Mechanical systems with frictional contact: Geometric theory and time discretization methods

Von der Fakultät Konstruktions-, Produktions- und Fahrzeugtechnik
der Universität Stuttgart zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von
Giuseppe Capobianco
aus Vaduz (FL)

Hauptberichter:	Prof. Dr. ir. habil. Remco I. Leine
Mitberichter:	Prof. Dr. Olivier Brüls
Tag der mündlichen Prüfung:	20. August 2021

Institut für Nichtlineare Mechanik
der Universität Stuttgart

2021

Alla mia famiglia

Preface

The work presented in this thesis has been carried out at the Institute for Nonlinear Mechanics of the University of Stuttgart. I would like to thank Prof. Dr. Remco I. Leine for giving me the possibility and the freedom to pursue my research interests during the last six years.

I am sincerely grateful to Dr. Simon R. Eugster for supporting and guiding my research. Given your effort and invaluable contribution to this work, you would have deserved to be an additional co-referee.[1] Thank you for your confidence in me and for asking me to join you in Stuttgart. I loved working closely with you and enjoyed our extensive discussions, which have shaped my understanding of mechanics in great measure. I am very grateful to my former colleague Dr. Tom Winandy. I could not have imagined a better companion for tackling the differential geometric description of mechanics. A special thanks goes to Jonas Harsch for sharing his knowledge during our numerous joint programming sessions.

I want to thank my co-referee Prof. Dr. Olivier Brüls for his interest in my work and for reviewing my thesis. It was your talk at the symposium of the European Network for Nonsmooth Dynamics in Stuttgart that sparked my interest in the nonsmooth generalized-α schemes.

I am very grateful to Prof. Dr. Christoph Glocker for his interest in my work, his suggestions and the many interesting discussions. Your well structured and comprehensive lectures provided a solid basis for my research.

For sharing good times together and for their support I thank my colleagues of the Institute for Nonlinear Mechanics. Further thanks go to all the students who did their bacholor or master thesis under my supervision.

Last but not least, I want to thank my family and friends for all their love, support and encouragement.

Stuttgart, May 2021 *Giuseppe Capobianco*

[1] Irrespective of his expertise and his efforts, the prevailing doctoral regulations (Promotionsordnung 2011) of the University of Stuttgart do not allow him to serve as a co-referee (Mitberichter) of the present work.

Contents

Preface **i**

Abstract **v**

Zusammenfassung **vii**

1 Introduction **1**
1.1 Motivation 1
1.2 Objectives and Literature Review 2
1.3 Main Contributions and Outline 11

2 Miscellaneous Topics from Linear Algebra **15**
2.1 Basic Notations 15
2.2 Euclidean Vector Space $\mathbb{E}^n$ 16
2.3 Tensors 19
2.4 Alternating Forms 21
2.5 Convex Cones 23

3 Introduction to Differential Geometry **27**
3.1 Preliminaries from Multivariate Calculus 27
3.2 Smooth Manifolds 28
3.3 Tangent and Cotangent Space 30
3.4 Vector Bundles 36
3.5 Vector and Tensor Fields 37
3.6 Vector Fields as Dynamical Systems 41
3.7 Lie Derivative 42
3.8 Differential Forms 43

4 Geometric Foundations of Mechanics **47**
4.1 Distributions 47
4.2 Connections and Covariant Derivatives 53
4.3 Structures on the Tangent Bundle 57

4.4 Galilean Manifolds 59

5 Dynamics of Mechanical Systems on Galilean Manifolds 69
5.1 On Space, Time and Space-time 69
5.2 Classical Space-time 71
5.3 Configuration Space-time and State Space 77
5.4 Virtual Displacements 81
5.5 Inertial Motion 87
5.6 Forces 89
5.7 Principle of Virtual Work 97
5.8 Lagrange's Equations of the Second Kind 100
5.9 Principle of Virtual Action and Hamilton's Principle 102
5.10 Bilateral Constraints 104

6 Generalization to Nonsmooth Mechanics 113
6.1 Nonsmooth Motions and the Principle of Virtual Action 113
6.2 Unilateral Constraints 117
6.3 Mechanical Systems with Frictional Contact 123

7 Time Finite Element based Moreau-type Integrators 129
7.1 Time Finite Element Discretization 129
7.2 Velocity Level Constraints 132
7.3 Discrete Contact Laws 133
7.4 Numerical Implementation 134

8 Nonsmooth Generalized-α Scheme 137
8.1 Acceleration Level Constraints and Stabilization 137
8.2 Nonsmooth Generalized-α Discretization 139
8.3 Discrete Normal Contact Laws 142
8.4 Discrete Friction Laws 146
8.5 Extension to General Velocities 148
8.6 Numerical Implementation 149
8.7 Examples 155

9 Conclusions and Outlook 165

A Normal Cone Inclusion Problems 169

Bibliography 173

Abstract

This dissertation deals with the mathematical description and the simulation of mechanical systems with frictional contact. First, a geometric theory for the description of smooth mechanical systems is developed, which is then extended to allow for nonsmooth motions, i.e., motions with discontinuous velocities. The developed nonsmooth theory of mechanics is used to describe mechanical systems with frictional contact. Finally, two numerical schemes for the simulation of such systems are derived by using a time finite element method and the generalized-α approach, respectively.

The presented geometric theory describes the dynamics of time-dependent finite-dimensional mechanical systems and is based on Galilean manifolds, which are used to model both the physical and the configuration space-time of the system. The theory for smooth systems is based on three postulates, which resemble Newton's laws of motion. Starting from these postulates, the principles of mechanics are derived. A particular effort is put into retrieving the principles and formalisms known in engineering mechanics as local representations of the fundamental principles of the geometric theory. In that spirit, it is shown that the developed theory for smooth systems for instance comprises the principle of virtual work, Lagrange's equations of the first and second kind, and the principle of virtual action. Moreover, the geometric treatment of bilateral constraints is used to derive conditions, which allow to assert if a set of time-dependent bilateral constraints is holonomic. These conditions are stated in vector-matrix notation and are hence directly accessible to engineering mechanics.

The generalization to a nonsmooth mechanical theory is achieved by relaxing the continuity assumptions of the motion and by generalizing the definition of forces. It is shown that the equality of measures, which is typically used for the description of nonsmooth mechanical systems, results as a local representation of the geometric theory. Finally, the geometric treatment of unilateral constraints as well as the introduction of set-valued Coulomb-type friction laws give access to the description of mechanical systems with frictional contact, which is then discretized.

The first numerical scheme, derived using a time finite element method, extends the existing symmetric Moreau-type integrator by including also bilateral constraints. The second scheme represents an extension of the established nonsmooth generalized-α method for the simulation of mechanical systems with frictionless contacts. In contrast to the existing method, the unilateral constraints are complemented with discrete set-valued Coulomb-type friction laws such that the scheme can cope with mechanical systems with frictional contact.

This dissertation presents the theoretical foundations for the description of mechanical systems with and without frictional contact and provides a solid basis for the development of novel time integration schemes.

Zusammenfassung

Diese Dissertation befasst sich mit der mathematischen Beschreibung sowie der Simulation von mechanischen Systemen mit Reibkontakt. Zuerst wird eine geometrische Theorie für glatte mechanische Systeme entwickelt, welche dann so erweitert wird, dass auch Bewegungen mit Unstetigkeiten in den Geschwindigkeiten beschrieben werden können. Diese nichtglatte Theorie der Mechanik wird für die Beschreibung von mechanischen Systemen mit Reibkontakt verwendet. Schließlich werden zwei numerische Methoden für die Simulation von solchen Systemen hergeleitet. Dazu wird jeweils die Methode der finiten Elemente in der Zeit und der generalized-α Ansatz verwendet.

Die vorgestellte geometrische Theorie beschreibt die Dynamik von explizit zeitabhängigen endlich-dimensionalen mechanischen Systemen. Dabei wird sowohl die physikalische Raumzeit als auch die Konfigurationsraumzeit als Galilei-Mannigfaltigkeit modelliert. Die Theorie für glatte Systeme beruht auf drei Postulaten, welche formal den drei Newtonschen Axiomen entsprechen und als Startpunkt für die Entwicklung der Prinzipien der Mechanik dienen. Es ist ein besonderes Anliegen dieser Arbeit zu zeigen, dass die Prinzipien und Formalismen der technischen Mechanik als lokale Repräsentationen der fundamentalen Prinzipien der geometrischen Theorie gewonnen werden können. In diesem Bestreben wird gezeigt, dass die vorgestellte geometrische Theorie für glatte Systeme zum Beispiel das Prinzip der virtuellen Arbeit, die Lagrangeschen Gleichungen erster und zweiter Art, und das Prinzip der virtuellen Wirkung beinhaltet. Basierend auf der geometrischen Behandlung von bilateralen Bindungen, werden Bedingungen hergeleitet mit welchen man feststellen kann ob ein Bindungssatz holonom ist. Diese Bedingungen sind im Vektor-Matrix Kalkül formuliert, sodass sie direkt in der technischen Mechanik angewendet werden können.

Der Übergang von der glatten zur nichtglatten Theorie wird sowohl durch das Aufweichen der Kontinuitätsanforderung der Bewegung als auch durch eine Verallgemeinerung der Definition von Kräften erreicht. Es wird gezeigt, dass die bekannte Maßgleichung, welche typischerweise zur Beschreibung der Dynamik von nichtglatten mechanischen Systemen verwendet wird, als lokale Darstellung der geometrischen Theorie folgt. Nach der geometrischen Behandlung von ein-

seitigen Bindungen und nach dem Einführen von mengenwertigen Coulombschen Reibgesetzen wird die Theorie für die Beschreibung von mechanischen Systemen mit Reibkontakt verwendet. Schließlich erfolgt eine Zeitdiskretisierung auf zwei Arten.

Das erste numerische Schema, hergeleitet mit der Methode der finiten Elemente in der Zeit, erweitert den "symmetric Moreau-type integrator" um bilaterale Bindungen. Das zweite Schema hingegen ist eine Erweiterung der bestehenden nichtglatten generalized-α Methoden, welche für die Simulation von mechanischen Systemen mit reibungsfreiem Kontakt entwickelt wurden. Damit die hier vorgestellte Methode auch mechanische Systeme mit Reibkontakt abdeckt, werden diskrete mengenwertige Coulombsche Reibgesetze eingeführt.

— 1 —

Introduction

> It is wrong to think that the task of physics is to find out how nature is. Physics concerns what we can say about nature.
>
> —*Niels Bohr*

This dissertation is concerned with the dynamics of mechanical systems with frictional contact. First, a geometric theory for smooth mechanical systems is presented, which is then extended to include nonsmooth systems. Finally, the nonsmooth theory is applied for the description of mechanical systems with frictional contact. The derived equations describing the dynamics of a system with frictional contact are discretized using the time finite element method and the generalized-α approach, respectively, leading to two novel numerical schemes. These can be for instance employed for the simulation of flexible multibody systems with frictional contact.

1.1. Motivation

In the life cycle of most technological systems involving hardware, mechanical models play an important role. They are for example used during the engineering design process or are employed by the control algorithm used for the operation of the systems. For many applications, the modeling of frictional contact is crucial for the understanding of the system's dynamics. This is for example the case for legged robots, where the contact of the foots with the ground is used for locomotion. Another prominent example are industrial robots, which need to grasp objects for the fulfillment of their task.

From the perspective of engineering, the modeling process of a mechanical system starts with the choice of a mechanical theory, e.g., the theory of nonsmooth mechanics to model systems with frictional contact. Within the chosen theory, it is then the engineer's task to specify the kinematics of the system as well as to

appropriately select constitutive laws for the forces acting on the system. Finally, numerical schemes may be employed to simulate the system.

Since engineers are interested in efficient ways to systematically derive mechanical models for a broad range of technical applications, engineering mechanics has evolved in a diametrically opposed direction than theoretical mechanics, which aims at developing a mechanical theory to understand physical processes in nature and may be attributed to physicists. It can generally be observed, that the efficient calculus underlying engineering mechanics relies on the fact that many physically different objects, such as positions, velocities, accelerations and forces, are all regarded as elements of the same vector space. However, this comes at the price that it is mathematically possible to add for example positions and forces, which leads to physically meaningless vectors. In contrast to that, it is of particular interest for theoretical mechanics to precisely distinguish physically different objects also within the mathematical theory, which inevitably leads to the need of the more complex mathematical structure provided by differential geometry. The question that arises is, how the mathematically simpler formalism of engineering mechanics can be retrieved from a mathematically richer geometric theory of mechanics, which satisfies the requirements of theoretical mechanics.

1.2. Objectives and Literature Review

The principal topic of this dissertation is the mathematical description and the simulation of mechanical systems with frictional contact. The aim is to expound the complete line of thought which underlies the mechanical modeling process. This ranges from the development of a geometric mechanical theory, which can be applied to the description of mechanical systems with frictional contacts, to its discretization in order to simulate such systems. Moreover, it is a particular aim to forge a bridge between the geometric theory and engineering mechanics wherever possible. To that end, the following three objectives are identified.

O1 Development of a geometric theory for the dynamics of time-dependent finite-dimensional mechanical systems.

O2 Development of a theory for nonsmooth mechanics, which extends the theory resulting from objective **O1**, and apply it for the description of mechanical systems with frictional contact.

O3 Present different techniques for the discretization of the description of mechanical systems with frictional contact resulting from **O2** and thereby derive novel numerical schemes.

Clearly, the above objectives are rather general and need more specifications. In the following subsections, the three objectives **O1**–**O3** are refined along with a literature review given in the process.

O1 – Geometric theory for finite-dimensional mechanical systems

The aim of a mechanical theory is to establish a mathematical framework for the derivation of models for mechanical systems. Hence, it must contain the physics that is common to all mechanical systems while leaving unspecified the system-specific properties. In mechanics, all systems are regarded as a collection of particles, which move through space and time. Hence, a mechanical theory must introduce the concepts of space, time and motion of particles, while leaving unspecified the kinematics of the system, that is, the restriction of the motions of the particles due to the geometry of the specific system. To characterize the interaction of particles, the concept of force is introduced. The influence of a force on the motion of the particle is given by Newton's second law. Therefore, a mechanical theory must provide a definition of force as well as the concepts of mass and acceleration of the particle. Since a mechanical theory is written in the language of mathematics, it is necessary to translate all notions used to describe mechanical systems, that is, to every concept, such as position, acceleration, mass and force, mathematical objects have to be attributed. To account for the different "nature" of these concepts, from a scientific viewpoint, it is important that the different concepts are attributed to distinct mathematical objects, which then allows to study the subtleties of their interrelations.

For a fixed time, there are many possibilities to describe the position of a particle. For example, one can use Cartesian, spherical or cylindrical coordinates. Similarly, there are many ways to parametrize the velocity and acceleration of the particle as well as the forces acting on it. Hence, the mathematical objects attributed to these mechanical concepts must allow for different parametrizations, or in other words, they must be coordinate independent. The idea of coordinate independence lies at the basis of differential geometry, which is therefore identified as the mathematical discipline used to develop the mechanical theory presented in this thesis and motivates the adjective "geometric" in **O1**.

The very first question to be answered is: What geometric structure gives the best description of space and time? This question, which already tormented the ancient Greek philosophers, must be answered by any mechanical theory, because its answer sets the stage for the definitions of all other concepts, such as motion, acceleration and forces. In engineering mechanics, it is most prevalent to treat time and space separately. Typically, the notions of absolute space $\mathcal{S}$ and absolute time $\mathbb{R}$ are introduced, see e.g. [53, 63, 99]. While, for the purpose of engineering mechanics, this point of view is perfectly fine, the assumption of

an absolute space has caused some controversy in theoretical mechanics, which was finally resolved by general relativity.[1] Nowadays, due to the success of general relativity, it is well accepted that, at least in relativistic mechanics, space and time must be treated as an entity called space-time. Already shortly after Einstein's formulation of general relativity, i.e., the theory of relativistic gravity, Cartan [36, 37] and Friedrichs [41] independently formulated a geometric theory of Newtonian gravity, which is based on a four-dimensional nonrelativistic space-time. Based on these ideas, a lot of effort was put into the study of the differential geometric descriptions of nonrelativistic space-time and Newtonian gravity, which gave rise to the field of research known by the name of Newton–Cartan theory [31, 33, 61, 66, 108, 124, 125]. It was within this branch of research that Dombrowski and Horneffer [31, 32] came up with the concept of Galilean manifold. Ehlers [34][2] showed that of the space-time of general relativity is a Galilean manifold in the Newtonian limit, i.e., the limit where the speed of light tends to infinity. This result demands that, for the sake of consistency, a Galilean manifold is used to model the space-time of classical mechanics.

If the physical space-time is modeled as a four-dimensional Galilean manifold $\mathcal{E}$, it is natural to also model the configuration space-time of the system as a Galilean manifold $\mathcal{M}$, because the particles of a mechanical system move in physical space-time. In fact, not doing so would lead to a contradiction since the configuration space-time of a system consisting of a single particle is the space-time $\mathcal{E}$ itself. Hence, the first aim related to **O1** is identified as:

O1.1 Use Galilean manifolds to model both the physical space-time and the configuration space-time.

Already Dombrowski and Horneffer [31, 32] developed a mechanical theory, which models the configuration space-time of a mechanical system as a Galilean manifold. However, their theory only includes forces which can be written as quadratic polynomials in the velocity coordinates. This comes from the fact that, within their theory, the dynamics of the system is described by a linear connection on the Galilean manifold. The generalization to arbitrary forces was achieved by Loos [82, 83], who employed nonlinear connections to introduce a differential two-form defined on the state space of the mechanical system. This two-form, called action form of the systems, is the cornerstone of the theory and encodes the dynamics of the system. The approach of using the action form to characterize the dynamics of a mechanical system was picked up by Eugster et al. [40] and Winandy [127], who developed a theory for finite-dimensional mechanical systems based on a single postulate, which states the relation of

[1] A more detailed discussion of the controversial role of absolute space is given in Section 5.1.
[2] See Ehlers [35] for an English translation.

the action form of the system to its inertia and to the forces acting on the system. In his PhD thesis [127], Winandy showed that the postulate comprises many classical equations, such as Lagrange's equations, Hamilton's equations and Hamel's equations. Moreover, Winandy provides the link between the action form and the principle of virtual work as well as between the action form and Hamilton's principle. In Capobianco et al. [22], a closer look is taken on these interconnections and it is shown that three equivalent versions of the principle of virtual work and Hamilton's principle can be formulated with the help of the action form of the mechanical system. It is important to point out that the above references all work on the configuration space-time and none of them presents a link between the action form and the motions of the particle of the system in physical space-time. In other words, the link between the description of the motion in terms of generalized coordinates and in terms of particle coordinates is missing, justifying the next aim related to **O1**:

O1.2 Link the motions of the particles of the system to the motion of the system in configuration space-time.

Before moving to the next refinement, a remark is due. The idea of using a differential two-form to characterize the dynamics of mechanical systems was known even before Loos' works. Indeed, it can be traced back to Cartan's lectures on integral invariants [23], in which however only systems whose forces result from a potential are considered. Cartan's strategy was pursued by Gallissot [42], who also linked the dynamics on configuration space-time to the dynamics of the particles. However, Gallissot, and later Souriau [117],[3] did not explicitly work on Galilean manifolds. Moreover, they based their theory on an action form, which, if defined on a Galilean manifold, would not lead to a coordinate invariant theory, see [40, 127].

As discussed above, it is shown by [22, 40, 127] that the most important classical equations describing the motion of a mechanical system in generalized coordinates are comprised in the mechanical theory on Galilean manifold. Similarly, once the link to the motions of the system's particles is established as demanded by **O1.2**, the classical equations for the particles have to be derived. Typically, e.g., in engineering mechanics, the dynamics of a particle is described in a Euclidean vector space $\mathbb{E}^3$. Specifically, the position, the velocity, the acceleration and the forces are all regarded as vectors in $\mathbb{E}^3$, which allows, for example, to define the velocity as the time derivative of the position or to establish Newton's second law stating that the force equals mass times acceleration. Since this formalism in $\mathbb{E}^3$ has proven to be very efficient for calculations and is the most

[3] See [118] for an English translation.

used approach to particle mechanics, it must be shown that the formalism can be established also in the context of Galilean manifolds.

O1.3 Retrieve the particle mechanics formalism in a Euclidean vector space $\mathbb{E}^3$ together with the notions of position, velocity, acceleration and force as vectors in $\mathbb{E}^3$.

Bilateral constraints form a fundamental concept for the formulation of mechanical models. In fact, using bilateral constraints, one can derive a mechanical model by a posteriori constraining a "free" system, whose kinematics is often simpler than the kinematics of the constrained system expressed in minimal coordinates. It is well known that bilateral constraints can be formulated on position, velocity and acceleration level. It can be shown, that if the set of constraints on a mechanical system can all be formulated on position level, the dynamics of the constrained system with configuration space-time $\mathcal{M}$ can be reduced to the dynamics of a free system on a submanifold $\mathcal{N} \subset \mathcal{M}$ of the configuration space-time. Such a set of constraints is called holonomic. For a specific mechanical model, it is often important to know if a constraint is holonomic or not, because being holonomic has farreaching implications. For example, if it is known that a set of constraints is holonomic, it is clear that the system's states are confined to a submanifold $\mathcal{N} \subset \mathcal{M}$ and that it can hence be described by a smaller set of generalized coordinates. In the context of Galilean manifolds, Winandy [127] introduced the concepts of holonomic and nonholonomic bilateral constraints and using Frobenius' theorem identified a geometric condition to assert if a given set of constraints is holonomic. For the purpose of engineering, Frobenius' theorem is cumbersome and it would be handier to have conditions in vector-matrix notation, which can directly be evaluated to check if a set of constraints is holonomic. For scleronomic constraints, that is, constraints that do not explicitly depend on time, such conditions can be found in [93, 129]. To cover for the more general case, the following aim is formulated.

O1.4 State conditions in vector-matrix notation, which can be used to assert if a set of time-dependent bilateral constraints is holonomic.

Note that **O1.3** and **O1.4** are very similar, as they both ask to derive an efficient calculus, which can be used in engineering, while being coherent with the Galilean space-time concept.

To complete the picture, it is at this point appropriate to give a rough overview of the literature on geometric mechanics. It can generally be observed, that most treatises formulate a geometric theory of mechanics by introducing a time-independent configuration manifold $\mathcal{Q}$, which correspond to the picture of configuration space-time of the form $\mathbb{R} \times \mathcal{Q}$. For time-independent mechanical systems with forces that can be derived from a potential, the theory of symplectic

mechanics has been developed. In that case, the state space and the phase space of the system are defined to be the tangent bundle $T\mathcal{Q}$ and the cotangent bundle $T^*\mathcal{Q}$ of the configuration manifold, respectively, see for example [1, 10, 14, 30, 81, 84, 85, 89, 98, 104, 110]. It might be due to the popularity of symplectic mechanics that many theories for time-dependent mechanical systems have evolved as generalizations of symplectic mechanics and incorporate explicit time-dependence by considering the extended state space $\mathbb{R} \times T\mathcal{Q}$, or to extend the phase space as $\mathbb{R} \times T^*\mathcal{Q}$, see [29, 89, 104] and [1, 10], respectively. However, the physical interpretation of these spaces is problematic as there is no natural way to relate them to the configuration space-time $\mathbb{R} \times \mathcal{Q}$. Geometric theories which include general forces are for example Godbillon [52] for time-independent systems as well as Gallissot [42], Souriau [118], Loos [82] and Winandy [127] for time-dependent systems.

O2 – Generalization to nonsmooth mechanics

In many technical applications, the modeling of the contact between parts of the system with each other and with the environment is crucial for the understanding of the system's dynamics. It was mainly the description of contact phenomena that led to the necessity of a nonsmooth theory of mechanics. To motivate this, consider that a part of the mechanical system at study is moving towards the ground at high speed. Then, as soon as the part touches the ground, it is braked hard in a very short period of time. Since this deceleration process is often much faster than the dynamics one aims to study with the mechanical model, it can make sense to assume that the deceleration takes place instantaneously, that is, to assume that the velocity of the system may jump when a collision occurs. The situation is similar, when dry friction is at play. In that case, the transition from slipping to sticking is also very fast with respect to the time scale of interest and is mostly assumed to be instantaneous. Hence, because this dissertation is concerned with the description of mechanical systems with frictional contact, the theory developed within objective **O1** needs to be extended such that it can also cope with nonsmooth motions. This objective is summarized by **O2**.

In their seminal works [64, 90, 91], Moreau and Jean established a nonsmooth theory of mechanics, which allows for velocity jumps and which is able to describe frictional impacts as well as unilateral constraints. Starting from their works, the branch of research known as nonsmooth mechanics has evolved. Within nonsmooth mechanics the theory of Moreau and Jean has been extended in many ways. For instance, Glocker [48] presents a nonsmooth theory for rigid body systems which are subjected to general set-valued force laws. Also methods for the study of nonsmooth mechanical systems have been developed, see for example Leine and Nijmeijer [76] or Leine and van de Wouw [77], which study

the stability, convergence and bifurcations of nonsmooth mechanical systems. Moreover, the theory of nonsmooth mechanics has been used to derive models for many technical applications, see [65, 80, 92, 95, 96, 105, 123] to name a few. It would go beyond the scope of this section to review all extensions, refinements and applications of the theory of nonsmooth mechanics. The interested reader is referred to the exhaustive book by Brogliato [16].

Traditionally, nonsmooth mechanics states the dynamics of a mechanical system as an equation of ($\mathbb{R}^n$-valued) measures, which paired with the set-valued force laws, for example employed to model frictional contact, constitute a measure differential inclusion. However, there exists also the variational approach to nonsmooth mechanics, which characterizes the motions of the system by an extended form of Hamilton's principle or more generally the principle of virtual action, see [20, 79]. It can be observed that within the field of nonsmooth mechanics, the focus lies on the nonsmoothness of the motion as well as the many conceptual and mathematical hurdles that come with it. As a result, the core mechanical concepts such as space-time, position, velocity and force often take a back seat in the expositions of nonsmooth mechanics. In fact, seldom a word is spent about the (configuration) space-time and mostly positions, velocities and forces are regarded as elements of $\mathbb{R}^n$ with suitable dimension. Since the description of the dynamics within $\mathbb{R}^n$ is exactly what must result as the coordinate representation of a geometric theory of mechanics, the following aim related to **O2** can be stated.

O2.1 Retrieve the principle of virtual action and the equality of measures in the form known by the nonsmooth mechanics community.

A geometric theory that satisfies **O2.1** and that is complemented by a geometric description of impacts finally yields a coordinate invariant characterization of the dynamics of mechanical systems with frictional contact. The missing bit, namely the geometric description of impacts, is discussed in [5, 6, 49, 128].

Within multibody dynamics, two ingredients are widely used to model frictional contact between parts of the system with each other and with the environment. The first ingredient is the concept of unilateral constraints, which are used to guarantee that the two parts which come into contact do not penetrate each other. As the second ingredient, set-valued force laws are used, which can often be formulated as normal cone inclusions, see Leine and Nijmeijer [76].

Since objective **O2** implies that the developed theory of nonsmooth mechanics is an extension of the smooth theory resulting from **O1**, the following aim is an immediate consequence.

O2.2 Introduce the unilateral constrains in the same fashion as the bilateral constraints treated within the smooth theory of **O1**.

Note, in nonsmooth mechanics ideal unilateral constraints are often introduced as normal cone inclusions, and hence as set-valued force laws, see [48, 79].[4] However, in classical mechanics, ideal bilateral constraints are generally not introduced as normal cone inclusions. However, it will be seen that the path pursued here is very natural from a geometric point of view and that it implies the normal cone inclusions known from the nonsmooth mechanics literature.

O3 – Numerical schemes for mechanical systems with frictional contact

The dynamics of nonsmooth mechanical systems, and hence of mechanical systems with frictional contact, can be characterized in two equivalent ways: Either using the principle of virtual action or by means of an equality of measures, see the discussion leading to **O2.1**. Both qualify as the starting point for the derivation of numerical schemes, which can be used for the simulation of nonsmooth mechanical systems. Clearly, there are many ways to discretize these descriptions and it would go far beyond the scope of this thesis to present them all. However, for either of the starting points, a respective main line of thought can be identified. Due to the variational nature of the principle of virtual action, the main discretization approach is the use of a finite element method in time. For instance, [111] uses the discontinuous Galerkin method together with nonlinear polynomial ansatz functions, whereas [20] employs the Bubnov-Galerkin as well as the weighted residuals approach. When it comes to the discretization of the description of a mechanical system in terms of an equality of measures, typically, the equality of measures is integrated over the time interval constituting a time step of the scheme. Then, quadrature rules are used to discretize the appearing integrals, see [76, 87, 91, 120] for instance. These two lines of thought, i.e., the time finite element method and the idea of using quadrature rules on the integrated equality of measures, accommodate most derivations of numerical schemes for nonsmooth mechanical systems. To live up to **O3**, one representative example for each of the two lines of thought shall be given in this dissertation. Moreover, since some novelty within the derived schemes is aspired, the following choice is made.

O3.1 Use the time finite element method to extend the symmetric Moreau-type integrator of [20] to include bilateral constraints.

O3.2 Derive a nonsmooth generalized-α scheme, which extends the scheme of [18] to account for set-valued Coulomb-type friction as well as a general kinematic equation. Hereby, the friction laws should be included at velocity and acceleration level.

[4] Sometimes, ideal constraints are also called perfect constraints.

To better motivate the content of **O3.1** and **O3.2**, an overview over the challenges arising in the numerical treatment of nonsmooth mechanical systems is given in what follows.

The mathematical description of mechanical systems with frictional contact resulting from **O2** captures the nonimpulsive as well as the impulsive dynamics of the system, where the impulsive dynamics is typically caused by impacts between parts of the system. For the simulation of such systems, two approaches can be distinguished, the event-driven and the event-capturing schemes. The event-driven schemes use standard ordinary differential equations (ODE) solvers, or differential algebraic equations (DAE) solvers, to compute the nonimpulsive motion. Every time an impulsive event, e.g., an impact, is detected, the algebraic impact equations and impact laws (often in the form of inequalities) are solved in order to find the post-impact velocities, which are then used to continue the integration with the ODE/DAE solver. The main strength of event-driven integration is that ODE/DAE solvers with high-order accuracy can be used. However, since every impact is resolved, these schemes are not suitable to find motions with accumulation points, i.e., motions with an infinite number of impacts occurring in a finite interval of time.

In contrast to event-driven schemes, the event-capturing schemes, also called time-stepping schemes, can overcome accumulation points, because they smear the effects of impacts over a time step. However, the most widespread schemes, such as Moreau's time-stepping scheme [50, 91, 120] and variants thereof, are of first-order accuracy, allow penetration of the contacting bodies and show a high level of numerical dissipation. These properties are problematic, especially for the simulation of mechanical systems containing elastic parts. Several publications present improved event-capturing schemes by addressing at least one of these drawbacks. To arrive at event-capturing schemes with higher accuracy order for the impact-free motion, Acary [2] combines high-order Runge–Kutta methods with Moreau's time-stepping scheme, Studer et al. [120, 121] relies on extrapolation methods and Schindler and Acary [111] use a discontinuous Galerkin method to discretize the dynamics of the mechanical system. The numerical constraint drift, which is also responsible for contact penetration, is generally solved by a stabilization in the sense of Gear–Gupta–Leimkuhler (GGL) [44] and/or a projection approach, see for example Acary [3] or Schoeder et al. [112]. In Möller [87], and Capobianco and Eugster [20], integrators similar to Moreau's time-stepping scheme with improved long-term energy behavior have been derived. Since the symmetric Moreau-type integrator presented in [20] cannot be applied for the simulation of mechanical systems which are subjected to bilateral constraints, its extension is identified as aim **O3.1**.

In [17, 18, 25, 27], the nonsmooth generalized-α methods were introduced, which alleviate many of the problems of the Moreau-type time-stepping schemes.

In particular, the generalized-α schemes are second-order accurate for the impact-free motion, use the GGL stabilization to avoid penetration of the contacting bodies and it is known form structural mechanics applications, see [8, 26, 62], that generalized-α schemes perform well for flexible multibody systems. However, the nonsmooth generalized-α methods of [17, 18, 25, 27] are only applicable to multibody systems with frictionless contact, i.e., unilateral constraints without Coulomb friction. Furthermore, these nonsmooth generalized-α methods are restricted to mechanical systems for which the kinematic equation states that the velocity of the system corresponds to the time derivative of the position coordinates. The schemes of [17, 18, 25, 27] have been extended to cope with systems with frictional contact and a more general kinematic equation in [28, 43]. In [43], the augmented Lagrangian approach together with discrete set-valued Coulomb friction laws on position and velocity level have been used to arrive at a nonsmooth generalized-α scheme, which can describe systems with frictional contact. The scheme of [43] has been extended in [28] to allow for different rotation parametrizations for rigid bodies. This represents a particular case of a general kinematic equation since the angular velocity is not the time derivative of the rotation matrix describing the orientation of a rigid body. In nonsmooth mechanics, the set-valued Coulomb friction law is naturally stated as a force law on velocity level. Since in the case of sticking this friction law acts like a bilateral constraint, it can in that case be brought to acceleration level through differentiation. Similar to [28, 43], it is the aim **O3.2** of this dissertation to extend the nonsmooth generalized-α schemes [17, 18, 25, 27] to account for friction as well as a general kinematic equation. However, the explained link between the friction law on velocity and acceleration level shall be exploited instead of formulating the discrete friction laws on position and velocity level as done in [28, 43].

1.3. Main Contributions and Outline

The outline of this thesis directly reflects the objectives **O1**-**O3**. After reviewing selected topics from linear algebra in Chapter 2, a concise introduction to differential geometry is given in Chapter 3. On that basis, Chapter 4 presents the mathematical foundations for the achievement of objective **O1**. In particular, a concise treatment of distributions, connections and covariant derivative is given. Following Loos [82], these concepts are used to introduce action forms on Galilean manifolds together with a collection of theorems, which play a central role for the development of the mechanical theory in Chapter 5. For the definition of the action form of a mechanical system, Loos employed a (nonlinear) connection. Hence, the role of connections is as important for the mechanical theory on

Galilean manifolds as the role of action forms. However, Loos barely writes about connections and focuses only on the action forms.

Contribution 1. It is the contribution of Chapter 4 that the interconnection between nonlinear connections, action forms and the coordinate independent description of non-autonomous second-order ordinary differential equations on Galilean manifolds is worked out in detail.

The geometric mechanical theory for the description of time-dependent finite-dimensional systems required by **O1** is developed in Chapter 5. The theory is based on the physically motivated concept of Galilean manifold, as required by aim **O1.1**.

