Differential-Algebraic Equations and Beyond:

From Smooth to Nonsmooth Constrained Dynamical Systems

Jan Kleinert Hochschule Bonn-Rhein-Sieg Grantham-Allee 20, D-53757 Sankt Augustin jan.kleinert@h-brs.de

and

German Aerospace Center Simulation and Software Technology Linder Höhe, D-51147 Cologne jan.kleinert@dlr.de

Bernd Simeon Felix-Klein-Zentrum, TU Kaiserslautern D-67663 Kaiserslautern, Germany simeon@mathematik.uni-kl.de

November 20, 2018

Abstract: The present article presents a summarizing view at differential-algebraic equations (DAEs) and analyzes how new application fields and corresponding mathematical models lead to innovations both in theory and in numerical analysis for this problem class. Recent numerical methods for nonsmooth dynamical systems subject to unilateral contact and friction illustrate the topicality of this development.

Keywords: Differential-algebraic equations; Historical remarks; Index notions; BDF methods; Runge-Kutta methods; Partial differential-algebraic equations; Constrained mechanical system; Electric circuit analysis; Differential Variational Inequality; Unilateral Constraints; Measure differential Inclusion; Augmented Lagrangian; Complementarity Problem; Moreau–Jean time stepping

Mathematics Subject Classification (2010): 34A09, 65L80, 65M20, 01-02, 34-03

1 Introduction

Differential-Algebraic Equations (DAEs) are a prominent example for application-driven research that leads to new concepts and methodology in mathematics. Until the early 1980s of the last century, this topic was widely unknown but the introduction of powerful simulation software in electrical and mechanical engineering created a strong demand for the analysis and numerical solution of dynamical systems with constraints. Mathematicians all over the world then started to work on DAEs, which resulted in an avalanche of research over the following decades. Meanwhile, most issues have been resolved and sophisticated numerical methods have been found, but despite this maturity, the field is constantly expanding due to the ongoing trend to model complex phenomena in science and engineering by means of differential equations and additional constraints that stem from network structures, boundary and coupling conditions or physical conservation properties.

In this paper, we strive for a survey of this development and even more also discuss recent extensions towards dynamical systems that are nonsmooth. The latter topic is quite timely and arises, e.g., in the modeling of granular material. Clearly, our approach is exemplary in nature and does not aim at completeness. There is, moreover, a strong bias on mechanical systems. Those readers who would like to know more about the topic of DAEs and the rich oeuvre that has accumulated over the years are referred to the monographs of Brenan, Campbell & Petzold [BCP96], Griepentrog & März [GM86], Hairer & Wanner [HW96], Kunkel & Mehrmann [KM06], Lamour, März & Tischendorf [LMT13], and to the survey of Rabier & Rheinboldt [RR02]. Refer to the textbooks by Acary and Brogliato [AB08, Bro16] and references therein for a more elaborate literature survey on nonsmooth dynamical systems.

The paper is organized as follows. Section 2 outlines the emergence of DAEs and shows how it was influenced by applications from electrical and mechanical engineering problems. Section 3 then highlights some major results and numerical methods that were developed in the DAE context. Some extensions to classical DAEs for partial and stochastic differential equations are discussed in Section 4. Section 5 discusses nonsmooth dynamical systems. An overview of the most important theoretical concepts is given in Section 5.1, the equations of motion for a unilaterally constrained mechanical system are motivated in Section 5.2 and finally, Section 5.3 provides some insights into numerical methods and recent developments in the field

of nonsmooth dynamical systems.

2 Differential-algebraic equations

In this section, we describe how the DAEs became a hot topic in the 1980s and 1990s and then go further back in time to the works of Kirchhoff [Kir47] and Lagrange [Lag88] who introduced differential equations with constraints in order to model electric circuits and mechanical systems.

2.1 How the topic of DAEs emerged

In the beginning of the 1980s, the term 'DAE' was widely unknown in mathematics. But this changed rapidly, due to an increasing demand in several engineering fields but also due to the pioneering work of Bill Gear. The first occurrence of the term *Differential-Algebraic Equation* can be found in the title of Gear's paper Simultaneous numerical solution of differential-algebraic equations [Gea71b] from 1971, and in the same year his famous book Numerical Initial Value Problems in Ordinary Differential Equations [Gea71a] appeared where he already considers examples from electric circuit analysis in the form

$$\boldsymbol{E}\,\dot{\boldsymbol{x}} = \boldsymbol{\phi}(\boldsymbol{x},t) \tag{1}$$

with possibly singular capacitance matrix $\boldsymbol{E} \in \mathbb{R}^{n_x \times n_x}$ and right hand side function $\boldsymbol{\phi}$. For a regular matrix \boldsymbol{E} , (1) can be easily converted into a system of ordinary differential equations (ODEs). Otherwise, however, a DAE arises that calls for new approaches, both in theory and in numerical analysis.

Two application fields, namely electric circuit analysis and constrained mechanical systems, have been the major driving forces for the development of DAEs. Below, this statement will be made more explicit by looking at the corresponding modeling concepts. Bill Gear had the farsightedness to perceive very early the importance of these modeling approaches for today's simulation software. During an Oberwolfach workshop in 1981, he suggested to study the *mathematical pendulum in Cartesian coordinates*

$$\ddot{q}_1 = -2q_1\lambda, \tag{2a}$$

$$\ddot{q}_2 = -\gamma - 2q_2\lambda, \tag{2b}$$

$$0 = q_1^2 + q_2^2 - 1 \tag{2c}$$



Figure 1: The mathematical pendulum.

that describes the motion of a mass point with coordinates (q_1, q_2) in the plane subject to a constraint. The constraint models the massless rod of length 1 that connects the mass point to a pivot placed in the origin of the coordinate system, Fig. 1. The motion of the mass point is then determined by the gravity (parameter γ) and by the constraint forces that are expressed in terms of the unknown Lagrange multiplier λ .

The DAE (2) is an example for the Lagrange equations of the first kind that we will discuss below. By introducing velocity variables, it can be easily converted to a system of first order that fits into the class of linear-implicit DAEs (1). The most general form of a DAE is a *fully implicit system*

$$\boldsymbol{F}(\dot{\boldsymbol{x}}, \boldsymbol{x}, t) = \boldsymbol{0} \tag{3}$$

with state variables $\boldsymbol{x}(t) \in \mathbb{R}^{n_x}$ and a nonlinear, vector-valued function \boldsymbol{F} of corresponding dimension. Clearly, if the $n_x \times n_x$ Jacobian $\partial \boldsymbol{F}/\partial \dot{\boldsymbol{x}}$ is invertible, then by the implicit function theorem, it is theoretically possible to transform (3), at least locally, to an explicit system of ODEs. If $\partial \boldsymbol{F}/\partial \dot{\boldsymbol{x}}$ is singular, however, (3) constitutes a DAE.

Linda Petzold, a student of Bill Gear, continued and extended his pioneering work in various directions. In particular, the development of the DASSL code (the 'Differential-Algebraic System SoLver') that she had started in the early 1980s [Pet82, BCP96] set a corner stone that still persists today. DASSL is based on the Backward Differentiation Formulas (BDF), which are also popular for solving systems of stiff ODEs. The extension of the BDF methods to implicit systems (3) is intriguingly simple. One replaces

```
SUBROUTINE DDASSL (RES,NEQ,T,Y,YPRIME,TOUT,INFO,RTOL,ATOL,
+ IDID,RWORK,LRW,IWORK,LIW,RPAR,IPAR,JAC)
C***BEGIN PROLOGUE DDASSL
C***PURPOSE This code solves a system of differential/algebraic
C equations of the form G(T,Y,YPRIME) = 0.
```

Figure 2: Calling sequence of the DASSL code [Pet82, BCP96] that has had an enormous impact on the subject of DAEs and that is still in wide use today. The original code is written in FORTRAN77 in double precision. A recent implementation in C is part of the SUNDIALS suite of codes [HBG⁺05].

the differential operator d/dt in (3) by the difference operator

$$\varrho \boldsymbol{x}_{n+k} := \sum_{i=0}^{k} \alpha_i \boldsymbol{x}_{n+i} = \tau \dot{\boldsymbol{x}}(t_{n+k}) + \mathcal{O}(\tau^{k+1})$$
(4)

where \boldsymbol{x}_{n+i} stands for the discrete approximation of $\boldsymbol{x}(t_{n+i})$ with stepsize τ and where the α_i , $i = 0, \ldots, k$, denote the method coefficients. Using the finite difference approximation $\rho \boldsymbol{x}_{n+k}/\tau$ of the time derivative, the numerical solution of the DAE (3) then boils down to solving the nonlinear system

$$\boldsymbol{F}\left(\frac{\varrho\boldsymbol{x}_{n+k}}{\tau},\boldsymbol{x}_{n+k},t_{n+k}\right) = \boldsymbol{0}$$
(5)

for x_{n+k} in each time step. This is exactly the underlying idea of the DASSL code, see Fig. 2.

Soon after the first release of the DASSL code, it became very popular among engineers and mathematicians. For quite some problems, however, the code would fail, which in turn triggered new research in numerical analysis in order to understand such phenomena. As it turned out, the notion of an *index* of the DAE (3) was the key to obtain further insight.

Gear [Gea88, Gea90] introduced what we call today the *differentiation* index. This non-negative integer k is defined by

k = 0: If $\partial F / \partial \dot{x}$ is non-singular, the index is 0.

k > 0: Otherwise, consider the system of equations

$$F(\dot{\boldsymbol{x}}, \boldsymbol{x}, t) = \boldsymbol{0},$$

$$\frac{\mathrm{d}}{\mathrm{d}t}F(\dot{\boldsymbol{x}}, \boldsymbol{x}, t) = \frac{\partial}{\partial \dot{\boldsymbol{x}}}F(\dot{\boldsymbol{x}}, \boldsymbol{x}, t) \boldsymbol{x}^{(2)} + \dots = \boldsymbol{0},$$

$$\vdots$$
(6)

$$\frac{\mathrm{d}^s}{\mathrm{d}t^s} \boldsymbol{F}(\dot{\boldsymbol{x}}, \boldsymbol{x}, t) = \frac{\partial}{\partial \dot{\boldsymbol{x}}} \boldsymbol{F}(\dot{\boldsymbol{x}}, \boldsymbol{x}, t) \, \boldsymbol{x}^{(s+1)} + \dots = \boldsymbol{0}$$

as a system in the separate dependent variables $\dot{\boldsymbol{x}}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(s+1)}$, with \boldsymbol{x} and t as independent variables. Then the index k is the smallest s for which it is possible, using algebraic manipulations only, to extract an ordinary differential equation $\dot{\boldsymbol{x}} = \boldsymbol{\psi}(\boldsymbol{x}, t)$ (the underlying ODE) from (6).

Meanwhile other notions of an index have emerged, but despite its ambiguity with respect to the algebraic manipulations, the differentiation index is still the most popular and widespread tool to classify DAEs.

In the next chapter, other index concepts and their relation to the differential index will be addressed, and also more protagonists will enter the stage. This first section on the early days of DAEs closes now with a look at the application fields that set the ball rolling.

2.2 Electric circuits

In 1847, Kirchhoff first published his *circuit laws* that describe the conservation properties of electric circuits [Kir47]. These laws consist of the current law and the voltage law, which both follow from Maxwell's equations of electro-dynamics. When these laws are applied to circuits with time-dependent behavior, the corresponding equations are typically given as a linear-implicit system (1). Often, the structure even turns out to be a linear constant coefficient DAE

$$E\dot{x} + Hx = c(t) \tag{7}$$

with matrices $\boldsymbol{E}, \boldsymbol{H} \in \mathbb{R}^{n_x \times n_x}$ and a time-dependent source term $\boldsymbol{c}(t) \in \mathbb{R}^{n_x}$.

An example for such an electric circuit is the *differentiator* [GHR00] shown in Fig. 3. It consists of a resistance R, an inductance L, an ideal operational amplifier $A = \infty$, and a given voltage source V(t). The $n_x = 6$



Figure 3: Differentiator circuit.

unknowns read here $\boldsymbol{x} = (V_1, V_2, V_3, I, I_L, I_V)$ with voltages V_i and currents I, I_L, I_V . From Kirchhoff's laws and the properties of the amplifier and the inductance one obtains the relations

$$I + (V_1 - V_2)/R = 0,$$

$$-(V_1 - V_2)/R + I_L = 0,$$

$$-I_L + I_V = 0,$$

$$V_1 = V(t),$$

$$V_2 = 0,$$

$$V_2 - V_3 = L \cdot \dot{I}_L.$$

This linear system has the form (7) with singular inductance matrix

$$\boldsymbol{E} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & L & 0 \end{pmatrix}.$$
 (8)

If the matrix \boldsymbol{E} was regular, it could be brought to the right hand side by formal inversion, ending up in a system of ODEs. Here however, it is singular and thus we face a DAE problem.

Weierstrass and Kronecker were in Berlin at the same time as Kirchhoff, and it is quite obvious to suppose that they knew his work ¹. Weierstrass and

¹The relation of the work of Weierstrass and Kronecker to Kirchhoff's circuit laws was pointed out to me by Volker Mehrmann when we met in September 2014 during a Summer School on DAEs in Elgersburg, Germany.

later Kronecker were thus inspired to study such singular systems and provided an elegant theory that is still fundamental today in order to understand the specific properties of DAEs.

We assume that the matrix pencil $\mu \mathbf{E} + \mathbf{H} \in \mathbb{R}^{n_x \times n_x}[\mu]$ is regular. I.e., there exists $\mu \in \mathbb{C}$ such that the matrix $\mu \mathbf{E} + \mathbf{H}$ is regular. Otherwise, the pencil is singular, and (7) has either no or infinitely many solutions. This latter case has been first studied by Kronecker [Kro90], see also [Gan59, Cam82].

If $\mu E + H$ is regular, there exist nonsingular matrices U and V such that

$$\boldsymbol{U}\boldsymbol{E}\boldsymbol{V} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{N} \end{pmatrix}, \quad \boldsymbol{U}\boldsymbol{H}\boldsymbol{V} = \begin{pmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}$$
(9)

where N is a nilpotent matrix, I the identity matrix, and C a matrix that can be assumed to be in Jordan canonical form. Note that the dimensions of these square blocks in (9) are uniquely determined. The transformation (9) is called the *Weierstrass canonical form* [Wei68]. It is a generalization of the Jordan canonical form and contains the essential structure of the linear system (7).

In the Weierstrass canonical form (9), the singularity of the DAE is represented by the nilpotent matrix N. Its degree of nilpotency, i.e., the smallest positive integer k such that $N^k = 0$, plays a key role when studying closedform solutions of the linear system (7) and is identical to the differentiation index of (7).

To construct a solution of (7), we introduce new variables and right hand side vectors

$$\boldsymbol{V}^{-1}\boldsymbol{x} =: \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix}, \quad \boldsymbol{U}\boldsymbol{c} =: \begin{pmatrix} \boldsymbol{\delta} \\ \boldsymbol{\theta} \end{pmatrix}.$$
 (10)

Premultiplying (7) by U then leads to the *decoupled system*

$$\dot{\boldsymbol{y}} + \boldsymbol{C}\boldsymbol{y} = \boldsymbol{\delta}, \qquad (11a)$$

$$N\dot{z} + z = \theta$$
. (11b)

While the solution of the ODE (11a) follows by integrating and results in an expression based on the matrix exponential $\exp(-\mathbf{C}(t-t_0))$, the equation (11b) for \mathbf{z} can be solved recursively by differentiating. More precisely, it holds

$$N\ddot{m{z}}+\dot{m{z}}=\dot{m{ heta}} agenremty N^2\ddot{m{z}}=-N\dot{m{z}}+N\dot{m{ heta}}=m{z}-m{ heta}+N\dot{m{ heta}}$$

Repeating the differentiation and multiplication by N, we can eventually exploit the nilpotency and get

$$\mathbf{0} = \mathbf{N}^{k} \mathbf{z}^{(k)} = (-1)^{k} \mathbf{z} + \sum_{\ell=0}^{k-1} (-1)^{k-1-\ell} \mathbf{N}^{\ell} \boldsymbol{\theta}^{(\ell)}.$$

This implies the explicit representation

$$\boldsymbol{z} = \sum_{\ell=0}^{k-1} (-1)^{\ell} \boldsymbol{N}^{\ell} \boldsymbol{\theta}^{(\ell)}.$$
 (12)

The above solution procedure illustrates several crucial points about DAEs and how they differ from ODEs. Remarkably, the linear constant coefficient case displays already these points, and thus the work of Weierstrass and Kronecker represents still the fundament of DAE theory today.

We highlight two crucial points:

- (i) The solution of (7) rests on k-1 differentiation steps. This requires that the derivatives of certain components of $\boldsymbol{\theta}$ exist up to $\ell = k-1$. Furthermore, some components of \boldsymbol{z} may only be continuous but not differentiable depending on the smoothness of $\boldsymbol{\theta}$.
- (ii) The components of \boldsymbol{z} are directly given in terms of the right hand side data $\boldsymbol{\theta}$ and its derivatives. Accordingly, the initial value $\boldsymbol{z}(t_0) = \boldsymbol{z}_0$ is fully determined by (12) and, in contrast to \boldsymbol{y}_0 , cannot be chosen arbitrarily. Initial values $(\boldsymbol{y}_0, \boldsymbol{z}_0)$ where \boldsymbol{z}_0 satisfies (12) are called *consistent*. The same terminology applies to the initial value \boldsymbol{x}_0 , which is consistent if, after the transformation (10), \boldsymbol{z}_0 satisfies (12).

Today, more than 150 years after the discoveries of Kirchhoff, electric circuit analysis remains one of the driving forces in the development of DAEs. The interplay of modeling and mathematical analysis is particularly important in this field, and the interested reader is referred to Günther & Feldmann [GF99] and März & Tischendorf [MT97] as basic works. The first simulation code that generated a model in differential-algebraic form was the SPICE package [NP73].



Figure 4: Sketch of a multibody system with rigid bodies and typical interconnections.

2.3 Constrained mechanical systems

Even older than the DAEs arising from Kirchhoff's laws are the Euler-Lagrange equations. They were first published in Lagrange's famous work *Mécanique analytique* [Lag88] from 1788.

Consider a mechanical system that consists of rigid bodies interacting via springs, dampers, joints, and actuators, Fig. 4. The bodies possess a certain geometry and mass while the interconnection elements are massless. Let $q(t) \in \mathbb{R}^{n_q}$ denote a vector that comprises the coordinates for position and orientation of all bodies in the system. Revolute, translational, universal, and spherical joints are examples for bondings in such a multibody system. They may constrain the motion q and hence determine its kinematics.

If constraints are present, we express the resulting conditions on q in terms of n_{λ} constraint equations

$$\mathbf{0} = \boldsymbol{g}(\boldsymbol{q}) \,. \tag{13}$$

Obviously, a meaningful model requires $n_{\lambda} < n_q$. The equations (13) that restrict the motion q are called *holonomic constraints*, and the rectangular matrix

$$oldsymbol{G}(oldsymbol{q}) := rac{\partial oldsymbol{g}(oldsymbol{q})}{\partial oldsymbol{q}} \ \in \mathbb{R}^{n_\lambda imes n_q}$$

is called the *constraint Jacobian*.

Using both the redundant position variables q and additional Lagrange multipliers λ to describe the dynamics leads to the equations of constrained mechanical motion, also called the Lagrange equations of the first kind or the Euler-Lagrange equations

$$\boldsymbol{M}(\boldsymbol{q}) \, \boldsymbol{\ddot{q}} = \boldsymbol{f}(\boldsymbol{q}, \boldsymbol{\dot{q}}, t) - \boldsymbol{G}(\boldsymbol{q})^T \boldsymbol{\lambda}, \qquad (14a)$$

$$\mathbf{0} = \mathbf{g}(\mathbf{q}), \qquad (14b)$$

where $M(q) \in \mathbb{R}^{n_q \times n_q}$ stands for the mass matrix and $f(q, \dot{q}, t) \in \mathbb{R}^{n_q}$ for the vector of applied and internal forces.

