeit. Math. Physik, 58 (1909), pp. 186–191.
- [69] M. KUNZE AND M. D. P. MONTEIRO MARQUES, *Yosida-Moreau regularization of sweeping processes with unbounded variation*, J. Differential Equations, 130 (1996), pp. 292–306.
- [70] M. KUNZE AND M. D. P. MONTEIRO MARQUES, *Existence of solutions for degenerate sweeping processes*, J. Convex Anal., 4 (1997), pp. 165–176.
- [71] V. LAKSHMIKANTHAM, D. BAINOV, AND P. SIMEONOV, *Theory of Impulsive Differential Equations*, World Scientific, Singapore, 1989.
- [72] S. LANG, *Real and Functional Analysis*, 2nd ed., Grad. Texts in Math. 142, Springer-Verlag, Berlin, Heidelberg, New York, 1993.
- [73] B. LEIMKÜHLER AND S. REICH, *Symplectic integration of constrained Hamiltonian systems*, Math. Comp., 63 (1994), pp. 589–605.
- [74] F. LEMPPIO, *Modified Euler methods for differential inclusions*, in Set-Valued Analysis and Differential Inclusions, A. B. Kurzhansky and V. M. Veliov, eds., Birkhäuser, Boston, Basel, Berlin, 1993.
- [75] F. LEMPPIO AND V. VELIOV, *Discrete approximations of differential inclusions*, Bayreuth. Math. Schr., 54 (1998), pp. 149–232.
- [76] P. LÖTSTEDT, *Coulomb friction in two-dimensional rigid-body systems*, Z. Angew. Math. Mech., 61 (1981), pp. 605–615.
- [77] P. LÖTSTEDT, *Mechanical systems of rigid bodies subject to unilateral constraints*, SIAM J. Appl. Math., 42 (1982), pp. 281–296.
- [78] P. LÖTSTEDT, *Time-dependent contact problems in rigid-body mechanics*, Math. Prog. Study, 17 (1982), pp. 103–110.
- [79] P. LÖTSTEDT, *Numerical simulation of time-dependent contact and friction problems in rigid body mechanics*, SIAM J. Sci. Statist. Comput., 5 (1984), pp. 370–393.
- [80] M. MABROUK, *Liaisons unilatérales et chocs élastiques quelconques: Un résultat d'existence*, C. R. Acad. Sci. Paris Sér. I Math., 326 (1998), pp. 1353–1357.
- [81] M. MABROUK, *A unified variational model for the dynamics of perfect unilateral constraints*, European J. Mech. A Solids, 17 (1998), pp. 819–842.
- [82] M. T. MASON AND Y. WANG, *On the inconsistency of rigid-body frictional planar mechanics*, in Proceedings, IEEE International Conference on Robotics Automation, IEEE, Piscataway, NJ, 1988, pp. 524–528.
- [83] M. D. P. MONTEIRO MARQUES, *Chocs inélastiques standards: un résultat d'existence*, Sémin. Anal. Convexe, 15 (1985), pp. 1–32.

- [84] M. D. P. MONTEIRO MARQUES, *Differential Inclusions in Nonsmooth Mechanical Problems: Shocks and Dry Friction*, Prog. Nonlinear Differential Equations Appl. 9, Birkhäuser-Verlag, Basel, Boston, Berlin, 1993.
- [85] J.-J. MOREAU, *Sur les mesures différentielles de fonctions vectorielles et certain problèmes d'évolution*, C. R. Acad. Sci. Paris Sér. A, 282 (1976), pp. 837–840.
- [86] J.-J. MOREAU, *Evolution problem associated with a moving convex set in a Hilbert space*, J. Differential Equations, 26 (1977), pp. 347–374.
- [87] J.-J. MOREAU, *Application of convex analysis to some problems of dry friction*, in Trends in Applications of Pure Mathematics to Mechanics, Vol. II (Second Sympos., Kozubnik, 1977), Monogr. Stud. Math. 5, Pitman, Boston, 1979, pp. 263–280.