Contribution 2. It is a particular contribution of Chapter 5 that the presented theory links the dynamics of the system to the dynamics of the particles constituting the system. In doing so, Chapter 5 hits aim **O1.2**. Moreover, this Chapter complements the works of Loos [82] and Winandy [127].

To meet the scientific needs of theoretical mechanics, the theory is completely formulated in the language of differential geometry. However, for the purpose of engineering, the abstract geometric approach is inconvenient and cumbersome. Hence, Chapter 5 makes an effort to retrieve the formulae and principles in the form known in engineering mechanics.

Contribution 3. In Section 5.2, the physical space-time of classical mechanics is modeled as a Galilean manifold. By introducing the concepts of observer and reference field, which identifies the particles assumed at rest, it is shown at the end of Section 5.2 that the notions of position, velocity and acceleration as vectors in a Euclidean vector space $\mathbb{E}^3$ can be retrieved as local representations of their geometric counterparts. Based on that, the aim **O1.3** is met by deriving the vector space formalism of engineering mechanics as a local representation of the geometric theory of Chapter 5. In particular, the well known representations of the kinetic energy, virtual displacements, the principle of virtual work, Lagrange's central equation and the principle of virtual action are derived. This can be regarded as the main contribution of Chapter 5.

The geometric theory presented in Chapter 5 not only contains the results of engineering mechanics, but also shows the interconnection between the principles of mechanics. Moreover, being formulated in the language of differential geometry, the full arsenal of this mathematical discipline can be used within the mechanical theory. For instance, it is thanks to the geometric view on bilateral constraints that the question whether a set of constraints is holonomic can be answered by Frobenius' theorem. It is the essence of aim **O1.4** to make this result accessible

to engineering mechanics by deriving conditions in vector-matrix notation, which can be used to assert whether a set of constraints is holonomic.

Contribution 4. In Section 5.10, which studies bilateral constraints, aim **O1.4** is met by stating even two equivalent conditions, formulated in vector-matrix notation, which each can be used to assert if a set of time-dependent bilateral constraints is holonomic.

Chapter 6 is concerned with the description of nonsmooth mechanical systems. As required by **O2**, a nonsmooth theory of mechanics, which contains the theory presented in Chapter 5 as a special case, is elaborated in Section 6.1.

Contribution 5. By relaxing the continuity assumptions of the motion in the way proposed by Moreau [91] and by generalizing the definition of forces, a theory for nonsmooth mechanics is developed in Section 6.1. It is shown, that the principle of virtual action and the equality of measures in the form known by the nonsmooth mechanics community, are local representations of the fundamental principle of the presented theory – as required by **O2.1**.

In agreement with aim **O2.2**, the concept of conic distribution is used to define unilateral constraints in Section 6.2.

Contribution 6. It is the merit of Section 6.2 to show that unilateral constraints can conceptually be treated in the same way as the bilateral constraints in the smooth theory of Chapter 5. In fact, the developed view on constraints for nonsmooth mechanical systems is shown to comprise both unilateral and bilateral constrains. This reveals that, within nonsmooth mechanics, not only the constraint forces of unilateral constraints may be impulsive, but that this is even the case for bilateral constraints – a fact that is often not explicitly addressed in the literature.

Chapter 6 is concluded by addressing the second part of **O2**, namely, the application of the developed theory for the description of mechanical systems with frictional contact. This is done in Section 6.3, where a hard frictional contact model is introduced and the relevant equations for the description of the dynamics of mechanical systems with frictional contact are summarized. It is the description presented in Section 6.3 that is discretized in time to address the two parts of objective **O3**.

Contribution 7. In Chapter 7, the symmetric Moreau-type integrator [20] is extended to include bilateral constraints, see **O3.1**. To achieve that, the time finite element method used in [20] is adopted to discretize the system's dynamics, while the bilateral constraints are contemplated at velocity level.

Contribution 8. In Chapter 8, a nonsmooth generalized-α method is derived, which extends the existing methods [17, 18, 25, 27] to account for:

- frictional contact modeled as unilateral constraints complemented by set-valued Coulomb-type friction laws formulated on velocity and acceleration level.
- a wider class of mechanical systems, for which the generalized velocities and the time derivative of the position coordinates are related by a general kinematic equation, and are hence not necessarily identical.

Moreover, a set of benchmark systems is devised and finally used to validate the performance of the presented scheme.

Concluding remarks as well as an outlook on further scientific questions are given in Chapter 9.

— 2 —

Miscellaneous Topics from Linear Algebra

> Mathematics is the key and door
> to the sciences.
>
> —*Galileo Galilei*

This chapter presents some concepts of linear algebra based on the basic properties of finite dimensional vector spaces and linear maps. It is assumed that the reader is familiar with the notions of vector spaces and linear maps. The primary aim of the chapter is to give the reader a good working basis for the understanding of the subsequent chapters and to introduce the required notations. Hence, not all the definitions are given and sometimes, especially when the definition is involved but straightforward, the concepts are introduced by example.

After establishing some basic notations in Section 2.1, Euclidean point and vector spaces are introduced in Section 2.2. Tensors and alternating forms, as special type tensors, are treated in Sections 2.3 and 2.4, respectively. The chapter is concluded with a section on convex cones. For the reader interested in more details on these topics, Audin [11] is recommended for an in depth look on Euclidean spaces. Moreover, Lee [74] or Winandy [127] can be consulted for tensors and alternating forms. Finally, as an encyclopedic book on the variational analysis of convex sets Rockafellar and Wets [106] is cited.

2.1. Basic Notations

Summation convention. In this work, Einstein's summation convention is adopted. This implies a summation from 1 to n whenever an index i appears once as an upper and once as a lower index, where n is clear form the context. For instance,

$$a^i b_i = \sum_{i=1}^{n} a^i b_i, \quad v^i \frac{\partial}{\partial x^i} = \sum_{i=1}^{n} v^i \frac{\partial}{\partial x^i} \quad \text{or} \quad \frac{\partial}{\partial u^i} \otimes \eta^i = \sum_{i=1}^{n} \frac{\partial}{\partial u^i} \otimes \eta^i \,.$$

Vector-matrix notation. Any object x^i with one index, where $i = 1, \dots, n$, is gathered as an n-tuple $\mathbf{x}$, e.g.,

$$\mathbf{x} = (x^1, \dots, x^n) \quad \text{or} \quad \mathbf{f} = (f_1, \dots, f_n)\,.$$

Moreover, the convention is made that lists of tuples are again regarded as tuples, e.g.,

$$(\mathbf{x}, \mathbf{y}) = (x^1, \dots, x^n, y^1, \dots, y^m)\,.$$

To objects A_{ij} with two indices the matrix $\mathbf{A}$ is associated, where the left index is the row-index and the other is the column-index, e.g.,

$$A_{ij} \mapsto \mathbf{A} = \begin{pmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \dots & A_{nn} \end{pmatrix} \quad \text{or} \quad B^i{}_j \mapsto \mathbf{B} = \begin{pmatrix} B^1{}_1 & \dots & B^1{}_n \\ \vdots & \ddots & \vdots \\ B^n{}_1 & \dots & B^n{}_n \end{pmatrix}.$$

To gain access to vector-matrix notation, n-tuples are regarded as column vectors, such that we can write $x^i y_i = \mathbf{x}^{\mathrm{T}}\mathbf{y} = \mathbf{y}^{\mathrm{T}}\mathbf{x}$ or

$$y_j = A_{ij} x^i \quad \Leftrightarrow \quad \mathbf{y} = \mathbf{A}^{\mathrm{T}}\mathbf{x} = \mathbf{x}^{\mathrm{T}}\mathbf{A}\,,$$

where $(\cdot)^{\mathrm{T}}$ denotes the transpose. Sometimes, it is convenient to not change between normal and bold symbol. For example, the j-th component of $\mathbf{A}^{\mathrm{T}}\mathbf{x}$ is denoted by $(\mathbf{A}^{\mathrm{T}}\mathbf{x})_j$, such that

$$A_{ij}\, x^i = (\mathbf{A}^{\mathrm{T}}\mathbf{x})_j \quad \text{or} \quad B^i{}_j\, x^j = (\mathbf{B}\mathbf{x})^i\,.$$

To compactly write a series of inequalities, $\mathbf{x} \geq \mathbf{y}$ is understood in a component wise manner, that is,

$$(x^1, \dots, x^n) \geq (y^1, \dots, y^n) \quad \Leftrightarrow \quad x^i \geq y^i \quad \forall i = 1, \dots, n\,.$$

Span of a set. Consider a subset $S \subseteq V$ of a real vector space. The *span (linear hull) of* S is the vector subspace of V defined by

$$\operatorname{span}(S) \coloneqq \left\{ \sum_{i=1}^{k} \lambda^i v_i \;\middle|\; v_i \in S,\ \lambda^i \in \mathbb{R},\ 1 \leq k < \infty \right\}. \tag{2.1}$$

2.2. Euclidean Vector Space $\mathbb{E}^n$

A typical mathematical abstraction of space encountered in classical mechanics is driven by the idea that we can draw an arrow between two points in space. The length of this arrow is then the distance between the points. Moreover, we can measure the angles of the triangle formed by the arrows connecting three points in space. Mathematically, the space of the arrows can be seen as a three-dimensional Euclidean vector space $\mathbb{E}^3$.

Definition 2.1. A *Euclidean vector space* $\mathbb{E}^n$ is an n-dimensional real vector space together with a map $\cdot : \mathbb{E}^n \times \mathbb{E}^n \to \mathbb{R}$, called *inner product*, that for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{E}^n$ and all scalars $a, b \in \mathbb{R}$ satisfies the conditions

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}, \quad (a\mathbf{u} + b\mathbf{v}) \cdot \mathbf{w} = a(\mathbf{u} \cdot \mathbf{w}) + b(\mathbf{v} \cdot \mathbf{w}) \quad \text{and} \quad \mathbf{u} \cdot \mathbf{u} > 0 \text{ if } \mathbf{u} \neq 0\,.$$

The inner product induces the *Euclidean norm* $||\mathbf{u}|| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$, which is interpreted as the length of the vector $\mathbf{u}$. The *angle* $\theta = \measuredangle(\mathbf{u}, \mathbf{v})$ between two non-zero vectors $\mathbf{u}$ and $\mathbf{v}$ is implicitly defined by $\mathbf{u} \cdot \mathbf{v} = ||\mathbf{u}|| \, ||\mathbf{v}|| \cos\theta$.

A vector $\mathbf{u}$ is a *unit vector* if $||\mathbf{u}|| = 1$. Moreover, two vectors $\mathbf{u}, \mathbf{v}$ are *orthogonal* if $\mathbf{u} \cdot \mathbf{v} = 0$. A basis of pairwise orthogonal unit vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is said to be *orthonormal*. Consequently an orthonormal basis is characterized by

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{2.2}$$

for all $i, j = 1, \ldots, n$, where the *Kronecker delta* δ_{ij} has been introduced. The Kronecker delta provides the components of the identity matrix $\mathbf{I}$, i.e., $(\mathbf{I})_{ij} = \delta_{ij}$.

Let $\mathbf{e}_1^I, \ldots, \mathbf{e}_n^I$ be an orthonormal basis of $\mathbb{E}^n$. Any vector $\mathbf{v} \in \mathbb{E}^n$ can be represented as $\mathbf{v} = {}_I v^i \mathbf{e}_i^I$. The components of the vector $\mathbf{v}$ with respect to the basis I are denoted by ${}_I v^i$, which are gathered as the n-tuple ${}_I\mathbf{v}$. It follows immediately from (2.2) that

$${}_I v^i = \mathbf{e}_i^I \cdot \mathbf{v}\,. \tag{2.3}$$

For a different orthonormal basis $\mathbf{e}_i^B$, the vector can also be written as $\mathbf{v} = {}_B v^i \mathbf{e}_i^B$ and it is a consequence of (2.3) that

$${}_B v^j = \mathbf{e}_j^B \cdot \mathbf{v} = {}_I v^i \, \mathbf{e}_j^B \cdot \mathbf{e}_i^I\,.$$

Introducing the *transformation matrix* $\mathbf{A}_{BI}$ *from* I *to* B as

$$(\mathbf{A}_{BI})^j{}_i = \mathbf{e}_j^B \cdot \mathbf{e}_i^I\,, \tag{2.4}$$

the coordinate transformation between orthonormal bases I and B can compactly be written as

$${}_B\mathbf{v} = \mathbf{A}_{BI} \, {}_I\mathbf{v}\,. \tag{2.5}$$

It follows from the definition (2.4) that the transformation matrix $\mathbf{A}_{BI}$ is an orthogonal matrix, i.e., $\mathbf{A}_{BI}^{\mathrm{T}} \mathbf{A}_{BI} = \mathbf{I}$. Moreover, $\mathbf{A}_{IB} = \mathbf{A}_{BI}^{\mathrm{T}}$.

Consider a curve $\mathbf{x} : \mathbb{R} \to \mathbb{E}^n$, then $\dot{\mathbf{x}}$ denotes the derivative of the curve, which is again a curve in $\mathbb{E}^n$. A *moving frame* is a set of curves $\mathbf{e}_i^A : \mathbb{R} \to \mathbb{E}^n,\ t \mapsto \mathbf{e}_i^A(t)$ which for every t forms an orthonormal basis for $\mathbb{E}^n$. Everything that has been said about vectors and bases is still true pointwise, i.e., for a fixed t. Hence, the transformation matrix between moving frames defined by (2.4) must be

regarded as depending on t. Consequently, the bases I and B can be interpreted as depending on t with $\dot{\mathbf{e}}_i^I = \dot{\mathbf{e}}_i^B = 0$ and are therefore called *resting frames*.

Consider a resting frame I. From $\mathbf{x} = {}_Ix^i\,\mathbf{e}_i^I$, it follows that $\dot{\mathbf{x}} = {}_I\dot{x}^i\,\mathbf{e}_i^I$. Gathering the ${}_I\dot{x}^i$ as ${}_I\dot{\mathbf{x}}$, we have that

$$ {}_I(\dot{\mathbf{x}}) = {}_I\dot{\mathbf{x}}\,, \tag{2.6} $$

which shows that the components of $\dot{\mathbf{x}}$ with respect to a resting frame I correspond to the derivative of the components ${}_I\mathbf{x}$ of $\mathbf{x}$ with respect to I. This changes drastically if we consider a moving frame K. In fact, we can use that $\mathbf{x}$ and $\dot{\mathbf{x}}$ are vectors and hence transform by (2.5) to compute

$$ {}_K(\dot{\mathbf{x}}) = \mathbf{A}_{KI}\,{}_I(\dot{\mathbf{x}}) = \mathbf{A}_{KI}(\mathbf{A}_{IK}\,{}_K\mathbf{x})^{\cdot} = {}_K\dot{\mathbf{x}} + \mathbf{A}_{KI}\dot{\mathbf{A}}_{IK}\,{}_K\mathbf{x}\,, \tag{2.7} $$

where ${}_K\dot{\mathbf{x}}$ is the derivative of ${}_K\mathbf{x}$ and where (2.6) has been used for the second equality. Taking the derivative of $\mathbf{A}_{KI}\mathbf{A}_{IK} = \mathbf{I}$ immediately shows that the matrix $\mathbf{A}_{KI}\dot{\mathbf{A}}_{IK}$ is skew-symmetric.

The three-dimensional Euclidean vector space $\mathbb{E}^3$ inherits the cross product from $\mathbb{R}^3$ by defining ${}_B(\mathbf{u}\times\mathbf{v}) = {}_B\mathbf{u}\times{}_B\mathbf{v}$ for every orthonormal basis B. Moreover, the skew-symmetric matrix $\tilde{\boldsymbol{\alpha}} \in \mathbb{R}^{3\times3}$ is uniquely associated to a vector $\boldsymbol{\alpha} \in \mathbb{R}^3$ by the condition $\tilde{\boldsymbol{\alpha}}\mathbf{y} = \boldsymbol{\alpha}\times\mathbf{y}$ for all $\mathbf{y} \in \mathbb{R}^3$. For two frames I and K the *angular velocity $\boldsymbol{\omega}_{IK}$ of K with respect to I* is defined by

$$ {}_K\tilde{\boldsymbol{\omega}}_{IK} = \mathbf{A}_{KI}\dot{\mathbf{A}}_{IK}\,. $$

The angular velocity can now be used to restate (2.7) as

$$ {}_K(\dot{\mathbf{x}}) = {}_K\dot{\mathbf{x}} + {}_K\boldsymbol{\omega}_{IK} \times {}_K\mathbf{x}\,. $$

Coming back to the idea that we can draw an arrow between two points in space. Interpreting the arrows as vectors in $\mathbb{E}^n$ leads to the following mathematical abstraction of space.

Definition 2.2. An *n-dimensional Euclidean point space* consists of a set of points $\mathcal{P}$ together with a map $\mathbf{r} : \mathcal{P}\times\mathcal{P} \to \mathbb{E}^n$ with the properties

(i) For any three points $O, P, Q \in \mathcal{P}$, we have $\mathbf{r}(O,P) + \mathbf{r}(P,Q) = \mathbf{r}(O,Q)$.

(ii) For every point $P \in \mathcal{P}$ and every $\mathbf{v} \in \mathbb{E}^n$ there is a unique $Q \in \mathcal{P}$ such that $\mathbf{r}(P,Q) = \mathbf{v}$.

We can interpret $\mathbf{r}(O,P)$ as the vector pointing from O to P. Then the first property of $\mathbf{r}$ corresponds to the parallelogram rule. It is convenient to lighten notation by introducing $\mathbf{r}_{OP} := \mathbf{r}(O,P)$. By property (ii), the map $\mathbf{r}(O,\cdot)$ is a bijection between $\mathcal{P}$ and $\mathbb{E}^n$ for a fixed O.

2.3. TENSORS

Definition 2.3. Let V be a real vector space of dimension n. The set V^* of all real-valued linear maps $\omega : V \to \mathbb{R}$ is called *dual space of* V. The dual space, equipped with pointwise addition and scalar multiplication is again a real vector space of dimension n. An element ω of V^* is called *linear form*, *one-form* or *covector*.

Consider a basis $e_1, \dots, e_n$ of V. The linear forms $\varepsilon^1, \dots, \varepsilon^n \in V^*$ defined by

$$\varepsilon^i(e_j) = \delta^i_j \,,$$

where δ^i_j is the Kronecker delta, constitute a basis for V^*, called the *dual basis* to $e_1, \dots, e_n$. A proof can be found in the Appendix A.1 of [38].

Definition 2.4. Let V be a real vector space of dimension n. A *tensor of type* (r, s) *on* V, also called (r, s)-tensor on V, is a map

$$T : \underbrace{V^* \times \cdots \times V^*}_{r \text{ copies}} \times \underbrace{V \times \cdots \times V}_{s \text{ copies}} \to \mathbb{R}$$

which is multilinear, i.e., linear in each argument. The sum $r + s$ is called the *rank* of the tensor and the set of all tensors of type (r, s) on V is denoted as $\otimes^r_s V$.

Example 2.1.

(i) Let $e_1, \dots, e_n$ be a basis of V and $\varepsilon^1, \dots, \varepsilon^n$ the corresponding dual basis. Hence, $v = v^i e_i \in V$ and $\omega = \omega_i \varepsilon^i \in V^*$. Then the following are tensors of rank one:

$$\begin{aligned} \varepsilon^i &: V \to \mathbb{R}, \ v \mapsto \varepsilon^i(v) = v^i \\ e_i &: V^* \to \mathbb{R}, \ \omega \mapsto e_i(\omega) = \omega(e_i) = \omega_i \,. \end{aligned} \tag{2.8}$$

The linear form e_i is an element of the dual space V^{**} of V^*. It is straightforward to see, that the maps e_i form a basis for the *bidual space* V^{**}. Hence, the map which associates to every $e_i \in V$ the bidual vector defined by (2.8) is a vector space isomorphism $V \cong V^{**}$, which justifies the overload of notation.

(ii) Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a matrix. The map

$$A : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \ (\mathbf{x}, \mathbf{y}) \mapsto A(\mathbf{x}, \mathbf{y}) = \mathbf{x}^{\mathrm{T}} \mathbf{A}\, \mathbf{y}$$

is a $(0, 2)$-tensor on $\mathbb{R}^n$. Hence, $A \in \otimes^0_2 \mathbb{R}^n$.

Even though, in Definition 2.4, the arguments of an (r, s)-tensor are ordered such that the first r arguments take covectors and the remaining take vectors,

also multilinear maps with other orderings of arguments are called *tensors*, e.g., the multilinear map $T : V \times V^* \times V \to \mathbb{R}$ is a tensor of rank three, also called three-tensor for short. The space of tensors of this type, i.e., multilinear maps from $V \times V^* \times V$ to $\mathbb{R}$, is denoted as $L(V, V^*, V; \mathbb{R})$. Moreover, it is customary to regard scalars as zero-tensors.

Definition 2.5. Let S and T be tensors on V with respective rank s and t. The *tensor product* $S \otimes T$ is the tensor of rank $s + t$ defined by

$$S \otimes T(u_1, \ldots, u_s, v_1, \ldots, v_t) := S(u_1, \ldots, u_s)T(v_1, \ldots, v_t) \tag{2.9}$$

for all u_i, v_j, where u_i and v_j are either vectors or covectors, depending on what the tensors S and T require.

To clarify this definition, consider a two-tensor $S \in L(V, V^*; \mathbb{R})$, that is, $S : V \times V^* \to \mathbb{R}$, and a three-tensor $T \in L(V, V^*, V; \mathbb{R})$, then

$$S \otimes T(u, \sigma, v, \eta, w) = S(u, \sigma)T(v, \eta, w)$$

for all $u, v, w \in V$ and for all $\sigma, \eta \in V^*$.

The tensor product allows to construct a higher-order tensor from two lower-order tensors. Hence, it is not surprising that bases for the tensor spaces can be constructed by taking multiple tensor products of the 1-tensors (2.8). The following proposition is formulated and proved for the tensor space $L(V, V^*, V; \mathbb{R})$ but extends analogously to general tensor spaces, see Proposition 12.4 in [74].

Proposition 2.1. *Let $e_1, \ldots, e_n$ be a basis of V and $\varepsilon^1, \ldots, \varepsilon^n$ the corresponding dual basis. The set*

$$B = \left\{ \varepsilon^i \otimes e_j \otimes \varepsilon^k \;\middle|\; 1 \leq i, j, k \leq n \right\}$$

is a basis of the space $L(V, V^*, V; \mathbb{R})$, which therefore has dimension n^3.

Any $T \in L(V, V^*, V; \mathbb{R})$ can consequently be written as a linear combination

$$T = T^j_{ik}\, \varepsilon^i \otimes e_j \otimes \varepsilon^k \tag{2.10}$$

with $T^j_{ik} = T(e_i, \varepsilon^j, e_k)$.

Proof. Let $u, v \in V$ and $\omega \in V^*$. It follows from the multilinearity of T that

$$T(u, \omega, v) = T(u^i e_i, \omega_j \varepsilon^j, v^k e_k) = T(e_i, \varepsilon^j, e_k)\, u^i \omega_j v^k\,. \tag{2.11}$$

Moreover, using the same u, v and ω as before, the evaluation of (2.10) gives

$$\begin{aligned} T(u, \omega, v) &= T^j_{ik}\, \varepsilon^i \otimes e_j \otimes \varepsilon^k(u, \omega, v) \\ &= T^j_{ik}\, \varepsilon^i(u) e_j(\omega) \varepsilon^k(v) \\ &= T^j_{ik}\, u^i \omega_j v^k\,, \end{aligned} \tag{2.12}$$

where (2.9) and (2.8) have been used. Setting $T^j_{ik} = T(e_i, \varepsilon^j, e_k)$, the comparison of (2.11) and (2.12) proofs that the set B spans $L(V, V^*, V; \mathbb{R})$. Moreover, setting the linear combination (2.10) to zero, by (2.12) immediately implies $T^j_{ik} = 0$ for all i, j, k, since (2.12) must be zero for arbitrary u, v and ω. Hence, the tensors in B are linearly independent, which together with the property that B spans the tensor space proves that B is a basis for $L(V, V^*, V; \mathbb{R})$. □

2.4. Alternating Forms

In this section, the fact that tensors of type $(0, k)$ on V, called *covariant k-tensors*, assign a real number to k elements of a vector space V is exploited. Hence, for such a tensor all arguments are equitable. This is similarly true for tensors of type $(l, 0)$ on V, which are called *contravariant l-tensors*. For simplicity, the notation $\otimes^k V^* = \otimes^0_k V$ for the space of covariant tensors and $\otimes^l V = \otimes^l_0 V$ for the space of contravariant tensors is introduced.

Since all arguments have equal rights, we can switch two arguments of a covariant tensor, which may or may not change the value of the tensor. A covariant k-tensor T on V is *symmetric* if its value remains unchanged by interchanging any pair of arguments, that is,

$$T(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) = T(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k)\,,$$

whenever $1 \leq i < j \leq k$ and for all $v_1, \ldots, v_k \in V$. Similarly, the tensor T is *alternating* (*antisymmetric* or *skew-symmetric*) if it changes sign whenever two arguments are interchanged, i.e.,

$$T(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) = -T(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k)\,,$$

whenever $1 \leq i < j \leq k$ and for all $v_1, \ldots, v_k \in V$.

The set of all alternating covariant k-tensor is denoted by $\bigwedge^k V^*$. Since the sum of two alternating tensors is again alternating and the same holds if a tensor is pre-multiplied by a real number, $\bigwedge^k V^*$ is a vector subspace of $\otimes^k V^*$. An alternating covariant k-tensors is said to be an *alternating form of degree k* or *k-form* for short. To have a notion of form for arbitrary degrees, we set $\bigwedge^0 V^* \coloneqq \mathbb{R}$ and $\bigwedge^1 V^* \coloneqq V^*$.

Since the k-forms are characterized by their behavior under permutations of arguments, the following concept is of value. Consider the set $I_k = \{1, \ldots, k\}$. A *permutation* s is a bijective map $s : I_k \to I_k$. The set of all permutations is called the *symmetric group* and is denoted by S_k. Since a permutation $s \in S_k$ is bijective, the k-tuple $(s(1), s(2), \ldots, s(k))$ is a reordering of the tuple $(1, 2, \ldots, k)$.

Therefore, s can be represented in the form

$$\begin{bmatrix} 1 & 2 & \dots & k \\ s(1) & s(2) & \dots & s(k) \end{bmatrix}.$$

A *transposition* is a permutation which interchanges only two numbers. An arbitrary permutation can be represented as a sequence of transpositions. While the representation of a permutation in terms of transpositions is not unique, the number p of invoked transpositions is.[1] Hence, we can define the *sign of the permutation* s as $\operatorname{sgn}(s) = (-1)^p$, which is 1 or -1 depending on whether p is even or odd, respectively.

Example 2.2. The three permutations in S_3 with positive sign are

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}.$$

The remaining three permutations in S_3 have negative sign and are given by

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}.$$

The *action of a permutation $s \in S_k$ on a covariant k-tensor* is defined as the map $\varphi_s : \otimes^k V^* \to \otimes^k V^*,\ f \mapsto \varphi_s f$ that is given by

$$\varphi_s f : (v_1, \dots, v_k) \mapsto f(v_{s(1)}, \dots, v_{s(k)}),$$

which formalizes the permutation of arguments for a covariant k-tensor.

Similar to the tensor product, it is possible to define a product between a k-form and an l-form yielding a $k+l$-form.

Definition 2.6. The *wedge product* (or *exterior product*) of two alternating forms $\eta \in \bigwedge^k V^*$ and $\omega \in \bigwedge^l V^*$ is defined as

$$\eta \wedge \omega = \frac{1}{k!\, l!} \sum_{s \in S_{k+l}} \operatorname{sgn}(s) \varphi_s(\eta \otimes \omega), \tag{2.13}$$

with $\eta \wedge \omega \in \bigwedge^{k+l} V^*$.

The *set of alternating forms of arbitrary degree*, denoted by $\bigwedge^\star V^*$ endowed with the wedge product is known as the *exterior algebra* or *Grassmann algebra*.

Proposition 2.2 (Lee [74], Proposition 14.11.)**.** *For all $\xi, \eta, \omega \in \bigwedge^\star V^*$ and all real numbers $a, b \in \mathbb{R}$, the wedge product satisfies*

[1] See Lang [71], Theorem 6.1.

(i) $\xi \wedge (\eta \wedge \omega) = (\xi \wedge \eta) \wedge \omega$,
(ii) $(a\xi + b\eta) \wedge \omega = a\xi \wedge \omega + b\eta \wedge \omega$,
(iii) $\eta \wedge \omega = (-1)^{kl}\omega \wedge \eta$, *when* $\eta \in \bigwedge^k V^*$ *and* $\omega \in \bigwedge^l V^*$,
(iv) $\omega \wedge \omega = 0$, *whenever the degree of* ω *is odd.*

Example 2.3. Let $\alpha, \beta \in \bigwedge^1 V^*$ be one-forms. Moreover, let $e_1, \dots, e_n$ be a basis of V and $\varepsilon^1, \dots, \varepsilon^n$ the corresponding dual basis. By (2.13), we have

$$\alpha \wedge \beta = \alpha \otimes \beta - \beta \otimes \alpha \,, \tag{2.14}$$

which can be used to check that, in accordance with Proposition 2.2 (iii) and (iv), it holds that $\alpha \wedge \beta = -\beta \wedge \alpha$ and $\alpha \wedge \alpha = 0$. Moreover, one can use (ii) and (iii) of the same proposition to find

$$\begin{aligned} \alpha \wedge \beta &= (\alpha_i \varepsilon^i) \wedge (\beta_j \varepsilon^j) = \alpha_i \beta_j \, \varepsilon^i \wedge \varepsilon^j \\ &= \alpha_i \beta_j (\varepsilon^i \otimes \varepsilon^j - \varepsilon^j \otimes \varepsilon^i) = (\alpha_i \beta_j - \alpha_j \beta_i) \, \varepsilon^i \otimes \varepsilon^j \,, \end{aligned}$$

where (2.14) has been invoked to transition from the first to the second row.

2.5. Convex Cones

A subset $C \subseteq V$ of an n-dimensional real vector space V is *convex*, if for all $u, v \in C$ it holds that

$$(1 - \lambda)u + \lambda v \in C \quad \forall \lambda \in [0, 1] \,.$$

A subset $K \subseteq V$ is a *cone*, if for any $v \in K$ also

$$\lambda v \in K \quad \forall \lambda > 0 \,. \tag{2.15}$$

Moreover, a cone is *closed* if condition (2.15) additionally holds for $\lambda = 0$. The unit ball $B \coloneqq \{\mathbf{x} \in \mathbb{R}^n \mid ||\mathbf{x}|| \leq 1\}$, for instance, is a convex set but not a cone. An example for a closed convex cone is the positive orthant $\mathbb{R}^{n+}_0 \coloneqq \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \geq 0\}$. The *conical hull* of a set $S \subseteq V$ is the closed convex cone

$$\operatorname{cone}(S) \coloneqq \left\{ \sum_{i=1}^{k} \lambda^i v_i \;\middle|\; v_i \in S,\ \lambda^i \geq 0,\ 1 \leq k < \infty \right\} . \tag{2.16}$$

A cone $K \subseteq V$ is said to be *finitely generated* if there are vectors $b_1, \dots, b_r \in K$ $(r < \infty)$ such that

$$K = \operatorname{cone}\{b_1, \dots, b_r\} \,.$$

Hence, for finitely generated cones, the conical hull plays the role of the span for vector (sub-)spaces and the vectors b_i can be used to generate the vectors

in K. In fact, it is a consequence of (2.16) that any $v \in K$ can be represented as $v = v^i b_i$ with $v^i \geq 0$. Moreover, comparing the definition (2.1) of the span to that of the conical hull, we find that

$$\operatorname{span}\{b_1, \ldots, b_r\} = \operatorname{cone}\{b_1, \ldots, b_r, -b_1, \ldots, -b_r\}\,, \tag{2.17}$$

which shows that every vector subspace of V is a finitely generated cone.

A cone K is *polyhedral*, if there exist linear forms $\alpha^1, \ldots, \alpha^m \in V^*$ such that

$$K = \left\{ v \in V \;\middle|\; \alpha^i(v) \geq 0,\ i = 1, \ldots, m \right\}.$$

Theorem 2.1 (Rockafellar and Wets [106], Theorem 5.52)**.** *A cone $K \subseteq V$ is polyhedral if and only if it is finitely generated. In particular, every polyhedral cone is closed and convex.*

For $\lambda \in \mathbb{R}$, the scalar multiple of a set $C \subseteq V$ is introduced as

$$\lambda C \coloneqq \{\lambda v \mid v \in C\}\,. \tag{2.18}$$

Moreover, $C + D$ denotes the Minkowski sum of two sets $C, D \subseteq V$, that is,

$$C + D \coloneqq \{v + w \mid v \in C,\ w \in D\}\,. \tag{2.19}$$

It follows immediately from the definitions (2.18) and (2.19) that

$$\lambda(C + D) = \lambda C + \lambda D \quad \text{and} \quad (\lambda + \mu)C = \lambda C + \mu C \tag{2.20}$$

where $C, D \subseteq V$ are two sets and $\lambda, \mu \in \mathbb{R}$.

Let $C \subseteq \mathbb{R}^n$ be a closed convex non-empty set. The *normal cone of C at* $\mathbf{x} \in C$ is defined as

$$\mathcal{N}_C(\mathbf{x}) = \{\mathbf{y} \in C \mid \mathbf{y}^{\mathrm{T}}(\mathbf{x}^* - \mathbf{x}) \leq 0, \forall \mathbf{x}^* \in C\}\,, \tag{2.21}$$

whereas $\mathcal{N}_C(\mathbf{x})$ is empty if $\mathbf{x} \notin C$. One can show that the normal cone $\mathcal{N}_C(\mathbf{x})$ is a closed convex cone, see [106]. It can easily be verified, that $\mathcal{N}_C(\mathbf{x}) = \{0\}$ if $\mathbf{x}$ is an interior point of the set C. Indeed, for an interior point $\mathbf{x}$ there exists a ball around $\mathbf{x}$ contained in C. This allows to choose $\mathbf{x}^*$ from this ball such that any direction $\mathbf{x}^* - \mathbf{x}$ can be produced in (2.21), which leaves only $\mathbf{y} = 0$ to fulfill the inequality.

The normal cone has a number of less known scaling properties, which are stated as propositions in the following. These may conveniently be proved using topics from Convex Analysis such as the subdifferential and the support function. However, in order to introduce as little machinery as possible, these properties are proved by using only definition (2.21).