The standard example for such a constrained mechanical system are the equations (2) of the mathematical pendulum. For a long time, it was common sense that the Euler-Lagrange equations should be transformed to the *state* space form, also called the Lagrange equations of the second kind. In case of the pendulum, this means that the Cartesian coordinates can be expressed as $q_1 = \sin \alpha$, $q_2 = -\cos \alpha$ with the angle α as minimal coordinate, Fig. 1. By inserting these relations into (2), the constraints and the Lagrange multiplier cancel, and one arrives at the second order ODE

$$\ddot{\alpha} = -\gamma \sin \alpha \tag{15}$$

as state space form.

It seems obvious that a state space form such as (15) constitutes a more appropriate and easier model than the differential-algebraic system (14), or (2), respectively, in redundant coordinates. In practice, however, the state space form suffers from serious drawbacks:

The analytical complexity of the constraint equations (13) makes it in various applications impossible to obtain a set of minimal coordinates that is valid for all configurations of the multibody system. Moreover, although we know from the theorem on implicit functions that such a set exists in a neighborhood of the current configuration, it might loose its validity when the configuration changes. This holds in particular for multibody systems with so-called *closed kinematic loops*.

Even more, the modeling of subsystems like electrical and hydraulic feedback controls, which are essential for the performance of modern mechanical systems, is limited. The differential-algebraic model, on the other hand, bypasses topological analysis and offers the choice of using a set of coordinates q that possess physical significance. This reasoning in favor of the differential-algebraic model (14) became more and more widespread in the 1980s, driven by the development of sophisticated software packages, so-called *multibody formalisms*. One of the first packages that fully exploited this new way of modeling is due to Haug [Hau89].

A look at the leading software tools in the field today shows a clear picture. Some of the codes generate a differential-algebraic model whenever a constraint is present, while others try to generate a state space form as long as it is convenient. But the majority of the commercial products relies on the differential-algebraic approach as the most general way to handle complex technical applications [GB94, Sch90].

The main difference between the DAEs arising from electric circuit analysis and the DAEs that model constrained mechanical systems is the richer structure of the latter. E.g., for *conservative multibody systems*, i.e., systems where the applied forces can be written as the gradient of a potential U, the Euler-Lagrange equations (14) result from Hamilton's principle of least action

$$\int_{t_0}^{t_1} \left(T - U - \boldsymbol{g}(\boldsymbol{q})^T \boldsymbol{\lambda} \right) \mathrm{d}t \to \text{ stationary } !$$
 (16)

where the kinetic energy possesses a representation as quadratic form

$$T(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \frac{1}{2} \dot{\boldsymbol{q}}^T \boldsymbol{M}(\boldsymbol{q}) \dot{\boldsymbol{q}}$$

In the least action principle (16), we observe the fundamental Lagrange multiplier technique for coupling constraints and dynamics [Bri08]. Extensions of the multiplier technique exist in various more general settings such as dissipative systems or even inequality constraints, see Section 5.

The pendulum equations (2) are the example for a constrained mechanical system. Though they simply describe the motion of a single mass point, several key properties of the Euler-Lagrange equations can already be studied: the differential equations are of second order, the constraint equations are mostly nonlinear, and one observes a clear semi-explicit structure with differential variables q and algebraic variables λ .

The Euler-Lagrange equations are of index 3 and form the prototype for a system of higher index. Index reduction techniques are thus required and in fact, already in 1972 this issue was addressed by Baumgarte [Bau72]. He observed that in (14), the Lagrange multipliers can be eliminated by differentiating the constraints twice. The first differentiation leads to the constraints at velocity level

$$\mathbf{0} = \frac{\mathrm{d}}{\mathrm{d}t} \, \boldsymbol{g}(\boldsymbol{q}) = \boldsymbol{G}(\boldsymbol{q}) \, \dot{\boldsymbol{q}}. \tag{17}$$

A second differentiation step yields the constraints at acceleration level

$$\mathbf{0} = \frac{\mathrm{d}^2}{\mathrm{d}t^2} \, \boldsymbol{g}(\boldsymbol{q}) = \boldsymbol{G}(\boldsymbol{q}) \, \ddot{\boldsymbol{q}} + \boldsymbol{\kappa}(\boldsymbol{q}, \dot{\boldsymbol{q}}) \,, \quad \boldsymbol{\kappa}(\boldsymbol{q}, \dot{\boldsymbol{q}}) := \frac{\partial \boldsymbol{G}(\boldsymbol{q})}{\partial \boldsymbol{q}} (\dot{\boldsymbol{q}}, \dot{\boldsymbol{q}}) \,, \qquad (18)$$

where the two-form κ comprises additional derivative terms. The combination of the dynamic equation

$$\boldsymbol{M}(\boldsymbol{q}) \ddot{\boldsymbol{q}} = \boldsymbol{f}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) - \boldsymbol{G}(\boldsymbol{q})^T \boldsymbol{\lambda}$$

with (18) results in a linear system for \ddot{q} and λ with the saddle point matrix

$$\begin{pmatrix} \boldsymbol{M}(\boldsymbol{q}) & \boldsymbol{G}(\boldsymbol{q})^T \\ \boldsymbol{G}(\boldsymbol{q}) & \boldsymbol{0} \end{pmatrix} \in \mathbb{R}^{(n_q+n_\lambda)\times(n_q+n_\lambda)}.$$
 (19)

For a well-defined multibody system, this matrix is invertible in a neighborhood of the solution, and in this way, the Lagrange multiplier can be computed as a function of \boldsymbol{q} and $\dot{\boldsymbol{q}}$.

However, the well-known drift-off phenomenon requires additional stabilization measures, and Baumgarte came up with the idea to combine original and differentiated constraints as

$$\mathbf{0} = \mathbf{G}(\mathbf{q}) \ddot{\mathbf{q}} + \boldsymbol{\kappa}(\mathbf{q}, \dot{\mathbf{q}}) + 2\alpha \mathbf{G}(\mathbf{q}) \dot{\mathbf{q}} + \beta^2 \mathbf{g}(\mathbf{q})$$
(20)

with scalar parameters α and β . The free parameters α and β should be chosen in such a way that

$$\mathbf{0} = \ddot{\boldsymbol{w}} + 2\alpha \dot{\boldsymbol{w}} + \beta^2 \boldsymbol{w} \tag{21}$$

becomes an asymptotically stable equation, with $\boldsymbol{w}(t) := \boldsymbol{g}(\boldsymbol{q}(t))$.

From today's perspective, the crucial point in Baumgarte's approach is the choice of the parameters. Nevertheless, it was the very beginning of a long series of works that tried to reformulate the Euler-Lagrange equations in such a way that the index is lowered while still maintaining the information of all constraint equations. For a detailed analysis of this stabilization and related techniques we refer to Ascher et al. [ACPR95, AL97]. Another – very early – stabilization of the Euler-Lagrange equations is due to Gear, Gupta & Leimkuhler [GGL85]. This formulation represents still the state-of-the-art in multibody dynamics. It uses a formulation of the equations of motion as system of first order with velocity variables $\boldsymbol{v} = \dot{\boldsymbol{q}}$ and simultaneously enforces the constraints at velocity level (17) and the position constraints (13), where the latter are interpreted as invariants and appended by means of extra Lagrange multipliers.

In this way, one obtains an enlarged system

$$\dot{\boldsymbol{q}} = \boldsymbol{v} - \boldsymbol{G}(\boldsymbol{q})^T \boldsymbol{\mu},$$

$$\boldsymbol{M}(\boldsymbol{q}) \, \dot{\boldsymbol{v}} = \boldsymbol{f}(\boldsymbol{q}, \boldsymbol{v}, t) - \boldsymbol{G}(\boldsymbol{q})^T \boldsymbol{\lambda},$$

$$\boldsymbol{0} = \boldsymbol{G}(\boldsymbol{q}) \, \boldsymbol{v},$$

$$\boldsymbol{0} = \boldsymbol{g}(\boldsymbol{q})$$

$$(22)$$

with additional multipliers $\boldsymbol{\mu}(t) \in \mathbb{R}^{n_{\lambda}}$. A straightforward calculation shows

$$\mathbf{0} = \frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{g}(\boldsymbol{q}) = \boldsymbol{G}(\boldsymbol{q}) \dot{\boldsymbol{q}} = \boldsymbol{G}(\boldsymbol{q}) \boldsymbol{v} - \boldsymbol{G}(\boldsymbol{q}) \boldsymbol{G}^{T}(\boldsymbol{q}) \boldsymbol{\mu} = -\boldsymbol{G}(\boldsymbol{q}) \boldsymbol{G}^{T}(\boldsymbol{q}) \boldsymbol{\mu}$$

and one concludes $\boldsymbol{\mu} = \boldsymbol{0}$ since $\boldsymbol{G}(\boldsymbol{q})$ is of full rank and hence $\boldsymbol{G}(\boldsymbol{q})\boldsymbol{G}^{T}(\boldsymbol{q})$ invertible. With the additional multipliers $\boldsymbol{\mu}$ vanishing, (22) and the original equations of motion (14) coincide along any solution. Yet, the index of the *GGL formulation* (22) is 2 instead of 3. Some authors refer to (22) also as stabilized index-2 system [FL91].

The last paragraphs on stabilized formulations of the Euler-Lagrange equations demonstrate that the development of theory and numerical methods for DAEs was strongly intertwined with the mathematical models. This holds for all application fields where DAEs arise. Even more, the application fields typically provide rich structural features of the model equations that are crucial for determining the index and for the numerical treatment.

3 Major results and numerical methods

Between 1989 and 1996, both theory and numerical analysis of DAEs were booming, and many groups in mathematics and engineering started to explore this new research topic. Driven by the development of powerful simulation packages in the engineering sciences, the demand for efficient and robust integration methods was growing steadily while at the same time, it had become apparent that higher index problems require stabilization measures or appropriate reformulations.

3.1 Perturbation index and implicit Runge-Kutta methods

The groundbreaking monograph on *The Numerical Solution of Differential-Algebraic Equations by Runge-Kutta Methods* [HLR89] by Ernst Hairer, Christian Lubich and Michel Roche presented several new method classes, a new paradigm for the construction of convergence proofs, a new index concept, and the new RADAU5 code. From then on, Hairer and Lubich played a very strong role in the further development of DAEs and corresponding numerical methods.

The *perturbation index* as defined in [HLR89] sheds a different light on DAEs and adopts the idea of a well-posed mathematical model. While the differential index is based on successively differentiating the original DAE (3) until the obtained system can be solved for $\dot{\boldsymbol{x}}$, the perturbation index measures the sensitivity of the solutions to perturbations in the equation:

The system $F(\dot{x}, x, t) = 0$ has perturbation index $k \ge 1$ along a solution x(t) on $[t_0, t_1]$ if k is the smallest integer such that, for all functions \hat{x} having a defect

$$F(\hat{x}, \hat{x}, t) = \delta(t)$$

there exists on $[t_0, t_1]$ an estimate

$$\|\hat{\boldsymbol{x}}(t) - \boldsymbol{x}(t)\| \le c \Big(\|\hat{\boldsymbol{x}}(t_0) - \boldsymbol{x}(t_0)\| + \max_{t_0 \le \xi \le t} \|\boldsymbol{\delta}(\xi)\| + \ldots + \max_{t_0 \le \xi \le t} \|\boldsymbol{\delta}^{(k-1)}(\xi)\|\Big)$$

whenever the expression on the right hand side is sufficiently small. Note that the constant c depends only on F and on the length of the interval, but not on the perturbation δ . The perturbation index is k = 0 if

$$\|\hat{\boldsymbol{x}}(t) - \boldsymbol{x}(t)\| \le c \Big(\|\hat{\boldsymbol{x}}(t_0) - \boldsymbol{x}(t_0)\| + \max_{t_0 \le \xi \le t} \|\int_{t_0}^{\xi} \boldsymbol{\delta}(\tau) \,\mathrm{d}\tau\|\Big),$$

which is satisfied for ordinary differential equations.

If the perturbation index exceeds k = 1, derivatives of the perturbation show up in the estimate and indicate a certain degree of ill-posedness. E.g., if $\boldsymbol{\delta}$ contains a small high frequency term $\epsilon \sin \omega t$ with $\epsilon \ll 1$ and $\omega \gg 1$, the resulting derivatives will induce a severe amplification in the bound for $\hat{\boldsymbol{x}}(t) - \boldsymbol{x}(t)$.

Unfortunately, the differential and the perturbation index are not equivalent in general and may even differ substantially [CG95].

The definition of the perturbation index is solely a prelude in [HLR89]. As the title says, most of the monograph deals with Runge-Kutta methods, in particular implicit ones. These are extended to linear-implicit systems $E\dot{x} = \phi(x, t)$ by discretizing $\dot{x} = E^{-1}\phi(x, t)$ and assuming for a moment that the matrix E is invertible. Multiplying the resulting scheme by E, one gets the method definition

$$\boldsymbol{E}\boldsymbol{X}_{i} = \boldsymbol{E}\boldsymbol{x}_{0} + \tau \sum_{j=1}^{s} a_{ij}\boldsymbol{\phi}(\boldsymbol{X}_{j}, t_{0} + c_{j}\tau), \quad i = 1, \dots, s; \quad (23a)$$

$$\boldsymbol{x}_{1} = \left(1 - \sum_{i,j=1}^{s} b_{i} \gamma_{ij}\right) \boldsymbol{x}_{0} + \tau \sum_{i,j=1}^{s} b_{i} \gamma_{ij} \boldsymbol{X}_{j}.$$
(23b)

Here, the method coefficients are denoted by $(a_{ij})_{i,j=1}^s$ and b_1, \ldots, b_s while $(\gamma_{ij}) = (a_{ij})^{-1}$ is the inverse of the coefficient matrix, with s being the number of stages. Obviously, (23) makes sense also in the case where **E** is singular.

Using *stiffly accurate methods* for differential-algebraic equations is advantageous, which becomes evident if we consider the discretization of the semi-explicit system

$$\dot{\boldsymbol{y}} = \boldsymbol{a}(\boldsymbol{y}, \boldsymbol{z}),$$
 (24a)

$$\mathbf{0} = \mathbf{b}(\mathbf{y}, \mathbf{z}) \tag{24b}$$

with differential variables \boldsymbol{y} and algebraic variables \boldsymbol{z} . The method (23) then reads

$$\boldsymbol{Y}_{i} = \boldsymbol{y}_{0} + \tau \sum_{j=1}^{s} a_{ij} \boldsymbol{a}(\boldsymbol{Y}_{j}, \boldsymbol{Z}_{j}), \quad i = 1, \dots, s,$$
 (25a)

$$\boldsymbol{0} = \boldsymbol{b}(\boldsymbol{Y}_i, \boldsymbol{Z}_i), \qquad (25b)$$

for the internal stages and

$$\boldsymbol{y}_1 = \boldsymbol{y}_0 + \tau \sum_{j=1}^s b_j \boldsymbol{a}(\boldsymbol{Y}_j, \boldsymbol{Z}_j),$$
 (26a)

$$\boldsymbol{z}_{1} = \left(1 - \sum_{i,j=1}^{s} b_{i} \gamma_{ij}\right) \boldsymbol{z}_{0} + \tau \sum_{i,j=1}^{s} b_{i} \gamma_{ij} \boldsymbol{Z}_{j}$$
(26b)

as update for the numerical solution after one step. For stiffly accurate methods, we have $\sum_{i,j=1}^{s} b_i \gamma_{ij} = 1$ and $\boldsymbol{y}_1 = \boldsymbol{Y}_s, \boldsymbol{z}_1 = \boldsymbol{Z}_s$. The update (26) is hence superfluous and furthermore, the constraint $\boldsymbol{0} = \boldsymbol{b}(\boldsymbol{y}_1, \boldsymbol{z}_1)$ is satisfied by construction.

It is not the purpose of this article to dive further into the world of Runge-Kutta methods, but like in numerical ODEs, the rivalry between multistep methods and Runge-Kutta methods also characterizes the situation for DAEs. While Linda Petzold's DASSL code is the most prominent multistep implementation, the RADAU5 and RADAU codes [HW96, HW99] represent the one-step counter parts and have also become widespread in various applications.

The competition for the best code was a major driving force in the numerical analysis of DAEs, and from time to time those in favor of multistep methods looked also at one-step methods, e.g., in [AP91], and vice versa.

Simultaneously to the joint work with Ernst Hairer and Michel Roche, Christian Lubich investigated a different class of discretization schemes, the *half-explicit methods* [Lub89]. These methods are tailored for semi-explicit DAEs and discretize the differential equations explicitly while the constraint equations are enforced in an implicit fashion. As example, consider the Euler-Lagrange equations (14) with velocity constraint (17). The half-explicit Euler method as generic algorithm for the method class reads

$$\begin{aligned} \boldsymbol{q}_{n+1} &= \boldsymbol{q}_n + \tau \boldsymbol{v}_n, \\ \boldsymbol{M}(\boldsymbol{q}_n) \, \boldsymbol{v}_{n+1} &= \boldsymbol{M}(\boldsymbol{q}_n) \, \boldsymbol{v}_n + \tau \boldsymbol{f}(\boldsymbol{q}_n, \boldsymbol{v}_n, t_n) - \tau \boldsymbol{G}(\boldsymbol{q}_n)^T \boldsymbol{\lambda}_n, \\ \boldsymbol{0} &= \boldsymbol{G}(\boldsymbol{q}_{n+1}) \, \boldsymbol{v}_{n+1}. \end{aligned}$$
 (27)

Only a linear system of the form

$$\begin{pmatrix} \boldsymbol{M}(\boldsymbol{q}_n) & \boldsymbol{G}(\boldsymbol{q}_n)^T \\ \boldsymbol{G}(\boldsymbol{q}_{n+1}) & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_{n+1} \\ \tau \boldsymbol{\lambda}_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{M}(\boldsymbol{q}_n) \boldsymbol{v}_n + \tau \boldsymbol{f}(\boldsymbol{q}_n, \boldsymbol{v}_n, t_n) \\ \boldsymbol{0} \end{pmatrix}$$

arises here in each step. The scheme (27) forms the basis for a class of extrapolation methods [Lub89, LENP95], and also for half-explicit Runge-Kutta



Figure 5: Manifold \mathcal{M} , tangent space $\mathcal{T}_{y}\mathcal{M}$, and local parametrization.

methods as introduced in [HLR89] and then further enhanced by Brasey & Hairer [BH93] and Arnold & Murua [AM98].

These methods have in common that only information of the velocity constraints is required. As remedy for the drift off, which grows only linearly but might still be noticeable, the following projection, which is also due to Lubich [Lub89], can be applied: Let \boldsymbol{q}_{n+1} and \boldsymbol{v}_{n+1} denote the numerical solution of the system, obtained by integration from consistent values \boldsymbol{q}_n and \boldsymbol{v}_n . Then, the projection consists of the following steps:

$$solve \begin{cases} \mathbf{0} = \mathbf{M}(\tilde{\mathbf{q}}_{n+1})(\tilde{\mathbf{q}}_{n+1} - \mathbf{q}_{n+1}) + \mathbf{G}(\tilde{\mathbf{q}}_{n+1})^{T} \boldsymbol{\mu}, & \text{for } \tilde{\mathbf{q}}_{n+1}, \boldsymbol{\mu}; (28a) \\ \mathbf{0} = \mathbf{g}(\tilde{\mathbf{q}}_{n+1}) & \\ solve \begin{cases} \mathbf{0} = \mathbf{M}(\tilde{\mathbf{q}}_{n+1})(\tilde{\mathbf{v}}_{n+1} - \mathbf{v}_{n+1}) + \mathbf{G}(\tilde{\mathbf{q}}_{n+1})^{T} \boldsymbol{\eta}, \\ \mathbf{0} = \mathbf{G}(\tilde{\mathbf{q}}_{n+1})\tilde{\mathbf{v}}_{n+1} & \text{for } \tilde{\mathbf{v}}_{n+1}, \boldsymbol{\eta}. (28b) \end{cases} \end{cases}$$

A simplified Newton method can be used to solve the nonlinear system (28a) while (28b) represents a linear system for $\tilde{\boldsymbol{v}}_{n+1}$ and $\boldsymbol{\eta}$ with similar structure.

The projection can also be employed for stabilizing the equations of motion with acceleration constraint (18) where the position and velocity constraints are invariants and not preserved by the time integration, see Eich [Eic93] and von Schwerin [Sch99]. Such projection methods are particularly attractive in combination with explicit ODE integrators.