- [88] J.-J. MOREAU, *Standard inelastic shocks and the dynamics of unilateral constraints*, in Unilateral Problems in Structural Mechanics, G. del Piero and F. Maceri, eds., C.I.S.M. Courses and Lectures 288, Springer-Verlag, Vienna, New York, 1985, pp. 173–221.
- [89] J.-J. MOREAU, *Une formulation du contact a frottement sec: application au calcul numérique*, C. R. Acad. Sci. Paris Sér. II, 302 (1986), pp. 799–801.
- [90] J.-J. MOREAU, *Bounded variation in time*, in Topics in Nonsmooth Mechanics, Birkhäuser, Basel, Boston, 1988, pp. 1–74.
- [91] J.-J. MOREAU, *Unilateral contact and dry friction in finite freedom dynamics*, in Nonsmooth Mechanics and Applications, J.-J. Moreau and P. D. Panagiotopoulos, eds., International Centre for Mechanical Sciences, Courses and Lectures 302, Springer-Verlag, Vienna, New York, 1988, pp. 1–82.
- [92] J.-J. MOREAU, *Numerical experiments in granular dynamics: Vibration-induced size segregation*, in Contact Mechanics, M. Raous, M. Jean, and J.-J. Moreau, eds., Plenum Press, Proceedings of the 2nd Contact Mechanics International Symposium, 1994, Carry-Le-Rouet, France, New York, 1995, pp. 347–158.
- [93] I. P. NATANSON, *Theory of Functions of a Real Variable*, rev. ed., Frederick Ungar, New York, 1961; original translated from the Russian by L. F. Boron and annotated by E. Hewitt.
- [94] I. NEWTON, *Philosophiæ Naturalis Principia Mathematica*, Reg. Soc. Phrases, London, 1686.
- [95] H.-D. NIEPAGE, *Inverse stability and convergence of difference approximations for boundary value problems for differential inclusions*, Numer. Funct. Anal. Optim., 9 (1987), pp. 761–778.
- [96] H.-D. NIEPAGE AND W. WENDT, *On the discrete convergence of multistep methods for differential inclusions*, Numer. Funct. Anal. Optim., 9 (1987), pp. 591–617.
- [97] P. PAINLEVÉ, *Sur le lois du frottement de glissement*, C. R. Acad. Sci. Paris, 121 (1895), pp. 112–115. Articles under the same title appeared in C. R. Acad. Sci. Paris, 141 (1905), pp. 401–405 and 546–552.
- [98] J.-S. PANG, *A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems*, Math. Programming, 51 (1991), pp. 101–131.
- [99] J.-S. PANG, *Complementarity problems*, in Handbook of Global Optimization, Nonconvex Optim. Appl. 2, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1995, pp. 271–338, in [54].
- [100] J.-S. PANG AND L. Q. QI, *Nonsmooth equations: Motivation and algorithms*, SIAM J. Optim., 3 (1993), pp. 443–465.
- [101] J.-S. PANG AND D. E. STEWART, *A unified approach to frictional contact problems*, Internat. J. Engrg. Sci., 37 (1999), pp. 1747–1768.
- [102] J.-S. PANG AND J. C. TRINKLE, *Complementarity formulations and existence of solutions of dynamic multi-rigid-body contact problems with Coulomb friction*, Math. Programming, 73 (1996), pp. 199–226.
- [103] J.-S. PANG, J. C. TRINKLE, AND G. LO, *A complementarity approach to a quasistatic multi-rigid-body contact problem*, Comput. Optim. Appl., 5 (1996), pp. 139–154.
- [104] J.-S. PANG AND J. A. TZITZOURIS, *private communication*, manuscript in preparation.
- [105] L. PAOLI AND M. SCHATZMAN, *Mouvement à un nombre fini de degrés de liberté avec contraintes unilatérales: Cas avec perte d'énergie*, Math. Model. Numer. Anal., 27 (1993), pp. 673–717.