Proposition 2.3. *Let $\alpha > 0$ and $C \subseteq \mathbb{R}^n$ be a closed convex non-empty set. It then holds that*

$$\mathbf{y} \in \mathcal{N}_{\alpha C}(\mathbf{x}) \quad \Longleftrightarrow \quad \mathbf{y} \in \mathcal{N}_C\left(\frac{1}{\alpha}\mathbf{x}\right).$$

Proof. Consider $\mathbf{x} \in \alpha C$. Applying definition (2.21) to the inclusion $\mathbf{y} \in \mathcal{N}_{\alpha C}(\mathbf{x})$ implies that $\mathbf{y}^{\mathrm{T}}(\mathbf{x}^* - \mathbf{x}) \leq 0, \forall \mathbf{x}^* \in \alpha C$, which if divided by $\alpha > 0$ corresponds to

$$\mathbf{y}^{\mathrm{T}}\left(\frac{1}{\alpha}\mathbf{x}^* - \frac{1}{\alpha}\mathbf{x}\right) \leq 0 \quad \forall \frac{1}{\alpha}\mathbf{x}^* \in C \quad \Longleftrightarrow \quad \mathbf{y} \in \mathcal{N}_C\left(\frac{1}{\alpha}\mathbf{x}\right).$$

Furthermore, $\mathbf{x} \notin \alpha C$ implies $\mathcal{N}_{\alpha C}(\mathbf{x}) = \emptyset$, which in turn yields $\mathcal{N}_C(\frac{1}{\alpha}\mathbf{x}) = \emptyset$. The other direction of the implication follows by taking the reciprocal value of α. □

Proposition 2.4. *Consider a closed convex non-empty set $C \subseteq \mathbb{R}^n$ and set $C(\alpha) \coloneqq \alpha C$. If*

$$\mathbf{y} \in \mathcal{N}_{C(\alpha_i)}(\mathbf{x}_i), \quad i = 1, 2, 3 \ldots,$$

where $\alpha_i \geq 0$, then it holds that

$$\mathbf{y} \in \mathcal{N}_{C\left(\sum_i \alpha_i\right)}\left(\sum_i \mathbf{x}_i\right).$$

Proof. Retaining the case $\alpha_i = 0$ for the end of the proof, first consider $\alpha_i > 0$ for all i. By using Proposition 2.3 together with $C(\alpha) = \alpha C$ we infer from $\mathbf{y} \in \mathcal{N}_{C(\alpha_i)}(\mathbf{x}_i)$, that

$$\mathbf{y} \in \mathcal{N}_C\left(\frac{1}{\alpha_i}\mathbf{x}_i\right), \quad i = 1, 2, 3 \ldots$$

Without loss of generality, consider $i = 1, 2$ and use definition (2.21) for either, that is,

$$\mathbf{y}^{\mathrm{T}}\left(\frac{1}{\alpha_1}\mathbf{x}_1^* - \frac{1}{\alpha_1}\mathbf{x}_1\right) \leq 0 \quad \forall \frac{1}{\alpha_1}\mathbf{x}_1^* \in C$$
$$\mathbf{y}^{\mathrm{T}}\left(\frac{1}{\alpha_2}\mathbf{x}_2^* - \frac{1}{\alpha_2}\mathbf{x}_2\right) \leq 0 \quad \forall \frac{1}{\alpha_2}\mathbf{x}_2^* \in C\,.$$

Summation of the inequalities, while multiplying by α_1 and α_2 respectively, yields

$$\mathbf{y}^{\mathrm{T}}(\mathbf{x}_1^* + \mathbf{x}_2^* - \mathbf{x}_1 - \mathbf{x}_2) \leq 0 \quad \forall \mathbf{x}^* = \mathbf{x}_1^* + \mathbf{x}_2^* \in (\alpha_1 + \alpha_2)C\,,$$

where property (2.20) has been used. Repeated summation gives

$$\mathbf{y}^{\mathrm{T}}\left(\sum_i \mathbf{x}_i^* - \sum_i \mathbf{x}_i\right) \leq 0 \quad \forall \mathbf{x}^* = \sum_i \mathbf{x}_i^* \in \sum_i \alpha_i C = C\left(\sum_i \alpha_i\right),$$

which is equivalent to

$$\mathbf{y} \in \mathcal{N}_{C\left(\sum_i \alpha_i\right)}\left(\sum_i \mathbf{x}_i\right).$$

If $\alpha_i = 0$ for some i, then it holds that $C(\alpha_i) = C(0) = \{0\}$ and $\mathbf{x}_i = 0$ whereas $\mathcal{N}_{\{0\}}(0) = \mathbb{R}^n$. One easily verifies that the proposition still holds. □

Proposition 2.4 may also be written in integral form as

$$\mathbf{y} \in \mathcal{N}_{C(\alpha(t))}\big(\mathbf{x}(t)\big)\ \forall t \in \mathcal{I} \subseteq \mathbb{R} \implies \mathbf{y} \in \mathcal{N}_{C\left(\int_{\mathcal{I}} \alpha(t)\mathrm{d}t\right)}\left(\int_{\mathcal{I}} \mathbf{x}(t)\,\mathrm{d}t\right), \quad (2.22)$$

where $\alpha(t) \geq 0$ and where $\mathbf{y}$ is a fixed value, i.e., does not depend on t.

— 3 —

Introduction to Differential Geometry

This chapter gives a brief introduction to differential geometry. The concepts of differential geometry are developed only to the extent needed to formulate the mechanical theory in the subsequent chapters. This exposition is mainly based on Jeffrey M. Lee [73], John M. Lee [74] and Abraham and Marsden [1], which are recommended as general references for differential geometry.

3.1. Preliminaries from Multivariate Calculus

For a function $f : \mathbb{R}^n \to \mathbb{R},\ \mathbf{x} \mapsto f(\mathbf{x})$, the i-th *partial derivative* at a point $\bar{\mathbf{x}}$ is defined as

$$\frac{\partial f}{\partial x^i}\Big|_{\mathbf{x}=\bar{\mathbf{x}}} := \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} f(\bar{\mathbf{x}} + t\,\mathbf{e}_i)\,, \tag{3.1}$$

where $\mathbf{e}_i \in \mathbb{R}^n$ is the vector having all components equal to zero except the i-th, which is one. Considering a second function $g : \mathbb{R}^n \to \mathbb{R}$, the partial derivative satisfies the *product rule*

$$\frac{\partial (fg)}{\partial x^i}\Big|_{\mathbf{x}=\bar{\mathbf{x}}} = g(\bar{\mathbf{x}})\frac{\partial f}{\partial x^i}\Big|_{\mathbf{x}=\bar{\mathbf{x}}} + f(\bar{\mathbf{x}})\frac{\partial g}{\partial x^i}\Big|_{\mathbf{x}=\bar{\mathbf{x}}}\,.$$

A vector-valued function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ can be seen as a collection of m functions $f^j : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbf{f}(\mathbf{x}) = \big(f^1(\mathbf{x}), \ldots, f^m(\mathbf{x})\big)$. The *Jacobian matrix* of $\mathbf{f}$ is the matrix

$$\Big(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big)^j{}_i := \frac{\partial f^j}{\partial x^i}\,. \tag{3.2}$$

For the composition of $\mathbf{f}$ with a vector-valued function $\boldsymbol{g} : \mathbb{R}^m \to \mathbb{R}^l,\ \mathbf{y} \mapsto \boldsymbol{g}(\mathbf{y})$, the *chain rule* for partial derivatives holds in the form

$$\frac{\partial (g^k \circ \mathbf{f})}{\partial x^i}\Big|_{\mathbf{x}=\bar{\mathbf{x}}} = \frac{\partial g^k}{\partial y^j}\Big|_{\mathbf{y}=\mathbf{f}(\bar{\mathbf{x}})}\frac{\partial f^j}{\partial x^i}\Big|_{\mathbf{x}=\bar{\mathbf{x}}}\,,$$

which in vector-matrix notation can be written as

$$\frac{\partial (\boldsymbol{g} \circ \mathbf{f})}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\bar{\mathbf{x}}} = \frac{\partial \boldsymbol{g}}{\partial \mathbf{y}}\Big|_{\mathbf{y}=\mathbf{f}(\bar{\mathbf{x}})}\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\Big|_{\mathbf{x}=\bar{\mathbf{x}}}\,.$$

3.2. Smooth Manifolds

In this section, the notion of a smooth manifold is introduced. These manifolds can be thought of as sets which locally look like $\mathbb{R}^n$ and which are equipped with a differentiable structure, such that concepts of multivariate calculus can be transferred to manifolds.

Definition 3.1. Let $\mathcal{M}$ be a set. A map $\phi : \mathcal{M} \supseteq U \to \mathbb{R}^n$ is a *local chart of* $\mathcal{M}$ if ϕ is a bijection from U to an open subset of $\mathbb{R}^n$. An *atlas* on $\mathcal{M}$ is a family $A = \{(U_i, \phi_i) \mid i = 1, 2, \dots\}$ of charts which covers $\mathcal{M}$ in the sense that $\mathcal{M} = \bigcup_i U_i$. Herein, the notation (U_i, ϕ_i) is used to emphasize the domain U_i of the local chart.

A chart (U, ϕ) is defined *around (or in the neighborhood of)* p if $p \in U$. Let (U, ϕ) be the chart

$$\phi : U \to \mathbb{R}^n,\ p \mapsto (q^1, \dots, q^n)\,.$$

The components q^i of $\phi(p)$ are called *coordinates of the point* p *with respect to* ϕ. The *i-th coordinate map of* ϕ is defined as $\phi^i : \mathcal{M} \to \mathbb{R},\ p \mapsto q^i$. Mostly, the i-th coordinate map is also denoted as q^i, such that $q^i(p) = q^i$.

If the domains of two charts (U_i, ϕ_i) and (U_j, ϕ_j) overlap, i.e., $U_i \cap U_j \neq \emptyset$, we can introduce the *change of coordinates* $\phi_i \circ \phi_j^{-1}|_{\phi_j(U_i \cap U_j)}$ and $\phi_j \circ \phi_i^{-1}|_{\phi_i(U_i \cap U_j)}$, where the restriction is necessary to guarantee that the image of the inverse of one chart lies in the domain of the other. However, the restriction is mostly dropped in the notation. It is easy to see, that $(\phi_i \circ \phi_j^{-1})^{-1} = \phi_j \circ \phi_i^{-1}$.

Definition 3.2. An atlas A on $\mathcal{M}$ is called *smooth*, if for any two charts in A, the change of coordinates is smooth (infinitely differentiable). Moreover, such an atlas A is called *maximal* if there is no other smooth atlas containing A.

Definition 3.3. An *n-dimensional smooth (differentiable) manifold* is a set $\mathcal{M}$ together with a smooth maximal atlas $A_{\max}$ containing charts mapping to $\mathbb{R}^n$, that is, $\phi : \mathcal{M} \supseteq U \to \mathbb{R}^n$ for all $(U, \phi) \in A_{\max}$. The maximal atlas $A_{\max}$ is said to define a *differentiable structure* on $\mathcal{M}$.

A maximal smooth atlas $A_{\max}$ is associated to a smooth atlas A by the following rule. A chart (U, ϕ) is defined to be in $A_{\max}$ if the change of coordinates between ϕ and any chart contained in A is smooth. Any two atlases on $\mathcal{M}$ are *equivalent* if they induce the same maximal atlas on $\mathcal{M}$. These considerations allow to specify a smooth manifold by introducing a set $\mathcal{M}$ together with a smooth atlas A on $\mathcal{M}$, where the differentiable structure on $\mathcal{M}$ is defined by the maximal atlas associated to A. For brevity, the differentiable structure is said to be defined by A.

Example 3.1. The following are examples of n-dimensional smooth manifolds.

(i) Any open subset $U \subseteq \mathbb{R}^n$, where the differentiable structure is defined by $A = \{(U, \mathrm{id})\}$ with "id" being the identity map on U.

(ii) The n-sphere $S^n = \{\mathbf{x} \in \mathbb{R}^{n+1} \mid \mathbf{x}^\mathsf{T}\mathbf{x} = 1\}$ with an atlas containing two different stereographic projections, see Example 1.39 in [73].

For convenience, the (maximal) smooth atlas defining the differentiable structure of a manifold $\mathcal{M}$ is often not mentioned explicitly and statements like "for all charts of $\mathcal{M}$" must be understood as "for all charts in the maximal atlas defining the differentiable structure on $\mathcal{M}$". The differentiable structure on a smooth manifold allows to transport the concept of differentiability of maps from $\mathbb{R}^n$ to manifolds.

Definition 3.4. Let $\mathcal{M}$ and $\mathcal{N}$ be smooth manifolds with respective dimensions m and n. A map $f : \mathcal{M} \to \mathcal{N}$ is *differentiable (of class C^k)*, if for every $p \in \mathcal{M}$ there exist charts (U, ϕ) around p and (V, ψ) around $f(p)$ such that the chart representation of f given by

$$\psi \circ f \circ \phi^{-1} : \mathbb{R}^m \supseteq \phi\big(U \cap f^{-1}(V)\big) \to \psi(f(U) \cap V) \subseteq \mathbb{R}^n \tag{3.3}$$

is differentiable (of class C^k) at $\phi(p)$.

This definition is well-defined thanks to the differentiable structures on the manifolds $\mathcal{M}$ and $\mathcal{N}$. Indeed, if (3.3) is differentiable, then the representation $\bar{\psi} \circ f \circ \bar{\phi}^{-1}$ using two different charts is still differentiable due to the smoothness of the change of coordinates and the fact that

$$\bar{\psi} \circ f \circ \bar{\phi}^{-1} = (\bar{\psi} \circ \psi^{-1}) \circ \psi \circ f \circ \phi^{-1} \circ (\phi \circ \bar{\phi}^{-1}) \,. \tag{3.4}$$

The *rank k of f at p* is defined as the rank of the Jacobian matrix of any chart representation of f, specifically

$$k = \operatorname{rank} \frac{\partial(\psi \circ f \circ \phi^{-1})}{\partial \mathbf{q}}\bigg|_{\phi(p)} \,. \tag{3.5}$$

Computing the Jacobian matrix of (3.4) reveals that the rank defined by (3.5) is independent of the charts involved. To see this, the chain rule is used. Moreover, recall that the change of coordinates are bijections and their Jacobian matrices consequently have full rank. A map f is said to have *constant rank k* if the rank of f equals k for all points of its domain. A *diffeomorphism* is a bijective map $f : \mathcal{M} \to \mathcal{N}$ for which both f and f^{-1} are smooth.

The set of smooth maps $f : \mathcal{M} \to \mathcal{N}$ is denoted by $\mathrm{C}^\infty(\mathcal{M}; \mathcal{N})$. Moreover, the abbreviation $\mathrm{C}^\infty(\mathcal{M})$ is adopted for the set of real-valued functions $\mathrm{C}^\infty(\mathcal{M}; \mathbb{R})$. In this work, whenever the differentiability of a map is not explicitly mentioned, the map is assumed to be smooth.

Definition 3.5. Let $\mathcal{M}$ be a smooth manifold of dimension n. A subset $\mathcal{N} \subset \mathcal{M}$ is *an $(n-k)$-dimensional submanifold of $\mathcal{M}$* if any of the two equivalent conditions is fulfilled.

(i) For every point $p \in \mathcal{N}$ there is a chart (U, ϕ) of $\mathcal{M}$ with $p \in U$ such that

$$\phi(U \cap \mathcal{N}) = \phi(U) \cap (\mathbb{R}^{n-k} \times \{\mathbf{c}\}) \quad \text{for some } \mathbf{c} \in \mathbb{R}^k . \tag{3.6}$$

(ii) For every point $p \in \mathcal{N}$ there is smooth map $\boldsymbol{g} : U \to \mathbb{R}^k$ of constant rank k with $p \in U$ such that

$$U \cap \mathcal{N} = \boldsymbol{g}^{-1}(\mathbf{c}) \quad \text{for some } \mathbf{c} \in \mathbb{R}^k . \tag{3.7}$$

Usually, $\mathbf{c}$ is chosen to be 0, which can always be accomplished by composition with a translation in $\mathbb{R}^n$.

To show the equivalence of (i) and (ii), first assume that the submanifold $\mathcal{N}$ is defined by (i). By condition (3.6), the coordinate maps q^i of ϕ have the property that $q^{n-k+i}(p) = c^i$ with $i = 1, \dots, k$ for any point $p \in U \cap \mathcal{N}$, where c^i denote the components of $\mathbf{c}$. Hence, it suffices to define $\boldsymbol{g} = (q^{n-k+1}, \dots, q^n)$ to satisfy condition (3.7). On the other hand, assume that the submanifold $\mathcal{N}$ is defined by (ii). Choose any chart $(\bar{U}, \bar{\phi})$ of $\mathcal{M}$, then by (3.7) it holds that $\boldsymbol{g} \circ \bar{\phi}^{-1}(\bar{q}^1, \dots, \bar{q}^n) - \mathbf{c} = 0$ for all $\bar{\mathbf{q}} \in \bar{\phi}(\bar{U} \cap \mathcal{N})$. By the implicit function theorem there exist local maps f^i such that on a subset of $V \subseteq \bar{\phi}(\bar{U} \cap \mathcal{N})$ $\bar{q}^{n-k+i} = f^i(\bar{q}^1, \dots, \bar{q}^{n-k})$ with $i = 1, \dots, k$. Hence, we can construct a chart (U, ϕ) with $\phi(U) \subseteq V$ and coordinate maps q^i that satisfy (3.6) by setting

$$\begin{aligned} q^i &= \bar{q}^i && \text{for } i = 1, \dots, n-k \\ q^{n-k+i} &= \bar{q}^{n-k+i} - f^i(\bar{q}^1, \dots, \bar{q}^{n-k}) && \text{for } i = 1, \dots, k . \end{aligned}$$

Example 3.2.

(i) The n-sphere $S^n = \left\{ \mathbf{x} \in \mathbb{R}^{n+1} \;\middle|\; \mathbf{x}^{\mathrm{T}}\mathbf{x} = 1 \right\}$ is an n-dimensional submanifold of $\mathbb{R}^n$ because $S^n = g^{-1}(1)$ for $g : \mathbb{R}^{n+1} \setminus \{0\} \to \mathbb{R},\ \mathbf{x} \mapsto g(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{x}$.

(ii) The graph of any smooth map $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^n$ defined as

$$\operatorname{graph}(\mathbf{f}) = \left\{ (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{m+n} \;\middle|\; \mathbf{y} = \mathbf{f}(\mathbf{x}) \right\}$$

is a submanifold of $\mathbb{R}^{m+n}$ because $\phi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y} - \mathbf{f}(\mathbf{x}))$ is a global chart of $\mathbb{R}^{m+n}$ fulfilling (3.6) for $\mathbf{c} = 0$.

3.3. Tangent and Cotangent Space

There are several approaches to defining tangent vectors and the tangent space to a manifold, see for example Chapter 3 in John M. Lee [74]. In this work, the

curve approach is used to define tangent vectors, as the author is convinced that it is the most intuitive definition. However, its relation to the *derivation approach* and the *coordinate approach* are discussed, since each of these approaches has its merits. Interpreting a tangent vector as a derivation, for instance, provides the most efficient approach for computations.

To arrive at the concept of tangent vector to a point on a general manifold, it is instructive to first consider the case of submanifolds of $\mathbb{R}^m$. The simplest case are one-dimensional submanifolds, that is, images of curves in $\mathbb{R}^m$. Consider a map $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathbb{R}^m$ and define

$$\dot{\gamma}(0) \coloneqq \left.\frac{\mathrm{d}\gamma}{\mathrm{d}t}\right|_{t=0} = \lim_{h \downarrow 0} \frac{\gamma(h) - \gamma(0)}{h} \,. \tag{3.8}$$

Imagining vectors as arrows, the vector $\Delta\gamma = \gamma(h) - \gamma(0)$ can be pictured as the arrow pointing from the tip of $\gamma(0)$ to the tip of $\gamma(h)$, see Figure 3.1. With that

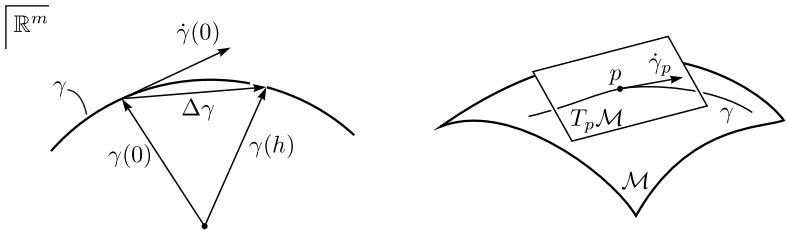


Figure 3.1: Pictorial interpretation of a tangent vector to a curve γ in $\mathbb{R}^m$ and to a manifold $\mathcal{M}$.

picture in mind, it can be seen that $\Delta\gamma/h$ in the limit $h \to 0$ is tangent to the curve at the point $\gamma(0)$. Hence, we can define $\dot{\gamma}(0)$ to be the tangent vector at the point $\gamma(0)$.

Now, an n-dimensional smooth submanifold $\mathcal{M} \subseteq \mathbb{R}^m$ is considered. A vector $X_p \in \mathbb{R}^m$ is a tangent vector to $\mathcal{M}$ at a point $p \in \mathcal{M}$ if we can find a curve $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ with $\gamma(0) = p$ such that $\dot{\gamma}(0) = X_p$. Clearly, there is more than one curve defining a tangent vector X_p because it suffices that two curves γ_1 and γ_2 with $\gamma_1(0) = \gamma_2(0) = p$ coincide in the vicinity of p to define the same tangent vector. Indeed, if there is an $\varepsilon > 0$ such that $\gamma_1(t) = \gamma_2(t)$ for all $t \in [-\varepsilon, \varepsilon]$, then $\dot{\gamma}_1(0) = \dot{\gamma}_2(0)$. Two curves are said to be equivalent, if they define the same tangent vector. Hence, the equivalence class $[\gamma]$, i.e., the set of curves which are equivalent to γ, is in one-to-one correspondence with the tangent vector $X_p = \dot{\gamma}(0)$ defined by γ.

The fact that the manifold $\mathcal{M}$ is a subset of $\mathbb{R}^m$ has been used to give meaning to the difference $\gamma(h) - \gamma(0)$ arising in the derivative $\dot{\gamma}$, see (3.8). Since on a

general manifold the difference of two points is not defined, the tangent vector to a curve cannot be defined by a difference quotient. However, two curves can still be defined to be equivalent if they coincide in the vicinity of a point, which is the same as the following.

Let $\mathcal{M}$ be a smooth manifold. Two curves $\gamma_i : \mathbb{R} \supseteq \mathcal{I}_i \to \mathcal{M}$ are *equivalent in* $p \in \mathcal{M}$ if $\gamma_1(0) = \gamma_2(0) = p$ and

$$\frac{\mathrm{d}(\phi \circ \gamma_1)}{\mathrm{d}t}\bigg|_{t=0} = \frac{\mathrm{d}(\phi \circ \gamma_2)}{\mathrm{d}t}\bigg|_{t=0} \tag{3.9}$$

for some chart (U, ϕ) defined in the neighborhood of p. To denote that the two curves are equivalent in p, the notation $\gamma_1 \sim_p \gamma_2$ is used. It remains to check that the equivalence relation established by (3.9) is chart independent. Therefore, a different chart $\bar{\phi}$ is used to compute

$$\frac{\mathrm{d}(\bar{\phi} \circ \gamma_i)}{\mathrm{d}t}\bigg|_{t=0} = \frac{\partial\big(\bar{\phi} \circ \phi^{-1}\big)}{\partial \mathbf{q}}\bigg|_{\phi(p)} \frac{\mathrm{d}(\phi \circ \gamma_i)}{\mathrm{d}t}\bigg|_{t=0},$$

where the change of coordinates $\bar{\phi} \circ \phi^{-1}$ as well as the chain rule have been used. It follows immediately that if two curves satisfy (3.9) for a chart ϕ, they do so also for any other chart $\bar{\phi}$. The *equivalence class* $[\gamma]$ denotes the set of curves which are equivalent to a curve γ in p. It follows directly, that $[\gamma_i] = [\gamma]$ for all $\gamma_i \in [\gamma]$. Hence, the same equivalence class can be represented by any curve it contains and an element of $[\gamma]$ is consequently called a *representative* of the class.

Definition 3.6. Let $\mathcal{M}$ be a manifold. A *tangent vector* to $\mathcal{M}$ at a point $p \in \mathcal{M}$ is an equivalence class with respect to $\sim_p$ of curves in $\mathcal{M}$ through p. The set of all tangent vectors at p is called the *tangent space* $T_p\mathcal{M}$ *of* $\mathcal{M}$ *at* p.

In other words, the tangent vector at a point p is the set containing all curves through p that have the same tangent vector in a chart, see (3.9).

The tangent space $T_p\mathcal{M}$ is equipped with a vector space structure over $\mathbb{R}$ by defining that the equivalence class $[\gamma] = a^1[\gamma_1] + a^2[\gamma_2]$ contains the curves γ satisfying

$$\frac{\mathrm{d}(\phi \circ \gamma)}{\mathrm{d}t}\bigg|_{t=0} = a^1 \frac{\mathrm{d}(\phi \circ \gamma_1)}{\mathrm{d}t}\bigg|_{t=0} + a^2 \frac{\mathrm{d}(\phi \circ \gamma_2)}{\mathrm{d}t}\bigg|_{t=0} \tag{3.10}$$

for some chart (U, ϕ) around p, where $a^1, a^2 \in \mathbb{R}$. Moreover, it can be shown that the tangent space $T_p\mathcal{M}$ at any point p of an n-dimensional smooth manifold has dimension n.[1]

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ denote the standard basis of $\mathbb{R}^n$. It is straightforward to see, that the curves

$$\beta_i^{\phi}(t) \coloneqq \phi^{-1}\big(\phi(p) + t\mathbf{e}_i\big) \tag{3.11}$$

[1] See John M. Lee [74], Proposition 3.10 or Winandy [127], Theorem 3.31.

induced by a chart ϕ around $p \in \mathcal{M}$ are representatives of n linearly independent tangent vectors $[\beta_i^\phi] \in T_p\mathcal{M}$, which consequently form a basis for $T_p\mathcal{M}$. This basis is called the *induced basis of* ϕ. Indeed, one can use (3.10) to see that for any $X_p = [\gamma] \in T_p\mathcal{M}$ it holds that

$$X_p = X_p^i[\beta_i^\phi] \quad \Leftrightarrow \quad \frac{\mathrm{d}(\phi \circ \gamma)}{\mathrm{d}t}\Big|_{t=0} = X_p^i \frac{\mathrm{d}(\phi \circ \beta_i^\phi)}{\mathrm{d}t}\Big|_{t=0} = X_p^i \, \mathbf{e}_i \,. \tag{3.12}$$

The derivative of a function $f \in \mathrm{C}^\infty(\mathcal{M})$ in the direction of the tangent vector $[\gamma] \in T_p\mathcal{M}$ is called the *Lie derivative at* p *of* f *in the direction* $[\gamma]$ and is defined by

$$L_{[\gamma]} f \coloneqq \frac{\mathrm{d}(f \circ \gamma)}{\mathrm{d}t}\Big|_{t=0} \,. \tag{3.13}$$

The Lie derivative, seen as a map $L_{[\gamma]} : \mathrm{C}^\infty(\mathcal{M}) \to \mathbb{R}$ has the properties of a *derivation in* p, which are that $L_{[\gamma]}$ is a linear operator which is local, i.e., only depends on the value of f in the vicinity of p, and $L_{[\gamma]}$ satisfies the product rule

$$L_{[\gamma]}(fg) = g(p) L_{[\gamma]} f + f(p) L_{[\gamma]} g \,, \tag{3.14}$$

where $f, g \in \mathrm{C}^\infty(\mathcal{M})$ and $fg : \mathcal{M} \to \mathbb{R},\ p \mapsto f(p)g(p)$.

It can be shown that $[\gamma] \mapsto L_{[\gamma]}$ is an isomorphism between the tangent space $T_p\mathcal{M}$ and the vector space of all derivations in $p \in \mathcal{M}$. Hence, a tangent vector $[\gamma]$ can be understood as a derivation that can act on a function f by $L_{[\gamma]} f$. Due to the aforementioned isomorphism, one can equivalently pursue the *derivation approach* to tangent vectors and define the tangent space as the vector space of all derivations in a point of the manifold.

It is notationally efficient to overlap the two notions of tangent vector. Specifically, for a tangent vector $X_p \in T_p\mathcal{M}$ it makes sense to write $\gamma \in X_p$ as it is an equivalence class of curves. Then γ is a representative of X_p and it holds that $X_p = [\gamma]$. However, the tangent vector X_p may also act as a derivation on a function $f \in \mathrm{C}^\infty(\mathcal{M})$, allowing to write $X_p(f)$, which must be understood as $X_p(f) \coloneqq L_{[\gamma]} f$ for any $\gamma \in X_p$. Moreover, the derivative-notation is resurrected for a tangent vector by setting

$$\dot{\gamma}_p \coloneqq \frac{\mathrm{d}\gamma}{\mathrm{d}t}\Big|_{t=0} \coloneqq [\gamma]. \tag{3.15}$$

This means that taking the derivative of γ is interpreted as the operation of assigning the tangent vector $[\gamma]$ to the curve γ, which closes the circle to the picture of a tangent vector created by looking at submanifolds of $\mathbb{R}^m$, see Figure 3.1.

By the isomorphism $[\gamma] \mapsto L_{[\gamma]}$, the basis $[\beta_i^\phi]$ induced by a chart ϕ around a point $p \in \mathcal{M}$ is uniquely associated to the basis derivation

$$\frac{\partial}{\partial q^i}\Big|_p \coloneqq L_{[\beta_i^\phi]} \,, \tag{3.16}$$

where q^i denotes the i-th coordinate map of ϕ. Again, rather than two different objects, $[\beta_i^\phi]$ and $\frac{\partial}{\partial q^i}\big|_p$ must be seen as two interpretations of the same tangent vector. The notation in (3.16) is motivated applying (3.16) to a function $f \in \mathrm{C}^\infty(\mathcal{M})$, which yields

$$\frac{\partial}{\partial q^i}\Big|_p (f) = \frac{\partial (f \circ \phi^{-1})}{\partial q^i}\Big|_{\phi(p)}, \tag{3.17}$$

where (3.11), (3.13) and the definition of the partial derivative (3.1) have been used. To shorten notation, it is convenient to introduce

$$\frac{\partial f}{\partial q^i}\Big|_p \coloneqq \frac{\partial}{\partial q^i}\Big|_p (f)\,.$$

By our notational overlap and in view of (3.12), a tangent vector $X_p \in T_p\mathcal{M}$ has the basis representation

$$X_p = X_p^i \frac{\partial}{\partial q^i}\Big|_p \quad \text{with} \quad X_p^i = X_p(q^i)\,, \tag{3.18}$$

where q^i denotes the i-th coordinate map of ϕ. Indeed, the component X_p^j can be computed by

$$X_p(q^j) = X_p^i \frac{\partial q^j}{\partial q^i}\Big|_p = X_p^i \frac{\partial (q^j \circ \phi^{-1})}{\partial q^i}\Big|_{\phi(p)} = X_p^i\, \delta_i^j = X_p^j\,,$$

where (3.18) as well as (3.17) have been invoked. The rule for the coordinate transformation of the components of a vector under the change of induced bases follows immediately. In fact, consider a different chart $\bar\phi$ around p with coordinate maps $\bar q^j$ and

$$X_p = X_p^i \frac{\partial}{\partial q^i}\Big|_p = \bar X_p^j \frac{\partial}{\partial \bar q^j}\Big|_p\,,$$

then the components with respect to the induced basis of $\bar\phi$ result from

$$\bar X_p^j = X_p(\bar q^j) = X_p^i\, \frac{\partial \bar q^j}{\partial q^i}\Big|_p = X_p^i\, \frac{\partial (\bar q^j \circ \phi^{-1})}{\partial q^i}\Big|_{\phi(p)}\,. \tag{3.19}$$

Hence, the components of a tangent vector transform linearly under a change of induced bases. It is the transformation rule (3.19) that gives access to the *coordinate approach* to tangent vectors. Herein, a tangent vector is defined as an equivalence class of chart dependent assignments of $\mathbb{R}^n$ tuples, where two tuples $(X_p^1, \dots, X_p^n)$ and $(\bar X_p^1, \dots, \bar X_p^n)$ assigned to the charts ϕ and $\bar\phi$, respectively, are considered equivalent, and hence represent the same tangent vector, if they fulfill (3.19).

Example 3.3. An example and a counter example for the coordinate approach to tangent vectors are given. For both examples consider a curve $\gamma : \mathbb{R} \to \mathcal{M}$ with $\gamma(0) = p$.

(i) For any two charts ϕ and $\bar{\phi}$, the $\mathbb{R}^n$ tuples with components $(q^i \circ \gamma)\dot{}(0)$ and $(\bar{q}^i \circ \gamma)\dot{}(0)$ are equivalent and hence define the same tangent vector. In fact,

$$(\bar{q}^j \circ \gamma)\dot{}(0) = (\bar{q}^j \circ \phi^{-1} \circ \phi \circ \gamma)\dot{}(0) = \frac{\partial(\bar{q}^j \circ \phi^{-1})}{\partial q^i}\bigg|_{\phi(p)} (q^i \circ \gamma)\dot{}(0)\,,$$

where the chain rule has been used. In view of (3.19) this example establishes the relationship to the curve approach to tangent vectors as $L_{[\gamma]}(q^i) = (q^i \circ \gamma)\dot{}(0)$ and $L_{[\gamma]}(\bar{q}^i) = (\bar{q}^i \circ \gamma)\dot{}(0)$.

(ii) Contrary to (i), for any two charts ϕ and $\bar{\phi}$, the $\mathbb{R}^n$ tuples with components $(q^i \circ \gamma)\ddot{}(0)$ and $(\bar{q}^i \circ \gamma)\ddot{}(0)$ generally define different tangent vectors. Indeed, by a similar reasoning as in (i) one arrives at

$$(\bar{q}^j \circ \gamma)\ddot{}(0) = \frac{\partial(\bar{q}^j \circ \phi^{-1})}{\partial q^i}(q^i \circ \gamma)\ddot{}(0) + \frac{\partial(\bar{q}^j \circ \phi^{-1})}{\partial q^k \partial q^l}(q^k \circ \gamma)\dot{}(0)\,(q^l \circ \gamma)\dot{}(0)\,,$$

where the partial derivatives are evaluated in $\phi(p)$. Hence, there is no natural way to define the "acceleration vector" of a curve on a manifold. This obstruction is resolved by introducing an additional structure on a manifold called connection, see Section 4.2.