3.2 DAEs and differential geometry

Already in 1984, Werner Rheinboldt had investigated DAEs from the viewpoint of differential geometry [Rhe84]. While the approaches discussed so far are mainly inspired by differential calculus and algebraic considerations, a fundamentally different aspect comes into play by his idea of *differential* equations on manifolds. Referring to [AMR88, Arn81] for the theoretical underpinnings, we shortly illustrate this approach by considering the semi-explicit system

$$\dot{\boldsymbol{y}} = \boldsymbol{a}(\boldsymbol{y}, \boldsymbol{z}),$$
 (29a)

$$\mathbf{0} = \mathbf{b}(\mathbf{y}) \tag{29b}$$

under the assumption

$$\frac{\partial \boldsymbol{b}}{\partial \boldsymbol{y}}(\boldsymbol{y}) \cdot \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{z}}(\boldsymbol{y}, \boldsymbol{z}) \in \mathbb{R}^{n_z \times n_z} \quad \text{is invertible}$$
(30)

in a neighborhood of the solution. Clearly, (29) is of index 2 where the constraint $\mathbf{0} = \mathbf{b}(\mathbf{y})$, assuming sufficient differentiability, defines the manifold

$$\mathcal{M} := \{ \boldsymbol{y} \in \mathbb{R}^{n_y} : \boldsymbol{b}(\boldsymbol{y}) = \boldsymbol{0} \}.$$
(31)

The full rank condition (30) for the matrix product $\partial \boldsymbol{b}/\partial \boldsymbol{y} \cdot \partial \boldsymbol{a}/\partial \boldsymbol{z}$ implies that the Jacobian $\boldsymbol{B}(\boldsymbol{y}) = \partial \boldsymbol{b}(\boldsymbol{y})/\partial \boldsymbol{y} \in \mathbb{R}^{n_z \times n_y}$ possesses also full rank n_z . Hence, for fixed $\boldsymbol{y} \in \mathcal{M}$, the tangent space

$$\mathcal{T}_{\boldsymbol{y}}\mathcal{M} := \{ \boldsymbol{v} \in \mathbb{R}^{n_{\boldsymbol{y}}} : \boldsymbol{B}(\boldsymbol{y})\boldsymbol{v} = \boldsymbol{0} \}$$
(32)

is the kernel of \boldsymbol{B} and has the same dimension $n_y - n_z$ as the manifold \mathcal{M} . Fig. 5 depicts \mathcal{M} , $\mathcal{T}_y \mathcal{M}$, and a solution of the DAE (29), which, starting from a consistent initial value, is required to proceed on the manifold.

The differential equation on the manifold \mathcal{M} that is equivalent to the DAE (29) is obtained as follows: The hidden constraint

$$\boldsymbol{0} = \boldsymbol{B}(\boldsymbol{y})\boldsymbol{a}(\boldsymbol{y}, \boldsymbol{z})$$

can be solved for $\boldsymbol{z}(\boldsymbol{y})$ according to the rank condition (30) and the implicit function theorem. Moreover, for $\boldsymbol{y} \in \mathcal{M}$ it holds $\boldsymbol{a}(\boldsymbol{y}, \boldsymbol{z}(\boldsymbol{y})) \in \mathcal{T}_{\boldsymbol{y}}\mathcal{M}$, which defines a vector field on the manifold \mathcal{M} . Overall,

$$\dot{\boldsymbol{y}} = \boldsymbol{a}(\boldsymbol{y}, \boldsymbol{z}(\boldsymbol{y})) \quad \text{for } \boldsymbol{y} \in \mathcal{M}$$
 (33)

represents then a differential equation on the manifold [Rhe84].

In theory, and also computationally [Rhe96], it is possible to transform the differential equation (33) from the manifold to an ordinary differential equation in a linear space of dimension $n_y - n_z$. For this purpose, one introduces a *local parametrization*

$$\boldsymbol{\psi}: E \to \mathcal{U} \tag{34}$$

where E is an open subset of $\mathbb{R}^{n_y-n_z}$ and $\mathcal{U} \subset \mathcal{M}$, see Fig. 5. Such a parametrization is not unique and holds only locally in general. It is, however, possible to extend it to a family of parametrizations such that the whole manifold is covered. For $\boldsymbol{y} \in \mathcal{U}$ and local coordinates $\boldsymbol{\xi} \in E$ we thus get the relations

$$oldsymbol{y} = oldsymbol{\psi}(oldsymbol{\xi}), \qquad \dot{oldsymbol{y}} = oldsymbol{\Psi}(oldsymbol{\xi}), \quad oldsymbol{\Psi}(oldsymbol{\xi}) := rac{\partialoldsymbol{\psi}}{\partialoldsymbol{\xi}}(oldsymbol{\xi}) \in \mathbb{R}^{n_y imes (n_y - n_z)}.$$

Premultiplying (33) by the transpose of the Jacobian $\Psi(\boldsymbol{\xi})$ of the parametrization and substituting \boldsymbol{y} by $\psi(\boldsymbol{\xi})$, we arrive at

$$\Psi(\boldsymbol{\xi})^T \Psi(\boldsymbol{\xi}) \dot{\boldsymbol{\xi}} = \Psi(\boldsymbol{\xi})^T \boldsymbol{a}(\boldsymbol{\psi}(\boldsymbol{\xi}), \boldsymbol{z}(\boldsymbol{\psi}(\boldsymbol{\xi}))) \,. \tag{35}$$

Since the Jacobian Ψ has full rank for a valid parametrization, the matrix $\Psi^T \Psi$ is invertible, and (35) constitutes the desired ordinary differential equation in the local coordinates $\boldsymbol{\xi}$. In analogy to a mechanical system in minimal coordinates, we call (35) a *local state space form*.

The process of transforming a differential equation on a manifold to a local state space form constitutes a *push forward* operator, while the reverse mapping is called a *pull back* operator [AMR88]. It is important to realize that the previously defined concept of an index does not appear in the theory of differential equations on manifolds. Finding hidden constraints by differentiation, however, is also crucial for the classification of DAEs from a geometric point of view.

The geometrical viewpoint was also considered very early by Sebastian Reich [Rei90], but its full potential became clear only a couple of years later when the topic of geometric numerical integration emerged, cf. [HLW02].

3.3 Singularly perturbed problems and regularization

In the early days of DAEs, regularization was a quite popular means to convert the algebraic part into a differential equation. Motivated by physical examples such as stiff springs or parasitic effects in electric circuits, a number of authors have looked into this topic. Furthermore, it is also interesting to start with a singularly perturbed ODE, discretize it, and then to analyze the behavior of exact and numerical solution in the limit case.

To study an example for a semi-explicit system, we consider Van der Pol's equation

$$\epsilon \ddot{q} + (q^2 - 1)\dot{q} + q = 0 \tag{36}$$

with parameter $\epsilon > 0$. This is an oscillator equation with a nonlinear damping term that acts as a controller. For large amplitudes $q^2 > 1$, the damping term introduces dissipation into the system while for small values $q^2 < 1$, the sign changes and the damping term is replaced by an excitation, leading thus to a self-exciting oscillator. Introducing Liénhard's coordinates [HNW93]

$$z := q, \quad y := \epsilon \dot{z} + (z^3/3 - z),$$

we transform (36) into the first order system

$$\dot{y} = -z, \qquad (37a)$$

$$\epsilon \dot{z} = y - \frac{z^3}{3} + z. \qquad (37b)$$

The case $\epsilon \ll 1$ is of special interest. In the limit $\epsilon = 0$, the equation (37b) turns into a constraint and we arrive at the semi-explicit system

$$\dot{y} = -z, \qquad (38a)$$

$$0 = y - \frac{z^3}{3} + z.$$
 (38b)

In other words, Van der Pol's equation (37) in Liénhard's coordinates is an example of a singularly perturbed system which tends to the semi-explicit (38) when $\epsilon \to 0$.

Such a close relation between a singularly perturbed system and a differential-algebraic equation is quite common and can be found in various application fields. Often, the parameter ϵ stands for an almost negligible physical quantity or the presence of strongly different time scales. Analyzing the *reduced system*, in this case (38), usually proves successful to gain a better understanding of the original perturbed equation [O'M74]. In the context of regularization methods, this relation is also exploited, but in reverse direction [Han90]. One starts with a DAE such as (38) and replaces it by a singularly perturbed ODE, in this case (37). In numerical analysis, the derivation and study of integration schemes via a singularly perturbed ODE has been termed the *indirect approach* [HLR89] and lead to much additional insight [HLR88, LP86, Lub93], both for the differential-algebraic equation as limit case and for the stiff ODE case. A particularly interesting method class for the indirect approach are Rosenbrock methods as investigated by Rentrop, Roche & Steinebach [RRS89].

3.4 General fully implicit DAEs

In contrast to the solution theory in the linear constant coefficient case, the treatment of fully implicit DAEs without a given internal structure is still challenging, even from today's perspective. For this purpose, Campbell [Cam93] introduced the derivative array as key concept that carries all the information of the DAE system. The derivative array is constructed from the definition of the differential index, i.e., one considers the equations

$$F(\dot{\boldsymbol{x}}, \boldsymbol{x}, t) = \boldsymbol{0},$$

$$\frac{\mathrm{d}}{\mathrm{d}t}F(\dot{\boldsymbol{x}}, \boldsymbol{x}, t) = \frac{\partial}{\partial \dot{\boldsymbol{x}}}F(\dot{\boldsymbol{x}}, \boldsymbol{x}, t) \boldsymbol{x}^{(2)} + \dots = \boldsymbol{0},$$

$$\vdots$$

$$\frac{\mathrm{d}^{k}}{\mathrm{d}t^{k}}F(\dot{\boldsymbol{x}}, \boldsymbol{x}, t) = \frac{\partial}{\partial \dot{\boldsymbol{x}}}F(\dot{\boldsymbol{x}}, \boldsymbol{x}, t) \boldsymbol{x}^{(k+1)} + \dots = \boldsymbol{0}$$
(39)

for a DAE of index k. Upon discretization, (39) becomes an overdetermined system that can be tackled by least squares techniques. The challenge in this procedure, however, is the in general unknown index k and its determination.

Algorithms based on the derivative array are a powerful means for general unstructured DAE systems, and this holds even for the linear constant coefficient case since the computation of the Weierstrass form or the Drazin inverse are very sensitive to small perturbations and thus problematic in finite precision arithmetic. For the derivative array, in contrast, so-called staircase algorithms have been developed that rely on orthogonal matrix multiplications and are much more stable [BLMV15].

3.5 Constrained Hamiltonian systems

In the conservative case, the Lagrange equations (14) of constrained mechanical motion can be reformulated by the transformation to Hamilton's canonical equations. This leads to a mathematical model that is typical for molecular dynamics simulations. Again, constraints come into play in this application field, and the time discretizations need to cope with the index and stability issues.

We define the Lagrange function

$$L(\boldsymbol{q}, \dot{\boldsymbol{q}}) := T(\boldsymbol{q}, \dot{\boldsymbol{q}}) - U(\boldsymbol{q})$$
(40)

as the difference of kinetic and potential energy. The conjugate momenta $p(t) \in \mathbb{R}^{n_q}$ are then given by

$$\boldsymbol{p} := \frac{\partial}{\partial \dot{\boldsymbol{q}}} L(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \boldsymbol{M}(\boldsymbol{q}) \dot{\boldsymbol{q}}, \qquad (41)$$

and for the Hamiltonian we set

$$H := \boldsymbol{p}^T \dot{\boldsymbol{q}} - L(\boldsymbol{q}, \dot{\boldsymbol{q}}) \,. \tag{42}$$

Since the velocity $\dot{\boldsymbol{q}}$ can be expressed as $\dot{\boldsymbol{q}}(\boldsymbol{p}, \boldsymbol{q})$ due to (41) if the mass matrix is invertible, we view the Hamiltonian as a function $H = H(\boldsymbol{p}, \boldsymbol{q})$. Moreover, we observe that H is the total energy of the system because

$$H = \boldsymbol{p}^T \boldsymbol{M}(\boldsymbol{q})^{-1} \boldsymbol{p} - \frac{1}{2} \dot{\boldsymbol{q}}^T \boldsymbol{M}(\boldsymbol{q}) \dot{\boldsymbol{q}} + U(\boldsymbol{q}) = T + U.$$

Using the least action principle (16) in the new coordinates p and q and applying the Lagrange multiplier technique as above in the presence of constraints, we can express the equations of motion as

$$\dot{\boldsymbol{q}} = \frac{\partial}{\partial \boldsymbol{p}} H(\boldsymbol{p}, \boldsymbol{q}),$$

$$\dot{\boldsymbol{p}} = -\frac{\partial}{\partial \boldsymbol{q}} H(\boldsymbol{p}, \boldsymbol{q}) - \boldsymbol{G}(\boldsymbol{q})^T \boldsymbol{\lambda},$$

$$\boldsymbol{0} = \boldsymbol{g}(\boldsymbol{q}).$$
(43)

The Hamiltonian equations (43) possess a rich mathematical structure that should be preserved by numerical methods. In the 1990s, this problem class led to the new field of geometric integration, see the monograph by Hairer, Lubich & Wanner [HLW02] for an extensive exposition. For brevity, we simply mention the SHAKE scheme as one of the established methods. In case of a Hamiltonian $H = \frac{1}{2} \boldsymbol{p}^T \boldsymbol{M}^{-1} \boldsymbol{p} + U(\boldsymbol{q})$ with constant mass matrix, it reads

$$\boldsymbol{q}_{n+1} - 2\boldsymbol{q}_n + \boldsymbol{q}_{n-1} = -\tau^2 \boldsymbol{M}^{-1} (\nabla U(\boldsymbol{q}_n) + \boldsymbol{G}(\boldsymbol{q}_n)^T \boldsymbol{\lambda}_n), \qquad (44)$$

$$\boldsymbol{0} = \boldsymbol{g}(\boldsymbol{q}_{n+1}).$$

In each time thus a nonlinear system for \boldsymbol{q}_{n+1} and $\boldsymbol{\lambda}_n$ needs to be solved, which is closely related to the projection step (28a).

4 Beyond classical DAEs

At the end of the last century, new topics emerged that were closely related to DAEs but also included important aspects from other fields. In particular, the topic of Partial Differential Algebraic Equations (PDAEs) became attractive by then, driven by time-dependent partial differential equations that were treated by the method of lines and that featured additional constraints.

4.1 Navier-Stokes incompressible

It requires convincing examples to demonstrate the benefits of a differentialalgebraic viewpoint in the PDE context. One such example is sketched next.

A classical example for a PDAE is given by the Navier-Stokes equations

$$\dot{\boldsymbol{u}} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} + \frac{1}{\rho}\nabla p = \nu \boldsymbol{\Delta} \boldsymbol{u} + \boldsymbol{l},$$
 (45a)

$$0 = \nabla \cdot \boldsymbol{u} \tag{45b}$$

for the velocity field $\boldsymbol{u}(x,t)$ and the pressure p(x,t) in a *d*-dimensional domain Ω , with mass density ρ , viscosity ν , and source term $\boldsymbol{l}(x,t)$. The second equation (45b) models the incompressibility of the fluid and defines a constraint for the velocity field. For simplification, the convection term $(\boldsymbol{u} \cdot \nabla)\boldsymbol{u}$ in (45a) can be omitted, which makes the overall problem linear and more amenable for the analysis. In an abstract notation, the resulting Stokes problem then reads

$$\dot{\boldsymbol{u}} + \boldsymbol{\mathcal{A}}\boldsymbol{u} + \boldsymbol{\mathcal{B}}'\boldsymbol{p} = \boldsymbol{l}, \qquad (46a)$$

$$\mathcal{B}\boldsymbol{u} = 0, \qquad (46b)$$

with differential operators \mathcal{A} and \mathcal{B} expressing the Laplacian and the divergence, respectively. The notation \mathcal{B}' stands for the conjugate operator of \mathcal{B} , which here is the gradient.

The discretization, e.g., by a Galerkin-projection

$$\boldsymbol{u}(x,t) \doteq \boldsymbol{N}(x)\boldsymbol{q}(t), \qquad p(x,t) \doteq \boldsymbol{Q}(x)\boldsymbol{\lambda}(t)$$

with ansatz functions N and Q in some finite element spaces, transforms the infinite-dimensional PDAE (46) to the DAE

$$\boldsymbol{M}\dot{\boldsymbol{q}} + \boldsymbol{A}\boldsymbol{q} + \boldsymbol{B}^{T}\boldsymbol{\lambda} = \boldsymbol{l},$$
 (47a)

$$Bq = 0. \tag{47b}$$

While the mass matrix \boldsymbol{M} and stiffness matrix \boldsymbol{A} are symmetric positive definite and symmetric positive semi-definite, respectively, and easy to handle, the constraint matrix \boldsymbol{B} is generated by mixing the discretizations for the velocity field and the pressure. It is well-known in mixed finite elements [BF91] that a bad choice for the discretization will either result in a rank-deficient matrix \boldsymbol{B} or in a situation where the smallest singular value of \boldsymbol{B} is approaching zero for a decreasing mesh size. This means that the DAE (47) may become singular or almost singular due to the spatial discretization. The famous LBB-condition by Ladyshenskaja, Babŭska, and Brezzi [BF91] gives a means to classify the discretization pairs for \boldsymbol{u} and p. If the matrix \boldsymbol{B} has full rank, the index of the DAE (47) is k = 2.

To summarize, PDEs with constraints such as the Navier-Stokes equations often feature a rich structure that should be exploited, and building on the available PDE methodology reveals interesting cross-connections with the differential-algebraic viewpoint. In this context, the abstract formulation (46) as *transient saddle point problem* defines a rather broad problem class where many application fields can be subsumed [Sim13].

By combining the state-of-the-art in DAEs with advanced PDE methodology and numerics, powerful algorithms can then be developed that break new ground. Time-space adaptivity for PDAEs is one such topic where many different aspects are put together in order to set up numerical schemes with sophisticated error control. The work by Lang [Lan13] defines a cornerstone in this field. In electrical circuit simulation, the inclusion of heating effects or semi-conductors results also in PDAE models where ODEs, DAEs, and PDEs are coupled via network approaches, see, e.g., [Gün01, ABGT03].

4.2 Stochastic DAEs

In the spring of 2006, Oberwolfach offered again the showcase for the latest developments in DAEs – 25 years after the workshop where Bill Gear had first investigated the mathematical pendulum (2) in Cartesian coordinates. The organizers were Stephen Campbell, Roswitha März, Linda Petzold, and Peter Rentrop. Among the participants from all over the world was Bill Gear himself, and during the week it became evident that DAEs were now well-established in many fields.

The same year, the book by Kunkel & Mehrmann [KM06] appeared, which shed new light on topics such as boundary value problems in differentialalgebraic equations and the numerical treatment of fully implicit systems (3).

Most talks at the meeting addressed the field of PDAEs, but among the other prominent topics were also optimization and optimal control problems with constraints described by DAEs, see, e.g., [CR08, KKBS04] and model order reduction for descriptor systems [RS07]. An emerging topic at the time were *stochastic differential-algebraic equations* or SDAEs in short. Since many models in science and engineering contain uncertain quantities, it is natural to extend the methodology for DAEs by corresponding random terms. This could either be a parameter or coefficient that is only known approximately or even an extra diffusion term in the differential equation that is expressed in terms of a Wiener process. For the constant coefficient system (7), such a diffusion term leads to the linear SDAE

$$\boldsymbol{E} \mathrm{d}\boldsymbol{x}(t) + \boldsymbol{H}\boldsymbol{x}(t) = \boldsymbol{c}(t) + \boldsymbol{C} \mathrm{d}\boldsymbol{W}(t)$$
(48)

with a Wiener process \boldsymbol{W} in \mathbb{R}^{n_x} and a square matrix \boldsymbol{C} . For work in this field and applications in electrical circuit analysis we refer to [HMS02, Win03].