- [106] L. PAOLI AND M. SCHATZMAN, *Vibrations avec contraintes unilatérales et perte d'énergie aux impacts, en dimension finie*, C. R. Acad. Sci. Paris Sér. I, 317 (1993), pp. 97–101.
- [107] M. F. O. S. PEREIRA AND J. A. C. AMBRÓSIO, EDS., *Computational dynamics in multibody systems*, Papers from the NATO Advanced Study Institute on Computer Aided Analysis of Rigid and Flexible Mechanical Systems, Portugal, 1994, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1995.

- [108] J. B. PERRY, P. HENNING, J. CREMER, AND G. VANĚČEK, JR., *Contact analysis in a virtual environment*, in Proceedings of the First Workshop on Simulation and Interaction in Virtual Environments (SIVE 95), University of Iowa, Iowa City, IA, 1995.
- [109] F. PFEIFFER AND C. GLOCKER, *Multibody Dynamics with Unilateral Contacts*, Wiley Ser. Nonlinear Sci., John Wiley, New York, 1996.
- [110] S. D. POISSON, *Traité de Mécanique*, 2nd ed., Bachelier, Paris, 1833.
- [111] F. A. POTRA, *Runge-Kutta integrators for multibody dynamics*, Mech. Structures Mach., 23 (1995), pp. 181–197.
- [112] L. Q. QI AND J. SUN, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993), pp. 353–367.
- [113] S. REICH, *Symplectic integrators for systems of rigid bodies*, in Integration Algorithms and Classical Mechanics, AMS, Providence, RI, 1996, pp. 181–191.
- [114] E. J. ROUTH, *Dynamics of a System of Rigid Bodies*, MacMillan, London, 1860.
- [115] A. L. RUINA, *Constitutive relations for frictional slip*, in Mechanics of Geomaterials: Rocks, Concrete, Soils, Z. P. Bažant, ed., John Wiley, Chichester, New York, Brisbane, Toronto, Singapore, 1985, pp. 169–188.
- [116] J. M. SANZ-SERNA, *Symplectic integrators for Hamiltonian problems: An overview*, in Acta Numerica 1992, A. Iserles, ed., Cambridge University Press, Cambridge, UK, 1992, pp. 273–279.
- [117] M. SCHATZMAN, *Sue une classe de problèmes hyperboliques non linéaires*, C. R. Acad. Sci. Paris Sér. A-B, 277 (1973), pp. A671–A674.
- [118] M. SCHATZMAN, *Le système différentiel $(d^2u/dt^2) + \partial\varphi(u) \ni f$ avec conditions initiales*, C. R. Acad. Sci. Paris Sér. A-B, 284 (1977), pp. A603–A606.
- [119] M. SCHATZMAN, *A class of nonlinear differential equations of second order in time*, Nonlinear Anal., 2 (1978), pp. 355–373.
- [120] M. SCHATZMAN, *A hyperbolic problem of second order with unilateral constraints: The vibrating string with a concave obstacle*, J. Math. Anal. Appl., 73 (1980), pp. 138–191.
- [121] M. SCHATZMAN AND M. BERCOVIER, *Numerical approximation of a wave equation with unilateral constraints*, Math. Comp., 53 (1989), pp. 55–79.
- [122] W. SCHIEHLEN, *Multibody system dynamics: Roots and perspectives*, Multibody Syst. Dyn., 1, Portugal, (1997), pp. 149–188.
- [123] M. F. O. SEABRA PEREIRA AND J. A. C. AMBRÓSIO, EDS., *Computational methods for rigid and flexible mechanical systems*, Papers from the NATO Advanced Study Institute on Computer Aided Analysis of Rigid and Flexible Mechanical Systems, Tróia, Portugal, 1993, Nonlinear Dynam. 9, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1996.
- [124] A. A. SHABANA, *Flexible multibody dynamics: Review of past and recent developments*, Multibody Syst. Dyn., 1 (1997), pp. 189–222.