Definition 3.7. Let $f : \mathcal{M} \to \mathcal{N}$ be a differentiable map between two manifolds $\mathcal{M}$ and $\mathcal{N}$. The *differential of f in $p \in \mathcal{M}$* is the map

$$\mathrm{D}f_p : T_p\mathcal{M} \to T_{f(p)}\mathcal{N},\ X_p \mapsto \mathrm{D}f_p X_p$$

equivalently defined by either

$$\mathrm{D}f_p X_p(g) \coloneqq X_p(g \circ f)\ \forall g \in \mathrm{C}^\infty(\mathcal{N}) \quad \text{or} \quad \mathrm{D}f_p X_p \coloneqq (f \circ \gamma)\dot{}_p,\ \gamma \in X_p\,, \tag{3.20}$$

depending on whether the tangent vector is interpreted as derivation in p or as equivalence class of curves. Recall that by convention $(f \circ \gamma)\dot{}_p = [f \circ \gamma]$.

Definition 3.8. Let $\mathcal{M}$ be a manifold. The *cotangent space $T_p^*\mathcal{M}$ of $\mathcal{M}$ at p* is the dual space of the tangent space, that is $T_p^*\mathcal{M} \coloneqq (T_p\mathcal{M})^*$.

The dual basis to the basis $\frac{\partial}{\partial q^1}\big|_p, \ldots, \frac{\partial}{\partial q^n}\big|_p$ of the tangent space is denoted as $\mathrm{d}q_p^1, \ldots, \mathrm{d}q_p^n$. It follows from Proposition 2.1 that the representation of a covector $\omega_p \in T_p^*\mathcal{M}$ is

$$\omega_p = \omega_i\, \mathrm{d}q_p^i \quad \text{with} \quad \omega_i = \omega_p\left(\frac{\partial}{\partial q^i}\bigg|_p\right). \tag{3.21}$$

3.4. Vector Bundles

The tangent space $T_p\mathcal{M}$ to a point p of a manifold $\mathcal{M}$ is often thought to be attached to the manifold at the point p. This picture is translated into mathematical language by the following.

Definition 3.9. The *tangent bundle* $T\mathcal{M}$ of a smooth n-dimensional manifold $\mathcal{M}$ is

$$T\mathcal{M} \coloneqq \bigcup_{p\in\mathcal{M}} \left(\{p\} \times T_p\mathcal{M}\right),$$

together with the *natural projection* $\pi : T\mathcal{M} \to \mathcal{M},\ (p, X_p) \mapsto p$. The tangent bundle is a smooth manifold of dimension $2n$ in virtue of the smooth atlas containing the *natural charts* defined by

$$\Phi : \pi^{-1}(U) \to \mathbb{R}^{2n},\ \left(p, X_p = u^i \tfrac{\partial}{\partial q^i}\big|_p\right) \mapsto (q^1, \dots, q^n, u^1, \dots, u^n), \tag{3.22}$$

which are induced by the charts (U, ϕ) of $\mathcal{M}$, where $\phi(p) = (q^1, \dots, q^n)$.

Instead of looking at a particular tangent space to a point of $\mathcal{M}$, the tangent bundle allows to treat all tangent spaces at once. For instance, the differential of a map f at p can be extended to a differential "at every p".

Definition 3.10. Let $f : \mathcal{M} \to \mathcal{N}$ be a differentiable map between two manifolds $\mathcal{M}$ and $\mathcal{N}$. The *differential of* f is the map

$$\mathrm{D}f : T\mathcal{M} \to T\mathcal{N},\ (p, X_p) \mapsto (f(p), \mathrm{D}f_p X_p)$$

Definition 3.11. Let $\mathcal{B}, \mathcal{F}$ and $\mathcal{M}$ be smooth manifolds and let $\pi : \mathcal{B} \to \mathcal{M}$ be a differentiable surjective map. The quadruple $(\mathcal{B}, \mathcal{M}, \pi, \mathcal{F})$ is called a *fiber bundle* if for each point $p \in \mathcal{M}$ there is a neighborhood $U \subseteq \mathcal{M}$ of p and a diffeomorphism $\theta : \pi^{-1}(U) \to U \times \mathcal{F}$ such that the diagram

$$\begin{array}{ccc} \pi^{-1}(U) & \xrightarrow{\theta} & U \times \mathcal{F} \\ {\scriptstyle\pi}\downarrow & \swarrow_{\mathrm{pr}_1} & \\ U & & \end{array}$$

commutes, that is, if $\mathrm{pr}_1 \circ \theta = \pi$, where pr_1 is the projection onto the first factor. The map θ is called *local trivialization* and π is the *bundle projection*. The manifolds are referred to as *total space* $\mathcal{B}$, *typical fiber* $\mathcal{F}$ and *base manifold* $\mathcal{M}$. The set $\mathcal{B}_p \coloneqq \pi^{-1}(p)$ is called the *fiber over* p.

Definition 3.12. A *section of a fiber bundle* $(\mathcal{B}, \mathcal{M}, \pi, \mathcal{F})$ is a smooth map $s : \mathcal{M} \supseteq U \to \mathcal{B}$ such that $\pi(s(p)) = p$ for all $p \in U$. The set of all sections of $\mathcal{B}$ is denoted by $\Gamma(\mathcal{B})$. Hence, a section selects an element $s(p) = (p, s_p)$ of the fiber $\mathcal{B}_p = \pi^{-1}(p)$ for every point $p \in U$.

Definition 3.13. A *vector bundle* of rank k over $\mathcal{M}$ is a fiber bundle $(\mathcal{B}, \mathcal{M}, \pi, \mathbb{R}^k)$ with typical fiber $\mathbb{R}^k$ such that all fibers $\mathcal{B}_p$ are k-dimensional vector spaces. Moreover, the restrictions of any trivialization θ to $\mathcal{B}_p$ is an isomorphism from $\mathcal{B}_p$ to $\{p\} \times \mathbb{R}^k \cong \mathbb{R}^k$.

Let (U, ϕ) be a chart of $\mathcal{M}$ and let $B_1, \ldots, B_k \in \Gamma(\mathcal{B})$ be local sections defined on $U \subseteq \mathcal{M}$, which pointwise form a basis for the fibers of the vector bundle $\mathcal{B}$. The associated *bundle chart* of $\mathcal{B}$ is

$$\Phi : \pi^{-1}(U) \to \mathbb{R}^{n+f}, \ \big(p, X_p = u^i (B_i)_p\big) \mapsto (q^1, \ldots, q^n, u^1, \ldots, u^f) \tag{3.23}$$

where π is the bundle projection and $\phi(p) = (q^1, \ldots, q^n)$. The collection of all such charts equips $\mathcal{B}$ with a differentiable structure.

It can be shown,[2] that the tangent bundle $T\mathcal{M}$ of an n-dimensional manifold $\mathcal{M}$ is a smooth vector bundle $(T\mathcal{M}, \mathcal{M}, \pi, \mathbb{R}^n)$. Using the same idea as for the construction of the tangent bundle, other vector spaces can be attached to the points of a smooth n-dimensional manifold $\mathcal{M}$ to create other smooth vector bundles over $\mathcal{M}$. The most important bundles are the *cotangent bundle*

$$T^*\mathcal{M} \coloneqq \bigcup_{p \in \mathcal{M}} \big(\{p\} \times T_p^*\mathcal{M}\big)$$

the *bundle of* (r, s)*-tensors*

$$\otimes_s^r T^*\mathcal{M} \coloneqq \bigcup_{p \in \mathcal{M}} \big(\{p\} \times \otimes_s^r T_p^*\mathcal{M}\big)$$

and the *bundle of alternating* k*-tensors*

$$\bigwedge^k T^*\mathcal{M} \coloneqq \bigcup_{p \in \mathcal{M}} \big(\{p\} \times \bigwedge^k T_p^*\mathcal{M}\big) .$$

Let $\mathcal{B}$ denote any of these bundles, then $\mathcal{B}$ comes together with the natural projection $\pi : \mathcal{B} \to \mathcal{M}, \ (p, X_p) \mapsto p$, where X_p is an element of the vector space attached to $p \in \mathcal{M}$.

3.5. Vector and Tensor Fields

For sections of vector bundles, the vector space structure of the fibers of $\mathcal{B}$ is transported to the sections of $\mathcal{B}$ in a pointwise fashion, i.e., for any $u, v \in \Gamma(\mathcal{B})$ and any $\alpha \in \mathbb{R}$, the vector space operations are defined by

$$(u + v)(p) \coloneqq (p, u_p + v_p) \quad \text{and} \quad (\alpha u)(p) \coloneqq (p, \alpha u_p) .$$

[2] See John M. Lee [74], Proposition 10.4.

By introducing the multiplication between $f \in \mathrm{C}^\infty(\mathcal{M})$ and $v \in \Gamma(\mathcal{B})$ as

$$(fv)(p) \coloneqq (p, f(p)v_p) \tag{3.24}$$

for all $p \in \mathcal{M}$, the section can be scaled differently in every point of the manifold.

Different names are used for sections of the bundles defined so far. A *vector field* is a section of $T\mathcal{M}$ and the notation $\mathrm{Vect}(\mathcal{M}) \coloneqq \Gamma(T\mathcal{M})$ is introduced. *Covector fields* are sections of $T^*\mathcal{M}$ and (k,l)*-tensor fields* sections of $\otimes_l^k T^*\mathcal{M}$. Finally, the sections of $\bigwedge^k T^*\mathcal{M}$ are called *differential* k*-forms* and $\Omega^k(\mathcal{M})$ is used to denote the set of all differential k-forms. By construction, the set of covector fields is $\Omega^1(\mathcal{M})$. Moreover, $\Omega^\star(\mathcal{M})$ is used to indicate the set of differential forms of arbitrary degree.

Since the notion of vectors and tensors have been extended to fields on a manifold, the objects and operations related to vectors and tensors encountered in Chapter 2 can be extended in a pointwise fashion.

The *duality pairing* of a vector field $X \in \mathrm{Vect}(\mathcal{M})$ and covector field $\omega \in \Omega^1(\mathcal{M})$ is pointwise defined by

$$\omega(X)(p) \coloneqq \omega_p(X_p)\,, \tag{3.25}$$

which yields a real number for every $p \in \mathcal{M}$ and therefore defines a scalar field on $\mathcal{M}$, i.e., $\omega(X) \in \mathrm{C}^\infty(\mathcal{M})$. Since a tangent vector in p can act on $f \in \mathrm{C}^\infty(\mathcal{M})$ as a derivation in p, we can pointwise define

$$X(f)(p) \coloneqq X_p(f) \tag{3.26}$$

implying $X(f) \in \mathrm{C}^\infty(\mathcal{M})$. Hence, a vector field X, interpreted as a map $X : \mathrm{C}^\infty(\mathcal{M}) \to \mathrm{C}^\infty(\mathcal{M})$, has the properties of a *derivation*, which by definition satisfies the product rule

$$X(fg) = g\,X(f) + f\,X(g) \tag{3.27}$$

for all $f, g \in \mathrm{C}^\infty(\mathcal{M})$. To see this, recall that $X_p(f) = L_{[\gamma]}$ for $\gamma \in X_p$ and use (3.14).

In view of (3.16), the *(local) basis fields* induced by a chart (U, ϕ) of $\mathcal{M}$ are defined to be the vector fields

$$\frac{\partial}{\partial q^i} : U \to T\mathcal{M},\ p \mapsto \left(p, \frac{\partial}{\partial q^i}\Big|_p\right).$$

Similarly, the *(local) dual basis fields* induced by the chart (U, ϕ) are introduced as the sections

$$\mathrm{d}q^i : U \to T^*\mathcal{M},\ p \mapsto \left(p,\, \mathrm{d}q^i_p\right),$$

where $\mathrm{d}q^1_p, \ldots, \mathrm{d}q^n_p$ is the dual basis of $\frac{\partial}{\partial q^1}\big|_p, \ldots, \frac{\partial}{\partial q^n}\big|_p$. Since sections can be multiplied by functions on $\mathcal{M}$, see (3.24), vector fields $X \in \mathrm{Vect}(\mathcal{M})$ and covector fields $\omega \in \Omega^1(\mathcal{M})$ can locally (on the domain U of the chart ϕ) be represented as

$$X = X^i \frac{\partial}{\partial q^i} \quad \text{and} \quad \omega = \omega_i \, \mathrm{d}q^i \,, \tag{3.28}$$

where the functions $X^i, \omega_i \in \mathrm{C}^\infty(\mathcal{M})$ are called the *component functions* of the (co-)vector field. These can be computed by

$$X^i = X(q^i) \quad \text{and} \quad \omega_i = \omega\left(\frac{\partial}{\partial q^i}\right), \tag{3.29}$$

where q^i is the i-th coordinate map of the chart ϕ. The evaluation of (3.28) and (3.29) in a point $p \in U$ is in harmony with (3.18) and (3.21).

The tensor product can be pointwise extended to fields. For instance,

$$T(p) = \left(T^i_j \frac{\partial}{\partial q^i} \otimes \mathrm{d}q^j\right)(p) = \left(p, T^i_j(p) \frac{\partial}{\partial q^i}\Big|_p \otimes \mathrm{d}q^j_p\right),$$

such that, in accordance with (3.25),

$$T(\eta, X) = T^i_j \, \eta\Big(\frac{\partial}{\partial q^i}\Big) \mathrm{d}q^j(X) \;\; \in \mathrm{C}^\infty(\mathcal{M})$$

for $X \in \mathrm{Vect}(\mathcal{M})$ and $\eta \in \Omega^1(\mathcal{M})$. The same procedure is used to define the wedge product of differential forms.

Consider the (1,1)-tensor field T from above. It is possible to see T as a *vector-valued one-form* if its first argument is "ignored" and T is interpreted as the map

$$T : T\mathcal{M} \to T\mathcal{M}, \;\; (p, X_p) \mapsto \left(p, T^i_j(p) \, \mathrm{d}q^j_p(X_p) \, \frac{\partial}{\partial q^i}\Big|_p\right).$$

For vector fields $X, Y \in \mathrm{Vect}(\mathcal{M})$ and $f \in \mathrm{C}^\infty(\mathcal{M})$ it makes sense to write $Y(X(f))$, since $X(f) \in \mathrm{C}^\infty(\mathcal{M})$. This allows for the following definition.

Definition 3.14. The map $[\cdot, \cdot] : \mathrm{Vect}(\mathcal{M}) \times \mathrm{Vect}(\mathcal{M}) \mapsto \mathrm{Vect}(\mathcal{M})$ defined by

$$[X, Y](f) = X(Y(f)) - Y(X(f)) \tag{3.30}$$

for all $f \in \mathrm{C}^\infty(\mathcal{M})$ is called the *Lie bracket* on $\mathcal{M}$.[3]

[3] To see that $[X, Y]$ is well defined by (3.30), that is, that $[X, Y]$ is indeed a vector field on $\mathcal{M}$, confer John M. Lee [74].

It is immediately clear from (3.30) that the Lie bracket is bilinear. Moreover, direct computations involving (3.30) reveal that the Lie bracket has the properties

$$[X,Y] = -[Y,X] \quad \text{and} \quad [X,fY] = X(f)Y + f[X,Y]\,, \tag{3.31}$$

where $f \in \mathrm{C}^\infty(\mathcal{M})$. Using the local representations $X = X^i \frac{\partial}{\partial q^i}$ and $Y = Y^i \frac{\partial}{\partial q^i}$ in (3.30) leads to the local description

$$[X,Y] = \left(X^i \frac{\partial Y^j}{\partial q^i} - Y^i \frac{\partial X^j}{\partial q^i} \right) \frac{\partial}{\partial q^j}\,, \tag{3.32}$$

where the product rule and the symmetry of second derivatives has been used.

Consider a differentiable map $f : \mathcal{M} \to \mathcal{N}$ between two manifolds $\mathcal{M}$ and $\mathcal{N}$. The differential of f as the map $Df : T\mathcal{M} \to T\mathcal{N}$ can be used to map vector fields $X \in \mathrm{Vect}(\mathcal{M})$, which as sections in $T\mathcal{M}$ are maps of the form $X : \mathcal{M} \to T\mathcal{M}$. The resulting map is $DfX : \mathcal{M} \to T\mathcal{N}$, which can be seen as a "vector field" on $\mathcal{N}$ depending however on the point on $\mathcal{M}$. In the case where f is a diffeomorphism, the differential DfX can be used to construct a vector field on $\mathcal{N}$ by defining

$$f_*X \coloneqq DfX \circ f^{-1}\,,$$

which is indeed a map from $\mathcal{N}$ to $T\mathcal{N}$. The vector field $f_*X \in \mathrm{Vect}(\mathcal{N})$ is called *pushforward of X with f*. Since as a diffeomorphism f is invertible, the *pullback of $Y \in \mathrm{Vect}(\mathcal{N})$ with f* can be defined as the vector field $f^*Y \coloneqq (f^{-1})_*Y$ on $\mathcal{M}$. Similarly, for a covector field $\eta \in \Omega^1(\mathcal{N})$ the *pullback of η with f* is introduced as the covector field $f^*\eta \in \Omega^1(\mathcal{M})$ which satisfies

$$f^*\eta(X) = \eta(f_*X) \circ f \tag{3.33}$$

for all $X \in \mathrm{Vect}(\mathcal{M})$. The *pushforward of $\omega \in \Omega^1(\mathcal{M})$ with f* is the covector field $f_*\omega \coloneqq (f^{-1})^*\omega$ on $\mathcal{N}$. With the notion of pushforward of (co-)vector fields at hand, the pullback of a tensor field is defined by a condition similar to (3.33). For any tensor field T on $\mathcal{N}$, the *pullback f^*T of T with f* is computed by evaluating T at the pushforwards of the arguments of f^*T. For instance, consider $T \in \Gamma(\otimes^1_2 T^*\mathcal{N})$, then

$$f^*T(\omega, X, Y) = T(f_*\omega, f_*X, f_*Y) \circ f$$

for all $X, Y \in \mathrm{Vect}(\mathcal{M})$ and all $\omega \in \Omega^1(\mathcal{M})$. The *pushforward* f_*S of a tensor S on $\mathcal{M}$ is given by $f_*S \coloneqq (f^{-1})^*S$. Finally, for tensor fields of rank zero, i.e., functions $g \in \mathrm{C}^\infty(\mathcal{N})$, the pullback is defined by $f^*g \coloneqq g \circ f$.

3.6. Vector Fields as Dynamical Systems

A curve $v : \mathbb{R} \supseteq \mathcal{I} \to T\mathcal{M}$ is a *vector field along* $\gamma : \mathcal{I} \to \mathcal{M}$, if $\pi \circ v = \gamma$, where $\pi : T\mathcal{M} \to \mathcal{M}$ is the natural projection of the tangent bundle $T\mathcal{M}$. Hence, a vector field along a curve γ looks like a vector field on $\mathcal{M}$ evaluated along γ. Indeed, for any $X \in \mathrm{Vect}(\mathcal{M})$, the curve $X \circ \gamma$ is a vector field along γ.

It is natural to agree that objects defined on $\mathcal{M}$ are evaluated along the curve γ on $\mathcal{M}$ whenever they interact with a vector field along γ. For instance, let $\theta \in \Omega^1(\mathcal{M})$ be a covector field on $\mathcal{M}$, then $\theta(v)$ denotes the scalar function defined on $\mathcal{I} \subseteq \mathbb{R}$ for which $\theta(v)(t) = \theta_{\gamma(t)}(v_{\gamma(t)})$. Similarly, the differential of a map $f : \mathcal{M} \to \mathcal{N}$ acting on v yields a vector field along $f \circ \gamma$, i.e.,

$$\mathrm{D}fv : \mathcal{I} \to T\mathcal{N},\ t \mapsto \left(f \circ \gamma(t), \mathrm{D}f_{\gamma(t)} v_{\gamma(t)}\right). \tag{3.34}$$

For any curve $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ the *tangent field along* γ is defined as

$$\dot\gamma : \mathcal{I} \to T\mathcal{M},\ t \mapsto \dot\gamma(t) = (\gamma(t), \dot\gamma_{\gamma(t)}),$$

where we recall that $\dot\gamma_{\gamma(t)}$ denotes the tangent vector at $\gamma(t)$ defined by the curve γ, that is, $\dot\gamma_{\gamma(t)}$ is the equivalence class $[\gamma] \in T_{\gamma(t)}\mathcal{M}$. In accordance with (3.15), also the notation $\frac{\mathrm{d}\gamma}{\mathrm{d}t}$ is used for $\dot\gamma$. This very suggestive notation is further justified by looking at the basis representation of $\dot\gamma$ with respect to a chart ϕ of $\mathcal{M}$. As a first step in that direction, consider $\dot\gamma$ acting as derivation on $f \in \mathrm{C}^\infty(\mathcal{M})$, then using (3.26) and (3.13)

$$\dot\gamma(f) = \dot\gamma_\gamma(f) = \frac{\mathrm{d}(f \circ \gamma)}{\mathrm{d}t} = (f \circ \gamma)\dot{}. \tag{3.35}$$

Let $q^i(t) \coloneqq q^i \circ \gamma(t)$ denote the components of the chart representation of γ with respect to ϕ, then by (3.28) and (3.29) the basis representation of $\dot\gamma$ with respect to the basis induced by ϕ is

$$\dot\gamma = (q^i \circ \gamma)\dot{} \frac{\partial}{\partial q^i}\bigg|_\gamma = \dot q^i \frac{\partial}{\partial q^i}\bigg|_\gamma, \tag{3.36}$$

where also (3.35) has been used.

Every vector field $X \in \mathrm{Vect}(\mathcal{M})$ defines a dynamical system, that is, a first order differential equation on $\mathcal{M}$, which is given by

$$\dot\gamma = X \circ \gamma. \tag{3.37}$$

A curve $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ is an *integral curve of* X if it satisfies (3.37). Let $X^i \in \mathrm{C}^\infty(\mathcal{M})$ be the component functions of X with respect to the basis fields induced by the chart ϕ on $\mathcal{M}$, see (3.28). The components of (3.36)

satisfying (3.37) must solve the differential equations $(q^i \circ \gamma)^{\cdot} = X^i \circ \gamma$, which can be written as the system of first order ordinary differential equations

$$\dot{\mathbf{q}} = \mathbf{X} \circ \phi^{-1}(\mathbf{q}) \,, \tag{3.38}$$

where again $q^i(t) \coloneqq q^i \circ \gamma(t)$, which gathered as a tuple reads as $\mathbf{q} \coloneqq \phi \circ \gamma$. Hence, a curve γ is an integral curve if and only if its chart representation solves the first order ODE (3.38).

By the Cauchy–Lipschitz theorem,[4] for every $\mathbf{q}_0 \in \mathbb{R}^n$ there exists a unique (local) solution $\mathbf{q}(t)$ of (3.38) with $\mathbf{q}(0) = \mathbf{q}_0$. Consequently, for every $p \in \mathcal{M}$ there exists a unique locally defined integral curve γ of X going through p. Since by uniqueness there is exactly one integral curve of X going through p, the integral curves of a vector field cannot intersect.

Let $\varphi^X(t, p)$ denote the unique (local) integral curve of X satisfying the initial condition $\varphi^X(0, p) = p$, then the map

$$\varphi_t^X : \mathcal{M} \supseteq U \to \mathcal{M}, \; p \mapsto \varphi^X(t, p)$$

is called the *(local) flow of* X. Since there is exactly one integral curve of X going through a specific point of $\mathcal{M}$, the flow of X is a (local) diffeomorphism and has the property

$$\varphi_{t_1+t_2}^X = \varphi_{t_1}^X \circ \varphi_{t_2}^X = \varphi_{t_2}^X \circ \varphi_{t_1}^X \,,$$

see Winandy [127], Theorem 3.27.

3.7. Lie Derivative

The flow φ_t^X of a vector field $X \in \mathrm{Vect}(\mathcal{M})$ is a (local) diffeomorphism. Hence, vector fields can be pulled back with the flow of X, which allows to inspect how a vector field Y changes along the flow of a vector field X. The same idea can be used for tensor fields.

Definition 3.15. Let $X, Y \in \mathrm{Vect}(\mathcal{M})$. The *Lie derivative of* Y *along* X is defined as the vector field $\mathcal{L}_X Y \in \mathrm{Vect}(\mathcal{M})$ pointwise given by

$$\mathcal{L}_X Y(p) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \big((\varphi_t^X)^* Y\big)(p) \,,$$

where ϕ_t^X is the flow of X through the point p with $\varphi_0^X(p) = p$. Likewise, for any tensor field T the *Lie derivative of* T *along* X is defined pointwise by

$$\mathcal{L}_X T(p) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \big((\varphi_t^X)^* T\big)(p) \,. \tag{3.39}$$

[4] A version of the Cauchy–Lipschitz theorem for smooth functions is given in John M. Lee [74], Theorem D.1.

For a function $f \in \mathrm{C}^\infty(\mathcal{M})$, seen as a tensor of rank zero, it is a consequence of (3.39) that

$$\mathcal{L}_X f(p) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \big(f \circ \varphi_t^X(p)\big) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \big(f \circ \gamma_p(t)\big) = X(f)(p)\,, \tag{3.40}$$

since $(\varphi_t^X)^* f = f \circ \varphi_t^X$ by definition of the pullback and where $\gamma_p(t) = \varphi_t^X(p)$ has been used to denote the integral curve of X with $\gamma_p(0) = p$. For a covariant tensor field F of rank k, the Lie derivative satisfies

$$\mathcal{L}_X F(Y_1, \ldots, Y_k) = \mathcal{L}_X(F(Y_1, \ldots, Y_k)) - \sum_{i=1}^{k} F(Y_1, \ldots, \mathcal{L}_X Y_i, \ldots, Y_k)\,,$$

where X, Y_i are vector fields on $\mathcal{M}$. This and other properties of the Lie derivative can be found in Proposition 12.32 of John M. Lee [74].

Theorem 3.1 (John M. Lee [74], Theorem 9.38)**.** *Let $X, Y \in \mathrm{Vect}(\mathcal{M})$ be two vector fields on a smooth manifold $\mathcal{M}$, then it holds that*

$$\mathcal{L}_X Y = [X, Y]\,,$$

where $[\cdot, \cdot]$ is the Lie bracket introduced in (3.30)*.*

3.8. Differential Forms

The concept of the total differential $\mathrm{d}f$ of a function $f : \mathbb{R}^n \to \mathbb{R}$ can be transferred to functions $f : \mathcal{M} \to \mathbb{R}$ and is then called the *exterior derivative* $\mathrm{d}f$ *of* f. It is defined as the covector field on M satisfying

$$\mathrm{d}f(X) = X(f) \tag{3.41}$$

for all $X \in \mathrm{Vect}(\mathcal{M})$. The representation of the exterior derivative of f with respect to the dual basis induced by a chart ϕ of $\mathcal{M}$ by (3.29) is

$$\mathrm{d}f = \frac{\partial f}{\partial q^i}\,\mathrm{d}q^i\,, \tag{3.42}$$

which looks like the total differential of a function on $\mathbb{R}^n$. From (3.42), it follows that for the coordinate map $q^i : \mathcal{M} \supseteq U \to \mathbb{R}$ of the chart ϕ it holds that $\mathrm{d}(q^i) = \mathrm{d}q^i$, which justifies the notation $\mathrm{d}q^i$ for the dual basis fields.

A differential form of degree zero is by definition a smooth function. Hence, the exterior derivative is a map $\mathrm{d} : \Omega^0(\mathcal{M}) \to \Omega^1(\mathcal{M})$ by (3.41). The following theorem extends the exterior derivative to differential forms of arbitrary degree.

Theorem 3.2 (Abraham and Marsden [1], Theorem 2.4.5)**.** *Let $\mathcal{M}$ be a smooth manifold. For any $k \in \mathbb{N}$, there exists a unique operator $\mathrm{d} : \Omega^k(\mathcal{M}) \to \Omega^{k+1}(\mathcal{M})$, called the* exterior derivative*, such that*

(i) for $k = 0$, the map d is defined by (3.41)*.*

(ii) it holds that $\mathrm{d}^2 = \mathrm{d} \circ \mathrm{d} = 0$.

(iii) d is an antiderivation, that is, for $\alpha \in \Omega^r(\mathcal{M})$ and $\beta \in \Omega^\star(\mathcal{M})$

$$\mathrm{d}(\alpha \wedge \beta) = \mathrm{d}\alpha \wedge \beta + (-1)^r \alpha \wedge \mathrm{d}\beta \,.$$

(iv) d is local, that is, let $\alpha \in \Omega^\star(\mathcal{M})$, then $\mathrm{d}(\alpha|_U) = \mathrm{d}\alpha|_U$ for all $U \subset \mathcal{M}$.

Consider the local representation $\omega = \omega_{i_1 \dots i_k} \mathrm{d}q^{i_1} \wedge \cdots \wedge \mathrm{d}q^{i_k}$ of a k-form. It follows by properties (ii) and (iii) of Theorem 3.2 that

$$\mathrm{d}\omega = \mathrm{d}\omega_{i_1 \dots i_k} \wedge \mathrm{d}q^{i_1} \wedge \cdots \wedge \mathrm{d}q^{i_k} \,. \tag{3.43}$$

Example 3.4. Let $\alpha = \alpha_i \, \mathrm{d}q^i$ be a one-form, then by (3.43) and (3.42) it follows that

$$\mathrm{d}\alpha = \mathrm{d}\alpha_i \wedge \mathrm{d}q^i = \frac{\partial \alpha_i}{\partial q^j} \, \mathrm{d}q^j \wedge \mathrm{d}q^i = \left(\frac{\partial \alpha_i}{\partial q^j} - \frac{\partial \alpha_j}{\partial q^i} \right) \mathrm{d}q^j \otimes \mathrm{d}q^i \,, \tag{3.44}$$

where the last equality involves (2.14).

A differential form $\omega \in \Omega^k(\mathcal{M})$ is said to be *closed* if $\mathrm{d}\omega = 0$ and *exact* if there is an $\alpha \in \Omega^{k-1}(\mathcal{M})$ such that $\omega = \mathrm{d}\alpha$. Because $\mathrm{d} \circ \mathrm{d} = 0$ it follows immediately that every exact form is closed. Indeed, $\mathrm{d}\omega = \mathrm{d}^2\alpha = 0$. The converse is not true in general, however there exists the following local result.

Theorem 3.3 (Poincaré lemma[5])**.** *If a differential k-form ω on $\mathcal{M}$ is closed, then for each $p \in \mathcal{M}$, there is a neighborhood $U \subseteq \mathcal{M}$ of p for which the restriction $\omega|_U$ is exact.*

Definition 3.16. Let $\mathcal{M}$ be a manifold, $X \in \mathrm{Vect}(\mathcal{M})$ and $\omega \in \Omega^{k+1}(\mathcal{M})$, then the *interior product of X and ω* is the k-form $X \lrcorner \omega \in \Omega^k(\mathcal{M})$ defined by

$$X \lrcorner \omega(Y_1, \dots, Y_k) = \omega(X, Y_1, \dots, Y_k) \,.$$

For a zero-form $\omega \in \Omega^0(\mathcal{M}) = \mathrm{C}^\infty(\mathcal{M})$, the interior product is $X \lrcorner \omega \coloneqq 0$. In the literature also the notation $i_X \omega = X \lrcorner \omega$ is encountered.

Proposition 3.1. *Let $X, Y \in \mathrm{Vect}(\mathcal{M})$ be vector fields and let $\omega \in \Omega^1(\mathcal{M})$ be a one-form, then*

$$\mathrm{d}\omega(X, Y) = \mathcal{L}_X\big(\omega(Y)\big) - \mathcal{L}_Y\big(\omega(X)\big) - \omega\big([X, Y]\big) \,. \tag{3.45}$$

[5] Abraham and Marsden [1], Theorem 2.4.17

Proof. The claim is proved by computing the terms in coordinates. Let q^i be the coordinate maps of a chart ϕ of $\mathcal{M}$, then using $\omega = \omega_i \, \mathrm{d}q^i$ it holds that

$$\mathrm{d}\omega(X,Y) = \mathrm{d}\omega_i \wedge \mathrm{d}q^i(X,Y) = (\mathrm{d}\omega_i \otimes \mathrm{d}q^i - \mathrm{d}q^i \otimes \mathrm{d}\omega_i)(X,Y)\,,$$

where (3.43) as well as (2.14) has been used. By the definition of the tensor product and equation (3.41) the above can be recast to

$$\begin{aligned} \mathrm{d}\omega(X,Y) &= \mathrm{d}\omega_i(X)\mathrm{d}q^i(Y) - \mathrm{d}q^i(X)\mathrm{d}\omega_i(Y) \\ &= X(\omega_i)Y(q^i) - X(q^i)Y(\omega_i)\,. \end{aligned} \tag{3.46}$$

Furthermore, by (3.40), the product rule (3.27) and (3.41), we have that

$$\begin{aligned} \mathcal{L}_X\big(\omega(Y)\big) &= X\big(\omega_i \, \mathrm{d}q^i(Y)\big) = X(\omega_i)Y(q^i) + \omega_i X\big(Y(q^i)\big) \\ \mathcal{L}_Y\big(\omega(X)\big) &= Y\big(\omega_i \, \mathrm{d}q^i(X)\big) = Y(\omega_i)X(q^i) + \omega_i Y\big(X(q^i)\big)\,. \end{aligned} \tag{3.47}$$

Finally, by the definition of the Lie bracket (3.30), it holds that

$$\omega\big([X,Y]\big) = \omega_i \, \mathrm{d}q^i\big([X,Y]\big) = \omega_i X\big(Y(q^i)\big) - \omega_i Y\big(X(q^i)\big)\,. \tag{3.48}$$

Using (3.46), (3.47) and (3.48) in (3.45) completes the proof. □

— 4 —

Geometric Foundations of Mechanics

> ... any physical theory would gain reputation if it could be stated in terms of notions used in geometry, the queen of all sciences.[1]
>
> —*Manfred Trümper*

This chapter is concerned with the differential geometric foundations for the description of finite-dimensional mechanical systems. The concepts presented here are mostly specialized topics, which go beyond the scope of introductory differential geometry textbooks.

4.1. Distributions

This section introduces a distribution on $\mathcal{M}$, which results by restricting every tangent space $T_p\mathcal{M}$ to a subspace $\Delta_p \subseteq T_p\mathcal{M}$. When studying such restrictions, the question arises if there is a submanifold $\mathcal{N} \subset \mathcal{M}$ such that $T_p\mathcal{N} = \Delta_p$, that is, if there exists a submanifold of $\mathcal{M}$ which is everywhere tangent to the subspaces Δ_p of the distribution. This question is answered by Frobenius' theorem, which is briefly discussed here. Finally, the more general concept of conic distribution is introduced, which arises from the restriction of every tangent space $T_p\mathcal{M}$ to a finitely generated cone.