5 Nonsmooth dynamical systems

Differential equations for dynamical systems become particularly challenging when nonsmooth functions must be considered in one way or the other. Economical and financial mathematics have been a driving force for nonsmooth dynamical systems, where finding equilibria in market situations can be solved using variational inequalities that are closely related to nonlinear programming and convex optimization. Nonsmooth differential equations play a role in optimal control, robotic path planning and perturbation analysis. They occur also in electrical circuits with ideal diodes, see Figure 6a. An



(a) A simple electrical cir- (b) A simulation of granular material with perfect cuit with an ideal diode D. unilateral contacts and friction

Figure 6: Examples from electrical and mechanical engineering that include nonsmooth phenomena.

ideal diode transmits electrical current only in one direction and blocks it in the other. If the electrical current I is reversed, the conductivity of the ideal diode suddenly changes from 0 to ∞ . In other words, the electric current at the diode must be positive at all times, $I(t) \ge 0$. If the current is nonzero, I(t) > 0 the voltage must be zero, V(t) = 0. If there is a nonzero voltage V(t) > 0 at the diode, this means, that the diode is blocking and does not transmit any current, I(t) = 0. This translates into the complementarity condition

$$I(t) \ge 0, \quad V(t) \ge 0, \quad I(t) \cdot V(t) = 0.$$

at the diode. Which formalism can we use to incorporate inequalities and complementarity conditions into Kirchhoff's circuit laws? Another example are multibody dynamical systems with impacts, where contacts are modeled using a positivity constraint $g(\mathbf{q}) \geq 0$ of a signed distance function, see for instance Figure 6b. Without long range and adhesive forces, the contact force f_c between two contacting bodies must be nonnegative. In addition, it must be zero if the contact gap is positive. This again yields a complementarity problem

$$g(\boldsymbol{q}) \ge 0, \quad f_c \ge 0, \quad g(\boldsymbol{q}) \cdot f_c = 0.$$

It is easy to construct examples, where the velocity of a rigid body must be discontinuous to enforce an inequality constraint. Think of a point mass accelerating under gravity towards the ground. At the moment of impact the velocity of the particle must jump from a negative to a nonnegative value instantaneously. This toy example demonstrates the two main difficulties that must be tackled in nonsmooth dynamical systems.

- 1. DAEs give us a mathematical structure to describe mechanical systems subject to *algebraic* constraints. How can we incorporate *inequality* constraints $g(\mathbf{q}) \ge 0$?
- 2. Newton's second axiom

$$\boldsymbol{M}\ddot{\boldsymbol{q}} = \boldsymbol{f}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) \tag{49}$$

implies that the trajectory q(t) is at least once continuously differentiable, which is obviously not the case here. Clearly, Equation (49) still has validity almost everywhere, but how can we handle the instances in time with impacts?

In the following, we will take a closer look on the treatment of dynamical systems that are constrained to a feasible set:

$$F(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) = \boldsymbol{0},$$

$$\boldsymbol{q} \in K = \{ \boldsymbol{q} \mid \boldsymbol{g}(\boldsymbol{q}) \ge \boldsymbol{0} \}$$
(50)

for some appropriate function $\boldsymbol{g}: \mathbb{R}^{n_q} \to \mathbb{R}^{n_{\lambda}}$.

5.1 A short zoology

This section provides the most important concepts with regard to nonsmooth dynamical systems, including Moreau's sweeping process, differential inclusions, projected dynamical systems, variational inequalities, complementarity dynamic systems, differential variational inequalities and finally measure differential equations and measure differential inclusions. The list is far from complete and many more authors than the ones cited here have made important contributions. In addition, the overview is biased as the subject is approached from the perspective of nonsmooth mechanical problems. For brevity, nonsmooth phenomena due to friction are excluded from this introduction all together.

5.1.1 Unilateral constraints and Moreau's sweeping process

Jean Jacques Moreau studied mechanical systems subject to unilateral constraints since the mid 1960s [Mor66, Mor77, Mor88b, MP90, Mor93, Mor99]. A *unilateral constraint* restricts the motion of a dynamical system only in one direction, e.g. by imposing inequality constraints. Imagine a particle's position x(t) is restricted to a moving closed set K(t). Consider a particular point in time t^* , where the particle is located on the boundary of its feasible set, $x(t^*) \in \partial K(t^*)$. Assume for simplicity that the boundary can locally be described as a smooth manifold \mathcal{M} . Then we can find a tangent space $T_{x(t^*)}\mathcal{M}$ and a normal direction **n** to $T_{x(t^*)}\mathcal{M}$. The particle is only allowed to move freely in the half space to one side of the tangent space towards the interior of $K(t^*)$. In other words, the constraint restricts the motion unilat*erally.* In contrast to this, if the particle were to be restricted to the manifold \mathcal{M} , the particle would not be allowed to move in either direction along the normal n, i.e. it would be *bilaterally constrained*. The space spanned by the normal vector is often called the annihilator of the tangent space. The assumption that the boundary of the moving set K(t) can locally be described as a manifold is too restrictive in general. It suffices, if we can describe the boundary of the set through so-called tangent cones² and normal cones rather than tangent spaces and their associated annihilators, see Figure 7.

A function x is said to be a solution to the *first order sweeping process* for the time-dependent set K, if

$$x(0) \in K(0), \tag{51a}$$

$$x(t) \in K(t), \tag{51b}$$

$$-\dot{x}(t) \in N_{x(t)}K(t),\tag{51c}$$

where $N_{x(t)}K(t)$ denotes the normal cone to the set K(t) at x(t). It is defined as the polar of the tangent cone $T_{x(t)}K(t)$,

$$N_{x(t)}K(t) = \left\{ \xi \mid \langle \xi, y \rangle \le 0 \quad \forall y \in T_{x(t)}K(t) \right\}.$$

The differential inclusion (DI) in (51c) means that if x(t) is on the boundary of the set, the derivative must point inward, see Figure 7b. In the interior of the set K(t), it holds $N_{x(t)}K(t) = \{0\}$, and thus $\dot{x}(t) = 0$. In other words, the particle is *swept* with the moving boundary.

 $^{^2 \}mathrm{See}$ for instance [RW98] for a definition of tangent cones of sets.



(a) The tangent space $T_x \mathcal{M}$ and its annihilator $N_x \mathcal{M}$ at a point xof a manifold \mathcal{M} , compare Figure 5.



(b) The tangent cones $T_{x_i}K$ and normal cones $N_{x_i}K$ at three points x_i , i = 1, 2, 3 in a set K. Note that $N_{x_3}K = \{\mathbf{0}\}$ and $T_{x_3}K$ takes up the whole space.

Figure 7: Comparison of tangent space and annihilator of manifolds to tangent and normal cones of sets.

5.1.2 Projected dynamical systems

In [NZ95], a projected dynamical system (PDS) is defined as

$$\dot{x} = \Pi_K(x, -F(x)) \tag{52}$$

where

$$\Pi_K(x,v) = \lim_{\delta \to 0} \frac{\operatorname{prox}_K(x+\delta v) - x}{\delta}$$

and K is a closed convex set defined by constraints on the system. Here, $\operatorname{prox}_K(x)$ denotes the projection operator

$$\operatorname{prox}_{K}(x) = \arg \min_{z \in K} ||x - z||.$$
 (53)

Because it holds

$$\Pi_K(x,v) = \operatorname{prox}_{T_xK}(v),$$



Figure 8: Two example scenarios for projected dynamical systems.

Equation (52) can be rewritten as a projection onto the tangent cone $T_x K$ [BDLA06]:

$$\dot{x} = \operatorname{prox}_{T_x K}(-F(x)). \tag{54}$$

Equations (52) and (54) are first order ordinary differential equations with nonsmooth right hand sides. In Moreau's sweeping process, the direction of the derivative is prescribed only by the moving set K, while here, the direction of the derivative is mainly prescribed by the right hand side -F(x). In contrast to Moreau's sweeping process, it holds $\dot{x} = -F(x) \neq 0$ in the interior of the set. The projection operator makes sure, that the system never moves towards the exterior of K if x is on the boundary, see Figure 8. According to [BDLA06], any solution to Equation (54) is also a solution of the differential inclusion

$$-\dot{x}(t) \stackrel{\text{a.e.}}{\in} F(x(t)) + N_{x(t)}K,$$
(55)

which is Moreau's first order sweeping process, if F(x) = 0.

Standard works on differential equations with discontinuous right hand side and differential inclusions are [AC84, Fil88]. Existence and uniqueness to such problems depends on properties of the set-valued map K(t) and its tangent and normal cones. Nagurney and Zhang motivate their definition of projected dynamical systems as a means to unify the theory of dynamical systems and variational inequalities.

5.1.3 Variational inequalities

Stuart Antman recites the emergence of variational inequalities in [Ant83]. In 1959, Antonio Signorini posed the problem of finding the material displacements of a heavy deformable body resting on a rigid frictionless flat ground. This problem, now called the *Signorini contact problem*, is particularly difficult, as the geometry of the contact region between the body and the ground is a priori unknown. Gaetano Fichera, a student of Signorini, studied existence and uniqueness of solutions to this problem using the calculus of variations and published his results in 1963 and 1964. In hindsight, Fichera's solution of Signorini's contact problem relies on the solution to a *variational inequality* (VI). In a general setting, a variational inequality is defined as follows. Let X be a Banach space, $K \subset X$ and $g: K \to X^*$ be a mapping from K to the dual space X^* of X. A variational inequality denotes the problem of finding $u \in K$, such that

$$\langle g(u), u' - u \rangle \ge 0 \quad \forall u' \in K.$$
 (56)

Figure 9a contains a geometric representation of a variational inequality in finite dimensions. Note that (56) translates to $-g(u) \in N_u K$. The term variational inequality was first introduced by Stampacchia in 1965 and a theory around variational inequalities quickly grew during the second half of the 1960s. Most researchers on the subject at that time heavily cited Fichera's work. For more details on how Fichera's solution to Signorini's problem can be solved using variational inequalities and how his work influenced their analysis in the years to come, see [Ant83].

5.1.4 Complementarity dynamic systems

A convex cone K is a set that is closed with respect to positive linear combinations, i.e. if x and y are in the cone K and $\alpha, \beta \ge 0$, it must hold that $\alpha x + \beta y \in K$. The variational inequality (56) is equivalent to a *cone complementarity problem* (CCP)

$$g(u) \in K^*, \quad u \in K, \quad \langle g(u), u \rangle = 0, \tag{57}$$



Figure 9: Geometric interpretations of a variational inequality and a cone complementarity problem in finite dimensions.

if the set K is a convex cone, which is proved in [Kar71]. Here,

$$K^* = \{ y \in X^* \mid \langle y, x \rangle \ge 0 \ \forall x \in K \} \subset X^*$$

is the *dual cone* to K. Notice that the dual cone is the negative normal cone to K at 0. Because of the complementarity condition in (57) it is clear, that if either g(u) or u is in the interior of their respective cones K^* and K, the other of the two must be zero (see Figure 9b), hence the term *complementarity*.

The positive orthant \mathbb{R}^n_+ is such a convex cone and it holds $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$, i.e. it is self-dual. For this specific cone the CCP (57) resembles in part the well-known Karush-Kuhn-Tucker necessary conditions for optimization problems with inequality constraints $g(u) \geq 0$:

$$\boldsymbol{g}(\boldsymbol{u}) \geq \boldsymbol{0}, \quad \boldsymbol{u} \geq \boldsymbol{0}, \quad \boldsymbol{u}^T \boldsymbol{g}(\boldsymbol{u}) = 0,$$

where the inequality " \geq " is to be understood componentwise.

Complementarity dynamic systems are differential equations coupled to a complementarity problem. They were considered by several authors, including J. J. Moreau himself, to tackle mechanical systems with inequality constraints [Mor66, Löt82, AP97, Jea99, Pfe03, PFU06]. Their works were heavily influenced by results from nonlinear programming and optimization.

5.1.5 Differential variational inequalities

In [AC84] a special version of a differential inclusion is mentioned, called a *differential variational inequality*. Given a closed convex subset $K \subset X$ of a vector space X and a set-valued map $F: K \to X$, a differential variational inequality consists of finding a function $x: [0, T] \to X$ satisfying

$$x(t) \in K \qquad \qquad \forall t \in [0, T], \tag{58a}$$

$$\dot{x}(t) \in F(x(t)) - N_{x(t)}K$$
 for a.a. $t \in [0, T]$. (58b)

Notice that the differential inclusion (58b) coincides with (55). The relation to variational inequalities becomes clear, if we write $\dot{x}(t) = f(x(t)) + g(t)$ for $f(x(t)) \in F(x(t))$ and $g(t) \in -N_{x(t)}K$. With this, (58) reads

$$x(t) \in K \qquad \qquad \forall t \in [0, T], \quad (59a)$$

$$\langle \dot{x}(t) - f(x(t)), y - x(t) \rangle \ge 0$$
 for a.a. $t \in [0, T]$ and all $y \in K$. (59b)

A different, slightly more general, definition of a differential variational inequality comes from Pang and Stewart [PS08]. They define a finite dimensional differential variational inequality as a differential equation coupled to a variational inequality

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(t, \boldsymbol{x}(t), \boldsymbol{u}(t)), \tag{60a}$$

$$\langle \boldsymbol{F}(t, \boldsymbol{x}(t), \boldsymbol{u}(t)), \boldsymbol{u}' - \boldsymbol{u} \rangle \ge 0 \quad \forall \boldsymbol{u}' \in K,$$
 (60b)

$$\Gamma(\boldsymbol{x}(0), \boldsymbol{x}(T)) = 0. \tag{60c}$$

The problem consists of finding functions $\boldsymbol{x} : [0, T] \to \mathbb{R}^n$ and $\boldsymbol{u} : [0, T] \to K \subset \mathbb{R}^m$ satisfying Equation (60) given the continuous functions $\boldsymbol{f}, \boldsymbol{F}, \Gamma$ and a subset $K \subset \mathbb{R}^m$. In this finite dimensional setting the dual pairing (60b) is the Euclidean scalar product and Equation (60c) is a prescribed initial or boundary condition. If K is a convex cone, the variational inequality (60) can be expressed as a cone complementarity problem. To differentiate their definition (60) from the definition (58) given by Aubin and Cellina, Pang and Stewart call problems of the type (58) variational inequalities of evolution (VIE), a convention that is adapted here. Pang and Stewart show in the same article, that the VIE is a special case of a DVI and suggest a unified version of the two, that essentially consists of a differential equation, an algebraic equation and a variational inequality. Finally, it should be noted that if the set K is a convex cone, a DVI and a VIE are conceptually equivalent.

Let \mathcal{U} denote the set of solutions to the variational inequality (60b). Then, the DVI (60) can be rewritten as differential inclusion

$$\dot{\boldsymbol{x}} \in \boldsymbol{f}(t, \boldsymbol{x}(t), \mathcal{U}),$$

The DVI (60) can also be rewritten as a DAE: The function \boldsymbol{u} solves the variational inequality (60b) if and only if

$$\mathbf{0} = \boldsymbol{g}(t, \boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{u} - \operatorname{prox}_{K}(\boldsymbol{u} - \boldsymbol{F}(t, \boldsymbol{x}, \boldsymbol{u})),$$

where prox_K is the orthogonal projection onto the set K as defined in (53). Thus, the variational inequality can be replaced by a nonsmooth algebraic constraint. Pang and Stewart plea their case why DVIs deserve a special treatment, even if they can be considered as either DIs or DAEs. They argue, that the DI theory is too general to be of practical use, since many of its assumptions are cumbersome to prove for specific situations. While the theory of DAEs appeals to them, it is problematic as the differentiability of the algebraic constraints play an important role in it. If the algebraic constraint however is nonsmooth, as is the case if it is constructed from a variational inequality, many results cannot be used.

Still, it is obvious that the authors of [PS08] were inclined to adopt as much terminology from and analogy to DAEs for their analysis as possible. As for DAEs, the unknowns of a DVI are separated into differential variables \boldsymbol{x} and algebraic variables \boldsymbol{u} . While a DAE can be seen as an ODE coupled to an algebraic equation through the algebraic variables, a DVI is a differential equation coupled to a variational inequality. Even the concept of the index of a DVI is adopted. A DVI of index zero is just an ordinary differential equation. Under certain conditions for the set-valued map \boldsymbol{F} , the algebraic unknown \boldsymbol{u} can be written as a function of \boldsymbol{x} and t using the implicit function theorem. One differentiation then reveals the DVI as system of differential equations in \boldsymbol{x} and \boldsymbol{u} . This is called a DVI of index one. According to the authors, higher index DVIs can be considered as well, but they restrict their analysis to DVIs of index one with absolutely continuous solutions.

5.1.6 Derivatives of functions of bounded variation

Differential variational inequalities and the related concepts help us to describe dynamical systems subject to inequality constraints. We opened the section on nonsmooth dynamical systems with another problem. If the velocity of a mechanical system subject to unilateral constraints is discontinuous, how do we interpret accelerations and forces? The results in the next two sections are taken from [Nat77, Mor88a, Glo00, AB08, Ger12].

For absolutely continuous functions $x : [0, T] \to \mathbb{R}$, the derivative $v = \dot{x}$ is of bounded variation. This entails that v is smooth almost everywhere, but can have a countable number of finite jump discontinuities, at which $v = \dot{x}$ is not defined per se³. Still, we can write

$$x(t) = x(0) + \int_0^t \dot{x}(\tau) \,\mathrm{d}\tau,$$

and just jump over the discontinuity points in the interval [0, t], as they are a zero-set of the Lebesgue measure $d\tau$ anyway.

How do we define the derivative of a function of bounded variation v? It turns out, that measure theory helps. If the function can have jumps, it changes infinitely fast from its left limit to its right limit at such a discontinuity point. The discontinuity points comprise only a zero-set of the Lebesgue measure dt, but surely the infinitely fast changes in v cannot simply be ignored. We must capture the changes of the function using measures, for which the set of discontinuity points is not a zero set.

We can construct the so-called Lebesgue–Stieltjes measure dv from a function of bounded variation v, so that it holds

$$v(t) = v(0) + \int_0^t \mathrm{d}v.$$

In this sense, the Lebesgue–Stieltjes measure dv plays the role of the derivative of v. The measure captures the changes in v regardless of whether they happen smoothly or abruptly. Using the Lebesgue decomposition and Radon–Nikodym theorems, the Lebesgue–Stieltjes measures of v can be decomposed into three parts,

$$\mathrm{d}v = a(t)\mathrm{d}t + \mathrm{d}\delta_p + \mathrm{d}\mu_s.$$

Here, a(t) is the so-called Radon-Nikodym density of the continuous part of the measure dv with respect to the Lebesgue measure dt, $d\delta_p$ is a sum

³We could easily define the function $v(t_c)$ at the discontinuity points t_c as either the left limit $v^-(t_c) = \lim_{h \to 0} v(t_c - h)$ or the right limit $v^+(t_c) = \lim_{h \to 0} v(t_c + h)$, with h > 0, as convention. While this potentially helps the analysis, it does not change the structure of the problem.
of Dirac delta measures that captures all jumps of v and $d\mu_s$ is a singular measure associated with the Cantor part of v. If v has no Cantor part, it is called a function of specially bounded variation and the last term vanishes. This is a valid assumption for a large class of problems. If in addition it is absolutely continuous, the second term vanishes and it holds dv = a(t)dtwith $a = \dot{v}$.

Note that the Lebesgue–Stieltjes measure dv corresponds to the distributional derivative of v:

$$\langle \mathrm{D}v, \varphi \rangle = -\int_0^T v \dot{\varphi} \, \mathrm{d}t = -\int_0^T v \, \mathrm{d}\varphi = \int_0^T \varphi \, \mathrm{d}v \quad \forall \varphi \in C_0^\infty(0, T),$$

because all functions in $C_0^{\infty}(0,T)$ are absolutely continuous and thus have a density w.r.t. dt. Here we use integration by parts for Lebesgue–Stieltjes integrals:

$$\int_{0}^{T} v \, \mathrm{d}w = \left[vw\right]_{0}^{T} - \int_{0}^{T} w \, \mathrm{d}v \tag{61}$$

for all v, w with bounded variation.