- [125] P. SHI, *The restitution coefficient of a linear elastic rod*, Math. Comput. Modelling, 28 (1998), pp. 427–435.
- [126] D. E. STEWART, *A high accuracy method for solving ODEs with discontinuous right-hand side*, Numer. Math., 58 (1990), pp. 299–328.
- [127] D. E. STEWART, *Numerical methods for friction problems with multiple contacts*, J. Austral. Math. Soc. Ser. B, 37 (1996), pp. 288–308.
- [128] D. E. STEWART, *Existence of solutions to rigid body dynamics and the paradoxes of Painlevé*, C. R. Acad. Sci. Paris Sér. I Math., 325 (1997), pp. 689–693.
- [129] D. E. STEWART, *Convergence of a time-stepping scheme for rigid body dynamics and resolution of Painlevé's problems*, Arch. Rational Mech. Anal., 145 (1998), pp. 215–260.
- [130] D. E. STEWART, *Reformulations of measure differential inclusions and their closed graph property*, J. Differential Equations, submitted; also available online from <http://www.math.uiowa.edu/faculty/comp-math-1998.htm>.
- [131] D. E. STEWART AND J. C. TRINKLE, *An implicit time-stepping scheme for rigid body dynamics with inelastic collisions and Coulomb friction*, Internat. J. Numer. Methods Engrg., 39 (1996), pp. 2673–2691.
- [132] D. E. STEWART AND J. C. TRINKLE, *An implicit time-stepping scheme for rigid body dynamics with inelastic collisions and Coulomb friction*, Internat. J. Numer. Methods Engrg., 39 (1996), pp. 2673–2691.
- [133] D. STOIANOVICI AND Y. HURMUZLU, *A critical study of the applicability of rigid-body collision theory*, ASME J. Appl. Mech., 63 (1996), pp. 307–316.
- [134] K. R. STROMBERG, *An Introduction to Classical Real Analysis*, Wadsworth, Belmont, CA, 1981.

- [135] W. J. STRONGE, *Rigid body collisions with friction*, Proc. Roy. Soc. London Ser. A, 431 (1990), pp. 168–181.
- [136] K. TAUBERT, *Differenzverfahren für gewöhnliche Anfangswertaufgaben mit unstetiger rechte Seite*, in Numerische Behandlung Nichtlineare Integrodifferential- und Differentialgleichungen, R. Ansorge and W. Törnig, eds., Lecture Notes in Mathematics 395, Springer-Verlag, Berlin, New York, 1974, pp. 137–148.
- [137] K. TAUBERT, *Differenzverfahren für Schwingungen mit trockener und zäher Reibung und für Regelungssysteme*, Numer. Math., 26 (1976), pp. 379–395.
- [138] K. TAUBERT, *Converging multistep methods for initial value problems involving multivalued maps*, Computing, 27 (1981), pp. 123–136.
- [139] J. C. TRINKLE, J.-S. PANG, S. SUDARSKY, AND G. LO, *On dynamic multi-rigid-body contact problems with Coulomb friction*, Z. Angew. Math. Mech., 77 (1997), pp. 267–279.
- [140] L. T. WATSON, S. C. BILLUPS, AND A. P. MORGAN, *Algorithm 652: HOMPACT: A suite of codes for globally convergent homotopy algorithms*, ACM Trans. Math. Software, 13 (1987), pp. 281–310.
- [141] J. M. WENDLANDT AND J. E. MARSDEN, *Mechanical integrators derived from a discrete variational principle*, Phys. D, 106 (1997), pp. 223–246.
- [142] J. YEN, *Constrained equations of motion in multibody dynamics as ODEs on manifolds*, SIAM J. Numer. Anal., 30 (1993), pp. 553–568.
- [143] J. YEN AND L. R. PETZOLD, *An efficient Newton-type iteration for the numerical solution of highly oscillatory constrained multibody dynamic systems*, SIAM J. Sci. Comput., 19 (1998), pp. 1513–1534.