Definition 4.1. A *distribution* Δ *of rank* f *on* $\mathcal{M}$ is a smooth vector subbundle of $T\mathcal{M}$. This means that for each $p \in \mathcal{M}$, an f-dimensional subspace $\Delta_p \subseteq T_p\mathcal{M}$ is given such that

$$\Delta = \bigcup_{p\in\mathcal{M}} \left(\{p\} \times \Delta_p\right).$$

Moreover, for each point of $\mathcal{M}$, there exists a neighborhood $U \subseteq \mathcal{M}$ on which there are vector fields $B_1, \ldots, B_f : U \to T\mathcal{M}$ such that the tangent vectors

[1] This quote is an excerpt of the first paragraph of the introduction of Trümper [125].

$(B_1)_p, \dots, (B_f)_p$ form a basis of Δ_p for each $p \in U$. The differentiable structure on Δ is then given by all bundle charts, see (3.23). A vector field $X \in \mathrm{Vect}(\mathcal{M})$ is said to *lie in the distribution* Δ if $X_p \in \Delta_p$ for each p in the domain of X, that is, if X takes values in Δ.

There are different ways of generating a distribution of rank f on a manifold $\mathcal{M}$. For example, one can choose subsets U_k that cover $\mathcal{M}$, that is, $\mathcal{M} = \cup_k U_k$, and for every U_k give vector fields $B_1, \dots, B_f \in \mathrm{Vect}(\mathcal{M})$ defined on U_k. The distribution is then characterized by

$$\begin{aligned} \Delta_p &= \mathrm{span}\big\{(B_1)_p, \dots, (B_f)_p\big\} \\ &= \big\{X_p \in T_p\mathcal{M} \;\big|\; X_p = v^i (B_i)_p \text{ for some } (v^1, \dots, v^f) \in \mathbb{R}^f\big\} \end{aligned} \tag{4.1}$$

for all $p \in U_k$. Alternatively, for every U_k, one-forms $\alpha^1, \dots, \alpha^{n-f} \in \Omega^1(\mathcal{M})$ defined on U_k may be used to locally characterize the distribution by

$$\begin{aligned} \Delta_p &= \ker \alpha_p^1 \cap \dots \cap \ker \alpha_p^{n-f} \\ &= \big\{X_p \in T_p\mathcal{M} \;\big|\; \alpha_p^\nu(X_p) = 0,\ \nu = 1, \dots, n-f\big\} \end{aligned} \tag{4.2}$$

for all $p \in U_k$. Clearly, the *(locally) defining vector fields* B_i and the *(locally) defining one-forms* α^ν must be chosen such that they define the same subspace in the overlap of two sets U_k of the covering of $\mathcal{M}$. To keep notation short, the notation

$$\Delta\Big|_{U_k} = \mathrm{span}\{B_1, \dots, B_f\} = \ker \alpha^1 \cap \dots \cap \ker \alpha^{n-f} \tag{4.3}$$

is used instead of the pointwise expressions (4.1) and (4.2).

A distribution Δ is said to be *involutive*, if for any two vector fields X, Y lying in Δ, also their Lie bracket $[X, Y]$ lies in Δ. Consider the local description (4.3) of a distribution, then it suffices to check if $[B_i, B_j]$ lies in Δ for all $i, j = 1, \dots, f$ to prove that the distribution is involutive. In fact, any X, Y lying in Δ is spanned by the basis B_i and by the properties (3.31) of the Lie bracket it holds that

$$[X, Y] = [X^i B_i, Y^j B_j] = X^i B_i(Y^j) B_j - Y^j B_j(X^i) B_i + X^i Y^j [B_i, B_j]$$

lies in the distribution if and only if $[B_i, B_j]$ lies in Δ. Using the locally defining one-forms α^ν, see (4.3), it can hence be stated that the distribution Δ is involutive if and only if

$$\alpha^\nu([B_i, B_j]) = 0 \quad \forall i, j = 1, \dots, f \text{ and } \forall \nu = 1, \dots, n-f\,. \tag{4.4}$$

Let the representations $\alpha^\nu = \alpha_k^\nu \mathrm{d}q^k$ and $B_i = B_i^k \frac{\partial}{\partial q^k}$ be given with respect to a chart with coordinate maps q^i. Hence, the local formula (3.32) for the Lie

bracket can be used to find that condition (4.4) locally reads as

$$\alpha_k^\nu \left(B_i^l \frac{\partial B_j^k}{\partial q^l} - B_j^l \frac{\partial B_i^k}{\partial q^l} \right) = 0 \,, \tag{4.5}$$

where the summation over k, l ranges from 1 to n, which is the dimension of $\mathcal{M}$. To get a more lightweight notation, the component functions of α^ν and B_i are gathered as n-tuples $\boldsymbol{\alpha}^\nu$ and $\mathbf{B}_i$, respectively. Then by the convention (3.2) for the Jacobian matrix, condition (4.5) translates to

$$[\mathbf{B}_i, \mathbf{B}_j]^{\mathrm{T}} \boldsymbol{\alpha}^\nu = 0 \quad \text{with} \quad [\mathbf{B}_i, \mathbf{B}_j] \coloneqq \frac{\partial \mathbf{B}_j}{\partial \mathbf{q}} \mathbf{B}_i - \frac{\partial \mathbf{B}_i}{\partial \mathbf{q}} \mathbf{B}_j \,, \tag{4.6}$$

where the bracket $[\mathbf{B}_i, \mathbf{B}_j]$ is just a symbol to remind of the origin of the n-tuple.

A k-form $\omega \in \Omega^k(\mathcal{M})$ annihilates the distribution Δ if

$$\omega(X_1, \ldots, X_k) = 0$$

whenever $X_1, \ldots, X_k \in \mathrm{Vect}(\mathcal{M})$ lie in the distribution. It follows immediately, that defining one-forms for Δ annihilate the distribution, see (4.2).

Proposition 4.1 (Jeffrey M. Lee [73], Lemma 11.15)**.** *For a distribution Δ on $\mathcal{M}$, the following conditions are equivalent:*

(i) The distribution is involutive.

(ii) For every (local) one-form $\omega \in \Omega^1(\mathcal{M})$ that annihilates Δ, also $\mathrm{d}\omega$ annihilates the distribution.

Proof. Suppose that (i) holds and let ω be a one-form that annihilates the distribution, then by (3.45) it holds that $\mathrm{d}\omega(X, Y) = 0$ for all vector fields X, Y lying in Δ. Hence, $\mathrm{d}\omega$ annihilates Δ. Conversely, suppose that (ii) holds. Choose defining one-forms α^i of Δ, which by definition annihilate the distribution. Consequently, by (ii), also the two-forms $\mathrm{d}\alpha^i$ annihilate the distribution. Hence, it follows from (3.45) that $\alpha^i([X, Y]) = 0$ for all vector fields X, Y lying in Δ, which by (4.2) implies that $[X, Y]$ lies in the distribution, proving that Δ is involutive. □

This proposition gives an alternative condition to (4.4) to check if a distribution is involutive. One has to verify if the exterior derivatives of all annihilating one-forms themselves annihilate the distribution. Fortunately, it can be exploited that any annihilating one-form must be a linear combination of the locally defining one-forms α^ν of the distribution Δ. Such that it suffices to check if $\mathrm{d}\alpha^\nu$ annihilates the distribution. Using the linearity of $\mathrm{d}\alpha^\nu$ and the fact that any

vector field lying in Δ is spanned by the basis B_i, see (4.3), it holds that the distribution Δ is involutive if and only if

$$\mathrm{d}\alpha^\nu(B_i, B_j) = 0 \quad \forall i, j = 1, \dots, f \text{ and } \forall \nu = 1, \dots, n-f\,. \tag{4.7}$$

With the representations $\alpha^\nu = \alpha^\nu_k \mathrm{d}q^k$ and $B_i = B^k_i \frac{\partial}{\partial q^k}$, where q^i are the coordinate maps of a chart of $\mathcal{M}$, it follows form (3.44) that condition (4.7) locally reads as

$$\left(\frac{\partial \alpha^\nu_k}{\partial q^l} - \frac{\partial \alpha^\nu_l}{\partial q^k}\right) B^l_i B^k_j = 0\,, \tag{4.8}$$

where the summation over k, l ranges from 1 to the dimension n of $\mathcal{M}$. Gathering the component functions of α^ν and B_i as n-tuples $\boldsymbol{\alpha}^\nu$ and $\mathbf{B}_i$, respectively, condition (4.8) translates to

$$\mathbf{B}^{\mathrm{T}}_j \mathrm{d}\boldsymbol{\alpha}^\nu \mathbf{B}_i = 0 \quad \text{with} \quad \mathrm{d}\boldsymbol{\alpha}^\nu \coloneqq \frac{\partial \boldsymbol{\alpha}^\nu}{\partial \mathbf{q}} - \left(\frac{\partial \boldsymbol{\alpha}^\nu}{\partial \mathbf{q}}\right)^{\mathrm{T}}, \tag{4.9}$$

where the d is just a symbol to remember the origin of the matrix $\mathrm{d}\boldsymbol{\alpha}^\nu$.

In addition to (4.3), there is yet another way to generate a distribution on $\mathcal{M}$, that is, by defining a foliation on $\mathcal{M}$. A *foliation* $\mathcal{F} = \{\mathcal{F}_\alpha\}_{\alpha \in A}$ of dimension f on an n-dimensional manifold $\mathcal{M}$ is a decomposition of $\mathcal{M}$ into disjoint connected subsets $\mathcal{F}_\alpha \subset \mathcal{M}$, called the *leaves of the foliation*, such that there is a chart (U, ψ) with coordinate maps g^i around every $p \in \mathcal{M}$ with the property

$$\mathcal{F}_\alpha \cap U = \{p \in \mathcal{M} \mid g^i(p) = c^i, \quad i = 1, \dots, n-f\} \tag{4.10}$$

for some $\mathbf{c} = (c^1, \dots, c^{n-f}) \in \mathbb{R}^{n-f}$. Such a chart is called *flat for* $\mathcal{F}$. It is immediately clear from the definition (3.6) of a submanifold of $\mathcal{M}$, that every leaf $\mathcal{F}_\alpha$ is an f-dimensional submanifold of $\mathcal{M}$. Since the leaves of a foliation are disjoint, there is exactly one leaf going through every point $p \in \mathcal{M}$. Let $\mathcal{N} \in \mathcal{F}$ be the leaf containing p, then it holds that $T_p\mathcal{N}$ is an f-dimensional subspace of $T_p\mathcal{M}$ for every $p \in \mathcal{M}$. Hence, a foliation $\mathcal{F}$ induces a distribution Δ on $\mathcal{M}$ by setting

$$\Delta_p = T_p\mathcal{N} \quad \forall p \in \mathcal{M}\,,$$

where $\mathcal{N} \in \mathcal{F}$ is the leaf for which $p \in \mathcal{N}$. Moreover, since curves in $\mathcal{N}$ going through p are mapped to a constant by the coordinate maps g^i of a flat chart (U, ψ) around p, see (4.10), we have that $X_p(g^i) = \mathrm{d}g^i_p(X_p) = 0$ for $i = 1, \dots, n-f$ and for all $X_p \in T_p\mathcal{N}$. Hence, for all $p \in U$ the distribution Δ defined by a foliation $\mathcal{F}$ is given by

$$\Delta_p = T_p\mathcal{N} = \left\{X_p \in T_p\mathcal{M} \,\middle|\, \mathrm{d}g^j_p(X_p) = 0,\ j = 1, \dots, n-f\right\} \tag{4.11}$$

and the $\mathrm{d}g^i$ with $i = 1, \dots, n-f$ are local defining one-forms for Δ.

Every foliation $\mathcal{F}$ on $\mathcal{M}$ defines a distribution Δ on $\mathcal{M}$ given by the tangent spaces to the leaves of $\mathcal{F}$, see (4.11). It follows, that every vector field lying in the distribution Δ is a vector field on some leaf $\mathcal{N} \in \mathcal{F}$, which is a submanifold of $\mathcal{M}$. Since, the Lie bracket of two vector fields on $\mathcal{N}$ is again a vector field on $\mathcal{N}$, the distribution Δ defined by a foliation is involutive. It is the merit of Frobenius' theorem to show that the converse is also true.

Theorem 4.1 (Frobenius[2])**.** *Let Δ be a distribution of rank f on $\mathcal{M}$. The following statements are equivalent:*

(i) There exists a foliation $\mathcal{F}$ of dimension f on $\mathcal{M}$ such that

$$\Delta_p = T_p\mathcal{N} \quad \forall p \in \mathcal{M}\,,$$

where $\mathcal{N} \in \mathcal{F}$ is the leaf for which $p \in \mathcal{N}$.

(ii) The distribution Δ is involutive.

By Frobenius' theorem, it is appropriate to call an involutive distribution *integrable* and the leaves of $\mathcal{F}$ *integral manifolds* of the distribution. From a practical point of view, the integrability of a distribution can be checked either by condition (4.4) or equivalently by (4.7).

Example 4.1. On a three-dimensional manifold $\mathcal{M}$, let Δ be the distribution of rank two locally defined by the one-form $\alpha = \mathrm{d}q^3 - a(q^1)\mathrm{d}q^2$, where q^i are the coordinate maps of a chart of $\mathcal{M}$ and $a : \mathbb{R} \to \mathbb{R}$ is a given smooth function. It is straightforward to check that Δ is spanned by the two vector fields

$$B_1 = \frac{\partial}{\partial q^1} \quad \text{and} \quad B_2 = \frac{\partial}{\partial q^2} + a(q^1)\frac{\partial}{\partial q^3}\,.$$

Hence, the n-tuples gathering the component functions of α as well as B_1 and B_2, respectively, are $\boldsymbol{\alpha} = (0, -a(q^1), 1)$, $\mathbf{B}_1 = (1, 0, 0)$ and $\mathbf{B}_2 = (0, 1, a(q^1))$. Using the notation of (4.6), the component tuple of the Lie bracket is

$$[\mathbf{B}_1, \mathbf{B}_2] = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \dot{a}(q^1) & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \dot{a}(q^1) \end{pmatrix}\,,$$

where $\dot{a}$ denotes the derivative of a. Consequently, $\boldsymbol{\alpha}^{\mathrm{T}}[\mathbf{B}_1, \mathbf{B}_2] = \dot{a}(q^1)$ and it follows from (4.6) that Δ is not involutive except a is constant. Hence, by Frobenius' theorem, Δ is not integrable. The same conclusion can be drawn from (4.9). Indeed,

$$\mathrm{d}\boldsymbol{\alpha} = \begin{pmatrix} 0 & \dot{a}(q^1) & 0 \\ -\dot{a}(q^1) & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

[2]See Abraham and Marsden [1], Theorem 2.2.26.

from which it can be concluded that $\mathbf{B}_2^{\mathrm{T}}\mathrm{d}\boldsymbol{\alpha}\mathbf{B}_1 = -\dot{a}(q^1)$.

Following Langerock [72], the notion of a distribution can be generalized by restricting every tangent space $T_p\mathcal{M}$ of a manifold $\mathcal{M}$ to a cone instead of a subspace. In this work, only a restriction to finitely generated cones is regarded.

Definition 4.2. A *conic distribution* $\mathfrak{D}$ *on* $\mathcal{M}$ is a subset of $T\mathcal{M}$ of the form

$$\mathfrak{D} = \bigcup_{p\in\mathcal{M}} \big(\{p\} \times \mathfrak{D}_p\big)\,,$$

where $\mathfrak{D}_p \subseteq T_p\mathcal{M}$ is a finitely generated cone, which by definition is closed and convex, see Section 2.5. Moreover, $\mathfrak{D}$ is smooth in the sense that for each point $p \in \mathcal{M}$ there exist (local) vector fields $B_1, \dots, B_f \in \mathrm{Vect}(\mathcal{M})$ such that

$$\mathfrak{D}_p = \mathrm{cone}\big\{(B_1)_p, \dots, (B_f)_p\big\}\,. \tag{4.12}$$

A vector field $X \in \mathrm{Vect}(\mathcal{M})$ is said to *lie in the conic distribution* $\mathfrak{D}$ if $X_p \in \mathfrak{D}_p$ for all p in the domain of X.

Since by Theorem 2.1, a finitely generated cone is polyhedral, it is a consequence of (4.12) that for each point $p \in \mathcal{M}$ there exist (local) one-forms $\alpha^1, \dots, \alpha^l$ such that

$$\mathfrak{D}_p = \big\{X_p \in T_p\mathcal{M} \,|\, \alpha_p^i(X_p) \geq 0,\ i = 1, \dots, l\big\}\,. \tag{4.13}$$

Let $U \subseteq \mathcal{M}$ be a neighborhood of p, then for some other point $q \in U$ there also are one-forms that define $\mathfrak{D}_q$ in the sense of (4.13). However, these one-forms are generally different from $\alpha^1, \dots, \alpha^l$. Hence, a collection of one-forms $\{\alpha^1, \dots, \alpha^k\}$ $(k \geq l)$, all defined on U, can be found such that for every point $p \in U$ a subset of this collection defines $\mathfrak{D}_p$ in the sense of (4.13). Specifically, the fibers of the conic distribution can locally be represented by

$$\mathfrak{D}_p = \big\{X_p \in T_p\mathcal{M} \,|\, \alpha_p^\nu(X_p) \geq 0,\ \forall \nu \in I(p)\big\} \tag{4.14}$$

for every $p \in U$, where I is a map assigning a subset of $\{1, \dots, k\}$ to every p. Similar to the case of (vector) distributions, the one-forms α^ν are called *(locally) defining one-forms of* $\mathfrak{D}$.

Since every subspace of a finite-dimensional vector space is a finitely generated cone, see (2.17), the (vector) distributions are special cases of conic distributions. In fact, a (vector) distribution Δ generated by k one-forms α^ν can be seen as a conic distribution generated by the $2k$ one-forms α^ν and $-\alpha^\nu$, i.e.,

$$\begin{aligned}\Delta_p &= \big\{X_p \in T_p\mathcal{M} \;\big|\; \alpha_p^\nu(X_p) = 0,\ \nu = 1, \dots, k\big\} \\ &= \big\{X_p \in T_p\mathcal{M} \;\big|\; \alpha_p^\nu(X_p) \geq 0,\ -\alpha_p^\nu(X_p) \geq 0,\ \nu = 1, \dots, k\big\}\,.\end{aligned} \tag{4.15}$$

Moreover, since the intersection of finitely generated cones is again a finitely generated cone, the intersection of two conic distributions is again a conic distribution. Especially, the intersection of a (vector) distribution Δ and a conic distribution $\mathfrak{C}$ is a conic distribution $\mathfrak{D} = \Delta \cap \mathfrak{C}$.

Definition 4.3. Consider a finitely generated subset $\mathcal{C} \subset \mathcal{M}$ locally defined by k functions $g^\nu : \mathcal{M} \supseteq U \to \mathbb{R}$ as

$$\mathcal{C} \cap U = \left\{ p \in U \,|\, g^\nu(p) \geq 0, \quad \nu = 1, \ldots, k \right\}.$$

The *tangent cone to* $\mathcal{C}$ *at* $p \in \mathcal{C}$ is the subset $T_{\mathcal{C}}(p) \subset T_p\mathcal{M}$ defined by

$$T_{\mathcal{C}}(p) = \left\{ X_p \in T_p\mathcal{M} \,\middle|\, \mathrm{d}g^\nu_p(X_p) \geq 0, \ \forall \nu \in A(p) \right\}, \tag{4.16}$$

where $A(p) \coloneqq \{\nu \mid g^\nu(p) \leq 0\}$ denotes the *index set of active constraints* at p. For points $p \notin \mathcal{C}$ the tangent cone is empty.

Note, that C is not necessarily convex. However, the tangent cone (4.16) is closed and convex, because it is pointwise defined as a polyhedral cone, cf. Theorem 2.1. Moreover, the above definition of tangent cone relies on the fact, that the set C is finitely generated, which implies that C is tangentially regular. For more general sets, there are several notions of tangent cone, see Rockafellar and Wets [106].

With the concept of tangent cone, one could study the integrability of conic distributions. Similar to the case of vector distributions, a set $\mathcal{C}$ can be interpreted as an "integral manifold" of $\mathfrak{D}$ if it holds that $\mathfrak{D}_p = T_{\mathcal{C}}(p)$ for all $p \in \mathcal{C}$. More details on integrability of conic distributions can be found in Langerock [72].

4.2. Connections and Covariant Derivatives

The covariant derivative is a generalization of the directional derivative in $\mathbb{R}^n$ to vector bundles $\mathcal{B}$ and can only be defined if $\mathcal{B}$ is equipped with additional structure. The needed structure is provided by a connection on $\mathcal{B}$. To motivate its definition, first the situation in $\mathbb{R}^n$ is studied.

Consider the trivial vector bundle $\mathcal{B} = \mathcal{M} \times \mathbb{R}^k$, where $\mathcal{M} \subseteq \mathbb{R}^n$ is an open subset. The natural projection $\pi : \mathcal{B} \to \mathcal{M}$ is given by the projection onto the first factor, that is $\pi = \mathrm{pr}_1$. Let $Y \in \Gamma(\mathcal{B})$ be a section of $\mathcal{B}$ and let $X \in \mathrm{Vect}(\mathcal{M})$. The directional derivative of Y with respect to X is defined as the section $\overline{\nabla}_X Y : \mathbb{R}^n \supseteq \mathcal{M} \to \mathcal{B}$ which is pointwise given by

$$\left(\overline{\nabla}_X Y\right)_p \coloneqq \frac{\mathrm{d}Y_{\gamma(t)}}{\mathrm{d}t}\bigg|_{t=0} = \lim_{h \downarrow 0} \frac{Y_{\gamma(h)} - Y_{\gamma(0)}}{h}, \tag{4.17}$$

where $\gamma : \mathbb{R} \to \mathcal{M}$ is any integral curve of X with $\gamma(0) = p \in \mathcal{M}$. Since the concatenation $Y \circ \gamma$ is a curve in $\mathcal{B}$, it holds that

$$(Y \circ \gamma)^{\cdot}(0) = \big(\gamma, Y_\gamma\big)^{\cdot}(0) = \big(\dot\gamma, (Y_\gamma)^{\cdot}\big)(0) = \Big(X_p, \big(\overline{\nabla}_X Y\big)_p\Big)\,, \tag{4.18}$$

where (4.17) and $\dot\gamma(0) = X_p$ has been used. By the definition (3.20) of the differential, it holds that $(DY_p)X_p = (Y \circ \gamma)^{\cdot}(0)$. Consequently, it follows from (4.18) that the directional derivative (4.17) can equivalently be defined as

$$\big(\overline{\nabla}_X Y\big)_p \coloneqq \eta\big((DY_p)X_p\big)\,, \tag{4.19}$$

where the notation $\eta = \mathrm{pr}_2$ is introduced for the projection onto the second factor. Either definition for $\overline{\nabla}_X Y$ can equivalently be generalized to a directional derivative for general vector bundles. However, for the purpose of this thesis, the generalization of (4.19) is more suitable and is discussed in the following.[3]

The definition (4.19) exploits that for the trivial bundle $\mathcal{B} = \mathcal{M} \times \mathbb{R}^k$, besides $\pi = \mathrm{pr}_1$, there is an additional projection η onto the second factor. For general vector bundles, there is no canonical way to define such a projection. Hence, if one wants to give meaning to a directional derivative on a general vector bundle, such a projection η must be chosen as an additional structure on the bundle. The question that immediately arises is: On what space should such an η project? Observe that for the trivial bundle $\mathcal{B}$ it holds that

$$\mathrm{D}\pi_{Y(p)}(Y \circ \gamma)^{\cdot}(0) = (\pi \circ Y \circ \gamma)^{\cdot}(0) = \dot\gamma(0) = X_p\,,$$

where $\pi = \mathrm{pr}_1$. Hence, a comparison with (4.18) and (4.19) reveals that η projects on $\ker \mathrm{D}\pi_{Y(p)}$.

Definition 4.4. Consider a fiber bundle $(\mathcal{B}, \mathcal{M}, \pi, \mathcal{F})$. The *vertical bundle* of $\mathcal{B}$ is the distribution $\mathrm{Ver}(\mathcal{B})$ on $\mathcal{B}$ defined by the subspaces

$$\mathrm{Ver}_a(\mathcal{B}) \coloneqq \Big\{ V_a \in T_a\mathcal{B} \;\Big|\; \mathrm{D}\pi_a V_a = 0 \Big\}$$

for all $a \in \mathcal{B}$. A section $V \in \Gamma(\mathrm{Ver}(\mathcal{B}))$ is called *vertical vector field*.

Definition 4.5. A *connection* on a fiber bundle $(\mathcal{B}, \mathcal{M}, \pi, \mathcal{F})$ is a vector-valued one form $\eta : T\mathcal{B} \to T\mathcal{B}$ on $\mathcal{B}$, such that the image of η is $\mathrm{Ver}(\mathcal{B})$ and $\eta \circ \eta = \eta$, i.e., a connection is a projection on the vertical bundle of $\mathcal{B}$. Hence, pointwise the connection defines a projection $\eta_a : T_a\mathcal{B} \to T_a\mathcal{B}$ onto $\mathrm{Ver}_a(\mathcal{B})$.

[3] For the generalization of (4.17), a diffeomorphism $P_{\gamma,t} : \mathcal{B}_{\gamma(0)} \to \mathcal{B}_{\gamma(t)}$, called *parallel transport*, between the fibers of $\mathcal{B}$ is introduced such that the covariant derivative can be defined by

$$\big(\nabla_X Y\big)_p = \frac{\mathrm{d}\big(P^{-1}_{\gamma,t}(Y_{\gamma(t)})\big)}{\mathrm{d}t}\Big|_{t=0} = \lim_{h \downarrow 0} \frac{P^{-1}_{\gamma,h} Y_{\gamma(h)} - Y_{\gamma(0)}}{h}\,.$$

Every connection induces a parallel transport, see [86], Theorem 17.8, or [119], Addendum 3.

The projection property of a connection assures that $\ker\eta$ is of constant rank implying that $\ker\eta$ is a distribution on $\mathcal{B}$, that is, $\ker\eta$ is a vector subbundle of $T\mathcal{B}$. Since this subbundle is complementary to the vertical bundle of $\mathcal{B}$, every connection on $\mathcal{B}$ induces a splitting of $T\mathcal{B}$ as

$$T\mathcal{B} = \mathrm{Hor}(\mathcal{B}) \oplus \mathrm{Ver}(\mathcal{B})\,,$$

where the *horizontal bundle* $\mathrm{Hor}(\mathcal{B}) \coloneqq \ker\eta$ is introduced.

If a connection is defined on a vector bundle $(\mathcal{B}, \mathcal{M}, \pi, \mathbb{R}^k)$ of rank k, the vector space structure of the fibers $\mathcal{B}_p$ can be exploited. Let $a = (p, a_p) \in \mathcal{B}$, then for every $Y = (p, Y_p) \in \mathcal{B}_{\pi(a)}$ it makes sense to define the curve $a + tY = (p, a_p + tY_p)$ in $\mathcal{B}$. It is clear by $\pi(a + tY) = p$ that the tangent vector $(a + tY)\dot{}(0)$ is vertical. Hence, we may define the *vertical lift* as the vector space isomorphism

$$v_a : \mathcal{B}_{\pi(a)} \to \mathrm{Ver}_a(\mathcal{B}),\ Y \mapsto Y^v = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}(a + tY)\,. \tag{4.20}$$

With that, the *vertical lift of a section* $Y \in \Gamma(\mathcal{B})$ is the vertical vector field $Y^v \in \mathrm{Vect}(\mathcal{B})$ satisfying $Y^v_a = v_a\big(Y(\pi(a))\big)$ for all $a \in \mathcal{B}$.

Definition 4.6. Let η be a connection on a vector bundle $(\mathcal{B}, \mathcal{M}, \pi, \mathbb{R}^k)$. Moreover let $X \in \mathrm{Vect}(\mathcal{M})$ be a vector field on $\mathcal{M}$ and $Y \in \Gamma(\mathcal{B})$ be a section of $\mathcal{B}$. The *covariant derivative of Y with respect to X* is defined as the section $\nabla_X Y \in \Gamma(\mathcal{B})$ given by

$$(\nabla_X Y)_p = (v_{Y(p)})^{-1}\big(\eta_{Y(p)}\big((\mathrm{D}Y_p)X_p\big)\big) \quad \forall p \in \mathcal{M}\,. \tag{4.21}$$

Since all the maps involved in (4.21) are linear, it is readily checked that ∇Y is linear, i.e.,

$$\nabla_{X_1+X_2} Y = \nabla_{X_1} Y + \nabla_{X_2} Y \quad \text{and} \quad \nabla_{fX} Y = f\nabla_X Y$$

for all $X_1, X_2 \in \mathrm{Vect}(\mathcal{M})$ and all $f \in \mathrm{C}^\infty(\mathcal{M})$.

Let Φ be the bundle chart of $\mathcal{B}$ defined by the chart (U, ϕ) and basis sections $B_1, \ldots, B_k \in \Gamma(\mathcal{B})$ defined on U, see (3.23). Since the vertical lift is a vector space isomorphism, it follows that the vertical bundle is locally spanned by the vertical lifts $\frac{\partial}{\partial u^i}$ of the basis fields B_i, that is,

$$\mathrm{Ver}(\mathcal{B})\Big|_{\pi^{-1}(U)} = \mathrm{span}\left\{\frac{\partial}{\partial u^1}, \ldots, \frac{\partial}{\partial u^k}\right\}.$$

Introducing the one-forms $\eta^i = \mathrm{d}u^i + \Gamma^i_\alpha \mathrm{d}q^\alpha \in \Omega^1(\mathcal{B})$ with the component functions $\Gamma^i_\alpha \in \mathrm{C}^\infty(\mathcal{B})$, the connection is locally given by

$$\eta = \frac{\partial}{\partial u^i} \otimes \eta^i = \frac{\partial}{\partial u^i} \otimes (\mathrm{d}u^i + \Gamma^i_\alpha \mathrm{d}q^\alpha)\,, \tag{4.22}$$

which follows from $\eta \circ \eta = \eta$ and the fact that the image of η is $\mathrm{Ver}(\mathcal{B})$. For the remainder of this section, that Roman indices range from 1 to k and Greek indices from 1 to n. Since the horizontal bundle is defined as $\mathrm{Hor}(\mathcal{B}) = \ker \eta$, it follows from (4.22) that the η_i are defining one-forms for the horizontal bundle. Moreover, due to the particular structure of η_i, locally the connection is completely characterized by the *connection forms* $\omega^i = \Gamma^i_\alpha \mathrm{d}q^\alpha \in \Omega^1(\mathcal{B})$.

To arrive at the chart representation of the covariant derivative, let $X = X^\alpha \frac{\partial}{\partial q^\alpha} \in \mathrm{Vect}(\mathcal{M})$ be a vector field on $\mathcal{M}$ and $Y = Y^i B_i$ be a section of $\mathcal{B}$. Using (3.29) and the definitions of the involved objects to compute the components of the vector field $\mathrm{D}Y(X)$, one finds that

$$\mathrm{D}Y(X) = X^\alpha \frac{\partial}{\partial q^\alpha} + X(Y^i)\frac{\partial}{\partial u^i}\,.$$

Consequently, with (4.22) and (4.20), the covariant derivative (4.21) locally is

$$\nabla_X Y = \big(X(Y^i) + \Gamma^i_\alpha(Y) X^\alpha\big) B_i\,. \tag{4.23}$$

Definition 4.7. Let η be a connection on a vector bundle $(\mathcal{B}, \mathcal{M}, \pi, \mathbb{R}^k)$. The connection is called *linear*, if for every bundle chart of the form (3.23), the connection forms ω_i have the form

$$\omega^i = (\Gamma^i_{\alpha j} \circ \pi) u^j \mathrm{d}q^\alpha$$

with $\Gamma^i_{\alpha j} \in \mathrm{C}^\infty(\mathcal{M})$. The functions $\Gamma^i_{\alpha j} \in \mathrm{C}^\infty(\mathcal{M})$ are called the *Christoffel symbols* of the connection.

In the literature, the terms "covariant derivative" and "connection" are sometimes treated as synonymous. In fact, a covariant derivative associated to a linear connection is often called a "Koszul (or linear) connection".

For a linear connection, due to $\Gamma^i_\alpha = (\Gamma^i_{\alpha j} \circ \pi) u^j$, the covariant derivative (4.23) can be written as

$$\nabla_X Y = \big(X(Y^i) + \Gamma^i_{\alpha j} Y^j X^\alpha\big) B_i\,. \tag{4.24}$$

It readily follows that for a linear connection, ∇_X has the properties

$$\nabla_X(Y_1 + Y_2) = \nabla_X Y_1 + \nabla_X Y_2 \quad \text{and} \quad \nabla_X(fY) = X(f)Y + f\nabla_X Y$$

for all $Y_1, Y_2 \in \Gamma(\mathcal{B})$ and all $f \in \mathrm{C}^\infty(\mathcal{M})$.

Consider a vector field β on $\mathcal{M}$ and a section σ of $\mathcal{B}$, both defined along a curve $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$, i.e., $\pi \circ \sigma = \gamma$ and $\pi_\mathcal{M} \circ \beta = \gamma$, where $\pi_\mathcal{M}$ is the natural projection of $T\mathcal{M}$. The covariant derivative $\nabla_\beta \sigma$ is the section of $\mathcal{B}$ defined along γ given by

$$(\nabla_\beta \sigma)_{\gamma(t)} = (v_{\sigma(t)})^{-1}\big(\eta_{\sigma(t)}\big((\mathrm{D}\sigma_{\gamma(t)})\beta_{\gamma(t)}\big)\big) \quad \forall t \in \mathcal{I}\,.$$

Hence, by (4.21), $\nabla_\beta \sigma(t) = \nabla_X Y \circ \gamma(t)$, where any vector field X and any section Y satisfying $X \circ \gamma = \beta$ and $Y \circ \gamma = \sigma$ can be taken. A section σ is said to be *parallel along* γ if $\nabla_{\dot\gamma}\sigma = 0$.

This section is concluded with the remark that there is no consensus on what a connection is. However, all notions of connection in some way or another define a projection onto the vertical bundle and are therefore equivalent to the definition of connection employed here. That being said, de León and Rodrigues [30], Jeffrey M. Lee [73] as well as Bullo and Lewis [19] provide further details on connections.

4.3. Structures on the Tangent Bundle

In this section, some aspects of connections on the tangent bundle to a manifold $\mathcal{M}$ are discussed. Moreover, the notion of second-order fields is introduced and finally some results of Riemannian geometry are reviewed.