5.1.7 Measure differential equations and measure differential inclusions

A measure differential equation (MDE) takes the form

$$dx = F(t, x) dt + G(t, x) du,$$
(62)

where dx and du denote the Lebesgue–Stieltjes measures of the functions of bounded variation x and u and dt denotes the Lebesgue measure of time [Lee74, PD82]. The corresponding initial value problem consists of finding $x : [0,T] \to \mathbb{R}$ given an initial value $x(0) = x_0$. Measure differential equations were motivated from Optimal Control Theory and the theory of perturbed systems. Here, the function u takes the role of inputs or external perturbations. In control theory, u drives the state x of the system to a certain target. In certain applications it might be useful to allow jumps in the input, which in turn result in jumps in the state and vice-versa. In perturbation theory, impulsive perturbations are studied. See [Pic17] for a recent, slightly different approach towards measure differential equations in the context of differential geometry. One way to read Equation (62) is as a short-form for the variational condition

$$\int_0^T \varphi \, \mathrm{d}x = \int_0^T \varphi \, F(t,x) \, \mathrm{d}t + \int_0^T \varphi \, G(t,x) \, \mathrm{d}u \quad \forall \varphi \in \mathcal{T}, \qquad (63)$$

for some appropriate space of test functions \mathcal{T} . If u were absolutely continuous, we could write $du = \dot{u} dt$. Then it is obvious from (62), that dxhas a density $F(t, x) + G(t, x)\dot{u}$ with respect to dt and thus we can write $dx = \dot{x} dt$. Finally, using the fundamental lemma of calculus of variations, we could derive a differential equation

$$\dot{x} = F(t, x) + G(t, x)\dot{u}$$

from (63).

Moreau was well aware that the sweeping process results in discontinuous velocities. He introduced *measure differential inclusions* (MDI) as an extension to the differential inclusion in (51) [Mor88b, MM93, Glo00, Ste01, Pao05]. In [MM93] a *measure differential* inclusion is defined as

$$\mathrm{d}u \in N_u K,$$

for a convex subset $K \subset X$ of a Hilbert space X. It should be noted here, that this definition would have just as much meaning in a Banach space, which will be exploited in the next section. In this abstract setting, the normal cone is a subset of the dual space X^* of X. In the same book, existence and uniqueness results are given for second order measure differential inclusions that result from mechanical systems with unilateral constraints and impacts. See also [ABG08] for higher order MDIs in the context of Moreau's sweeping process.

Up to this point all considerations concerning normal and tangent cones could be directly translated to the finite dimensional spaces in which the functions take their values. In other words, if we consider functions taking values in \mathbb{R}^n , the dual pairings $\langle \cdot, \cdot \rangle$ can be interpreted as the Euclidean scalar product in \mathbb{R}^n and the negativity condition in the definition of the normal cone must hold in a pointwise sense. This gives us a geometric interpretation of tangent and normal cones, complementarity problems and variational inequalities. As soon as distributional derivatives play a role, this interpretation becomes more and more difficult and the problems are usually considered in infinite dimensional function spaces.

5.1.8 A very diverse field

In the past 60 years, many researchers devoted themselves to dynamical systems that encounter nonsmooth phenomena. These researchers borrowed results from convex and functional analysis, numerical analysis, nonlinear programming and optimization, measure theory and differential geometry to produce a rich structure and a diverse selection of concepts to describe "nonsmoothness" of differential equations. DAEs did not play a major role in the early development of the field. It is only recently that the DAE theory has become an important influence on the field of nonsmooth dynamical systems.

Many of introduced concepts are very similar and under mild assumptions some can be shown to be equivalent [BDLA06]. The diversity of the field comes at a small price however, which is the lack of a common language. It can be very hard to judge which of these concepts is most appropriate for the application at hand. It is extremely difficult to find the vocabulary that captures the relevant phenomena of a specific problem, but still leaves enough generality to be useful to a large class of problems. While this can be said about any mathematical theory, it is especially apparent in the field of nonsmooth differential equations.

5.2 Nonsmooth mechanical systems with impacts

We have already established that the trajectory \boldsymbol{q} of a mechanical system with inequality constraints cannot be the solution of Newton's second axiom (49), since we must allow $\dot{\boldsymbol{q}}$ to have jump discontinuities. To understand the physics behind the system, we have to fall back to more general mechanical principles.

5.2.1 Hamilton's principle as a differential inclusion

Classical mechanics offers a few physical principles as governing equations, that are formulated in a variational setting rather than as a second order ordinary differential equation. Hamilton's principle of least action (16) is an example. While there have been several attempts to extend the principle for discontinuous solutions since its development, most of them were undertaken quite recently with the help of the new tools provided by the growing field of nonsmooth dynamical systems, see [Erd77, PG00, FMOW03, LAG09] and

the references therein. It is unclear whether William Hamilton had nonsmooth mechanical systems on his mind in the 1830s. But his principle is formulated with just enough generality to be extended to the nonsmooth case.

To begin, let us consider the problem in a function space setting. While the (generalized) coordinates \boldsymbol{q} are not continuously differentiable, they must still be continuous: A material point of a mechanical system cannot just disappear and reappear elsewhere. The velocity $\dot{\boldsymbol{q}}$ is continuous almost everywhere, except for a countable number of time points, where the velocity undergoes finite jumps. In mathematical terms, $\boldsymbol{q} : \mathbb{R} \supset [0, T] \rightarrow \mathbb{R}^N$ is an absolutely continuous function and $\dot{\boldsymbol{q}} : [0, T] \rightarrow \mathbb{R}^N$ is a function of specially bounded variation,

$$\boldsymbol{q} \in \mathrm{AC}\left([0,T]\right)$$
 and $\dot{\boldsymbol{q}} \in \mathrm{SBV}\left([0,T]\right)$

Introducing inequality constraints $g(q) \ge 0$ into the system translates to finding a feasible physical solution q in the set

$$\mathcal{S} = \{ \boldsymbol{q} \in \mathrm{AC}\left([0,T]\right) \mid \boldsymbol{g}(\boldsymbol{q}(t)) \ge 0 \; \forall t \in [0,T] \}.$$

The classical principle of virtual work and the closely related principle of D'Alembert can be used to examine systems, where the feasible set is a manifold

$$\mathcal{M} = \{ \boldsymbol{q} \in \mathrm{AC}\left([0,T]\right) \mid \boldsymbol{g}(\boldsymbol{q}(t)) = 0 \; \forall t \in [0,T] \}$$

with a sufficiently smooth function \boldsymbol{g} . The principle requires the constraint forces $\boldsymbol{f}_c(\boldsymbol{q})$ to be orthogonal to the tangent space $T_{\boldsymbol{q}}\mathcal{M}$ of the manifold at \boldsymbol{q} at all times, i.e.

$$-\boldsymbol{f}_{c}(\boldsymbol{q}) \in N\boldsymbol{q}\mathcal{M},$$

where $N_{\boldsymbol{q}}\mathcal{M}$ is the annihilator space of $T_{\boldsymbol{q}}\mathcal{M}$, see Figure 7a. This concept can be extended for the inequality case, $\boldsymbol{q} \in \mathcal{S}$, simply by replacing the tangent space with the tangent cone and the annihilator with the normal cone to \mathcal{S} at \boldsymbol{q} , see Figure 7b [LAG09]:

$$-\boldsymbol{f}_{c}(\boldsymbol{q}) \in N_{\boldsymbol{q}}\mathcal{S}.$$
(64)

Notice, that the normal cone

$$N_{\boldsymbol{q}}\mathcal{S} = \left\{ \boldsymbol{\xi} \in \mathrm{AC}\left([0,T]\right)^* \mid \langle \boldsymbol{\xi}, \boldsymbol{d} \rangle \le 0 \; \forall \boldsymbol{d} \in T_{\boldsymbol{q}}\mathcal{S} \right\}$$

is a subset of the *dual space* of the space of absolutely continuous functions.

Leine, Aeberhard and Glocker use the nonsmooth version (64) of the principle of virtual work and some further mild assumptions⁴ to derive a version of Hamilton's principle of least action that is valid in the presence of inequality constraints and discontinuous velocities [LAG09]. The classical version of this principle states that the Fréchet-derivative of the action vanishes at q,

$$\delta S = \delta \int_0^T L(\boldsymbol{q}, \dot{\boldsymbol{q}}) \mathrm{d}t = 0$$
(65)

where $L(\mathbf{q}, \dot{\mathbf{q}}) = T(\mathbf{q}, \dot{\mathbf{q}}) - U(\mathbf{q})$ denotes the difference between kinetic and potential energy. In the nonsmooth case, the principle reads

$$\delta S \in N_{\boldsymbol{q}} \mathcal{S}. \tag{66}$$

If the feasible set takes up the whole space, S = AC([0, T]), the normal cone to S at any $q \in S$ contains only the zero, $N_q S = \{0\}$. In this case, (66) reappears as its classical pendant (65).

5.2.2 Forces and Accelerations are Measures

It would be convenient to have a form similar to the Lagrange equations of the first kind (14) for inequality constraints that makes the constraint forces explicit. To reach this goal at the end of Section 5.2.3, we must first accept that the constraint forces of inequality constraints are not necessarily classical functions anymore. If the velocity has a jump, its derivative at that point in time does not exist classically. We can find a weak derivative at the discontinuity point, that is a Dirac delta distribution. This Dirac delta distribution is a measure rather than a function.

Recall that the constraint forces for inequality constraints are from the dual space of absolutely continuous functions

$$-\boldsymbol{f}_{c}(\boldsymbol{q}) \in N_{\boldsymbol{q}}\mathcal{S} \subset (\mathrm{AC}\left([0,T]\right))^{*}.$$

The space of absolutely continuous functions equipped with the weak norm

$$\|\boldsymbol{q}\| = \max\left\{\sup_{t\in[0,T]}\|\boldsymbol{q}(t)\|,\sup_{t\in[0,T]}\|\dot{\boldsymbol{q}}(t)\|
ight\}$$

⁴Particularly, the feasible set S does not have to be convex, but it is required to be tangentially regular.

is a Banach space. Its dual space is given by the signed Radon measures [Nat77, Ger12], in the sense that any functional $\boldsymbol{f} \in (AC([0,T]))^*$ can be written as

$$\langle \boldsymbol{f}, \boldsymbol{x}
angle = \int_0^T \boldsymbol{x} \, \mathrm{d} \boldsymbol{p},$$

where $d\boldsymbol{p}$ is the Lebesgue–Stieltjes measure of a function of bounded variation \boldsymbol{p} . It is remarkable how impulsive forces are a direct consequence only of the principle of virtual work applied to the correct function space.

5.2.3 Existence of Lagrangian multipliers

Up to this point, S could be any tangentially regular subset of AC ([0, T]). We haven't exploited yet that the set S is defined by the inequality constraints $g(q) \geq 0$. In doing so we can characterize the problem using Lagrangian multipliers.

Before the polish mathematician Stanislav Kurcyusz died in a tragic accident at a young age in 1978, he made important contributions to optimization in Banach spaces subject to operator inequality constraints. His theory translates very nicely to the variational formulation of nonsmooth mechanics. Using the results published in [Kur76, ZK79] the following theorem can be formulated, that rewrites (66) in terms of Lagrangian multipliers [Kle15] and a complementarity condition:

Theorem 1 Let $g : \mathbb{R}^{n_q} \to \mathbb{R}^{n_{\lambda}}$ be continuously differentiable and let

$$\mathcal{S} = \{ \boldsymbol{q} \in AC([0,T]) \mid \boldsymbol{g}(\boldsymbol{q}(t)) \ge 0 \; \forall t \in [0,T] \}$$

denote the set of admissible trajectories. Let $q \in S$, let S be tangentially regular and assume that the Robinson regularity condition

$$\operatorname{im}(\boldsymbol{G}(\boldsymbol{q})) + \operatorname{span}(\boldsymbol{g}(\boldsymbol{q})) - \mathbb{R}^{n_{\lambda}}_{+} = \mathbb{R}^{n_{\lambda}}$$

holds, where $\operatorname{im}(\boldsymbol{G}(\boldsymbol{q}))$ denotes the image of the Jacobian of \boldsymbol{g} at \boldsymbol{q} , $\operatorname{span}(\boldsymbol{g}(\boldsymbol{q}))$ the space spanned by $\boldsymbol{g}(\boldsymbol{q})$ and $\mathbb{R}^{n_{\lambda}}_{+}$ the positive orthant in $\mathbb{R}^{n_{\lambda}}$. Then there exists a non-negative measure $d\boldsymbol{\lambda}$, such that

$$0 = \delta S(\delta \boldsymbol{q}) + \int_0^T \delta \boldsymbol{q}^T \boldsymbol{G}(\boldsymbol{q})^T d\boldsymbol{\lambda}$$
 (67a)

$$0 = \int_0^T \boldsymbol{g}(\boldsymbol{q})^T \mathrm{d}\boldsymbol{\lambda}.$$
 (67b)

for all variations $\delta q \in AC([0,T])$ with $\delta q(0) = \delta q(T) = 0$.

We can now bring Equation (67a) in a more recognizable form that resembles the Lagrange equations of first kind (14). Using $\delta \dot{\boldsymbol{q}} dt = d(\delta \boldsymbol{q})$, integration by parts for Lebesgue–Stieltjes measures (61) and $\delta \boldsymbol{q}(0) = \delta \boldsymbol{q}(T) = 0$, we can rewrite the Fréchet derivative $\delta S(\delta \boldsymbol{q})$ of the action S as

$$\delta S(\delta \boldsymbol{q}) = \int_{0}^{T} \delta \boldsymbol{q}^{T} \frac{\partial L}{\partial \boldsymbol{q}} dt + \int_{0}^{T} \delta \dot{\boldsymbol{q}}^{T} \frac{\partial L}{\partial \dot{\boldsymbol{q}}} dt$$
$$= \int_{0}^{T} \delta \boldsymbol{q}^{T} \frac{\partial L}{\partial \boldsymbol{q}} dt + \int_{0}^{T} \frac{\partial L}{\partial \dot{\boldsymbol{q}}}^{T} d(\delta \boldsymbol{q})$$
$$= \int_{0}^{T} \delta \boldsymbol{q}^{T} \frac{\partial L}{\partial \boldsymbol{q}} dt + \underbrace{\left[\delta \boldsymbol{q}^{T} \frac{\partial L}{\partial \dot{\boldsymbol{q}}}\right]_{0}^{T}}_{=0} - \int_{0}^{T} \delta \boldsymbol{q}^{T} d\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}\right)$$
$$= \int_{0}^{T} \delta \boldsymbol{q}^{T} \frac{\partial L}{\partial \boldsymbol{q}} dt - \int_{0}^{T} \delta \boldsymbol{q}^{T} d\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}\right). \tag{68}$$

Plugging (68) into (67a) yields

$$0 = \int_0^T \delta \boldsymbol{q}^T \frac{\partial L}{\partial \boldsymbol{q}} dt - \int_0^T \delta \boldsymbol{q}^T d\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}\right) + \int_0^T \delta \boldsymbol{q}^T \boldsymbol{G}(\boldsymbol{q})^T d\boldsymbol{\lambda}$$
(69)

or, equivalently expressed as an equality of measures,

$$d\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}\right) = \frac{\partial L}{\partial \boldsymbol{q}}dt + \boldsymbol{G}(\boldsymbol{q})^{T}d\boldsymbol{\lambda}.$$
(70)

These are the nonsmooth version of the Euler–Lagrange equations. As long as the momentum $\partial L/\partial \dot{q}$ has discontinuous jumps, i.e. is not absolutely continuous, it holds

$$\mathrm{d}\left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}}\right) \neq \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\boldsymbol{q}}} \,\mathrm{d}t.$$

Therefore we cannot derive an ordinary differential equation from the variational problem (69) using the fundamental lemma of calculus, as is usually done to arrive at the Euler-Lagrange equations. The physics can only be described variationally, in this case in the form of a measure differential equation (70).

Finally, by plugging in the Lagrange function of a multibody system

$$L(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \frac{1}{2} \dot{\boldsymbol{q}}^T \boldsymbol{M}(\boldsymbol{q}) \dot{\boldsymbol{q}} - U(\boldsymbol{q})$$

and using

$$\begin{split} \frac{\partial L}{\partial \boldsymbol{q}} &= \frac{1}{2} \dot{\boldsymbol{q}}^T \frac{\partial \boldsymbol{M}(\boldsymbol{q})}{\partial \boldsymbol{q}} \dot{\boldsymbol{q}} - \frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}}, \\ \mathrm{d} \left(\frac{\partial L}{\partial \dot{\boldsymbol{q}}} \right) &= \boldsymbol{M}(\boldsymbol{q}) \mathrm{d} \dot{\boldsymbol{q}} + \dot{\boldsymbol{q}}^T \mathrm{d} \boldsymbol{M}(\boldsymbol{q}) \\ &= \boldsymbol{M}(\boldsymbol{q}) \mathrm{d} \dot{\boldsymbol{q}} + \dot{\boldsymbol{q}}^T \dot{\boldsymbol{M}}(\boldsymbol{q}) \mathrm{d} t \\ &= \boldsymbol{M}(\boldsymbol{q}) \mathrm{d} \dot{\boldsymbol{q}} + \dot{\boldsymbol{q}}^T \frac{\partial \boldsymbol{M}(\boldsymbol{q})}{\partial \boldsymbol{q}} \dot{\boldsymbol{q}} \mathrm{d} t, \\ \boldsymbol{f}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) := -\frac{1}{2} \dot{\boldsymbol{q}}^T \frac{\partial \boldsymbol{M}(\boldsymbol{q})}{\partial \boldsymbol{q}} \dot{\boldsymbol{q}} - \frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}}, \end{split}$$

Equation (70) becomes

$$\boldsymbol{M}(\boldsymbol{q}) \,\mathrm{d}\boldsymbol{\dot{q}} = \boldsymbol{f}(\boldsymbol{q}, \boldsymbol{\dot{q}}, t) \,\mathrm{d}t + \boldsymbol{G}(\boldsymbol{q})^T \,\mathrm{d}\boldsymbol{\lambda}. \tag{71}$$

Replacing Equation (67a) in Theorem 1 with Equation (71) yields the Lagrange equations of first kind for the nonsmooth case with impacts

$$\boldsymbol{M}(\boldsymbol{q}) \,\mathrm{d}\dot{\boldsymbol{q}} = \boldsymbol{f}(\boldsymbol{q}, \dot{\boldsymbol{q}}, t) \,\mathrm{d}t + \boldsymbol{G}(\boldsymbol{q})^T \,\mathrm{d}\boldsymbol{\lambda},$$
 (72a)

$$\mathbf{0} \le \boldsymbol{g}(\boldsymbol{q}),\tag{72b}$$

$$0 \le d\lambda$$
, (72c)

$$0 = \int_0^T \boldsymbol{g}(\boldsymbol{q})^T \, \mathrm{d}\boldsymbol{\lambda}. \tag{72d}$$

The differential equation (14a) in the smooth case is replaced by the MDE (72a) in the nonsmooth version and the equality constraint (14b) turns into the complementarity conditions (72b)–(72d). Due to the positivity (72b) and (72c), (72d) means that the measure $d\lambda$ is nonzero only for subsets of [0, T] on which g is zero.

5.3 Numerical solution strategies

Nonsmooth dynamical systems have been researched rigorously since the late 1950s and many theoretical results were developed during the 1970s and 1980s. However, based on the number of publications on the subject, the development of numerical methods only slowly took up pace in the 1990s and led to a boom in the field just recently during the 2000s and 2010s.

In the 1990s, DAEs were a hot topic and many multibody simulation tools were developed. Contacts in multibody dynamical systems were usually modeled by penalizing interpenetration of rigid bodies with a stiff spring-damper element at the contact point. While this is a simple and useful model for many applications, it can be somewhat restrictive for mechanical systems with many contacts. At least two additional parameters per contact, a stiffness and a damping coefficient, have an influence on the overall behavior of the model and must be chosen appropriately. Small time steps must be used, so that a switch in contact state is not missed. Very stiff spring-damper elements can only be used together with very small time steps. As more computational power became available, industrial scale simulations of granular material became viable. The classical Discrete Element Method of Cundall and Strack [CS79] is based on a penalized contact model. Very small time steps must be used to maintain a stable simulation. Nonsmooth dynamical systems, where contacts are modeled using inequality constraints, became an attractive alternative to deal with contacts and collisions in multibody simulations. With the new approach, it is not necessary to capture the change in contact states exactly and the "infinitely stiff" character of the contact laws do not yield stiff differential systems. Much larger time steps can be used.

5.3.1 Even-driven and even-capturing methods

There is a very intuitive method for dealing with nonsmooth events when solving the equations of motion of a dynamical system. Remember, that the velocity is a function of bounded variation, and as such is smooth almost everywhere except at a countable number of discontinuity points. *Event-driven* integrators calculate the smooth trajectory of the system using available ODE or DAE solvers, until such an event is detected. The time integration stops, the event can be handled, e.g. by evaluating a contact model, and the integrator is restarted with new initial conditions until the next nonsmooth event is detected. Today, many ODE and DAE solvers integrate event-detection features. These techniques are useful if the overall number of nonsmooth events is not too large and the time points of the events can be predicted effectively.

But already the simple bouncing ball problem with a restitution coefficient 0 < e < 1 displays Zeno behavior, that is it has an accumulation point of discontinuous events in time. A rigorous event-driven time integrator would not be able to pass this accumulation point without any additional

trickery. In addition to this, there are numerous applications where the frequency of nonsmooth events in time is high and the numerical overhead of event-detection becomes the bottleneck of the entire simulation. This is especially true for dynamic simulations of granular matter, where each rigid particle is subject to frequent changes in the contact state. For this reason, this section focuses solely on time integration schemes that can step over nonsmooth events and capture the net movement during time steps, regardless of whether this movement is due to smooth motion or a nonsmooth phenomenon. Integration schemes of this type are often denoted as *event-capturing* or *time-stepping* methods.

Time–Stepping methods are commonly split into two categories, namely the Paoli–Schatzman [PS02a, PS02b] and the Moreau–Jean time–stepping schemes [JAJ98, Jea99]. The main difference between the two is that the first aims at finding a solution on position level while the latter tackle the problem on a velocity level. In both methods projections onto certain sets must be performed. In the first, projections onto arbitrary convex sets are applied, while in the latter only projection onto the normal and tangent cones to these sets play a role, which are easier to compute in general [AB08]. In the following, we will concentrate the discussion on Moreau–Jean time–stepping.

5.3.2 Nonsmooth time-stepping

Consider a discrete version of Equation (72a), that can be obtained by integrating over the time interval $[t_k, t_{k+1}]$:

$$\boldsymbol{M}(\boldsymbol{q}_k)\left(\boldsymbol{v}_{k+1} - \boldsymbol{v}_k\right) = \boldsymbol{k}_k + \boldsymbol{G}(\boldsymbol{q}_k)\boldsymbol{p}_{k+1}$$
(73)

where $h = t_{k+1} - t_k$, $\boldsymbol{q}_k \approx \boldsymbol{q}(t_k)$ and $\boldsymbol{v}_k \approx \boldsymbol{v}(t_k) := \dot{\boldsymbol{q}}(t_k)$,

$$oldsymbol{p}_{k+1} = \int_{t_k}^{t_{k+1}} \mathrm{d}oldsymbol{\lambda}$$

is an impulse that appears as a new unknown and

$$\boldsymbol{k}_k \approx \int_{t_k}^{t_{k+1}} \boldsymbol{f}(\boldsymbol{q}(\tau), \boldsymbol{v}(\tau), \tau) \,\mathrm{d}\tau$$

is an approximation of the force integral that can be obtained using a suitable quadrature rule. The generalized position of the mechanical system can be calculated for instance using a Θ -method

$$\boldsymbol{q}_{k+1} = \boldsymbol{q}_k + h(\Theta \boldsymbol{v}_{k+1} + (1 - \Theta)\boldsymbol{v}_k).$$
(74)

With $\Theta = 0$, Equation (74) corresponds to the explicit Euler, $\Theta = \frac{1}{2}$ yields the mid-point rule and $\Theta = 1$ the implicit Euler method.

For simplicity, it is assumed here, that the mass matrix \boldsymbol{M} and constraint Jacobian \boldsymbol{G} remain almost constant during a time step and that $\boldsymbol{f}(\boldsymbol{q}(\tau), \boldsymbol{v}(\tau))$ can be approximated or extrapolated for the time interval $[t_k, t_{k+1}]$. For applications in linear elasticity, \boldsymbol{f} is often linear in \boldsymbol{q} and \boldsymbol{v} , e.g.

$$\boldsymbol{f} = \boldsymbol{f}_{\text{ext}} - \boldsymbol{K}\boldsymbol{q} - \boldsymbol{D}\boldsymbol{v},$$

with stiffness and damping matrices K and D. If in this case the Θ -method is used for the approximation of the force integral k_k , no further extrapolation is needed, see for instance [AB08]. Convergence results for this time-stepping scheme are provided for instance in [MM93, Ste98, AB08].

In [SA12, SA14, KSD17] a more rigorous derivation of the discrete timestepping equations is provided. Similarly to the way finite element methods are constructed from a weak form of a partial differential equation, the authors consider equation (72) as a variational problem and apply a discontinuous Galerkin approximation in time to obtain the above mentioned time stepping equations.

The Lagrangian multiplier appears as an impulse p_{k+1} , that has the unit of momentum. No approximations for the accelerations and forces exist. The dynamics are solved directly on velocity level. The new unknown p_{k+1} does not distinguish between smooth or nonsmooth parts of the acceleration. This can potentially be a problem in applications, where particularly the loads due to contact forces are of interest [KOB13].

In order to advance the time-stepping scheme, we need to solve for p_{k+1} . The complementarity problem (72b) – (72c) for $d\lambda$ translates directly to a complementarity problem for the new unknown p_{k+1} .

$$g(q_{k+1}) \ge 0, \quad p_{k+1} \ge 0, \quad p_{k+1}^T g(q_{k+1}) = 0.$$
 (75)

In general, this is a nonlinear complementarity problem (NCP). For $\Theta = 1$, Equations (73),(74) and the NCP (75) amount to a nonsmooth version of the SHAKE integrator (44) [HLW02, HLW03]. The position \boldsymbol{q}_{k+1} is updated using the velocity \boldsymbol{v}_{k+1} , which is calculated in such a way, that the constraint is satisfied at the end of the time step from t_k to t_{k+1} .

Figure 10 shows a simulation of two-dimensional rotationless circular particles. In this simulation, the SHAKE integrator was used. The velocity is piecewise constant and the position is continuous and piecewise linear.



(e) The continuous, piecewise linear position and discontinuous, piecewise constant velocity of a single particle in vertical direction.

Figure 10: A simulation of a collapsing pile of 1000 rotationless disks in two dimensions. Figures 10a–10d show snapshots at different time points in the simulation. The images contain visualizations of the so–called *force chains* in the granular material. The vertical position and velocity of the highlighted particle is shown in Figure 10e.

5.3.3 Dealing with collisions

How can we make sure, that the constraints are satisfied at the end of the time step, if we do not use any event detection in our numerical method? There are two common strategies to do this at least approximately.

In the first strategy, we allow the constraints to be violated and assume that the time step size is small enough, so that these violations remain small. Let $\mathcal{I} = \{i_1, ..., i_n \in \mathbb{N}\}$ denote the index set of the active and the violated inequality constraints at time t_k ,

$$g_s(\boldsymbol{q}_k) \leq 0, \quad s \in \mathcal{I}.$$

For this *active set* of constraints it it clear, that the contact velocity must be non-negative by the next time step, so that the constraint is not violated any further,

$$\dot{g}_{s}(\boldsymbol{q}_{k+1}) = \frac{\partial g_{s}(\boldsymbol{q}_{k+1})}{\partial \boldsymbol{q}}^{T} \cdot \boldsymbol{v}_{k+1} \ge 0 \quad s \in \mathcal{I}.$$
(76)

Thus, with the additional assumption that the constraint Jacobian

$$oldsymbol{G}_{k,s} = rac{\partial g_s(oldsymbol{q}_k)}{\partial oldsymbol{q}} pprox rac{\partial g_s(oldsymbol{q}_{k+1})}{\partial oldsymbol{q}}$$

remains almost constant during one time step we can replace the NCP (75) with the following complementarity problem on velocity level, that is solved only for the active set of inequality constraints

$$\boldsymbol{G}_{k,s} \cdot \boldsymbol{v}_{k+1} \ge 0, \quad p_{k+1,s} \ge 0, \quad p_{k+1,s} \boldsymbol{G}_{k,s} \cdot \boldsymbol{v}_{k+1} = 0 \quad \forall s \in \mathcal{I}.$$

Here $p_{k+1,s}$ denotes the component of p_{k+1} associated to the inequality constraint with index s. Note that initial violations remain, as the reaction impulse $p_{k+1,s}$ is just large enough to not decrease the value of g_s any further.

The other strategy consists of linearizing the inequality constraint around \boldsymbol{q}_k ,

$$\boldsymbol{g}(\boldsymbol{q}_{k+1}) = \boldsymbol{g}\left(\boldsymbol{q}_{k} + h\left(\Theta\boldsymbol{v}_{k+1} + (1-\Theta)\boldsymbol{v}_{k}\right)\right)$$
$$\approx \boldsymbol{g}(\boldsymbol{q}_{k}) + h\frac{\partial\boldsymbol{g}(\boldsymbol{q}_{k})}{\partial\boldsymbol{q}}^{T}\left(\Theta\boldsymbol{v}_{k+1} + (1-\Theta)\boldsymbol{v}_{k}\right) \ge 0.$$
(77)

In this strategy we are looking ahead: We can calculate the velocity v_{k+1} in such a way, that the inequality constraint is satisfied at the end of the next time step. We do not need to separate the constraints into an active and inactive set. If the value of a linearized inequality constraint is positive and far away from zero, the NCP (75) implies that he corresponding contact impulse must be zero. Dividing the linearized inequality constraint (77) by the time step size h reveals it as an inequality constraint on velocity level with an additional stabilization term [AH04a]

$$\boldsymbol{u}_{k+1} = \frac{\boldsymbol{g}(\boldsymbol{q})}{h} + \frac{\partial \boldsymbol{g}(\boldsymbol{q}_k)}{\partial \boldsymbol{q}}^T \left(\Theta \boldsymbol{v}_{k+1} + (1-\Theta)\boldsymbol{v}_k\right) \ge 0, \quad (78)$$

very similar to Baumgarte stabilization of velocity based bilateral constraints [Bau72].

The NCP (75) is replaced by

$$u_{k+1} \ge 0$$
, $p_{k+1} \ge 0$, $p_{k+1}^T u_{k+1} = 0$.

As the constraints are considered on velocity level, these two strategies are very similar to index reduction methods for DAEs.

Both approaches amount to perfectly inelastic collisions. The unilateral constraint can be modified so that partially or perfectly elastic collisions are modeled, e.g. by incorporating Newton's impact law. In general however, we must separate the collision into a compression and decompression phase to accomplish this without introducing additional errors. This requires the solution of two complementarity problems per time step [AP97]. And even then, this only models elastic collisions between two bodies. modeling inelastic shocks through a network of contacting rigid bodies, such as for example Newton's cradle, is still a challenge in event–capturing strategies [NB18].

5.3.4 Dealing with complementarity

Since the early days of numerical methods for nonsmooth dynamical systems and regardless of which concept from Section 5.1 is used for the description of the problem, most authors formulate the discrete numerical method in terms of a complementarity problem similar to the one from the previous section [Mor66, Löt82, Löt84]. Posing the inequality constraints (76) or (78) on velocity level has a big advantage: they are linear in the velocity \boldsymbol{v}_{k+1} and therefore also in the unknown impulse p_{k+1} . Solving (73) for v_{k+1} and inserting the result into (78) reveals this,

$$egin{aligned} oldsymbol{u}_{k+1} =& rac{oldsymbol{g}(oldsymbol{q}_k)}{h} + oldsymbol{G}_k^T \left(\Thetaoldsymbol{v}_{k+1} + (1 - \Theta)oldsymbol{v}_k
ight) \ &= & rac{oldsymbol{g}(oldsymbol{q}_k)}{h} + oldsymbol{G}_k^T oldsymbol{M}_k^T oldsymbol{G}_k + \Thetaoldsymbol{M}_k^{-1}oldsymbol{G}_k oldsymbol{p}_{k+1} \ &= & oldsymbol{b}_k + oldsymbol{A}_k oldsymbol{p}_{k+1}, \end{aligned}$$

where $G_k = G(q_k)$ is the constraint Jacobian, $M_k = M(q_k)$ is the mass matrix,

$$oldsymbol{b}_k = rac{oldsymbol{g}(oldsymbol{q}_k)}{h} + oldsymbol{G}_k^T \left(oldsymbol{v}_k + \Theta oldsymbol{M}_k^{-1}oldsymbol{k}_k
ight)$$

and

$$\boldsymbol{A}_k = \Theta \boldsymbol{G}_k^T \boldsymbol{M}^{-1} \boldsymbol{G}_k$$

is the *Delassus* matrix. The NCP (75) is transformed to a *linear complementarity problem* (LCP)

$$\boldsymbol{u}_{k+1} = \boldsymbol{A}_k \boldsymbol{p}_{k+1} + \boldsymbol{b}_k \ge 0, \quad \boldsymbol{p}_{k+1} \ge 0, \quad \boldsymbol{p}_{k+1}^T \boldsymbol{u}_{k+1} = 0,$$
 (79)

which is equivalent to the quadratic program (QP)

$$\min_{\boldsymbol{p}_{k+1}} \boldsymbol{p}_{k+1}^T \boldsymbol{A}_k \boldsymbol{p}_{k+1} + \boldsymbol{p}_{k+1}^T \boldsymbol{b}_k$$

s.t. $0 \leq \boldsymbol{A}_k \boldsymbol{p}_{k+1} + \boldsymbol{b}_k,$
 $0 \leq \boldsymbol{p}_{k+1}.$ (80)

These LCPs or QPs were usually solved using Lemke's algorithm [Mur88], see for example [Löt84, Bar92, ST96, AP97, AH04b]. Lemke's algorithm a pivoting strategy for LCPs similar in spirit to Gaußian elimination. Lemke's algorithm also has a similar complexity bound as Gaußian elimination, which makes it impractical as soon as the complementarity problem becomes quite large. This is the case when many bodies are connected through a network of contacts, as for instance in granular assemblies.

5.3.5 Coulomb friction and the Painlevé paradox

Coulomb friction can be formulated as a nonlinear complementarity problem, see for instance [DF98]. The nonlinearity can be removed by approximating

the Coulomb friction cone by a polyhedral shape. The NCP for friction turns into an LCP that can be solved together with the LCP for unilateral contacts using Lemke's algorithm. Within this polyhedral approximation, the frictional force can only be exerted along a finite number of predefined directions that fan out from the contact point. This has two major disadvantages. Firstly, we introduce artificial anisotropy if we choose coarse approximations of the Coulomb friction cone. Secondly, for every possible direction of the frictional force a new unknown is introduced. As a consequence, for fine approximations of the Coulomb friction cone, the LCP becomes very large.

David Stewart uses a convergence proof of the numerical methods published in [ST96, AP97] that use a polyhedral approximation of the friction cone to resolve the much discussed Painlevé paradox [Ste98]. A detailed analysis of the paradox, first introduced in 1895, is provided in [CV16]. It consists of a simple example of a rigid body in frictional contact with a flat rigid surface. As the rigid body slides over the surface, there are instances in time where the system seems to have infinite solutions and instances where it does not seem to have any solutions. This seeming inconsistency of rigid body dynamics with Coulomb friction can be resolved by allowing shocks, i.e. impulsive forces, even when there are no impacts in the system. Stewart shows that the numerical methods of [ST96, AP97] applied to the Painlevé example converge to an impulsive solution of a measure differential inclusion that describes rigid body dynamics subject to Coulomb friction.

5.3.6 Augmented Lagrangian and projected Gauß–Seidel

It has already been established that the complexity bound for Lemke's algorithm is too restrictive for large problems. The publication [LT86] includes an alternative approach to deal with frictionless contact problems in the context of finite element methods based on an augmented Lagrangian method. This ideas was later adapted by Alart and Curnier in [AC91] to frictional contact problems. The authors formulate a variational minimization problem at each time step. A energy functional $\varphi(\mathbf{q})$ must be minimized by a trajectory \mathbf{q} without violating inequality constraints $g_i(\mathbf{q}) \geq 0, i = 1, ..., n_{\lambda}$. This variational inequality–constrained minimization problem is solved using the augmented Lagrangian method. The basic idea is to recast this as an unconstrained optimization problem. For the frictionless case, this leads to

$$\min \Phi_r(\boldsymbol{q}, \boldsymbol{\gamma}) = \varphi(\boldsymbol{q}) - \frac{1}{2r} \sum_{i=1}^{n_{\lambda}} \left(\|\gamma_i\|^2 - \operatorname{dist}^2(\gamma_i + rg_i(\boldsymbol{q}), \mathbb{R}_+) \right), \quad (81)$$

where

$$\operatorname{dist}(x, C) = x - \operatorname{prox}_C(x)$$

denotes the Euclidean distance of a point x from a set C and r > 0 is a numerical factor controlling the steepness of Φ_r . The complementarity conditions and inequality constraints are eliminated from (81). The unknown γ is an approximation of the Lagrangian multiplier associated with the inequality-constrained variational problem. A solution to the minimization problem (81) is sought by solving the saddle point problem $\nabla \Phi_r(\boldsymbol{q}, \boldsymbol{\gamma}) =$ 0. In the context of time-stepping methods, the saddle point problem is solved together with (73), where $\boldsymbol{\gamma}$ takes on the role of the reaction impulse \boldsymbol{p}_{k+1} . Alart and Curnier suggest a generalized Newton method for the solution, that boils down to the following iterative projection method [AC91, LN04, PFU06, Stu08b]. Let i = 0. Until convergence is achieved in $\boldsymbol{p}_{k+1}^{(i)}$, repeat the following steps:

- 1. $\boldsymbol{v}_{k+1} = \boldsymbol{v}_k + \boldsymbol{M}_k^{-1}(\boldsymbol{k}_k + \boldsymbol{G}_k \boldsymbol{p}_{k+1}^{(i)})$, where $\boldsymbol{p}_{k+1}^{(i)}$ is given from a previous iteration or from an initial guess.
- 2. Calculate \boldsymbol{u}_{k+1} from \boldsymbol{v}_{k+1} using (78).

3.
$$\boldsymbol{p}_{k+1}^{(i+1)} = \operatorname{prox}_{\mathbb{R}^{n_{\lambda}}_{+}}(\boldsymbol{p}_{k+1}^{(i)} - r\boldsymbol{u}_{k+1}).$$

4.
$$i \leftarrow i + 1$$
.

Note, that for any r > 0 the following equivalence holds

$$\boldsymbol{p} = \operatorname{prox}_{\mathbb{R}^{n_{\lambda}}_{\perp}}(\boldsymbol{p} - r\boldsymbol{u}) \quad \Leftrightarrow \quad \boldsymbol{p} \ge \boldsymbol{0}, \ \boldsymbol{u} \ge \boldsymbol{0}, \ \boldsymbol{p}^{T}\boldsymbol{u} = 0.$$
 (82)

In other words, if the iteration converges, it converges to a solution of the complementarity problem (79). For brevity, friction has not been considered here, details can be found in the before mentioned publications. Equation (82) is written as a vector equation. We can rewrite the projection step as

$$\boldsymbol{p}_{k+1}^{(i+1)} = \operatorname{prox}_{\mathbb{R}^{n_{\lambda}}_{+}} \left(\boldsymbol{p}_{k+1}^{(i)} - r\left(\boldsymbol{A}_{k} \boldsymbol{p}_{k+1}^{(i)} + \boldsymbol{b}_{k} \right) \right),$$

which reveals it as the Gauß–Jacobi method for the solution of linear systems of equations with an additional projection, called the projected Gauß–Jacobi method method (PGJ). Of course, given the previous iterate $\mathbf{p}_{k+1}^{(i)}$, the equation can be evaluated component–wise and the newly calculated components $j = 1, ..., l < n_{\lambda}$ can directly be used in the right–hand side of the equation for the calculation for the component with index l + 1. This method is called the projected Gauß–Seidel method (PGS). It has slightly better convergence properties than PGJ, but PGJ lends itself for a parallel implementation. The convergence of both PGJ and PGS can be controlled with a successive overrelaxation parameter α ,

$$\tilde{\boldsymbol{p}}_{k+1}^{(i+1)} = \operatorname{prox}_{\mathbb{R}_{+}^{n_{\lambda}}} \left(\boldsymbol{p}_{k+1}^{(i)} - r \left(\boldsymbol{A}_{k} \boldsymbol{p}_{k+1}^{(i)} + \boldsymbol{b}_{k} \right) \right)$$

$$\boldsymbol{p}_{k+1}^{(i+1)} = \alpha \cdot \tilde{\boldsymbol{p}}_{k+1}^{(i+1)} + (1 - \alpha) \cdot \boldsymbol{p}_{k+1}^{(i)},$$
(83)

yielding the class of *projected successive overrelaxation* (PSOR) methods.