The tangent bundle of an n-dimensional manifold $\mathcal{M}$, is a vector bundle of rank n over $\mathcal{M}$ and is itself a manifold of dimension $2n$. Since a vector field defines a system of first-order ordinary differential equations, and the dimension of $T\mathcal{M}$ is even, there are vector fields Z on $T\mathcal{M}$ which characterize second-order differential equations in first-order form. Such a vector field $Z \in \mathrm{Vect}(T\mathcal{M})$, called *second-order field*, needs to obey the second-order condition

$$\mathrm{D}\pi Z = \mathrm{id}_{T\mathcal{M}}\,, \tag{4.25}$$

where $\mathrm{D}\pi : T(T\mathcal{M}) \to T\mathcal{M}$ denotes the differential of the natural projection of $T\mathcal{M}$. An integral curve β of Z, that is, a curve satisfying $\dot\beta = Z \circ \beta$, by (4.25) has the property $\beta = (\pi \circ \beta)^{\cdot}$ and is called *second-order curve*. Indeed,

$$\beta = \mathrm{id}_{T\mathcal{M}}\beta = \mathrm{D}\pi Z \circ \beta = \mathrm{D}\pi\dot\beta = (\pi \circ \beta)^{\cdot}. \tag{4.26}$$

Hence, a second-order curve β can be seen as tangent field along the curve $\gamma = \pi \circ \beta$ on $\mathcal{M}$, i.e., $\beta = \dot\gamma$. With this notation, it can be seen that the integral curves of a second-order field Z indeed define a second-order differential equation, as $\beta = \dot\gamma$ must satisfy

$$\ddot\gamma = Z \circ \dot\gamma\,.$$

With respect to a natural chart Φ of $T\mathcal{M}$, see (3.22), it is a consequence of (4.25), that a second-order field Z has the local form

$$Z = u^i \frac{\partial}{\partial q^i} + Z^i \frac{\partial}{\partial u^i}\,, \tag{4.27}$$

where u^i denotes the $(n{+}i)$-th coordinate map of Φ and $Z^i \in \mathrm{C}^\infty(T\mathcal{M})$. Using $q^i(t) \coloneqq q^i \circ \beta(t)$ and $u^i \coloneqq u^i \circ \beta(t)$ to denote the components of a curve β in $T\mathcal{M}$,

the condition $\dot{\beta} = Z \circ \beta$ for β being an integral curve of the second-order field (4.27) locally reads as

$$\dot{\mathbf{q}}(t) = \mathbf{u}(t), \quad \dot{\mathbf{u}}(t) = \mathbf{Z} \circ \Phi^{-1}\big(\mathbf{q}(t), \mathbf{u}(t)\big)\,, \tag{4.28}$$

which is an autonomous second-order differential equation in first-order form.

It can be shown, that a connection can be associated to every second-order field. To state this, the *vertical endomorphism* $J \coloneqq v \circ \mathrm{D}\pi$ on $T\mathcal{M}$ is introduced, where $v : T\mathcal{M} \to \mathrm{Ver}(T\mathcal{M})$ is the vertical lift, see (4.20). It is straightforward to see, that J is a vector-valued one form and is locally given by

$$J = \frac{\partial}{\partial u^i} \otimes \mathrm{d}q^i\,.$$

Theorem 4.2 (Morandi et al. [89], Proposition 2.14)**.**[4] *Let Z be a second-order field on $T\mathcal{M}$ and J the vertical endomorphism on $T\mathcal{M}$, then η defined by*

$$\eta(X) = \frac{1}{2}\big(X + \mathcal{L}_Z J(X)\big) = \frac{1}{2}\big(X + [Z, J(X)] - J([Z, X])\big)$$

for all $X \in \mathrm{Vect}(T\mathcal{M})$ is a connection on $T\mathcal{M}$. With respect to a natural chart, the connection reads as

$$\eta = \frac{\partial}{\partial u^i} \otimes \left(\mathrm{d}u^i - \frac{1}{2}\frac{\partial Z^i}{\partial u^j}\mathrm{d}q^j\right). \tag{4.29}$$

The comparison of (4.29) with (4.22) reveals, that the connection forms of $\mathrm{Hor}(T\mathcal{M}) = \ker\eta$ are

$$\omega^i = -\frac{1}{2}\frac{\partial Z^i}{\partial u^j}\mathrm{d}q^j\,.$$

Hence, e.g., the connection is linear if $Z^i = -(\Gamma^i_{jk} \circ \pi)u^j u^k$ with $\Gamma^i_{jk} \in \mathrm{C}^\infty(\mathcal{M})$.

It follows from straightforward computations in coordinates that for such a Z it holds that

$$(\nabla_{\dot\gamma}\dot\gamma)^v = \ddot\gamma - Z \circ \dot\gamma\,,$$

for any curve $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$, where v denotes the vertical lift and ∇ is the covariant derivative defined by the connection associated to Z. Hence, the projection of the integral curves of Z to $\mathcal{M}$ are exactly the curves γ which are *autoparallel*, i.e., γ satisfies $\nabla_{\dot\gamma}\dot\gamma = 0$.

For a linear connection, ∇_X can be extended to a covariant derivative on one-forms by defining

$$\nabla_X f = X(f) \quad \text{and} \quad \nabla_X\omega(Y) = \nabla_X(\omega(Y)) - \omega(\nabla_X Y) \quad \forall Y \in \mathrm{Vect}(\mathcal{M})\,, \tag{4.30}$$

[4] See also de León and Rodrigues [30], Chapter 4.5.

where $f \in \mathrm{C}^\infty(\mathcal{M})$ and $\omega \in \Omega^1(\mathcal{M})$. Let F be a general tensor field, then $\nabla_X F$ is defined as the tensor field of the same type as F given by

$$\nabla_X F(A_1, \ldots, A_k) = \nabla_X(F(A_1, \ldots, A_k)) - \sum_{i=1}^{k} F(A_1, \ldots, \nabla_X A_i, \ldots, A_k)\,,$$

where the arguments A_i are either vector or covector fields, depending on what the tensor requires. For instance, consider a tensor field $F \in \Gamma(\otimes^1_1 T^*\mathcal{M})$, then

$$\nabla_X F(\omega, Y) = \nabla_X(F(\omega, Y)) - F(\nabla_X \omega, Y) - F(\omega, \nabla_X Y)\,,$$

where $X, Y \in \mathrm{Vect}(\mathcal{M})$ and $\omega \in \Omega^1(\mathcal{M})$.

The *torsion* of a linear connection is the rank-two tensor field T defined by

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]$$

for all $X, Y \in \mathrm{Vect}(\mathcal{M})$, where $[\cdot, \cdot]$ denotes the Lie-bracket on $\mathcal{M}$. Computations in coordinates show, that a linear connection is torsion-free, that is $T = 0$, if and only if the Christoffel symbols satisfy $\Gamma^k_{ij} = \Gamma^k_{ji}$.

Definition 4.8. A *Riemannian manifold* is a pair $(\mathcal{M}, g)$, where $\mathcal{M}$ is a smooth manifold and $g \in \Gamma(\otimes^0_2 T^*\mathcal{M})$ is a covariant tensor field of rank 2 on $\mathcal{M}$ that endows the tangent spaces $T_p\mathcal{M}$ with an inner product for all $p \in \mathcal{M}$. The tensor field g is called *Riemannian metric*.

Theorem 4.3 (Jeffrey M. Lee [73], Theorem 13.9.)**.** *For a given Riemannian manifold $(\mathcal{M}, g)$, there is a unique connection which is torsion-free and metric, i.e., satisfies $T = 0$ and $\nabla g = 0$. This unique connection is called* Levi-Civita connection *for $(\mathcal{M}, g)$. Its Christoffel symbols in some chart are given by*

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl} \left(\frac{\partial g_{jl}}{\partial q^i} + \frac{\partial g_{li}}{\partial q^j} - \frac{\partial g_{ij}}{\partial q^l} \right), \tag{4.31}$$

where $g = g_{ij}\,\mathrm{d}q^i \otimes \mathrm{d}q^j$ and g^{kl} denote the components of the inverse matrix of g_{ij}, that is, $g^{kl} g_{li} = \delta^k_i$.

4.4. Galilean Manifolds

In the previous section, it has been shown that second-order fields on the tangent bundle of an n-dimensional manifold define a system of n autonomous second-order differential equations. To describe n non-autonomous second-order differential equations with independent variable t, called *time*, one can work on an $(n+1)$-dimensional manifold and include the independent variable as a

coordinate with dynamics $\dot{t} = 1$. This idea leads to the notion of second-order fields on manifolds with time structure, which equipped with a (bundle) metric give rise to the concept of Galilean manifolds.

A *time structure*[5] ϑ on an $(n+1)$-dimensional manifold $\mathcal{M}$ is a non-vanishing closed one-form on $\mathcal{M}$. Hence, the Poincaré lemma[6] guarantees the existence of local *time functions* $t : \mathcal{M} \supseteq U \to \mathbb{R}$ with $\mathrm{d}t = \vartheta|_U$.

A chart (U, ϕ) of $\mathcal{M}$ given by

$$\phi : \mathcal{M} \supseteq U \to \mathbb{R}^{n+1},\ p \mapsto \phi(p) = (t, q^1, \dots, q^n) \tag{4.32}$$

is *adapted* to the time structure if its first coordinate map $q^0 = t$ is a time function. The existence of adapted charts is guaranteed by the existence of time functions and the fact that ϑ is nowhere zero. Therefore, the adapted charts provide an atlas of $\mathcal{M}$ such that it suffices to look at adapted charts in what follows. The change of coordinates

$$\bar{\phi} \circ \phi^{-1} \colon \phi(U \cap \bar{U}) \to \bar{\phi}(U \cap \bar{U}),\ (t, q^1, \dots, q^n) \mapsto (\bar{t}, \bar{q}^1, \dots, \bar{q}^n)$$

between two adapted charts (U, ϕ) and $(\bar{U}, \bar{\phi})$ of $\mathcal{M}$ with $U \cap \bar{U} \neq \emptyset$ is given by

$$\begin{aligned} \bar{t} &= t + \text{const.} \\ \bar{q}^i &= \bar{q}^i \circ \phi^{-1}\big(t, q^1, \dots, q^n\big),\ i = 1, \dots, n, \end{aligned}$$

where $\bar{q}^i : \bar{U} \to \mathbb{R}$ denotes the i-th coordinate map of the chart $\bar{\phi}$.

The time structure ϑ is used to introduce the *spacelike bundle* $A^0\mathcal{M}$ of $\mathcal{M}$ as the distribution of rank n on $\mathcal{M}$ resulting from the pointwise restriction of the tangent spaces $T_p\mathcal{M}$ to the *spaces of spacelike vectors in* p given by

$$A^0_p\mathcal{M} \coloneqq \ker \vartheta_p = \big\{ v_p \in T_p\mathcal{M} \;\big|\; \vartheta_p(v_p) = 0 \big\} \tag{4.33}$$

for every $p \in M$. As $\mathrm{d}\vartheta = 0$ annihilates the distribution trivially, it is integrable and therefore defines a foliation $\mathcal{F}$ of $\mathcal{M}$ by the Frobenius theorem.[7] It is clear, that a leaf of $\mathcal{F}$ is mapped to a constant value by any time function of $\mathcal{M}$. Hence, a leaf of $\mathcal{F}$ is called a *submanifold of simultaneous events* of $\mathcal{M}$. It follows from the definition (4.33) that the spacelike bundle is locally spanned by the induced basis fields associated to the coordinates q^i of an adapted chart (4.32), i.e.,

$$A^0\mathcal{M}\Big|_{\pi^{-1}(U)} = \operatorname{span}\Big\{ \frac{\partial}{\partial q^1}, \dots, \frac{\partial}{\partial q^n} \Big\}. \tag{4.34}$$

[5] See Dombrowski and Horneffer [31], and Loos [82].
[6] See Theorem 3.3.
[7] See Theorem 4.1.

Similar to $A^0\mathcal{M}$, the *time-normalized bundle* $A^1\mathcal{M}$ of $\mathcal{M}$ is defined as the affine subbundle of $T\mathcal{M}$ resulting from its pointwise restriction to the *affine space of time-normalized vectors in p* given by

$$A^1_p\mathcal{M} \coloneqq \left\{ v_p \in T_p\mathcal{M} \;\middle|\; \vartheta_p(v_p) = 1 \right\}$$

for every $p \in M$. In contrast to $A^0\mathcal{M}$, the bundle $A^1\mathcal{M}$ is not a distribution on $\mathcal{M}$. However, it is a fiber bundle of rank n over $\mathcal{M}$ in virtue of the *natural projection*

$$\pi : A^1\mathcal{M} \to \mathcal{M},\ (p, v_p) \mapsto p \tag{4.35}$$

being the restriction to $A^1\mathcal{M}$ of the natural projection of $T\mathcal{M}$. The differentiable structure is defined by the *natural charts*

$$\Phi : A^1\mathcal{M} \supseteq \pi^{-1}(U) \to \mathbb{R}^{2n+1},\ \left(p, X_p\right) \mapsto \left(t, q^1, \ldots, q^n, u^1, \ldots, u^n\right) \tag{4.36}$$

induced by the adapted charts (4.32), where $X_p = \frac{\partial}{\partial t}\big|_p + u^i \frac{\partial}{\partial q^i}\big|_p$. The affine bundle $A^1\mathcal{M}$ is canonically endowed with the time structure

$$\hat{\vartheta} \coloneqq \pi^* \vartheta\,, \tag{4.37}$$

which is the pullback of the time structure of $\mathcal{M}$ by the natural projection (4.35). The natural chart (4.36) is an adapted chart with respect to the time structure (4.37) of $A^1\mathcal{M}$ because it holds that $\hat{\vartheta} = \mathrm{d}t$ on $\pi^{-1}(U)$.

A curve $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ is *time-parametrized* if the tangent field along the curve is time-normalized, that is, if $\vartheta(\dot{\gamma}) = 1$. Since $\hat{\vartheta}$ is a time structure on $A^1\mathcal{M}$, curves β in $A^1\mathcal{M}$ are time-parametrized if $\hat{\vartheta}(\dot{\beta}) = 1$.

A vector field $Z \in \mathrm{Vect}(A^1\mathcal{M})$ is a *second-order field* if it is time-normalized, that is $\hat{\vartheta}(Z) = 1$ and obeys the second-order condition

$$\mathrm{D}\pi Z = \mathrm{id}_{A^1\mathcal{M}}\,, \tag{4.38}$$

where $\mathrm{D}\pi : T(A^1\mathcal{M}) \to T\mathcal{M}$ denotes the differential of the natural projection (4.35). An integral curve β of a second-order field Z is called a *second-order curve*. By $\hat{\vartheta}(Z) \circ \beta = \hat{\vartheta}(\dot{\beta}) = 1$, a second-order curve is time-parametrized and satisfies $\beta = \dot{(\pi \circ \beta)}$ because of (4.38), which can be seen by a computation similar to (4.26).

With respect to a natural chart Φ of $A^1\mathcal{M}$, it follows from (4.38), that a second-order field Z has the local form

$$Z = \frac{\partial}{\partial t} + u^i \frac{\partial}{\partial q^i} + Z^i \frac{\partial}{\partial u^i}\,, \tag{4.39}$$

where u^i denotes the $(n{+}1{+}i)$-th coordinate map of Φ and $Z^i \in \mathrm{C}^\infty(A^1\mathcal{M})$. Using $q^i(\tau) \coloneqq q^i \circ \beta(\tau)$ and $u^i \coloneqq u^i \circ \beta(\tau)$ to denote the components of a curve β

in $A^1\mathcal{M}$, the condition for β to be an integral curve of the second-order field (4.39) locally reads as

$$\dot{t}(\tau) = 1, \quad \dot{\mathbf{q}}(\tau) = \mathbf{u}(\tau), \quad \dot{\mathbf{u}}(\tau) = \mathbf{Z} \circ \Phi^{-1}\big(t(\tau), \mathbf{q}(\tau), \mathbf{u}(\tau)\big)\,,$$

which is a non-autonomous second-order differential equation in first-order form.

For any $p \in \mathcal{M}$ the choice of a time-normalized vector field $R : \mathcal{M} \supseteq U \to A^1\mathcal{M}$ induces a local *spacelike projector*

$$P^R : \pi^{-1}(U) \to A^0\mathcal{M},\ (p, X_p) \mapsto (p, X_p - \vartheta_p(X_p)R_p)\,, \tag{4.40}$$

where π is the natural projection of $T\mathcal{M}$ and $R(p) = (p, R_p)$. Overloading notation, the pointwise projection is denoted by $P^{R_p}(X_p) = X_p - \vartheta_p(X_p)R_p$.

It can be exploited that the fibers of $A^1\mathcal{M}$ are affine spaces, i.e., the sum of a time-normalized vector and a spacelike vector is a time-normalized vector. Let $a = (p, a_p) \in A^1\mathcal{M}$, then for every $Y = (p, Y_p) \in A^0\mathcal{M}$ it makes sense to define the curve $a + tY = (p, a_p + tY_p)$ in $A^1\mathcal{M}$. It is clear by $\pi(a + tY) = p$ that the tangent vector $(a + tY)\dot{}(0)$ is vertical. Hence, the *vertical lift* can be defined as the linear map

$$v_a : A^0_{\pi(a)}\mathcal{M} \to \mathrm{Ver}_a(A^1\mathcal{M}),\ Y \mapsto Y^v = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}(a + tY)\,. \tag{4.41}$$

Since by definition $\mathrm{Ver}_a(A^1\mathcal{M}) = \ker \mathrm{D}\pi_a$, the vertical subspace is spanned by induced basis vectors of a natural chart Φ of $A^1\mathcal{M}$, that is,

$$\mathrm{Ver}_a(A^1\mathcal{M}) = \mathrm{span}\Big\{ \frac{\partial}{\partial u^1}\bigg|_a, \ldots, \frac{\partial}{\partial u^n}\bigg|_a \Big\}\,. \tag{4.42}$$

Moreover, it can be seen from (4.41) that for the basis vectors induced by ϕ it holds that

$$v_a\left(\frac{\partial}{\partial q^i}\bigg|_{\pi(a)} \right) = \frac{\partial}{\partial u^i}\bigg|_a\,. \tag{4.43}$$

Hence, a comparison of (4.34) with (4.42) shows that the vertical lift is a vector space isomorphism. The *vertical lift of a spacelike vector field* $Y \in \Gamma(A^0\mathcal{M})$ is the vertical vector field $Y^v \in \mathrm{Vect}(A^1\mathcal{M})$ satisfying $Y^v_a = v_a\big(Y(\pi(a))\big)$ for all $a \in \mathcal{B}$. Hence, the vertical lift of a spacelike vector field $Y = Y^i \frac{\partial}{\partial q^i}$ is given by

$$Y^v = (Y^i \circ \pi)\frac{\partial}{\partial u^i}\,. \tag{4.44}$$

The *vertical homomorphism* $\mu : T(A^1\mathcal{M}) \to \mathrm{Ver}(A^1\mathcal{M})$ is defined as the vector-valued one-form on $A^1\mathcal{M}$ which is pointwise given by $\mu_a = v_a \circ P^a \circ \mathrm{D}\pi_a$ for all $a \in A^1\mathcal{M}$. Its local expression with respect to the natural chart is

$$\mu = \frac{\partial}{\partial u^i} \otimes \big(\mathrm{d}q^i - u^i \mathrm{d}t\big)\,. \tag{4.45}$$

Theorem 4.4. *Let Z be a second-order field on $A^1\mathcal{M}$ and μ the vertical homomorphism on $A^1\mathcal{M}$, then η defined by*

$$\begin{aligned}\eta(X) &= \frac{1}{2}\big(X + \mathcal{L}_Z\mu(X) - \hat{\vartheta}(X)Z\big)\\ &= \frac{1}{2}\big(X + [Z,\mu(X)] - \mu([Z,X]) - \hat{\vartheta}(X)Z\big)\end{aligned} \tag{4.46}$$

for all $X \in \mathrm{Vect}(A^1\mathcal{M})$ is a connection on $A^1\mathcal{M}$. With respect to a natural chart, the connection reads as

$$\eta = \frac{\partial}{\partial u^i} \otimes \left(\mathrm{d}u^i - Z^i\mathrm{d}t - \frac{1}{2}\frac{\partial Z^i}{\partial u^j}\big(\mathrm{d}q^j - u^j\mathrm{d}t\big)\right).$$

The connection given in Theorem 4.4 was proposed in Loos [83] and a proof can be found in Eugster et al. [40]. One can easily convince oneself, that $\ker\mu \cap \ker\eta$ is a line bundle that is spanned by the second-order field Z.

Vertical vector fields can be used to define *semi-basic* forms on $A^1\mathcal{M}$ as the differential forms ω on $A^1\mathcal{M}$ for which $V \lrcorner \omega$ vanishes for any vertical vector field V. An equivalent statement is that the local representation of ω with respect to the dual basis $\mathrm{d}t, \mathrm{d}q^1, \ldots, \mathrm{d}q^n, \mathrm{d}u^1, \ldots, \mathrm{d}u^n$ induced by the natural chart (4.36) does not contain terms in $\mathrm{d}u^1, \ldots, \mathrm{d}u^n$. A differential form ω on $A^1\mathcal{M}$ is *basic*, if there is a differential form α on M such that $\omega = \pi^*\alpha$, where π^* denotes the pullback by the natural projection (4.36). Hence, basic forms are semi-basic. However, in contrast to semi-basic forms, the chart representations of the coefficients of basic forms do not depend on $u^1, \ldots, u^n$.

The vertical homomorphism μ can be used to define an operator ∂ on the exterior algebra on $A^1\mathcal{M}$ that increases the degree of a form by one. For a smooth function $f \in \Omega^0(A^1\mathcal{M})$, the operator is defined by $\partial f = \mathrm{d}f \circ \mu$ and the definition for differential forms of arbitrary degree is found in [40]. The operator ∂ is an antiderivation, that is, for $\alpha \in \Omega^r(\mathcal{M})$ and $\beta \in \Omega^\star(\mathcal{M})$, it satisfies

$$\partial(\alpha \wedge \beta) = \partial\alpha \wedge \beta + (-1)^r \alpha \wedge \partial\beta\,.$$

Moreover, in local coordinates induced by a natural chart the rules

$$\partial f = \frac{\partial f}{\partial u^i}(\mathrm{d}q^i - u^i\mathrm{d}t), \quad \partial(\mathrm{d}q^i) = \partial(\mathrm{d}t) = 0 \quad \text{and} \quad \partial(\mathrm{d}u^i) = \mathrm{d}u^i \wedge \mathrm{d}t \tag{4.47}$$

apply, where f denotes a smooth function on $A^1\mathcal{M}$.

Definition 4.9. A *Galilean manifold* is a triple $(\mathcal{M}, \vartheta, g)$, where $\mathcal{M}$ is a smooth manifold, ϑ a time structure on $\mathcal{M}$ and $g \in \Gamma(\otimes^0_2 T^*\mathcal{M})$ is a covariant tensor field of rank 2 on $\mathcal{M}$ which endows the subspaces $A^0_p\mathcal{M}$ with an inner product for all $p \in \mathcal{M}$. The tensor field g is called *Galilean metric*.

In local coordinates induced by an adapted chart (U, ϕ) of $\mathcal{M}$, the Galilean metric reads as

$$g = g_{ij}\, \mathrm{d}q^i \otimes \mathrm{d}q^j \,,$$

where the component functions g_{ij} form a symmetric, positive definite matrix $\boldsymbol{g}(p)$ for all $p \in U$. It follows immediately that the Galilean metric g on $\mathcal{M}$ is not Riemannian since $g\big(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\big) = 0$. The vertical lift is used to equip the vertical bundle $\mathrm{Ver}(A^1\mathcal{M})$ with a bundle metric $\hat{g}$, which is defined by

$$\hat{g}(X^v, Y^v) = g(X, Y) \circ \pi \quad \forall X, Y \in \Gamma(A^0\mathcal{M}) \,. \tag{4.48}$$

Using (4.44), it follows immediately from (4.48) that the bundle metric on $\mathrm{Ver}(A^1\mathcal{M})$ in coordinates is

$$\hat{g} = (g_{ij} \circ \pi)\, \mathrm{d}u^i \otimes \mathrm{d}u^j \,,$$

where g_{ij} are the component functions of g. To keep notation short, the component functions $(g_{ij} \circ \pi)$ are often identified with g_{ij}.

Let Z be a second-order field, η denote the associated connection defined by (4.46) and μ be the vertical homomorphism. The *action form of* Z is defined as the two-form $\Omega \in \Omega^2(A^1\mathcal{M})$ given by

$$\Omega(X, Y) = \hat{g}\big(\eta(X), \mu(Y)\big) - \hat{g}\big(\eta(Y), \mu(X)\big)$$

for all $X, Y \in \mathrm{Vect}(A^1\mathcal{M})$. As a consequence of the local representations (4.45) and (4.22) of the vertical homomorphism and the connection, the action form of Z locally reads as

$$\Omega = g_{ij}\left(\mathrm{d}u^i - Z^i \mathrm{d}t - \frac{1}{2}\frac{\partial Z^i}{\partial u^k}\big(\mathrm{d}q^k - u^k \mathrm{d}t\big)\right) \wedge \big(\mathrm{d}q^j - u^j \mathrm{d}t\big) \,, \tag{4.49}$$

where Z^i are the component functions of the second-order field Z. The following theorem gives necessary and sufficient conditions for a two-form to be the action form of a second-order field.

Theorem 4.5 (Eugster et al. [40], Theorem 1)**.** *Let $(\mathcal{M}, \vartheta)$ be a manifold with time structure. A two-form Ω on $A^1\mathcal{M}$ is the action form of a second-order field Z if and only if it satisfies the following conditions:*

(i) Ω vanishes on $\ker \mu$*, i.e.,*

$$\Omega(X, Y) = 0$$

for all $X, Y \in \mathrm{Vect}(A^1\mathcal{M})$ with $\mu(X) = \mu(Y) = 0$.

(ii) Ω *defines a bundle metric* g *on* $A^0\mathcal{M}$*, i.e., the matrix*

$$g_{ij} = \Omega\Big(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial q^j}\Big)$$

is symmetric and positive definite for all charts.

(iii) $\partial\Omega = 0$.

The second-order field Z *is the only vector field on* $A^1\mathcal{M}$ *for which*

$$Z \lrcorner \Omega = 0, \;\; \hat{\vartheta}(Z) = 1.$$

Since the second-order field Z is uniquely defined by its action form Ω, Theorem 4.5 allows to transfer the study of second-order fields, and by Theorem 4.4 ultimately the study of special classes of connections, to the study of action forms.

On a Galilean manifold, the spacelike projector P^R associated to a time-normalized vector field $R : \mathcal{M} \supseteq U \to A^1\mathcal{M}$ is used to introduce the scalar function

$$T_R : \pi^{-1}(U) \to \mathbb{R}, a \mapsto \frac{1}{2} g\big(P^R(a), P^R(a)\big)\,, \tag{4.50}$$

where g is the Galilean metric. With the help of (4.50), a particular action form, and hence second-order field, can be associated to any R.

Proposition 4.2 (Eugster et al. [40], Proposition 1)**.** *Let* $(\mathcal{M}, \vartheta, g)$ *be a Galilean manifold and let* T_R *be the function* (4.50) *associated to a time-normalized vector field* R*. The differential two-form*

$$\Omega_R = \mathrm{d}\big(T_R\hat{\vartheta} + \partial T_R\big) \tag{4.51}$$

is an action form that induces the bundle metric g *on* $A^0\mathcal{M}$*. The difference between an (arbitrary) action form* Ω *and* Ω_R *is a semibasic two-form* $\Phi_R := \Omega - \Omega_R$ *with the property* $\partial\Phi_R = 0$*.*

Without loss of generality, consider a chart (U, ϕ) for which $R = \frac{\partial}{\partial t}$. Since a time-normalized vector $a = (p, a_p)$ has the representation $a_p = \frac{\partial}{\partial t}\big|_p + u^i \frac{\partial}{\partial q^i}\big|_p$, we have

$$T_R = \frac{1}{2} g_{ij} u^i u^j\,, \tag{4.52}$$

where u^i are the coordinate maps of (4.36) and $g_{ij} = g_{ij} \circ \pi$ the component functions of the Galilean metric. The rules (4.47) can be used to arrive at the local expression

$$T_R\hat{\vartheta} + \partial T_R = T_R\,\mathrm{d}t + \frac{\partial T_R}{\partial u^i}(\mathrm{d}q^i - u^i\mathrm{d}t).$$

which inserted in the definition (4.51) leads to

$$\begin{aligned}\Omega_R &= \frac{\partial T_R}{\partial q^i}\mathrm{d}q^i \wedge \mathrm{d}t + \mathrm{d}\left(\frac{\partial T_R}{\partial u^i}\right) \wedge \big(\mathrm{d}q^i - u^i \mathrm{d}t\big) \\ &= \left(\mathrm{d}\left(\frac{\partial T_R}{\partial u^i}\right) - \frac{\partial T_R}{\partial q^i}\mathrm{d}t\right) \wedge \big(\mathrm{d}q^i - u^i \mathrm{d}t\big).\end{aligned} \tag{4.53}$$

Let Z_R denote the unique second-order field on $A^1\mathcal{M}$ associated to Ω_R, then the condition $Z_R \lrcorner \Omega_R = 0$ by (4.53) is equivalent to

$$Z_R\left(\frac{\partial T_R}{\partial u^i}\right) - \frac{\partial T_R}{\partial q^i} = \frac{\partial^2 T_R}{\partial t \partial u^i} + u^j \frac{\partial^2 T_R}{\partial q^j \partial u^i} + Z_R^j \frac{\partial^2 T_R}{\partial u^j \partial u^i} - \frac{\partial T_R}{\partial q^i} = 0\,, \tag{4.54}$$

where (2.14), (3.41) and (4.39) have been used. It follows from straightforward computations using (4.52), that the components of Z_R are

$$Z_R^i = -(\Gamma_{ij}^k \circ \pi) u^i u^j - g^{ij} \frac{\partial g_{jk}}{\partial t} u^k\,, \tag{4.55}$$

where g^{ij} are the components of the inverse matrix of g_{ij}, i.e., $g^{ij} g_{jk} = \delta_k^i$. Hereby, we have defined the functions

$$\Gamma_{ij}^k \coloneqq \frac{1}{2} g^{kl}\left(\frac{\partial g_{jl}}{\partial q^i} + \frac{\partial g_{li}}{\partial q^j} - \frac{\partial g_{ij}}{\partial q^l}\right). \tag{4.56}$$

It is no coincidence, that the functions (4.56) look like the Christoffel symbols (4.31) of the Levi-Civita connection on a Riemannian manifold. To see this, consider a linear connection on $T\mathcal{M}$, that is, on the tangent bundle of a Galilean manifold $(\mathcal{M}, \vartheta, g)$. The connection is *Galilean*, if it is torsion free, compatible with ϑ and metric, that is, if $\Gamma_{\lambda\mu}^\nu = \Gamma_{\mu\lambda}^\nu$, $\nabla\vartheta = 0$ and $\nabla g = 0$, where $\Gamma_{\lambda\mu}^\nu$ are the Christoffel symbols of the connection and Greek indices range from 0 to n. Let $B_\nu \coloneqq \frac{\partial}{\partial q^\nu}$ denote the basis fields induced by an adapted chart of $\mathcal{M}$, where $q^0 = t$ is set. Using (4.24) and (4.30), $\nabla\vartheta = 0$ is equivalent to

$$\nabla_{B_\mu}\vartheta\big(B_\lambda\big) = \nabla_{B_\mu}\big(\vartheta\big(B_\lambda\big)\big) - \vartheta\big(\nabla_{B_\mu} B_\lambda\big) = -\Gamma_{\mu\lambda}^0 = 0 \quad \forall \mu, \lambda\,. \tag{4.57}$$

Similarly, the representation $g = g_{ij}\,\mathrm{d}q^i \otimes \mathrm{d}q^j$ may be used to find that $\nabla g = 0$ is equivalent to

$$\nabla_{B_0} g\big(B_i, B_j\big) = \frac{\partial g_{ij}}{\partial t} - g_{kj}\Gamma_{0i}^k - g_{ik}\Gamma_{0j}^k = 0 \quad \forall i, j \tag{4.58}$$

and

$$\nabla_{B_i} g\big(B_j, B_k\big) = \frac{\partial g_{jk}}{\partial q^i} - g_{lk}\Gamma_{ij}^l - g_{jl}\Gamma_{ik}^l = 0 \quad \forall i, j, k\,. \tag{4.59}$$

By permutation of the indices, condition (4.59) can be rewritten as

$$\begin{aligned} 0 &= \nabla_{B_i} g\big(B_j, B_k\big) + \nabla_{B_j} g\big(B_k, B_i\big) - \nabla_{B_k} g\big(B_i, B_j\big) \\ &= \frac{\partial g_{jk}}{\partial q^i} + \frac{\partial g_{ki}}{\partial q^j} - \frac{\partial g_{ij}}{\partial q^k} - 2\Gamma^l_{\ ij} g_{lk} \end{aligned}$$

from which it is deduced that the Christoffel symbols Γ^k_{ij} of a Galilean connection are given by (4.56). However, since only the symmetric (with respect to g) part of the Christoffel symbols Γ^j_{0i} are fixed by condition (4.58) and the Γ^i_{00} are unconstrained, there is no unique Galilean connection on $T\mathcal{M}$. This is in contrast to the Riemannian case, where the Levi-Civita connection is the only linear connection which is torsion-free and metric, see Theorem 4.3.

Let Z_∇ be the second-order field on $T\mathcal{M}$ associated to a Galilean connection by Theorem 4.4. Then by (4.57), the component function Z^0_∇ of Z vanishes, such that

$$Z_\nabla = u^0 \frac{\partial}{\partial t} + u^i \frac{\partial}{\partial q^i} - (\Gamma^i_{\mu\nu} \circ \pi) u^\mu u^\nu \frac{\partial}{\partial u^i} \,. \tag{4.60}$$

Since $Z^0_\nabla = 0$, it follows from (4.28) that for integral curve $\dot\gamma : \mathbb{R} \supseteq \mathcal{I} \to T\mathcal{M}$ of Z_∇ it holds that $u^0 \circ \dot\gamma = \vartheta(\dot\gamma)$ is constant. Hence, by choice of an appropriate initial condition, one can single out time-parametrized integral curves of Z_∇, and it makes sense to speak of time-parametrized integral curves of Z_∇. By definition, a time-parameterized integral curve of Z_∇ is a second-order curve on $A^1\mathcal{M}$. Moreover, such a second-order curve is an integral curve of the second-order field on $A^1\mathcal{M}$ gotten from setting $u^0 = 1$ in (4.60). Consequently, to every Galilean connection, the second-order field Z on $A^1\mathcal{M}$ defined by the coefficients $Z^i = -(\Gamma^i_{\mu\nu} \circ \pi) u^\mu u^\nu$ can be associated, where $u^0 = 1$. Conversely, if the coefficients of a second-order field on $A^1\mathcal{M}$ have this form and the functions $\Gamma^i_{\mu\nu}$ satisfy (4.58) and (4.56), the functions $\Gamma^\lambda_{\mu\nu}$ define a Galilean connection on $T\mathcal{M}$ if, in line with (4.57), $\Gamma^0_{\mu\nu} = 0$ is set.

Proposition 4.3 (Loos [82], Lemma 3)**.** *Let R be a time-normalized vector field on a Galilean manifold. The second-order field Z_R on $A^1\mathcal{M}$ defined by Ω_R is associated to a Galilean connection with ∇^R on $T\mathcal{M}$. Moreover, for any time-parametrized curve on $\mathcal{M}$, it holds that*

$$\big(\nabla^R_{\dot\gamma} \dot\gamma\big)^v = \ddot\gamma - Z_R \circ \dot\gamma \,., \tag{4.61}$$

where v is the vertical lift from $A^0\mathcal{M}$ to $\mathrm{Ver}(A^1\mathcal{M})$.

Hence, the projection of the integral curves of Z_R to $\mathcal{M}$ are exactly the curves γ which are autoparallel, i.e., γ satisfies $\nabla^R_{\dot\gamma}\dot\gamma = 0$.

Proof. Without loss of generality, consider a chart (U, ϕ) for which $R = \frac{\partial}{\partial t}$. By (4.55), the components of Z_R then have the form

$$Z_R^i = -(\Gamma^i_{\mu\nu} \circ \pi) u^\mu u^\nu \quad \text{with} \quad \Gamma^i_{j0} = \frac{1}{2} g^{ik} \frac{\partial g_{kj}}{\partial t} \quad \text{and} \quad \Gamma^i_{00} = 0\,, \tag{4.62}$$

which satisfy (4.57) and (4.56). Hence, setting $\Gamma^0_{\mu\nu} = 0$ defines a Galilean connection on $T\mathcal{M}$. Since the Galilean connection satisfies $\nabla^R \vartheta = 0$ and for a time-parametrized curve γ we have $\vartheta(\dot\gamma) = 1$, it follows from (4.30) that

$$0 = \nabla^R_{\dot\gamma}\big(\vartheta(\dot\gamma)\big) = \nabla^R_{\dot\gamma}\vartheta(\dot\gamma) + \vartheta\big(\nabla^R_{\dot\gamma}\dot\gamma\big) = \vartheta\big(\nabla^R_{\dot\gamma}\dot\gamma\big)$$

implying that $\nabla^R_{\dot\gamma}\dot\gamma$ is spacelike and hence can be lifted by v. Moreover, from (4.24) and (4.62), we respectively have that

$$\begin{aligned} \nabla^R_{\dot\gamma}\dot\gamma &= \big(\ddot q^i + (\Gamma^i_{\mu\nu} \circ \gamma)\dot q^\mu \dot q^\nu\big) \frac{\partial}{\partial q^i}\Big|_\gamma \\ \ddot\gamma - Z_R \circ \dot\gamma &= \big(\ddot q^i + (\Gamma^i_{\mu\nu} \circ \gamma)\dot q^\mu \dot q^\nu\big) \frac{\partial}{\partial u^i}\Big|_{\dot\gamma}\,, \end{aligned} \tag{4.63}$$

where the notational overlap $q^\mu = q^\mu \circ \gamma$ implying $\dot q^0 = 1$ is used. It immediately follows from (4.43), that (4.61) is satisfied. □

The study of Galilean manifolds is a rather specialized topic, which has emerged from the objective of treating Newtonian gravity in the same fashion as relativistic gravity. In general relativity, the gravity of matter determines the metric as well as a linear connection on a four-dimensional space-time. This is why the literature on nonlinear connections on Galilean manifolds, and therefore on general second-order fields, is scarce. A comprehensive treatise on general second-order fields on Galilean manifolds is Loos [82], which is written in German. A large part of the exposition of Loos can be found in Eugster et al. [40].

— 5 —

Dynamics of Mechanical Systems on Galilean Manifolds

> And how the one of time, of space
> the three, might in the chain of
> symbols girdled be.
>
> —*William R. Hamilton*

In this chapter, a geometric theory for the description of smooth motions of mechanical systems is developed. The focus lies on the development of a consistent theory as well as bridging the gap between theoretical mechanics and engineering mechanics by showing that the equations known in engineering mechanics result as local representations of the presented geometric theory. This chapter pursues the way taken by Loos [82] and Winandy [127]. However, a different golden thread is chosen, such that the link between the dynamics of the system and the dynamics of the particles constituting the system can be addressed, which represents an extension to [82, 127].

5.1. On Space, Time and Space-time

In classical mechanics, and particularly in engineering mechanics, space and time are mostly treated as independent concepts. Particles are regarded to be moving in space, where the motion is a function of time. This section briefly motivates that the separation of space and time is very effective but questionable from the viewpoint of theoretical mechanics.

A typical mathematical abstraction of space is the three-dimensional Euclidean point space $\mathcal{P}$, see Definition 2.2.[1] Using the real numbers to model time, the

[1] See for example [24, 53, 59, 60, 63, 97, 99, 103, 107, 115]. Note, that even if the Euclidean point space is not mentioned explicitly, the motion is illustrated by a point moving in a "space" for which it is meaningful to draw arrows between points. The mathematical construction shown in this section is nothing but an abstraction of this picture.

motion of a particle P is regarded as a curve $\xi_P : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{P}$, which expresses the idea that the point in space occupied by the particle P may change with time. Since motion can only be perceived relatively, a reference particle[2] O is chosen and the position vector $\mathbf{r}_{OP}(t) \coloneqq \mathbf{r}(\xi_O(t), \xi_P(t)) \in \mathbb{E}^3$ of P at time t is defined. Thus, for a fixed time instant t, $\mathbf{r}_{OP}(t)$ represents the vector pointing from the point $\xi_O(t)$ in space $\mathcal{P}$ occupied by the reference particle O to the point $\xi_P(t)$ in space occupied by the particle P. Often, the reference particle O is regarded as a fixed origin, which might be expressed by $\xi_O = \text{const}$. The velocity and the acceleration of the particle are defined by $\mathbf{v}_P \coloneqq \dot{\mathbf{r}}_{OP}$ and $\mathbf{a}_P \coloneqq \ddot{\mathbf{r}}_{OP}$, respectively. With these definitions, Newton's second law is formulated as $\mathbf{F} = m\,\mathbf{a}_P$, where the scalar m denotes the mass of the particle and $\mathbf{F} \in \mathbb{E}^3$ is the force acting on the particle P. Representing Newton's law with respect to a basis of $\mathbb{E}^3$ results in second-order differential equations describing the motion of the particle, i.e., the equations of motion of the particle.

This approach to particle mechanics is very efficient as it involves only a few mathematical concepts and it is therefore the theory of choice if one is only interested in the equations of motion. This explains why the Euclidean space approach is prevalent in engineering mechanics. From the perspective of theoretical mechanics however, this approach is not satisfactory, since concepts which are physically different are lumped together, making it impossible to mathematically tell the difference between for example a force and an acceleration. This has the consequence, that one could find mathematically allowed relations between e.g. force, position and velocity, which have no physical meaning. To prevent this, and hence capture the "nature" of the physical concepts, it is one aim of theoretical mechanics, to associate a unique mathematical object to every physical quantity, such that every mathematically allowed expression is physically meaningful and can be validated experimentally. In turn, this also means that no mathematical object or structure should be introduced, which has no physical meaning or cannot be validated experimentally. In that spirit, the use of an absolute space $\mathcal{P}$ has been a controversial issue in mechanics. Newton, for instance, was an advocate for the existence of an absolute space, whereas Leibniz was of the opinion that space made no sense except as the relative location of particles. A preliminary solution to the controversy has been provided by general relativity, which led to the widely accepted idea that space and time cannot be distinguished and must be seen as two aspects of the same entity called

[2] One may also choose a reference point which is not a particle, e.g. the midpoint of the laboratory in which the described experiment is performed. However, such a point can only be characterized in relation to material points in the first place. Hence, a reference particle can be used without loss of generality.

space-time[3], modeled as a four-dimensional semi-Riemannian[4] manifold. For the concern of classical mechanics, it has been shown by Ehlers [34][5] that space-time can be modeled as a four-dimensional Galilean manifold in the Newtonian limit of general relativity, i.e., the limit where the speed of light tends to infinity.

5.2. Classical Space-time

In this section, the physical space-time of classical mechanics is modeled as a four-dimensional Galilean manifold. It is shown that the Galilean structure models the acts of measuring time as well as distances and angles. Moreover, the concept of rest and observer are introduced, which are invoked to retrieve the notions of position, velocity and acceleration as vectors in $\mathbb{E}^3$.

Let space-time be represented by a four-dimensional manifold $\mathcal{E}$. A chronometer can be used to measure the temporal distance between two events. This is abstracted by introducing a time structure θ on the space-time $\mathcal{E}$. The time showed by a chronometer is then represented as a time function $t : \mathcal{E} \supseteq U \to \mathbb{R}$, which by definition fulfills $\theta\big|_U = \mathrm{d}t$. The temporal distance between two events $e_1, e_2 \in U$ is

$$|t(e_2) - t(e_1)| \,. \tag{5.1}$$

We say that e_1 happened before e_2 if $t(e_2) > t(e_1)$ and that both events are *simultaneous* if $t(e_2) = t(e_1)$. Two different time functions t and $\bar{t}$ defined on the subsets U and $\bar{U}$ of $\mathcal{E}$, respectively, fulfill $\mathrm{d}t = \mathrm{d}\bar{t}$ on $U \cap \bar{U}$. Consequently, they agree on the temporal distance between two events as they can only differ by a constant c, that is,

$$\bar{t} = t + c \,. \tag{5.2}$$

The spacelike bundle $A^0\mathcal{E} = \ker\theta \subset T\mathcal{E}$, see (4.33), is integrable and its integral manifolds define a foliation $\mathcal{F}$ of the space-time $\mathcal{E}$. The leaves of $\mathcal{F}$ are submanifolds of simultaneous events, because by construction, any time function is constant when restricted to a leaf $\mathcal{S} \in \mathcal{F}$.

The idea that distances and angles can be instantly measured is represented in the model by assuming that every leaf of $\mathcal{F}$ locally looks like a three-dimensional Euclidean point space $\mathcal{P}$. Specifically, for any event there exists a neighborhood U such that every nonempty intersection of U with a leaf of $\mathcal{F}$ is diffeomorphic to a subset of $\mathcal{P}$. Hence, let $t : \mathcal{E} \supseteq U \to \mathbb{R}$ be a time function and let $\mathcal{S}_\tau \in \mathcal{F}$

[3] For a discussion of different space-time models encountered in physics, see Toupin [122], and Bekaert and Morand [12].

[4] A semi-Riemannian manifold is a Riemannian manifold for which the positive-definiteness of the metric is relaxed to a non-degeneracy condition.

[5] Ehlers [34] has been republished in English, see [35].

denote the leaf on which t takes the value τ, i.e., $t(\mathcal{S}_\tau \cap U) = \tau$, then there is a diffeomorphism $\varphi_\tau : \mathcal{S}_\tau \cap U \to V \subseteq \mathcal{P}$. Consequently, for any two simultaneous events $e, f \in \mathcal{S}_\tau \cap U$ there is a unique vector

$$\mathbf{r}_{ef} \coloneqq \mathbf{r}\big(\varphi_\tau(e), \varphi_\tau(f)\big) \in \mathbb{E}^3$$

pointing from $\varphi_\tau(e)$ to $\varphi_\tau(f)$, see Definition 2.2. Loosely speaking, $\mathcal{S}_\tau$ denotes the space at time $t = \tau$. For different time instants, the spaces are different, but locally diffeomorphic. However, there is no canonical way to identify points on spaces at different times, which is in contrast to the concept of absolute space.

Definition 5.1. For a particle P, also called material point P, the *motion* is defined as a time-parametrized curve $\xi_P : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{E}$. Hence, the *(absolute) velocity* $\dot{\xi}_P$ of P has the property $\theta(\dot{\xi}_P) = 1$.

For any time function t, a motion ξ_P fulfills

$$\mathrm{d}t(\dot{\xi}_P) = (t \circ \xi_P)\dot{} = 1 \quad \Rightarrow \quad t \circ \xi_P(\tau) = \tau + \tau_0 \tag{5.3}$$

by Definition 5.1, where $\tau_0 \in \mathbb{R}$ is a constant. The motions of two particles P and Q are said to be *synchronized*, if $\xi_P(\tau)$ and $\xi_Q(\tau)$ are simultaneous for one and hence for any τ. Loosely speaking, it is assumed that time passes equally for every particle and that time can be used to parametrize the motion, because, up to a constant, t equals τ along the motion. For this reason, henceforth t is used instead of τ to parametrize the motion as the constant τ_0 in (5.3) can always be eliminated by changing the time function, see (5.2). For the same reason, two motions can always be synchronized.

Typically, one observes an event e relative to the motion $\xi_O : \mathcal{I} \to U \subseteq \mathcal{E}$ of some chosen *reference (material) point* O. Using an appropriate time function $t : \mathcal{E} \supseteq U \to \mathbb{R}$ with $t(U) \subseteq \mathcal{I}$, an event $e \in U$ is simultaneous to $\xi_O(t(e))$, such that e can uniquely be associated with the position vector $\mathbf{x} = \mathbf{r}_{\xi_O(t(e))e} \in \mathbb{E}^3$ from the event $\xi_O(t(e))$ to e. We can say, that the event e happens at time $t = t(e)$ at the position $\mathbf{x} = \mathbf{r}_{\xi_O(t(e))e}$ with respect to O. Hence, the choice of a reference point O defines the map

$$\phi_O : E \supseteq U \mapsto \mathbb{R} \times \mathbb{E}^3,\ e \mapsto (t, \mathbf{x}) = \big(t(e), \mathbf{r}_{\xi_O(t(e))e}\big)\,. \tag{5.4}$$

Introducing a *frame of reference* consisting of three time-dependent vectors $\mathbf{e}_i^I : \mathbb{R} \supseteq \mathcal{I} \to \mathbb{E}^3$ $(i = 1, 2, 3)$, which form an orthonormal basis of $\mathbb{E}^3$, i.e., $\mathbf{e}_i^I(t) \cdot \mathbf{e}_j^I(t) = \delta_{ij}$ for all $t \in \mathcal{I}$, allows to define the chart

$$K_I : \mathcal{I} \times \mathbb{E}^3 \to \mathbb{R}^4,\ \big(t, \mathbf{x} = {}_I x^i \mathbf{e}_i^I(t)\big) \mapsto \big(t, {}_I x^1, {}_I x^2, {}_I x^3\big) = (t, {}_I\mathbf{x}) \tag{5.5}$$

of $\mathbb{R} \times \mathbb{E}^3$. This chart reads out the components with respect to the basis I. The concatenation

$${}_I\phi_O = K_I \circ \phi_O : \mathcal{E} \supseteq U \mapsto \mathbb{R}^4,\ e \mapsto (t, {}_I\mathbf{x}) = \big(t(e), {}_I\mathbf{r}_{\xi_O(t(e))e}\big) \tag{5.6}$$

is an adapted chart of $\mathcal{E}$ and is referred to as *Cartesian chart of* $\mathcal{E}$. Obviously, other charts of $\mathcal{E}$ can be constructed, e.g., $\mathbf{x} = \mathbf{r}_{\xi_O(t(e))e}$ can be addressed by its length and two appropriately chosen angles with respect to the frame of reference I, which would lead to a chart of $\mathcal{E}$ representing spherical coordinates. Associating a space-time event $e \in \mathcal{E}$ with four real numbers, i.e. choosing a chart of $\mathcal{E}$, may be regarded as an observation of the event e. Since the construction of a chart relies on a point of reference O as well as on a frame of reference $(\mathbf{e}_1^I, \mathbf{e}_2^I, \mathbf{e}_3^I)$, their combination, that is, $(O, \mathbf{e}_1^I, \mathbf{e}_2^I, \mathbf{e}_3^I)$, is called *observer*.

To derive the change of Cartesian coordinates, consider the motion of a different reference point $\bar{O}$ and a time function $\bar{t} : \mathcal{E} \supseteq \bar{U} \to \mathbb{R}$. Following the construction above, $e \in U \cap \bar{U}$ is mapped to the position vector $\bar{\mathbf{x}} = \mathbf{r}_{\xi_{\bar{O}}(\bar{t}(e))e} \in \mathbb{E}^3$. Hence, e, $\xi_{\bar{O}}(\bar{t}(e))$ and $\xi_O(t(e))$ are simultaneous. Consequently, introducing $\phi_{\bar{O}}(e) = (\bar{t}, \bar{\mathbf{x}})$ and using (5.2), the change of reference point $\phi_{\bar{O}} \circ \phi_O^{-1}$ is given by

$$\begin{aligned} \bar{t} &= t + c \\ \bar{\mathbf{x}} &= \mathbf{r}_{\xi_{\bar{O}}(\bar{t}(e))e} = \mathbf{r}_{\xi_O(t(e))e} + \mathbf{r}_{\xi_O(t(e)+c)\xi_O(t(e))} = \mathbf{x} + \mathbf{r}_{\bar{O}O}(t(e)) \,, \end{aligned}$$

where the notation $\mathbf{r}_{\bar{O}O}(t(e)) = \mathbf{r}_{\xi_{\bar{O}}(t(e)+c)\xi_O(t(e))}$ has been introduced. It then follows from the representation of $\bar{\mathbf{x}}$ with respect to a basis B, that the change of Cartesian coordinates is given by

$$\begin{aligned} {}_B\phi_{\bar{O}} \circ {}_I\phi_O^{-1} : \; & {}_I\phi(U \cap \bar{U}) \to {}_B\phi(U \cap \bar{U}), \\ & \begin{pmatrix} t \\ {}_I\mathbf{x} \end{pmatrix} \mapsto \begin{pmatrix} \bar{t} \\ {}_B\bar{\mathbf{x}} \end{pmatrix} = \begin{pmatrix} t + c \\ \mathbf{A}_{BI}(t) {}_I\mathbf{x} + {}_B\mathbf{r}_{\bar{O}O}(t) \end{pmatrix} , \end{aligned} \tag{5.7}$$

where the transformation matrix is defined as

$$(\mathbf{A}_{BI})^j{}_i(t) = \mathbf{e}_j^B(t + c) \cdot \mathbf{e}_i^I(t) \tag{5.8}$$

for all t, see (2.4). This change of Cartesian coordinates induces the change of induced basis fields

$$\begin{aligned} \frac{\partial}{\partial t} &= \frac{\partial}{\partial \bar{t}} + \Big(\dot{\mathbf{A}}_{BI}(t) \mathbf{A}_{BI}^{\mathrm{T}}(t) \big({}_B\bar{\mathbf{x}} - {}_B\mathbf{r}_{\bar{O}O}(t) \big) + {}_B\dot{\mathbf{r}}_{\bar{O}O}(t) \Big)^j \frac{\partial}{\partial {}_B\bar{x}^j} \\ \frac{\partial}{\partial {}_Ix^i} &= (\mathbf{A}_{BI})^j{}_i(t) \frac{\partial}{\partial {}_B\bar{x}^j} \,, \end{aligned} \tag{5.9}$$

which follows from (5.7) by (3.19). Note, that $\dot{\mathbf{A}}_{BI}$ and ${}_B\dot{\mathbf{r}}_{\bar{O}O}$ denote the derivative of $\mathbf{A}_{BI}$ and ${}_B\mathbf{r}_{\bar{O}O}$, respectively. Moreover, the evaluation at t must be read as the concatenation with the coordinate map t of ${}_I\phi_O$.

Consider the map $\phi_O : \mathcal{E} \supseteq U \to \mathbb{R} \times \mathbb{E}^3$ stemming from the choice of a reference point O and a time function $t : U \to \mathbb{R}$, see (5.4). For $e \in U$, the differential $(\mathrm{D}\phi_O)_e : T_e\mathcal{E} \to T_{\phi_O(e)}(\mathbb{R} \times \mathbb{E}^3) \cong \mathbb{R} \times \mathbb{E}^3$ by (3.20) maps the basis

vectors of $T_e\mathcal{E}$ induced by the Cartesian chart ${}_I\phi_O$ to

$$(\mathrm{D}\phi_O)_e \frac{\partial}{\partial t}\bigg|_e = \big(1, {}_Ix^j(e)\,\dot{\mathbf{e}}_j^I(t(e))\big) \quad \text{and} \quad (\mathrm{D}\phi_O)_e \frac{\partial}{\partial {}_Ix^i}\bigg|_e = \big(0, \mathbf{e}_i^I(t(e))\big)\,, \tag{5.10}$$

where the chart (5.5) of $\mathbb{R}\times\mathbb{E}^3$ has been used. Hence, $\mathrm{D}\phi_O$ can be used to map vector fields on U yielding

$$\mathrm{D}\phi_O\left({}_Iv^0\frac{\partial}{\partial t} + {}_Iv^i\frac{\partial}{\partial {}_Ix^i}\right) = \big({}_Iv^0,\ {}_Iv^0\,{}_Ix^i\dot{\mathbf{e}}_i^I(t) + {}_Iv^i\mathbf{e}_i^I(t)\big)\,,$$

with t and ${}_Ix^i$ being the coordinate maps of the chart ${}_I\phi_O$ of $\mathcal{E}$.

It is immediately clear from (5.10) that the restriction $\mathrm{D}\phi_O|_{A_e^0\mathcal{E}}$ is a vector space isomorphism between $A_e^0\mathcal{E}$ and $\{0\}\times\mathbb{E}^3\cong\mathbb{E}^3$.

Proposition 5.1. *For two different reference points, let ϕ_O and $\phi_{\bar O}$ be defined on the subsets U and $\bar U$ of $\mathcal{E}$, respectively. For any $e\in U\cap\bar U$, it holds that*

$$(\mathrm{D}\phi_{\bar O})_e v_e = (\mathrm{D}\phi_O)_e v_e \quad \forall v_e\in A_e^0\mathcal{E}\,. \tag{5.11}$$

Proof. Let ${}_I\phi_O(e) = (t, {}_I\mathbf{x})$ and ${}_B\phi_{\bar O}(e) = (\bar t, {}_B\bar{\mathbf{x}})$. A spacelike vector

$$v_e = {}_Iv^i\frac{\partial}{\partial {}_Ix^i}\bigg|_e = {}_Bv^i\frac{\partial}{\partial {}_B\bar x^i}\bigg|_e$$

by (5.10) is mapped to

$$(\mathrm{D}\phi_O)_e v_e = (0, {}_Iv^i\mathbf{e}_i^I(t)) \quad \text{and} \quad (\mathrm{D}\phi_{\bar O})_e v_e = (0, {}_Bv^i\mathbf{e}_i^B(\bar t))\,. \tag{5.12}$$

It follows from the change of basis fields (5.9) that ${}_B\mathbf{v} = \mathbf{A}_{BI}(\bar t)\,{}_I\mathbf{v}$ and consequently

$${}_Bv^j\mathbf{e}_j^B(\bar t) = (\mathbf{A}_{BI})^j{}_i(\bar t)\,{}_Iv^i\mathbf{e}_j^B(\bar t) = {}_Iv^i(\mathbf{e}_j^B(\bar t)\cdot\mathbf{e}_i^I(t))\mathbf{e}_j^B(\bar t) = {}_Iv^i\mathbf{e}_i^I(t)\,,$$

because of (5.8) and (2.3). This implies (5.11) by (5.12). □

Proposition 5.1 assures that the space-time $\mathcal{E}$ can naturally be equipped with a Galilean metric defined by

$$g_e : A_e^0\mathcal{E}\times A_e^0\mathcal{E}\to\mathbb{R},\ (u_e, v_e)\mapsto g_e(u_e, v_e) = (\mathrm{D}\phi_O)_e u_e\cdot(\mathrm{D}\phi_O)_e v_e \tag{5.13}$$

for every event $e\in\mathcal{E}$ and arbitrary reference point O, where the "dot" denotes the inner product of $\{0\}\times\mathbb{E}^3\cong\mathbb{E}^3$. Hence, space-time is a Galilean manifold $(\mathcal{E},\theta,g)$. It follows from (5.10), that the Galilean metric of space-time with respect to a Cartesian chart ${}_I\phi_O$ reads as

$$g = \delta_{ij}\,\mathrm{d}\big({}_Ix^i\big)\otimes\mathrm{d}\big({}_Ix^j\big)\,. \tag{5.14}$$

The fact that time passes for every particle means that every particle necessarily moves through space-time. Moreover, one experiences that whenever motion is perceived, this is done relative to some reference. For instance, the motion of a car may be observed relative to the street or the motion of the sun relative to the horizon. Especially when it comes to observing velocities, only relative velocities can be observed and never the absolute velocity of a particle in space-time, because the velocity at which time itself changes is not observable. A *reference field* is a time-normalized vector field

$$r : \mathcal{E} \supseteq U_r \to A^1\mathcal{E} \,.$$

With that, the *velocity of a particle P with respect to the reference field r*, or *r-velocity* for short, is defined as the spacelike vector field

$$v_P \coloneqq P^r(\dot{\xi}_P) = \dot{\xi}_P - r \circ \xi_P \tag{5.15}$$

along the motion $\xi_P : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{E}$ of the particle, where P^r denotes the spacelike projector (4.40). Based on that, a particle P is regarded to be at *rest with respect to r* if its r-velocity is zero. It follows directly from the definition (5.15) that a particle is at rest w.r.t. r if and only if the motion ξ_P is an integral curve of the reference field, i.e., $\dot{\xi}_P = r \circ \xi_P$. An adapted chart ϕ of $\mathcal{E}$ with $\phi(e) = (t, x^1, x^2, x^3)$ is said to be a *resting chart with respect to r* if the reference field has the representation $r = \frac{\partial}{\partial t}$. This definition implies that the spatial coordinates $x^i \circ \xi_P$ of a resting material point P are constant.

The second-order field Z_r determined by the action form Ω_r, see (4.51), is used to define the *acceleration of P with respect to r*, or *r-acceleration*, as the vector field

$$a_P = \ddot{\xi}_P - Z_r \circ \dot{\xi}_P \tag{5.16}$$

along the velocity $\dot{\xi}_P : \mathbb{R} \supseteq \mathcal{I} \to A^1\mathcal{E}$ of the particle.

The fact that the leaves of simultaneous events locally look like an Euclidean point space can be exploited to retrieve the formalism on $\mathbb{E}^3$ typically used in textbooks about classical mechanics. As usual, the transition to $\mathbb{E}^3$ is achieved by choosing a reference point O, that is, by the map ϕ_O and its differential map in our case.

The motion of a particle P relative to the motion of O is described by

$$\phi_O \circ \xi_P(t) = (t, \mathbf{r}_{OP}(t)) \,,$$

where the *position vector* of P with respect to O is defined as

$$\mathbf{r}_{OP}(t) \coloneqq \mathbf{r}_{\xi_O(t)\xi_P(t)} \,.$$

Since the restriction of $(\mathrm{D}\phi_O)_e$ to $A^0_e\mathcal{E}$ is a vector space isomorphism between $A^0_e\mathcal{E}$ and $\{0\}\times\mathbb{E}^3\cong\mathbb{E}^3$, the *velocity vector* $\mathbf{v}_P$ of P can be defined by

$$(0,\ \mathbf{v}_P) :=\ \mathrm{D}\phi_O v_P\,, \tag{5.17}$$

where (3.34) has been used. By Proposition 5.1, the definition of the velocity vector is independent of the choice of reference point O. A similar procedure can be used for the acceleration. By Proposition 4.3, it holds for the r-acceleration a_p, defined in (5.16), that

$$a_P = \left(\nabla^r_{\dot\xi_P}\dot\xi_P\right)^v.$$

Since the vertical lift v establishes a vector space isomorphism between $A^0_{\pi(a)}\mathcal{E}$ and $\mathrm{Ver}_a(A^1\mathcal{E})$ for all $a\in A^1\mathcal{E}$, defining the *acceleration vector* by

$$(0,\ \mathbf{a}_P) :=\ \mathrm{D}\phi_O\left(\nabla^r_{\dot\xi_P}\dot\xi_P\right)$$

uniquely associates $\mathbf{a}_P$ to a_P.

With respect to an arbitrary Cartesian chart ${}_I\phi_O$, see (5.6), the motion of the particle P has the representation

$${}_I\phi_O\circ\xi_P(t) = K_I\circ\phi_O\circ\xi_P(t) = (t, {}_I\mathbf{r}_{OP}(t))\,,$$

where ${}_I\mathbf{r}_{OP}$ collects the components of the position vector $\mathbf{r}_{OP}$ with respect to the frame of reference $(\mathbf{e}^I_1,\mathbf{e}^I_2,\mathbf{e}^I_3)$. Hence, by (3.36), the (absolute) velocity $\dot\xi_P$ can be locally represented as

$$\dot\xi_P = \frac{\partial}{\partial t}\Big|_{\xi_p} + {}_I\dot r_{OP}\frac{\partial}{\partial {}_Ix^i}\Big|_{\xi_p}\,. \tag{5.18}$$

Suppose that ${}_I\phi_O$ is a resting chart for r, i.e., $r=\frac{\partial}{\partial t}$. It then follows from (5.15) and (5.18) that the r-velocity of the particle P is

$$v_P = {}_I\dot r_{OP}\frac{\partial}{\partial {}_Ix^i}\Big|_{\xi_p}\,. \tag{5.19}$$

Moreover, by definition (5.15) of the velocity vector $\mathbf{v}_P$ it can be seen that

$${}_I\mathbf{v}_P = {}_I\dot{\mathbf{r}}_{OP}\,, \tag{5.20}$$

where (5.10) and (5.19) are involved in the computations. Furthermore, since the components of the Galilean metric (5.14) are constant with respect to the resting chart ${}_I\phi_O$, the Christoffel symbols of ∇^r all vanish, see (4.62). Hence, it follows from (4.63) and (5.10) that

$${}_I\mathbf{a}_P = {}_I\ddot{\mathbf{r}}_{OP}\,. \tag{5.21}$$

Finally, a comparison of (5.20) and (5.21) with (2.6) shows that it suffices to identify the reference frames $\mathbf{e}_i^I$ leading to resting Cartesian charts ${}_I\phi_O$ with respect to r with resting frames of $\mathbb{E}^3$ to retrieve the typical formulae

$$\mathbf{v}_P = \dot{\mathbf{r}}_{OP} \quad \text{and} \quad \mathbf{a}_P = \ddot{\mathbf{r}}_{OP} \tag{5.22}$$

for the velocity and acceleration vectors of a particle P.

5.3. Configuration Space-time and State Space

Definition 5.2. A *mechanical system of dimension* n is a measurable set S of particles together with an $(n+1)$-dimensional configuration space-time manifold $\mathcal{M}$, as well as a map

$$\kappa :\ S \times \mathcal{M} \to \mathcal{E},\ (x,p) \mapsto \kappa(x,p)$$

defining the kinematics of the system. The map κ is assumed to be such that for a fixed $p \in \mathcal{M}$ all events contained in the image $\kappa(S,p)$ are simultaneous, i.e., every time function of $\mathcal{M}$ is constant on $\kappa(S,p)$. Moreover, with the notation $\kappa_x(\,.\,) = \kappa(x,\,.\,)$, the map κ is such, that through any $p \in \mathcal{M}$ there is a curve for which

$$\theta(\mathrm{D}\kappa_x\dot{\gamma}) \neq 0\,. \tag{5.23}$$

The event $\kappa(x,p)$ is occupied by the particle $x \in S$ in the configuration $p \in \mathcal{M}$ of the mechanical system. Using this terminology, the two conditions on κ can be restated in the following way. For a given configuration $p \in M$, the events occupied by the system S are simultaneous and there exist evolutions of configurations $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ such that time passes along the events $\kappa_x \circ \gamma$ occupied by every particle x. The latter can be seen from

$$\theta(\mathrm{D}\kappa_x\dot{\gamma}) = \mathrm{d}t_{\mathcal{E}}\big((\kappa_x \circ \gamma)\dot{}\big) = (t_{\mathcal{E}} \circ \kappa_x \circ \gamma)\dot{} \neq 0\,,$$

where $t_{\mathcal{E}} : \mathcal{E} \supset U \to \mathbb{R}$ is a time function on $\mathcal{E}$ and where (3.34) has been invoked.

For a time function $t_{\mathcal{E}}$, the value of $t_{\mathcal{E}} \circ \kappa(x,\,.\,)$ is independent of the choice of the system's particle $x \in S$, since for a given configuration, events occupied by the system S are simultaneous. This allows to associate to every time function $t_{\mathcal{E}}$ on $\mathcal{E}$ a time function

$$t :\ \mathcal{M} \supset \kappa_x^{-1}(U) \to \mathbb{R},\ p \mapsto t(p) = t_{\mathcal{E}} \circ \kappa(x,p) \tag{5.24}$$

on $\mathcal{M}$, where $x \in S$ can be chosen arbitrarily. Since time functions on $\mathcal{E}$ exist in the neighborhood of every event in $\mathcal{E}$, it follows by definition (5.24) that a time function of the form (5.24) exists in some neighborhood of every configuration in $\mathcal{M}$. Therefore, a global one-form $\vartheta \in \Omega^1(\mathcal{M})$ on $\mathcal{M}$ can be defined by setting

$$\vartheta = \mathrm{d}t \tag{5.25}$$

on the domain of any time function t of $\mathcal{M}$. The global one-form ϑ is well-defined, because for a different time function $\bar{t}$, it follows from (5.2) and (5.24) that $\mathrm{d}t = \mathrm{d}\bar{t}$ on the intersection of the domains of the two time functions. Hence, t and $\bar{t}$ define the same one-form ϑ for points on $\mathcal{M}$ lying in the domains of both time functions. Consider a curve $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$, then, independently of the chosen particle x,

$$\vartheta(\dot{\gamma}) = \mathrm{d}t(\dot{\gamma}) = (t \circ \gamma)^{\cdot} = (t_{\mathcal{E}} \circ \kappa_x \circ \gamma)^{\cdot} = \mathrm{d}t_{\mathcal{E}}\left((\kappa_x \circ \gamma)^{\cdot}\right) = \theta(\mathrm{D}\kappa_x \dot{\gamma})\,, \qquad (5.26)$$

where (5.25), (5.24) and the definition of the differential (3.34) for vector fields along γ have been used. Consequently, the one-form ϑ is non-vanishing by (5.23). Moreover, ϑ is closed by definition (5.25). Hence, the one-form ϑ defines a time structure on $\mathcal{M}$, implying the existence of adapted charts

$$\phi :\ \mathcal{M} \supseteq U \to \mathbb{R}^{n+1},\ p \mapsto \phi(p) = (t, q^1, \ldots, q^n) = (t, \mathbf{q}) \qquad (5.27)$$

of the configuration space-time $\mathcal{M}$, where the first coordinate function is a time function.