The augmented Lagrangian method is applied to granular simulations in [FHdS02]. In [TNA08, AT10, TA11] a numerical method based on the PGJ, PGS and PSOR schemes is proposed for general cone complementarity problems. The authors use this method to solve for the normal and tangential reaction impulses involved in Coulomb friction simultaneously by directly projecting the three-dimensional contact force onto the Coulomb friction cone. They demonstrate a large number of numerical examples including the simulation of large granular assemblies and provide implementation details in a HPC context.

5.3.7 Recent developments

Over the course of the last two decades several simulation codes were developed around nonsmooth mechanical systems. It is not our intention to compile a complete list, but a few of these codes deserve to be mentioned.

Jean and Dubois initiated the open source software LMGC90 [JAJ98, DJR⁺11, LMG18] which is currently being developed at the University of Montpellier. SICONOS [Sic18] is an open source software developed at IN-RIA in Grenoble by the TRITOP team led by Vincent Acary, following the previous works of Bernard Brogliato's BIPOP team. DynamY [Stu08a] is another example for a C++ library for nonsmooth mechanical systems that was developed during the course of the PhD thesis [Stu08b] at the ETH Zürich.

Algoryx [Alg18] is a spin-off software company from Umeå University, that develops commercial software for the simulation of nonsmooth mechanical systems. Another free simulation code is the PE Rigid Body Physics Engine [Pre14, Eng18] developed at the University of Erlangen. Finally, the open source software Chrono [Chr18] is currently being developed by a large team at the University of Wisconsin and the Università di Parma by Tasora Negrut, Serban and a team of their students.

Recent research in the field of nonsmooth mechanical systems concentrates either on solving large complementarity problems in a short time, or on increasing the order of the integration methods. The first is useful for applications with many rigid bodies in close contact, as is the case for the simulation of granular material. The latter is interesting for general flexible multibody simulation scenarios with a small number of contacts. Here, the presence of unilateral contact and friction decreases the overall order of the integration methods compared to standard DAE solvers, if the Moreau– Jean or Paoli–Schatzman integrators must be used to resolve the collisions consistently.

Fast solvers for complementarity problems So far, only two strategies have been discussed to deal with the complementarity conditions. The first was the direct solution of an LCP using a pivoting strategy such as Lemke's method. The other was an augmented Lagrangian approach that uses a fix point iteration for the Lagrangian multipliers, which involves a projection onto a feasible set.

Other numerical tools from the field of continuous optimization than Lemke's method can be borrowed to solve the complementarity problems in multibody simulations. For this purpose, it makes sense to introduce two categories of iterative solvers. Consider for this the QP (80), where the objective function takes on the form

$$f(\boldsymbol{p}) = \boldsymbol{p}^T \boldsymbol{A} \boldsymbol{p} + \boldsymbol{p}^T \boldsymbol{b}.$$

The first category consists of those methods, that only require the gradient $\nabla f(\mathbf{p}) = \mathbf{A}\mathbf{p} + \mathbf{b}$ in every iteration. The second category consists of those methods, that need to solve linear systems involving the Hessian \mathbf{A} in every iteration. The intuitive expectation is, that methods from the first category have numerically cheap iterations, but superlinear convergence at best. The second category on the other hand, that includes all Newton-type solution

strategies, requires the solution of several large linear systems but has the potential for quadratic convergence.

The PGJ, PGS and PSOR solvers can be seen as methods from the first category. Other methods have been suggested as well, such as projected gradient methods and spectral methods [RBDA05, HATN13]. The numerical results obtained with Nesterov's accelerated projected gradient descent method are very promising [MHNT15].

Methods from the second category typically use so-called complementarity functions [Man76, FK98, Wan13]. A complementarity function is a function that is zero, if and only if the associated NCP is solved. This way, the complementarity problem is recast into the form of a root finding problem. Complementarity functions are typically nonsmooth and the gradient is not guaranteed to exist in general. Therefore the root finding problem must be solved using a semismooth Newton method, where in place of the gradients an arbitrary element from the subgradient of the complementarity function is taken.

In [DBDB11] a hybrid strategy is proposed. The authors use a percontact version of the Fischer–Burmeister complementarity function to model exact Coulomb friction. A semismooth Newton method per contact is used in an inner loop, while the global NCP is solved using a Gauß–Seidel–like outer loop.

An alternative strategy to overcome the nonsmoothness of the complementarity function is to approximate it using a smooth function. This is done by introducing a smoothing parameter α in such a way, that as $\alpha \to 0$, the smoothed function converges to the original complementarity function. Then a series of smooth root finding problems can be solved with a classical Newton method. Each subproblem is solved approximately, possibly with just one Newton iteration, before the smoothing parameter is decreased and the previous iterate serves as an initial guess for the next iteration. This strategy, though motivated completely differently, is very similar in spirit to interior point methods and path-following algorithms [SQ99, CCK00].

The use of Interior Point Methods (IPM) is proposed in [KLHdS12, KSO14, Kle15, MTG18]. The strategy consists of using a logarithmic version of the objective function in the QP (80) and adding a logarithmic penalty term, also called potential. The penalty term is infinitely large at the boundary (except at the exact solution) and drives the current iterate towards the interior of the feasible set. The zero-set of the potential is a smooth curve, called the central path. It can be parametrized using a parameter α in such

a way, that the smooth curve passes through the solution of the complementarity problem at $\alpha = 0$. The strategy of interior point methods is to use a Newton-iteration to step towards the central path at a given α and then decrease the parameter iteratively. This way, the solution is approaches from within a close neighborhood of the central path. A big shortcoming of IPMs is, that the condition of the linear systems involved in the Newton-steps is known to diverge as $\alpha \to 0$. A remedy for this is preconditioning of the linear systems. Typically, a block-diagonal preconditioner is already very effective. IPMs are numerically costly, as a series of linear systems must be solved. But, especially in early iterations, the linear systems must not be solved to a high accuracy, so that inexact search directions can be calculated using Krylov subspace methods with generous tolerances. In [CRZ17] a compression based direct linear solver is proposed for the linear systems arising in IPMs in the context of nonsmooth rigid body dynamics.

Comparisons of some of the recent numerical methods for complementarity problems in the context of multibody simulations are provided in [Hey13, Mel16, MFJN17, ABH18].

Towards higher order integration The Moreau–Jean method discussed so far is attractive because it integrates nonsmooth motion at a fixed time step size for an arbitrary number of contacts, while maintaining robustness. Its main shortcoming is the global accuracy of order one, even in smooth phases of motion such as free flight without impacts. If no unilateral contacts are present in a multibody simulation, standard DAE solvers can be used with favorable properties such as second order accuracy or more, unconditional stability and controlled numerical damping.

Recently, some effort has been put into finding a compromise between both worlds. This is a mixed strategy that consists of a consistent time– stepping scheme to deal with unilateral contacts and a higher order integration method for the smooth phases of a simulation [SLG08, Aca12, CAVB13, SA14, SRKA15].

The general idea is to separate the time integration into two parts for smooth and nonsmooth motion. The smooth motion is integrated using a higher order DAE-solver on velocity level, i.e. using an index-2 formulation. If a unilateral constraint $g_i(\mathbf{q}) \geq 0$ switches from inactive, $g_i(t_{k-1}) > 0$, to active, $g_i(t_k) \leq 0$, during the time step $[t_{k+1}, t_k]$, the impact equations for this unilateral constraint are solved and the smooth motion is updated by the impulse contribution of the impact.

During all time steps that contain a discontinuous velocity jump, the order cannot be expected to be higher than one. Therefore the global order of accuracy of the mixed time-stepping approach cannot exceed one. Locally, during time intervals without impacts however, the accuracy order is that of the DAE solver used for the smooth motion.

The above mentioned literature contains numerical experiments with the Newmark and Hugh–Hilbert–Taylor integrators, the generalized α –method as well as several variants of the RADAU and Lobatto Runge–Kutta methods. In the context of higher order integration methods, constraint stabilization using the Gear–Gupta–Leimkuhler algorithm [GGL85], see (22), is discussed in [SUS14, BAC14, BAC18].

Acknowledgments

We would like to thank Alessandro Tasora, Dan Negrut, and Klaus Dreßler for fruitful discussions and valuable hints on constrained mechanical systems and nonsmooth extensions.

References

- [AB08] Acary V. and Brogliato B. (2008) Numerical Methods for Nonsmooth Dynamical Systems: Applications in Mechanics and Electronics. Springer Verlag.
- [ABG08] Acary V., Brogliato B., and Goeleven D. (May 2008) Higher order moreau's sweeping process: mathematical formulation and numerical simulation. *Mathematical Programming* 113(1): 133– 217.
- [ABGT03] Ali G., Bartel A., Günther M., and Tischendorf C. (2003) Elliptic partial differential-algebraic multiphysics models in electrical network design. *Mathematical Models and Methods in Applied Sciences* 13(09): 1261–1278.
- [ABH18] Acary V., Brémond M., and Huber O. (2018) On solving contact problems with coulomb friction: formulations and numerical comparisons. In Advanced Topics in Nonsmooth Dynamics, pages 375–457. Springer.

- [AC84] Aubin J. P. and Cellina A. (1984) Differential Inclusions: Set-Valued Maps and Viability Theory. Springer-Verlag, Berlin, Heidelberg.
- [AC91] Alart P. and Curnier A. (1991) A mixed formulation for frictional contact problems prone to Newton like solution methods. *Computer Methods in Applied Mechanics and Engineering* 92(3): 353 – 375.
- [Aca12] Acary V. (2012) Higher order event capturing time-stepping schemes for nonsmooth multibody systems with unilateral constraints and impacts. *Applied Numerical Mathematics* 62(10): 1259 - 1275.
- [ACPR95] Ascher U., Chin H., Petzold L., and Reich S. (1995) Stabilization of constrained mechanical systems with DAEs and invariant manifolds. J. Mech. Struct. Machines 23: 135–158.
- [AH04a] Anitescu M. and Hart G. D. (2004) A constraint-stabilized timestepping approach for rigid multibody dynamics with joints, contact and friction. *International Journal for Numerical Methods in Engineering* 60(14): 2335 – 2371.
- [AH04b] Anitescu M. and Hart G. D. (Aug 2004) A fixed-point iteration approach for multibody dynamics with contact and small friction. *Mathematical Programming* 101(1): 3 – 32.
- [AL97] Ascher U. and Lin P. (1997) Sequential regularization methods for nonlinear higher index DAEs. *SIAM J. Scient. Comput.* 18: 160–181.
- [Alg18] Algoryx (2018) Homepage. https://www.algoryx.se/. Accessed: 2018-11-01.
- [AM98] Arnold M. and Murua A. (1998) Non-stiff integrators for differential-algebraic systems of index 2. *Numerical Algorithms* 19(1-4): 25-41.
- [AMR88] Abraham R., Marsden J. E., and Ratiu T. (1988) Manifolds, Tensor Analysis, and Applications. Springer.

- [Ant83] Antman S. S. (11 1983) The influence of elasticity on analysis: Modern developments. *Bull. Amer. Math. Soc.* (N.S.) 9(3): 267 – 291.
- [AP91] Ascher U. M. and Petzold L. R. (1991) Projected implicit Runge-Kutta methods for differential-algebraic equations. SIAM J. Numer. Anal. 28: 1097–1120.
- [AP97] Anitescu M. and Potra F. A. (1997) Formulating dynamic multirigid-body contact problems with friction as solvable linear complementarity problems. *Nonlinear Dynamics* 14: 231 – 247.
- [Arn81] Arnold V. I. (1981) Ordinary Differential Equations. MIT Press, Cambridge.
- [AT10] Anitescu M. and Tasora A. (2010) An iterative approach for cone complementarity problems for nonsmooth dynamics. *Computational Optimization and Applications* 47(2): 207 – 235.
- [BAC14] Brüls O., Acary V., and Cardona A. (November 2014) Simultaneous enforcement of constraints at position and velocity levels in the nonsmooth generalized-α scheme. Computer Methods in Applied Mechanics and Engineering 281: 131 – 161.
- [BAC18] Brüls O., Acary V., and Cardona A. (June 2018) On the Constraints Formulation in the Nonsmooth Generalized-α Method. In Publishing S. I. (ed) Advanced Topics in Nonsmooth Dynamics. Transactions of the European Network for Nonsmooth Dynamics, pages 335 – 374.
- [Bar92] Baraff D. (1992) Dynamic simulation of non-penetrating rigid bodies. PhD thesis.
- [Bau72] Baumgarte J. (1972) Stabilization of constraints and integrals of motion in dynamical systems. Comp. Meth. in Appl. Mechanics 1: 1 - 16.
- [BCP96] Brenan K. E., Campbell S. L., and Petzold L. R. (1996) The Numerical Solution of Initial Value Problems in Ordinary Differential-Algebraic Equations. SIAM, Philadelphia.

- [BDLA06] Brogliato B., Daniilidis A., Lemarechal C., and Acary V. (2006) On the equivalence between complementarity systems, projected systems and unilateral differential inclusions. Systems and Control Letters 55(1): 45 – 51.
- [BF91] Brezzi F. and Fortin M. (1991) Mixed and Hybrid Finite Element Methods. Springer, New York.
- [BH93] Brasey V. and Hairer E. (1993) Half-explicit Runge–Kutta methods for differential-algebraic systems of index 2. *SIAM J. Numer. Anal.* 30: 538–552.
- [BLMV15] Benner P., Losse P., Mehrmann V., and Voigt M. (2015) Numerical linear algebra methods for linear differential-algebraic equations. In: Surveys in Differential-Algebraic Equations III, A. Ilchmann and T. Reis (Eds.), Springer DAE-Forum, pages 117–175.
- [Bri08] Brizard A. (2008) An Introduction to Lagrangian Mechanics. World Scientific Publishing.
- [Bro16] Brogliato B. (2016) Nonsmooth Mechanics Models, Dynamics and Control. Springer Verlag.
- [Cam82] Campbell S. L. (1982) Singular Systems of Differential Equations II. Research Notes in Mathematics 61. Pitman.
- [Cam93] Campbell S. L. (1993) Least squares completions for nonlinear differential-algebraic equations. *Numer. Math.* 65: 77–94.
- [CAVB13] Chen Q.-Z., Acary V., Virlez G., and Brüls O. (2013) A nonsmooth generalized- scheme for flexible multibody systems with unilateral constraints. *International Journal for Numerical Methods in Engineering* 96(8): 487 – 511.
- [CCK00] Chen B., Chen X., and Kanzow C. (2000) A penalized Fischer-Burmeister NCP-function. *Mathematical Programming* 88(1): 211 – 216.
- [CG95] Campbell S. and Gear C. (1995) The index of general nonlinear DAEs. *Numer. Math.* 72: 173–196.

- [Chr18] Chrono P. (2018) C++ library for multi-physics simulation. https://github.com/projectchrono/chrono. Accessed: 2018-11-01.
- [CR08] Callies R. and Rentrop P. (2008) Optimal control of rigid-link manipulators by indirect methods. *GAMM-Mitteilungen* 31(1): 27–58.
- [CRZ17] Corona E., Rahimian A., and Zorin D. (2017) A tensor-train accelerated solver for integral equations in complex geometries. *Journal of Computational Physics* 334: 145 – 169.
- [CS79] Cundall P. A. and Strack O. D. L. (1979) A discrete numerical model for granular assemblies. *Géotechnique* 29(1): 47 65.
- [CV16] Champneys A. R. and Várkonyi P. L. (2016) The painlevé paradox in contact mechanics. *IMA Journal of Applied Mathematics* 81(3): 538–588.
- [DBDB11] Daviet G., Bertails-Descoubes F., and Boissieux L. (2011) A hybrid iterative solver for robustly capturing Coulomb friction in hair dynamics. ACM Transactions on Graphics (TOG) Proceedings of ACM 30(6).
- [DF98] DeSaxcé G. and Feng Z.-Q. (1998) The bipotential method: A constructive approach to design the complete contact law with friction and improved numerical algorithms. *Mathematical and Computer Modelling* 28(4 8): 225 245. Recent Advances in Contact Mechanics.
- [DJR⁺11] Dubois F., Jean M., Renouf M., Mozul R., Martin A., and Bagnéris M. (May 2011) LMGC90. In *10e colloque national en* calcul des structures, page Clé USB. Giens, France.
- [Eic93] Eich E. (1993) Convergence results for a coordinate projection method applied to constrained mechanical systems. SIAM J. Numer. Anal. 30(5): 1467–1482.
- [Eng18] Engine P. P. (2018) Homepage. https://www.cs10.tf.fau.de/ research/software/pe/. Accessed: 2018-11-01.

- [Erd77] Erdmann G. (1877) Ueber unstetige lösungen in der variationsrechnung. Journal für die reine und angewandte Mathematik 82: 21–30.
- [FHdS02] Fortin J., Hjiaj M., and de Saxc G. (2002) An improved discrete element method based on a variational formulation of the frictional contact law. *Computers and Geotechnics* 29(8): 609 – 640.
- [Fil88] Filippov A. F. (1988) Differential equations with discontinuous righthand sides: control systems, volume 18. Springer Science & Business Media.
- [FK98] Ferris M. C. and Kanzow C. (1998) Complementarity and related problems: A survey. Mathematical Programming Technical Report 17.
- [FL91] Führer C. and Leimkuhler B. (1991) Numerical solution of differential-algebraic equations for constrained mechanical motion. Numer. Math., 59: 55–69.
- [FMOW03] Fetecau R. C., Marsden J. E., Ortiz M., and West M. (2003) Nonsmooth Lagrangian mechanics and variational collision integrators. SIAM Journal on Applied Dynamical Systems 2(3): 381 - 416.
- [Gan59] Gantmacher F. (1959) Matrizenrechnung, Teil 2. VEB Deutscher Verlag der Wissenschaften, Berlin.
- [GB94] Garcia de Jalón J. and Bayo E. (1994) *Kinematic and Dynamic* Simulation of Multibody Systems. Springer.
- [Gea71a] Gear C. (1971) Numerical Initial Value Problems in Ordinary Differential Equations. Prentice-Hall.
- [Gea71b] Gear C. (1971) Simultaneous numerical solution of differentialalgebraic equations. *IEEE Trans. Circuit Theory* CT-18(1): 89 – 95.
- [Gea88] Gear C. W. (1988) Differential-algebraic equation index transformation. SIAM J. Sci. & Statist. Comp. 9: 39–47.

- [Gea90] Gear C. W. (1990) Differential-algebraic equations, indices, and integral algebraic equations. *SIAM J. Numer. Anal.* 27: 1527– 1534.
- [Ger12] Gerdts M. (2012) Optimal Control of ODEs and DAEs. De Gruyter.
- [GF99] Günther M. and Feldmann U. (1999) CAD based electric circuit modeling in industry I: mathematical structure and index of network equations. *Surv. Math. Ind.* 8: 97–129.
- [GGL85] Gear C., Gupta G., and Leimkuhler B. (1985) Automatic integration of the Euler-Lagrange equations with constraints. J. Comp. Appl. Math. 12 & 13: 77–90.
- [GHR00] Günther M., Hoschek M., and Rentrop P. (2000) Differentialalgebraic equations in electric circuit simulation. Int. J. Electron. Commun. 54: 101–107.
- [Glo00] Glocker C. (2000) Set-Valued Force Laws in Rigid Body Dynamics. Habilitation, Lehrstuhr für Mechanik, Technische Universität München.
- [GM86] Griepentrog E. and März R. (1986) Differential-Algebraic Equations and Their Numerical Treatment. Teubner-Texte zur Mathematik No. 88. Teubner Verlagsgesellschaft, Leipzig.
- [Gün01] Günther M. (2001) Partielle differential-algebraische Systeme in der numerischen Zeitbereichsanalyse elektrischer Schaltungen. VDI-Verlag, Reihe 20, Düsseldorf.
- [Han90] Hanke M. (1990) On the regularization of index 2 differentialalgebraic equations. J. Math. Anal. Appl. 151(1): 236–253.
- [HATN13] Heyn T., Anitescu M., Tasora A., and Negrut D. (2013) Using Krylov subspace and spectral methods for solving complementarity problems in many-body contact dynamics simulation. International Journal for Numerical Methods in Engineering 95(7): 541 – 561.

- [Hau89] Haug E. (1989) Computer-Aided Kinematics and Dynamics of Mechanical Systems. Allyn and Bacon, Boston.
- [HBG⁺05] Hindmarsh A. C., Brown P. N., Grant K. E., Lee S. L., Serban R., Shumaker D. E., and Woodward C. S. (2005) Sundials: Suite of nonlinear and differential/algebraic equation solvers. ACM Transactions on Mathematical Software 31(3): 363–396.
- [Hey13] Heyn T. (2013) On the Modeling, Simulation, and Visualization of Many-Body Dynamics Problems with Friction and Contact. PhD thesis, Uiversity of Wisconsin-Madison.
- [HLR88] Hairer E., Lubich C., and Roche M. (1988) Error of Runge-Kutta methods for stiff problems studied via differential algebraic equations. *BIT Numerical Mathematics* 28(3): 678–700.
- [HLR89] Hairer E., Lubich C., and Roche M. (1989) The Numerical Solution of Differential-Algebraic Equations by Runge-Kutta Methods. Lecture Notes in Mathematics Vol. 1409. Springer, Heidelberg.
- [HLW02] Hairer E., Lubich C., and Wanner G. (2002) *Geometric Numerical Integration*. Springer-Verlag, Berlin.
- [HLW03] Hairer E., Lubich C., and Wanner G. (2003) Geometric numerical integration illustrated by the Störmer–Verlet method. *Acta Numerica* 12: 399 – 450.
- [HMS02] Higham D. J., Mao X., and Stuart A. M. (2002) Strong convergence of Euler-type methods for nonlinear stochastic differential equations. SIAM Journal on Numerical Analysis 40(3): 1041– 1063.
- [HNW93] Hairer E., Nørsett S., and Wanner G. (1993) Solving Ordinary Differential Equations I: Nonstiff Problems. Springer-Verlag, Berlin.
- [HW96] Hairer E. and Wanner G. (1996) Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems. Springer-Verlag, Berlin.

- [HW99] Hairer E. and Wanner G. (1999) Stiff differential equations solved by Radau methods. J. Comp. Appl. Math. 111: 93–111.
- [JAJ98] Jourdan F., Alart P., and Jean M. (1998) A gauss-seidel like algorithm to solve frictional contact problems. *Computer Methods* in Applied Mechanics and Engineering 155(1): 31 – 47.
- [Jea99] Jean M. (1999) The non-smooth contact dynamics method. Computer Methods in Applied Mechanics and Engineering 177(3 - 4): 235 - 257.
- [Kar71] Karamardian S. (1971) Generalized complementarity problem. Journal of Optimization Theory and Applications 8: 161 – 168.
- [Kir47] Kirchhoff G. (1847) Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird. Annalen der Physik 148(12): 497–508.
- [KKBS04] Körkel S., Kostina E., Bock H. G., and Schlöder J. P. (2004) Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. *Optimization Methods and Software* 19(3-4): 327–338.
- [Kle15] Kleinert J. (2015) Simulating Granular Material using Nonsmooth Time-Stepping and a Matrix-free Interior Point Method. PhD thesis, Technische Universität Kaiserslautern.
- [KLHdS12] Krabbenhoft K., Lyamin A. V., Huang J., and da Silva M. V. (2012) Granular contact dynamics using mathematical programming methods. *Computers and Geotechnics* 43(0): 165 – 176.
- [KM06] Kunkel P. and Mehrmann V. (2006) Differential-Algebraic Equations – Analysis and Numerical Solution. EMS Publishing House.
- [KOB13] Kleinert J., Obermayr M., and Balzer M. (2013) Modeling of large scale granular systems using the discrete element method and the non-smooth contact dynamics method: A comparison. In Proceedings of the ECCOMAS Multibody Dynamics Conference, Zagreb.

- [Kro90] Kronecker L. (1890) Algebraische Reduktion der Schaaren bilinearer Formen. Akademie der Wissenschaften Berlin III: 141–155.
- [KSD17] Kleinert J., Simeon B., and Dreßler K. (2017) Nonsmooth contact dynamics for the large-scale simulation of granular material. *Journal of Computational and Applied Mathematics* 316: 345 – 357. Selected Papers from NUMDIFF-14.
- [KSO14] Kleinert J., Simeon B., and Obermayr M. (2014) An inexact interior point method for the large-scale simulation of granular material. Computer Methods in Applied Mechanics and Engineering 278(0): 567 – 598.
- [Kur76] Kurcyusz S. (1976) On the existence and nonexistence of Lagrange multipliers in banach spaces. *Journal of Optimization Theory and Applications* 20(1): 81 – 110.
- [Lag88] Lagrange J. L. (1788) *Méchanique analytique*. Libraire chez la Veuve Desaint, Paris.
- [LAG09] Leine R. I., Aeberhard U., and Glocker C. (2009) Hamilton's principle as variational inequality for mechanical systems with impact. *Journal of Nonlinear Science* 19(6): 633 664.
- [Lan13] Lang J. (2013) Adaptive multilevel solution of nonlinear parabolic *PDE systems: theory, algorithm, and applications, volume 16.* Springer Science & Business Media.
- [Lee74] Leela S. (1974) Stability of measure differential equations. *Pa-cific J. Math.* 55(2): 489 498.
- [LENP95] Lubich C., Engstler C., Nowak U., and Pöhle U. (1995) Numerical integration of constrained mechanical systems using MEXX. *Mech. Struct. Mach.* 23: 473–495.
- [LMG18] LMGC90 (2018) Simulation platform. https://git-xen.lmgc. univ-montp2.fr/lmgc90/lmgc90_user. Accessed: 2018-11-01.
- [LMT13] Lamour R., März R., and Tischendorf C. (2013) Differential-Algebraic Equations: A Projector Based Analysis (Differential-Algebraic Equations Forum). Springer.

- [LN04] Leine R. I. and Nijmeijer H. (2004) Bifurcations of equilibria in non-smooth continuous systems. In *Dynamics and Bifurcations* of Non-Smooth Mechanical Systems, pages 125–176. Springer.
- [Löts2] Lötstedt P. (1982) Mechanical systems of rigid bodies subject to unilateral constraints. *SIAM J. Appl. Math.* 42: 281–296.
- [Löt84] Lötstedt P. (1984) Numerical simulation of time-dependent contact and friction problems in rigid body mechanics. *SIAM journal on scientific and statistical computing* 5(2): 370 – 393.
- [LP86] Lötstedt P. and Petzold L. (1986) Numerical solution of nonlinear differential equations with algebraic constraints i: Convergence results for BDF. *Math.Comp.* 46: 491–516.
- [LT86] Landers J. A. and Taylor R. L. (1986) An augmented alagrangian formulation for the finite element solution of contact problems. Technical report, CALIFORNIA UNIV BERKELEY DEPT OF CIVIL ENGINEERING.
- [Lub89] Lubich C. (1989) h^2 extrapolation methods for differentialalgebraic equations of index-2. *Impact Comp. Sci. Eng.* 1: 260– 268.
- [Lub93] Lubich C. (1993) Integration of stiff mechanical systems by Runge-Kutta methods. ZAMP 44: 1022–1053.
- [Man76] Mangasarian O. L. (1976) Equivalence of the complementarity problem to a system of nonlinear equations. *SIAM Journal on Applied Mathematics* 31(1): 89 – 92.
- [Mel16] Melanz D. J. (2016) Physics-based contact using the complementarity approach for discrete element applications in vehicle mobility and terramechanics. PhD thesis, The University of Wisconsin-Madison.
- [MFJN17] Melanz D., Fang L., Jayakumar P., and Negrut D. (2017) A comparison of numerical methods for solving multibody dynamics problems with frictional contact modeled via differential variational inequalities. *Computer Methods in Applied Mechanics and Engineering* 320: 668 – 693.

- [MHNT15] Mazhar H., Heyn T., Negrut D., and Tasora A. (May 2015) Using nesterov's method to accelerate multibody dynamics with friction and contact. *ACM Trans. Graph.* 34(3): 1 – 14.
- [MM93] Monteiro Marques M. (1993) Differential Inclusions in Nonsmooth Mechanical Problems: Shocks and Dry Friction. Progress in Nonlinear Differential Equations and Their Applications. Birkhäuser Basel.
- [Mor66] Moreau J. J. (1966) Quadratic Programming in Mechanics: Dynamics of One-Sided Constraints. SIAM Journal on Control and Optimization 4(1): 153 – 158.
- [Mor77] Moreau J. J. (1977) Evolution problem associated with a moving convex set in a Hilbert space. *Journal of Differential Equations* 26(3): 347 – 374.
- [Mor88a] Moreau J. J. (1988) Bounded variation in time. In Moreau J. J., Panagiotopoulos P. D., and Strang G. (eds) Topics in nonsmooth mechanics, pages 1 – 74. Birkhäuser Verlag.
- [Mor88b] Moreau J. J. (1988) Unilateral Contact and Dry Friction in Finite Freedom Dynamics, pages 1 – 82. Springer Vienna, Vienna.
- [Mor93] Moreau J. J. (1993) Some numerical methods in multibody dynamics: application to granular materials. *European Journal of Mechanics-A/Solids* 13(4-suppl): 93 – 114.
- [Mor99] Moreau J.-J. (1999) Numerical aspects of the sweeping process. Computer methods in applied mechanics and engineering 177: 329 - 349.
- [MP90] Moreau J. J. and Panagiotopoulos P. D. (1990) Nonsmooth mechanics and applications. ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik 70(10): 472 – 472.
- [MT97] März R. and Tischendorf C. (1997) Recent results in solving index-2 differential-algebraic equations in circuit simulation. *SIAM J. Sci. Comp.* 18: 139–159.

- [MTG18] Mangoni D., Tasora A., and Garziera R. (2018) A primaldual predictorcorrector interior point method for non-smooth contact dynamics. Computer Methods in Applied Mechanics and Engineering 330: 351 – 367.
- [Mur88] Murty K. G. (1988) Linear Complementarity, Linear and Nonlinear Programming. Sigma series in applied mathematics. Heldermann Verlag.
- [Nat77] Natanson I. P. (1977) Theorie der Funktionen einer reellen Veränderlichen, 4th edition. Harri Deutsch, Zürich.
- [NB18] Nguyen N. S. and Brogliato B. (2018) Comparisons of Multiple-Impact Laws For Multibody Systems: Moreau's Law, Binary Impacts, and the LZB Approach, pages 1 – 45. Springer International Publishing, Cham.
- [NP73] Nagel L. W. and Pederson D. (1973) Spice (simulation program with integrated circuit emphasis). Technical Report UCB/ERL M382, EECS Department, University of California, Berkeley.
- [NZ95] Nagurney A. and Zhang D. (1995) Projected Dynamical Systems and Variational Inequalities with Applications. Innovations in Financial Markets and Institutions. Springer US.
- [O'M74] O'Malley R. E. (1974) Introduction to Singular Perturbations. Academic Press, New York.
- [Pao05] Paoli L. (2005) An existence result for non-smooth vibro-impact problems. Journal of Differential Equations 211(2): 247 – 281.
- [PD82] Pandit S. and Deo S. (1982) Differential Systems Involving Impulses. Lecture Notes in Mathematics. Springer Berlin Heidelberg.
- [Pet82] Petzold L. (1982) A description of DASSL: A differential/algebraic system solver. In Proc. 10th IMACS World Congress, August 8-13 Montreal 1982.
- [Pfe03] Pfeiffer F. (Jun 2003) The idea of complementarity in multibody dynamics. Archive of Applied Mechanics 72(11): 807 816.

- [PFU06] Pfeiffer F., Foerg M., and Ulbrich H. (2006) Numerical aspects of non-smooth multibody dynamics. Computer Methods in Applied Mechanics and Engineering 195(50 - 51): 6891 - 6908.
- [PG00] Panagiotopoulos P. D. and Glocker C. (2000) Inequality constraints with elastic impacts in deformable bodies. the convex case. Archive of Applied Mechanics 70(5): 349 – 365.
- [Pic17] Piccoli B. (2017) Measure differential equations. arXiv preprint arXiv:1708.09738.
- [Pre14] Preclik T. (2014) Models and Algorithms for Ultrascale Simulations of Non-smooth Granular Dynamics. doctoralthesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU).
- [PS02a] Paoli L. and Schatzman M. (2002) A numerical scheme for impact problems i: The one-dimensional case. *SIAM Journal on Numerical Analysis* 40(2): 702–733.
- [PS02b] Paoli L. and Schatzman M. (2002) A numerical scheme for impact problems ii: The multidimensional case. *SIAM journal on numerical analysis* 40(2): 734–768.
- [PS08] Pang J.-S. and Stewart D. E. (2008) Differential variational inequalities. *Mathematical Programming* 113: 345 – 424.
- [RBDA05] Renouf M., Bonamy D., Dubois F., and Alart P. (2005) Numerical simulation of two-dimensional steady granular flows in rotating drum: On surface flow rheology. *Physics of Fluids* 17(10).
- [Rei90] Reich S. (1990) On a geometric interpretation of DAEs. *Circ.* Syst., Sig. Processing 9: 367–382.
- [Rhe84] Rheinboldt W. (1984) Differential algebraic systems as differential equations on manifolds. *Math. Comp.* 43(168): 2473–482.
- [Rhe96] Rheinboldt W. (1996) Manpak: A set of algorithms for computations on implicitly defined manifolds. *Computers and Math. Applic.* 32: 15–28.

- [RR02] Rabier P. and Rheinboldt W. (2002) Theoretical and numerical analysis of differential-algebraic equations. In Ciarlet P. and Lions J. (eds) Handbook of Numerical Analysis, Volume VIII. Elsevier, Amsterdam.
- [RRS89] Rentrop P., Roche M., and Steinebach G. (1989) The application of Rosenbrock-Wanner type methods with stepsize control in differential-algebraic equations. *Num. Math.* 55: 545–563.
- [RS07] Reis T. and Stykel T. (2007) Stability analysis and model order reduction of coupled systems. *Mathematical and Computer Modelling of Dynamical Systems* 13(5): 413–436.
- [RW98] Rockafellar R. T. and Wets R. J.-B. (1998) Variational Analysis. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg.
- [SA12] Schindler T. and Acary V. (2012) Application of timestepping schemes based on time discontinuous galerkin methods to multidimensional examples. In Euromech 514-New trends in Contact Mechanics.
- [SA14] Schindler T. and Acary V. (2014) Timestepping schemes for nonsmooth dynamics based on discontinuous Galerkin methods: Definition and outlook. *Mathematics and Computers in Simulation* 95(0): 180 – 199.
- [Sch90] Schiehlen W. (ed) (1990) *Multibody System Handbook*. Springer, Heidelberg.
- [Sch99] Schwerin R. (1999) Multibody System Simulation. Springer, Berlin.
- [Sic18] Siconos (2018) Simulation framework for nonsmooth dynamical systems. https://github.com/siconos/siconos. Accessed: 2018-11-01.
- [Sim13] Simeon B. (2013) Computational Flexible Multibody Dynamics: A Differential-Algebraic Approach. Springer.
- [SLG08] Studer C., Leine R., and Glocker C. (2008) Step size adjustment and extrapolation for time-stepping schemes in non-smooth dynamics. International Journal for Numerical Methods in Engineering 76(11).
- [SQ99] Sun D. and Qi L. (1999) On NCP-functions. Computational Optimization and Applications 13(1-3): 201-220.
- [SRKA15] Schindler T., Rezaei S., Kursawe J., and Acary V. (2015) Halfexplicit timestepping schemes on velocity level based on timediscontinuous galerkin methods. *Computer Methods in Applied Mechanics and Engineering* 290: 250 – 276.
- [ST96] Stewart D. E. and Trinkle J. C. (1996) An implicit time-stepping scheme for rigid body dynamics with inelastic collisions and Coulomb friction. *International Journal for Numerical Methods* in Engineering 39(15): 2673 – 2691.
- [Ste98] Stewart D. E. (1998) Convergence of a time-stepping scheme for rigid-body dynamics and resolution of Painlevé's problem. *Archive for Rational Mechanics and Analysis* 145(3): 215 – 260.
- [Ste01] Stewart D. (2001) Reformulations of measure differential inclusions and their closed graph property. Journal of Differential Equations 175(1): 108 - 129.
- [Stu08a] Studer C. (2008) C++ class library for the simulation of mechanical systems with unilateral contacts and friction. http: //www.zfm.ethz.ch/dynamY/. Accessed: 2018-11-01.
- [Stu08b] Studer C. W. (2008) Augmented time-stepping integration of non-smooth dynamical systems. PhD thesis, ETH Zürich.
- [SUS14] Schoeder S., Ulbrich H., and Schindler T. (2014) Discussion of the gear–gupta–leimkuhler method for impacting mechanical systems. *Multibody System Dynamics* 31(4): 477 – 495.
- [TA11] Tasora A. and Anitescu M. (2011) A matrix-free cone complementarity approach for solving large-scale, nonsmooth, rigid body dynamics. *Computer Methods in Applied Mechanics and Engineering* 200: 439 – 453.

- [TNA08] Tasora A., Negrut D., and Anitescu M. (2008) Large-scale parallel multi-body dynamics with frictional contact on the graphical processing unit. *Proceedings of the Institution of Mechanical En*gineers, Part K: Journal of Multi-body Dynamics 222(4): 315 – 326.
- [Wan13] Wanzke C. (2013) Partikelsimulation mit halbglatten Newton-Verfahren. Master's thesis, Technische Universität Kaiserslautern.
- [Wei68] Weierstrass K. (1868) Zur Theorie der bilinearen und quadratischen Formen. Monatsber. Akad. Wiss. Berlin pages 310–338.
- [Win03] Winkler R. (2003) Stochastic differential algebraic equations of index 1 and applications in circuit simulation. *Journal of computational and applied mathematics* 157(2): 477–505.
- [ZK79] Zowe J. and Kurcyusz S. (1979) Regularity and stability for the mathematical programming problem in Banach spaces. *Applied Mathematics and Optimization* 5(1): 49 – 62.

Index

augmented Lagrangian method, 52 stochastic, 26 BDF methods, 5 Euler-Lagrange equations, 11 closed kinematic loop, 11 Hamiltonian system, 22 complementarity dynamic system, 32 index complementarity problem differential, 5 linear, 51 nilpotency, 8 nonlinear, 47 perturbation, 15 cone convex, 33 kinetic energy, 12, 23 dual. 33 normal, 30 Lagrange polar, 30 function, 22 tangent, 29 multiplier, 12, 23 cone complementarity problem, 33 Lebesgue–Stieltjes measure, 36 conservative system, 22 local parametrization, 19 consistent initial value, 9 mass matrix, 11 constraint measure differential equation, 37 at acceleration level, 13 Moreau's sweeping process, 28 at position level, 10 multibody system, 12 Jacobian, 11 manifold, 19 potential energy, 23 coordinates projected dynamical system, 30 minimal, 11 projected Gauß–Seidel method, 54 projected Gaus–Jacobi method, 54 density projected successive overrelaxation, 54 Radon–Nikodym, 36 projection methods, 18 derivative array, 22 differential equation on manifold, 18 quadratic program, 51 differential inclusion, 30 differential, 38 Rosenbrock methods, 21 differential-algebraic equation, 2, 3, 5 Runge-Kutta methods, 16 implicit, 5, 6, 15 singularly perturbed system, 21 linear constant coefficient, 6 stabilized index-2 system, 14 semi-explicit, 19, 21

state space form, 20

tangent space, 19 transient saddle point problem, 25

unilateral constraint, 28

variational inequality, 31 differential, 33 of evolution, 34

Weierstrass canonical form, 8