Equation (5.26) reveals that for any particle x, time-parametrized curves in $\mathcal{M}$ are mapped to time-parametrized curves in $\mathcal{E}$. Moreover, $\mathrm{D}\kappa_x$ maps time-normalized and spacelike vectors on $\mathcal{M}$ to time-normalized and spacelike vectors on $\mathcal{E}$, respectively. Hence, if γ is a time-parametrized curve on $\mathcal{M}$ then the curve

$$\xi_x(t) = \kappa_x \circ \gamma(t) = \kappa(x, \gamma(t)) \qquad (5.28)$$

is a motion of a material point through space-time for every particle x, that is, ξ_x is a time-parametrized curve on $\mathcal{E}$ for every x.

Definition 5.3. A *motion* of the mechanical system in the configuration space-time $\mathcal{M}$ is a time-parametrized curve $\gamma : \mathbb{R} \supseteq I \to \mathcal{M}$, i.e., a curve with the property $\vartheta(\dot{\gamma}) = 1$.

Example 5.1. Consider a pendulum consisting of a thin bar attached to the ceiling through a pivot such that it can only perform planar motions. Since the bar is thin, the set S of all particles x can be modeled as an interval of $\mathbb{R}$. The position of the pendulum can be described using one coordinate only, e.g., an angle φ. Hence, the space-time manifold $\mathcal{M}$ is two-dimensional and $\psi : \mathcal{M} \supseteq U \to \mathbb{R}^2, p \mapsto (t, \varphi)$ is a local chart of $\mathcal{M}$. A motion $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ is time-parameterized and, therefore, $\psi \circ \gamma(t) = (t, \varphi(t))$. The kinematic map κ associates to every particle x its motion $\xi_x(t) = \kappa(x, \gamma(t)) = \kappa_x \circ \gamma(t)$. Let $\mathcal{S}_t$ be the submanifold of simultaneous events of $\mathcal{E}$ containing $\xi_x(t)$. As depicted in Figure 5.1, a snapshot of the pendulum can be drawn for a fixed time t, because the $\xi_x(t)$ are simultaneous for all $x \in S$ and hence all lie in $\mathcal{S}_t$, which locally looks like a Euclidean point space. Finally, $\mathcal{Q}_t$ denotes the submanifold of simultaneous

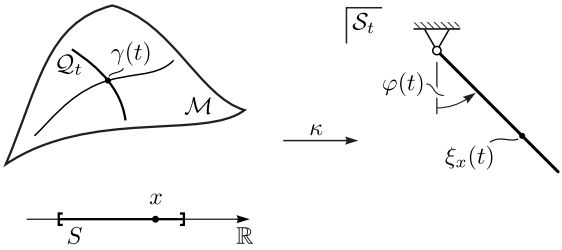


Figure 5.1: Model of a pendulum.

events of $\mathcal{M}$ which contains $\gamma(t)$. The manifold $\mathcal{Q}_t$ can be regarded as the configuration manifold of the pendulum at time t. Since $\mathcal{Q}_t$ is one-dimensional, two cases can occur. Either $\mathcal{Q}_t$ is homeomorphic to the real line $\mathbb{R}$ or to the unit circle S^1. In the first case the angle φ can attain arbitrary values and one can count the revolutions of the pendulum. This is for example important if a torsional spring is added to the pivot and deformations going beyond one circumvolution are to be studied. This is not possible in the other case, because the position described by φ is identified with $\varphi + 2\pi n$ for all $n \in \mathbb{Z}$.

The tangent field to a motion γ has the form $\dot{\gamma} = (\gamma, \dot{\gamma}_\gamma)$, where the tangent vector $\dot{\gamma}_\gamma$ is called *velocity* of the mechanical system. At time t, $\dot{\gamma}(t)$ therefore represents the configuration as well as the velocity of the system, which is why $\dot{\gamma}(t)$ denotes the state of the system at time t. Since the state $\dot{\gamma}(t)$ at any time t is time-normalized, the following is defined.

Definition 5.4. The *state space* of the mechanical system with configuration space-time $\mathcal{M}$ and time structure ϑ is the affine bundle of time-normalized vectors $A^1\mathcal{M}$. The state space is endowed with the time structure $\hat{\vartheta} := \pi^* \vartheta$, where π denotes the natural projection of $A^1\mathcal{M}$ to $\mathcal{M}$.

A motion γ in $\mathcal{M}$ induces a second-order curve $\dot{\gamma}$ in $A^1\mathcal{M}$. Any second-order curve β is called a *motion* in the state space $A^1\mathcal{M}$. By definition it then holds that $\gamma = \pi \circ \beta$ is a motion in $\mathcal{M}$ and $\beta = \dot{\gamma}$. As second-order curves, the motions of the mechanical system are integral curves of a second-order field Z on $A^1\mathcal{M}$.

Definition 5.5. The *second-order field Z of the mechanical system* is the second-order field having the motions of the system as integral curves, i.e., any motion γ fulfills $\ddot{\gamma} = Z \circ \dot{\gamma}$. The *action form Ω of the mechanical system* is the action form which uniquely defines Z by $\hat{\vartheta}(Z) = 1$ and $Z \lrcorner \Omega = 0$.

The natural chart of the state space $A^1\mathcal{M}$ induced by an adapted chart (5.27) is introduced as

$$\Phi\colon \pi^{-1}(U) \to \mathbb{R}^{2n+1},\ \left(p, \tfrac{\partial}{\partial t}\big|_p + u^i \tfrac{\partial}{\partial q^i}\big|_p\right) \mapsto (t, \mathbf{q}, \mathbf{u})\,. \tag{5.29}$$

Let $r : \mathcal{E} \supseteq U_r \to A^1\mathcal{E}$ be a reference field on space-time $\mathcal{E}$, then a time-normalized vector field

$$R : \mathcal{M} \supseteq U_R \to A^1\mathcal{M}$$

is a *reference field* on $\mathcal{M}$ if $\kappa_x(U_R) \subseteq U_r$ and

$$\mathrm{D}\kappa_x\, R = r \circ \kappa_x \tag{5.30}$$

for every particle $x \in S$. The *velocity of the mechanical system with respect to* R, or R*-velocity* for short, is defined as the vector field

$$V_R := P^R(\dot\gamma) = \dot\gamma - R \circ \gamma$$

along the motion γ of the system, where P^R is the spacelike projector. Condition (5.30) implies that

$$v_x = \dot\xi_x - r \circ \xi_x = \mathrm{D}\kappa_x(\dot\gamma - R \circ \gamma) = \mathrm{D}\kappa_x V_R\,, \tag{5.31}$$

where definition (5.15) of the r-velocity as well as (5.28) have been used. Consequently, for a motion γ with vanishing R-velocity, every particle is at rest with respect to r, as their r-velocities vanish. Based on that, a mechanical system is at *rest with respect to* R if its R-velocity is zero. An adapted chart ϕ of $\mathcal{M}$ with $\phi(p) = (t, q^1, \dots, q^n)$ is said to be a *resting chart with respect to* R if the reference field has the representation $R = \frac{\partial}{\partial t}$.

Definition 5.6. The *mass* is a positive measure m on S and its volume density ϱ is called *(volumetric) mass density*, that is, m and ϱ are related by $\mathrm{d}m = \varrho\,\mathrm{d}x$, where $\mathrm{d}x$ is the volume element of S.

The mass m and the Galilean metric g on $\mathcal{E}$ are used to define a Galilean metric on $\mathcal{M}$ as the bilinear, symmetric tensor field M, which is given by

$$M_p\colon A^0_p\mathcal{M}\times A^0_p\mathcal{M} \to \mathbb{R}, (u_p, v_p) \mapsto \int_S g_{\kappa_x(p)}\big((\mathrm{D}\kappa_x)_p u_p, (\mathrm{D}\kappa_x)_p v_p\big)\mathrm{d}m\,, \tag{5.32}$$

for every $p \in \mathcal{M}$, where the measure m is assumed to be such that M_p is positive definite. Thus, the configuration space-time $(\mathcal{M}, \vartheta, M)$ is a Galilean manifold. Using the basis fields induced by the adapted chart (5.27), the Galilean metric M can locally be written as

$$M = M_{ij}\,\mathrm{d}q^i \otimes \mathrm{d}q^j \quad \text{with} \quad M_{ij} = M\left(\frac{\partial}{\partial q^i}, \frac{\partial}{\partial q^j}\right). \tag{5.33}$$

The coefficient matrix M_{ij} is called the *mass matrix* of the mechanical system.

Definition 5.7. Let $(\mathcal{M}, \vartheta, M)$ be the configuration space-time of a mechanical system. The *kinetic energy* of the mechanical system with respect to a reference field $R : \mathcal{M} \supseteq U_R \to A^1\mathcal{M}$ is the scalar function

$$T_R : \ \pi^{-1}(U_R) \to \mathbb{R}, \ a \mapsto \frac{1}{2} M\big(P^R(a), P^R(a)\big) . \tag{5.34}$$

Using the definition of the R-velocity and (4.40), the kinetic energy along a motion $\dot\gamma$ is

$$T_R \circ \dot\gamma = \frac{1}{2} M(V_R, V_R) = \frac{1}{2} M_\gamma\big(\dot\gamma_\gamma - R_\gamma, \dot\gamma_\gamma - R_\gamma\big) . \tag{5.35}$$

To derive the kinetic energy in terms of the particle's motion, (5.31) and the definition (5.17) of the velocity vector are used to state

$$\mathrm{D}\phi_O(\mathrm{D}\kappa_x V_R) = \mathrm{D}\phi_O v_x = \ (0, \ \mathbf{v}_x) .$$

This, together with the respective definitions (5.13) and (5.32) of the Galilean metrics g and M allows to rewrite (5.35) in terms of the particle's velocities as

$$T_R \circ \dot\gamma = \frac{1}{2} \int_S \mathbf{v}_x \cdot \mathbf{v}_x \, \mathrm{d}m . \tag{5.36}$$

Let (U, ϕ) be an adapted chart of $\mathcal{M}$ and let us assume for simplicity that $U \subseteq U_R$. Moreover, let $R = \frac{\partial}{\partial t} + R^i \frac{\partial}{\partial q^i}$ be an arbitrary reference field. In the natural chart (5.29) induced by ϕ, the kinetic energy (5.34) locally reads as

$$\begin{aligned} T_R &= \frac{1}{2} M_{ij}(u^i - R^i)(u^j - R^j) = \frac{1}{2}(\mathbf{u} - \mathbf{R})^{\mathrm{T}} \mathbf{M} (\mathbf{u} - \mathbf{R}) \\ &= \frac{1}{2} \mathbf{u}^{\mathrm{T}} \mathbf{M} \mathbf{u} - \mathbf{R}^{\mathrm{T}} \mathbf{M} \mathbf{u} + \frac{1}{2} \mathbf{R}^{\mathrm{T}} \mathbf{M} \mathbf{R} , \end{aligned} \tag{5.37}$$

where the local expression of the metric (5.33) as well as its symmetry have been used. It can be recognized, that (5.37) has the form known from classical mechanics literature, see [100]. In the special case where ϕ is a resting chart with respect to R, i.e. $R = \frac{\partial}{\partial t}$, the local expression of the kinetic energy (5.37) reduces to

$$T_R = \frac{1}{2} M_{ij} u^i u^j = \frac{1}{2} \mathbf{u}^{\mathrm{T}} \mathbf{M} \mathbf{u}.$$

5.4. Virtual Displacements

The flow φ_ε^Y of a vector field $Y \in \mathrm{Vect}(\mathcal{M})$ on the configuration space-time $\mathcal{M}$ can be used to displace a configuration $p \in \mathcal{M}$ to a one-parameter family of configurations

$$\bar p(\varepsilon) = \varphi_\varepsilon^Y(p) , \tag{5.38}$$

where by definition $\varphi_0^Y = \mathrm{id}_{\mathcal{M}}$. Such a displacement occurs at fixed time if the displaced configuration $\bar{p}(\varepsilon)$ is simultaneous to the undisplaced configuration $p = \bar{p}(0)$. Using the definition of temporal distance (5.1), a displacement occurs at fixed time if for any time function t it holds that

$$t(\bar{p}(\varepsilon)) - t(p) = t \circ \varphi_\varepsilon^Y(p) - t(p) = (\varphi_\varepsilon^Y)^* t(p) - t(p) = 0\,,$$

where (5.38) has been used. Consequently, the vector field Y fulfills

$$\mathcal{L}_Y t = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[(\varphi_\varepsilon^Y)^* t - t \right] = 0\,,$$

implying that the vector field Y is spacelike by $\mathcal{L}_Y t = Y(t) = \mathrm{d}t(Y) = 0$.

Definition 5.8. A *virtual displacement field* on $\mathcal{M}$ is a vector field $Y \in \mathrm{Vect}(\mathcal{M})$ which is spacelike, i.e., a virtual displacement field satisfies the condition

$$\vartheta(Y) = 0 \tag{5.39}$$

and is therefore a section of the spacelike bundle $A^0\mathcal{M}$ of $\mathcal{M}$. The set of virtual displacement fields on $\mathcal{M}$ is denoted as $\mathrm{Virt}(\mathcal{M}) = \Gamma(A^0\mathcal{M})$.

The tangent vector δp to $\bar{p}$ at $\varepsilon = 0$ satisfies

$$\delta p \coloneqq \left.\frac{\mathrm{d}\bar{p}}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = Y_p\,, \tag{5.40}$$

because by definition $\bar{p}$ is an integral curve of Y. Since δp corresponds to the vector Y_p of the virtual displacement field Y at p, δp denotes the *virtual displacement* of (or at) p.

This definition of virtual displacement is coherent with classical mechanical texts such as Hamel [57, p. 75] or Pars [101, p. 15], where a virtual displacement is an imagined infinitesimal displacement occurring at fixed time. In fact, δp is the infinitesimal version of the displacement $\bar{p}$, which as a sequence of simultaneous configurations cannot be a real motion of the system, because $\vartheta(\dot{\bar{p}}) = 0 \neq 1$, cf. Definition 5.3.

For a motion $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$, a virtual displacement field $Y \in \mathrm{Virt}(\mathcal{M})$ by (5.38) induces a one-parameter family of motions

$$\bar{\gamma}(\varepsilon, t) = \bar{\gamma}_\varepsilon(t) = \varphi_\varepsilon^Y \circ \gamma(t)\,, \tag{5.41}$$

where $\bar{\gamma}_\varepsilon = \varphi_\varepsilon^Y \circ \gamma$ is defined implicitly. Using the definition of the virtual displacement of a configuration (5.40) for every time instant t allows to compute the virtual displacement of γ as the vector field along the motion given by

$$\delta\gamma = \left.\frac{\partial\bar{\gamma}}{\partial\varepsilon}\right|_{\varepsilon=0} = Y \circ \gamma\,, \tag{5.42}$$

where the partial derivative with respect to ε corresponds to the tangent vector to the curve $\bar{\gamma}(\cdot, t)$ at $\bar{\gamma}(0, t)$. The set of all virtual displacements of the motion $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ is denoted as $\mathrm{Virt}(\gamma(\mathcal{I}))$.

By (5.39), a virtual displacement field is locally represented as

$$Y = \delta q^i \frac{\partial}{\partial q^i} \,. \tag{5.43}$$

To agree with the notation used before and the notation used in the literature about classical mechanics, the coefficients of the virtual displacement fields are decorated with deltas. This is motivated by the following. Assume that the one-parameter family $\bar{\gamma}$ has the chart representation $\phi \circ \bar{\gamma}(\varepsilon, t) = \big(t, \bar{\mathbf{q}}(\varepsilon, t)\big)$ and define $\delta q^i(t) = \frac{\partial \bar{q}^i}{\partial \varepsilon}(0, t)$, then

$$\delta\gamma(t) = \frac{\partial \bar{\gamma}}{\partial \varepsilon}\Big|_{(0,t)} = \frac{\partial \bar{q}^i}{\partial \varepsilon}\Big|_{(0,t)} \frac{\partial}{\partial q^i}\Big|_{\gamma(t)} = \delta q^i(t) \frac{\partial}{\partial q^i}\Big|_{\gamma(t)} \,. \tag{5.44}$$

It is a consequence of (5.42) that

$$\delta q^i(t) = \delta q^i \circ \gamma(t) \,,$$

where the δq^i on the right hand side is the coefficient of Y introduced in (5.43).

The *variation of the function* $f : \mathcal{M} \to \mathbb{R}$ induced by a virtual displacement field Y is defined to be the function $Y(f)$, which, when evaluated along a motion γ, defines

$$\delta f = Y(f) \circ \gamma = \delta\gamma(f) = \frac{\partial (f \circ \bar{\gamma})}{\partial \varepsilon}\Big|_{\varepsilon=0} \,, \tag{5.45}$$

where (5.42) and (3.13) have been used.

By (5.28), the one-parameter family of motions $\bar{\gamma}$ induces the one-parameter family of motions $\bar{\xi}_x = \kappa_x \circ \bar{\gamma}$ of the particle x of the mechanical system. The virtual displacement of the particle x is then

$$\delta\xi_x = \frac{\partial \bar{\xi}_x}{\partial \varepsilon}\Big|_{\varepsilon=0} = \mathrm{D}\kappa_x \frac{\partial \bar{\gamma}}{\partial \varepsilon}\Big|_{\varepsilon=0} = \mathrm{D}\kappa_x \, \delta\gamma = \mathrm{D}\kappa_x Y \circ \gamma \,, \tag{5.46}$$

which is a spacelike vector field along $\xi_x(t) = \bar{\xi}_x(0, t)$. Hence, with respect to a Cartesian chart ${}_I\phi_O$, the virtual displacement of the particle has the representation $\delta\xi_x = \delta\xi_x^\nu \frac{\partial}{\partial {}_Ix^\nu}$ where ν ranges from 1 to 3. Using the definition of the differential (3.34) and (5.46), the components of the virtual displacement satisfy

$$\delta\xi_x^\nu = \delta\xi_x({}_Ix^\nu) = \mathrm{D}\kappa_x \, \delta\gamma({}_Ix^\nu) = \delta\gamma({}_Ix^\nu \circ \kappa_x) \,.$$

Introducing the map ${}_Ir_{Ox}^\nu \coloneqq {}_Ix^\nu \circ \kappa_x$, implying the usual notational overlap ${}_Ir_{Ox}^\nu(t) = {}_Ir_{Ox}^\nu \circ \gamma(t)$, it follows that

$$\delta\xi_x^\nu = \delta({}_Ir_{Ox}^\nu) = \delta q^i \frac{\partial {}_Ir_{Ox}^\nu}{\partial q^i}\Big|_\gamma \,, \tag{5.47}$$

where (5.45) and the local representation (5.44) of $\delta\gamma$ has been used. Since $\delta\xi_x$ is spacelike, the choice of a reference point O allows to locally define the *virtual displacement vector* of the particle x as

$$(0,\ \delta\mathbf{r}_x) = \mathrm{D}\phi_O \delta\xi_x\,. \tag{5.48}$$

Using (5.10) and (5.47), the virtual displacement vector has the representation

$$\delta\mathbf{r}_x = \delta({}_I r^\nu_{Ox})\mathbf{e}^I_\nu = \delta q^i \frac{\partial({}_I r^\nu_{Ox})}{\partial q^i}\bigg|_\gamma \mathbf{e}^I_\nu = \delta q^i \frac{\partial \mathbf{r}_{Ox}}{\partial q^i}\bigg|_\gamma\,, \tag{5.49}$$

where for the last equality it has been exploited that the basis $\mathbf{e}^I_\nu$ only depends on time. Moreover, the notation

$$\frac{\partial \mathbf{r}_{Ox}}{\partial q^i}\bigg|_p \coloneqq \frac{\partial(\mathbf{r}_{Ox}\circ\phi^{-1})}{\partial q^i}\bigg|_{\phi(p)}$$

has been introduced, where ϕ is the adapted chart of $\mathcal{M}$ having the q^i as coordinate maps.

The same reasoning that leads to the definition of virtual displacement fields on $\mathcal{M}$ is used to define a *virtual displacement field* on the state space $A^1\mathcal{M}$ as a vector field $\hat{Y} \in \mathrm{Vect}(A^1\mathcal{M})$ which is space-like, i.e., a virtual displacement field on $A^1\mathcal{M}$ satisfies

$$\hat{\vartheta}(\hat{Y}) = 0\,. \tag{5.50}$$

The set of virtual displacement fields on $A^1\mathcal{M}$ is denoted by

$$\mathrm{Virt}(A^1\mathcal{M}) = \Gamma\big(A^0(A^1\mathcal{M})\big)\,.$$

By (5.50), a virtual displacement field is locally given by

$$\hat{Y} = \delta q^i \frac{\partial}{\partial q^i} + \delta u^i \frac{\partial}{\partial u^i}\,,$$

where the basis vector fields are induced by the natural chart (5.29).

The flow $\varphi^{\hat{Y}}_\varepsilon$ of a virtual displacement field $\hat{Y} \in \mathrm{Virt}(A^1\mathcal{M})$ induces a one-parameter family of time-parametrized curves

$$\bar{\beta}(\varepsilon, t) = \bar{\beta}_\varepsilon(t) = \varphi^{\hat{Y}}_\varepsilon \circ \beta(t)\,,$$

by displacing the points along a time-parametrized curve $\beta : \mathbb{R} \supseteq I \to A^1\mathcal{M}$ in the state space. Again, $\bar{\beta}_\varepsilon = \varphi^{\hat{Y}}_\varepsilon \circ \beta$ has been implicitly defined, which corresponds to the curve of the family obtained for a fixed value of ε. The virtual displacement of β is then the vector field

$$\delta\beta = \frac{\partial\bar{\beta}}{\partial\varepsilon}\bigg|_{\varepsilon=0} = \hat{Y}\circ\beta \tag{5.51}$$

defined along the curve β.

In contrast to (5.41), not every one-parameter family of time-parametrized curves $\bar{\beta}$ is a family of motions, because the motions, as second-order curves in the state space, are only a subset of the time-parametrized curves. Hence, not every virtual displacement field induces a family of motions.

Since a motion β through $A^1\mathcal{M}$ is the tangent field $\beta = \dot{\gamma}$ to the motion $\gamma = \pi \circ \beta$ through $\mathcal{M}$, there is a natural way to associate a virtual displacement field $Y^c \in \mathrm{Virt}(A^1\mathcal{M})$ to every virtual displacement field $Y \in \mathrm{Virt}(\mathcal{M})$. The virtual displacement field Y on $\mathcal{M}$ induces the one-parameter family $\bar{\gamma}_\varepsilon = \varphi_\varepsilon^Y \circ \gamma$ of time-parametrized curves on $\mathcal{M}$. By construction, $\dot{\bar{\gamma}}_\varepsilon = (\varphi_\varepsilon^Y \circ \gamma)^{\cdot}$ is a one-parameter family of second-order curves on $A^1\mathcal{M}$. The virtual displacement field Y^c on $A^1\mathcal{M}$ is defined to be such that it induces the family of motions $\dot{\bar{\gamma}}_\varepsilon$, that is,

$$\varphi_\varepsilon^{Y^c} \circ \dot{\gamma} = (\varphi_\varepsilon^Y \circ \gamma)^{\cdot} = \dot{\bar{\gamma}}_\varepsilon \,.$$

This vector field Y^c is called the *complete lift of* Y and it is a consequence of (5.51) that

$$\delta\dot{\gamma} = \left.\frac{\partial \dot{\bar{\gamma}}_\varepsilon}{\partial \varepsilon}\right|_{\varepsilon=0} = Y^c \circ \dot{\gamma} \,,$$

from which immediately follows that $\delta\dot{\gamma} = (\delta\gamma)^{\cdot}$ because of the symmetry of second derivatives. As has been done in (5.45), the variation of a function $f : A^1\mathcal{M} \to \mathbb{R}$ induced by a virtual displacement field Y^c is defined to be the function $Y^c(f)$, which, when evaluated along a motion $\dot{\gamma}$, gives

$$\delta f = Y^c(f) \circ \dot{\gamma} = \delta\dot{\gamma}(f) = \left.\frac{\partial (f \circ \dot{\bar{\gamma}}_\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} . \tag{5.52}$$

With respect to a natural chart, the virtual displacement field Y^c on $A^1\mathcal{M}$ has the local form

$$Y^c = (\delta q^i \circ \pi)\frac{\partial}{\partial q^i} + \mathcal{L}_Z(\delta q^i \circ \pi)\frac{\partial}{\partial u^i} \,, \tag{5.53}$$

where δq^i are the component functions of (5.43), π denotes the natural projection of $A^1\mathcal{M}$ and Z is any second-order field, see [22]. Moreover, it holds that the Lie-bracket $[Z, Y^c]$ is vertical. In fact, using the coordinate maps t, q^i and u^i of a natural chart of the state space, one computes

$$\begin{aligned} [Z,Y^c](t) &= Z\big(Y^c(t)\big) - Y^c\big(Z(t)\big) = Z(0) - Y^c(1) = 0 \\ [Z,Y^c](q^i) &= Z\big(Y^c(q^i)\big) - Y^c\big(Z(q^i)\big) = Z(\delta q^i \circ \pi) - Y^c(u^i) = 0 \,, \end{aligned} \tag{5.54}$$

where (5.53) and the definition of the Lie-bracket (3.30) has been used. Consequently, $[Z, Y^c]$ is vertical, because its components with respect to $\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial q^i}$ are zero.

A virtual displacement $Y \in \mathrm{Virt}(\mathcal{M})$ is spacelike by definition and can hence be also lifted vertically. The vertical lift Y^v of Y is a virtual displacement field on $A^1\mathcal{M}$ since vertical vectors are spacelike. By (4.41), the virtual displacement field Y^v is associated to the one-parameter family of curves

$$\bar{\beta}_\varepsilon = \varphi_\varepsilon^{Y^v} \circ \dot{\gamma} = \dot{\gamma} + \varepsilon\,(Y \circ \gamma) = \dot{\gamma} + \varepsilon\,\delta\gamma = \ (\gamma, \dot{\gamma}_\gamma + \varepsilon\,\delta\gamma_\gamma)\,,$$

which only virtually displaces the velocity $\dot{\gamma}_\gamma$ of the system while leaving the position γ unchanged. Using this in (5.51), the virtual displacement of $\dot{\gamma}$ induced by Y^v is defined as

$$\underline{\delta}\dot{\gamma} := \ \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0}\big(\dot{\gamma} + \varepsilon\,\delta\gamma\big) = \delta\gamma^v = Y^v \circ \dot{\gamma}\,.$$

Similar to (5.52), the variation of a function $f : A^1\mathcal{M} \to \mathbb{R}$ induced by a virtual displacement field Y^v is defined to be the function $Y^v(f)$, which, when evaluated along a motion $\dot{\gamma}$, gives

$$\underline{\delta} f = Y^v(f) \circ \dot{\gamma} = \underline{\delta}\dot{\gamma}(f) = \delta\gamma^v(f) = \ \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} f(\dot{\gamma} + \varepsilon\,\delta\gamma)\,. \tag{5.55}$$

It follows from (4.44), that with respect to a natural chart, the virtual displacement field Y^v on $A^1\mathcal{M}$ has the local form

$$Y^v = (\delta q^i \circ \pi)\frac{\partial}{\partial u^i}\,, \tag{5.56}$$

where δq^i are the component functions of (5.43), and π denotes the natural projection of $A^1\mathcal{M}$. Comparing (5.56) with the local form (5.53) of the complete lift of Y shows, that the vertical lift of Y is related to the complete lift by

$$Y^v = \mu(Y^c)\,, \tag{5.57}$$

where μ denotes the vertical homomorphism on $A^1\mathcal{M}$, see (4.45).

Since the space-time $\mathcal{E}$ is a Galilean manifold, it is clear that everything valid for virtual displacement fields $Y \in \mathrm{Virt}(\mathcal{M})$ and virtual displacement of motions γ in $\mathcal{M}$ also holds for virtual displacement fields $y \in \mathrm{Virt}(\mathcal{E}) = \Gamma(A^0\mathcal{E})$ and virtual displacements of motions of particles ξ_x in $\mathcal{E}$. The virtual displacement of the particle x associated to a virtual displacement field y is therefore given by

$$\delta\xi_x = \ y \circ \xi_x = \ y \circ \kappa_x \circ \gamma\,. \tag{5.58}$$

The so defined virtual displacement $\delta\xi_x$ is more general then the virtual displacement of the particle x induced by a virtual displacement field $Y \in \mathrm{Virt}(\mathcal{M})$, cf. (5.46). Think of a planar rigid pendulum, whose configuration can be described by time and an angle for example, then the one-parameter family of motions

$\bar{\xi}_x = \kappa_x \circ \phi_\varepsilon^Y \circ \gamma$ induced by Y displaces all particles by instantaneously displacing the configuration, i.e., varying the angle. Hence, the displaced particles $\bar{\xi}_x$ respect the geometry and kinematics of the pendulum. In contrast, the one-parameter family of motions $\bar{\xi}_x = \phi_\varepsilon^y \circ \kappa_x \circ \gamma$ individually displaces the particle's motion $\xi_x = \kappa_x \circ \gamma$ for every $x \in S$. Consequently, the displaced particles must not necessarily respect the geometry and kinematics of the pendulum. With such a virtual displacement y, the pendulum can also be deformed or displaced by a translation which would not respect the pivot of the pendulum. However, in certain cases the virtual displacement (5.58) can be induced by a virtual displacement Y, namely if

$$\mathrm{D}\kappa_x Y = \; y \circ \kappa_x$$

for every $x \in S$. This becomes evident from the comparison of (5.58) with (5.46).

5.5. Inertial Motion

Any motion $\dot{\gamma} : \mathbb{R} \supseteq \mathcal{I} \to A^1\mathcal{M}$ of a mechanical system is a second-order curve and as such an integral curve of the second-order field Z of the mechanical system. The following definition introduces a particular second-order field Z_R, which defines a special kind of motion of the system.

Definition 5.9. With respect to a reference field $R : \mathcal{M} \supseteq U_R \to A^1\mathcal{M}$, the *inertial motions* of the mechanical system are integral curves of the second-order field Z_R, which is uniquely defined by the action form

$$\Omega_R = \mathrm{d}\big(T_R \hat{\vartheta} + \partial T_R\big) \tag{5.59}$$

induced by the kinetic energy T_R of the mechanical system.

Proposition 4.2 assures that Ω_R is indeed an action form. With respect to a natural chart of the state space, it is seen from (4.53) that

$$\Omega_R = \left(\mathrm{d}\left(\frac{\partial T_R}{\partial u^i}\right) - \frac{\partial T_R}{\partial q^i}\mathrm{d}t\right) \wedge \big(\mathrm{d}q^i - u^i \mathrm{d}t\big). \tag{5.60}$$

Moreover, it is a direct consequence of Proposition 4.3 that the inertial motions $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ correspond to the autoparallel curves of the Galilean connection with covariant derivative ∇^R, i.e., the inertial motions are characterized by

$$\nabla^R_{\dot{\gamma}} \dot{\gamma} = 0\,. \tag{5.61}$$

Interpreting the left-hand side of (5.61) as the acceleration of the mechanical system, it becomes evident that the inertial motions are defined as the motions of the system which are unaccelerated. Hence, Definition 5.9 is in line with the notion of inertial motion usually encountered in classical mechanics.

Definition 5.10. For a virtual displacement $Y \in \mathrm{Virt}(\mathcal{M})$, the *virtual work of the inertia* of the mechanical system with respect to the reference field R is the scalar function

$$\delta W_R = \Omega_R(Z, Y^c)\,, \tag{5.62}$$

where Z is the second-order field having motions $\dot\gamma$ of the system as integral curves.

With respect to a natural chart of the state space, the virtual work of the inertia locally reads as

$$\delta W_R = \Big(Z\Big(\frac{\partial T_R}{\partial u^i}\Big) - \frac{\partial T_R}{\partial q^i}\Big)(\delta q^i \circ \pi)\,, \tag{5.63}$$

which follows from (5.60) and the representation (5.53) of Y^c. Moreover, evaluating (5.63) along a motion yields

$$\delta W_R \circ \dot\gamma = \Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{\partial T_R}{\partial u^i}\Big) - \frac{\partial T_R}{\partial q^i}\Big)\delta q^i\,, \tag{5.64}$$

where the fact that $\dot\gamma$ is an integral curve of Z has been exploited. Moreover, the evaluations of the partial derivatives and of $\delta q^i \circ \pi$ along $\dot\gamma$ have been omitted.

To derive the virtual work of the inertia of the system in terms of the particle's motion $\xi_x = \kappa_x \circ \gamma$, the definition (5.59) of the action form Ω_R is used to evaluate the virtual work to

$$\begin{aligned} \delta W_R =\ & (\mathrm{d}T_R \wedge \hat\vartheta + \mathrm{d}\partial T_R)(Z, Y^c) \\ =\ & \mathrm{d}T_R(Z)\hat\vartheta(Y^c) - \mathrm{d}T_R(Y^c)\hat\vartheta(Z) + \mathrm{d}\partial T_R(Z, Y^c)\,. \end{aligned} \tag{5.65}$$

Using (3.45), the last term in (5.65) can be expanded to

$$\begin{aligned} \mathrm{d}\partial T_R(Z, Y^c) =\ & \mathcal{L}_Z\big(\partial T_R(Y^c)\big) - \mathcal{L}_{Y^c}\big(\partial T_R(Z)\big) - \partial T_R([Z, Y^c]) \\ =\ & \mathcal{L}_Z\big(\mathrm{d}T_R \circ \mu(Y^c)\big) - \mathcal{L}_{Y^c}\big(\mathrm{d}T_R \circ \mu(Z)\big) - \mathrm{d}T_R \circ \mu([Z, Y^c])\,, \end{aligned}$$

where $\partial T_R = \mathrm{d}T_R \circ \mu$ has been used. Since μ vanishes on both the second-order field Z and the vertical vector field $[Z, Y^c]$, see (5.54), this simplifies to

$$\mathrm{d}\partial T_R(Z, Y^c) = \mathcal{L}_Z\big(\mathrm{d}T_R \circ \mu(Y^c)\big) = \mathcal{L}_Z\big(\mathrm{d}T_R(Y^v)\big) = \mathcal{L}_Z\big(Y^v(T_R)\big)\,, \tag{5.66}$$

where (5.57) has been used. Since the virtual displacement field Y^c is spacelike, the first term in the second equation of (5.65) vanishes. Moreover, the second term reduces to $-\mathrm{d}T_R(Y^c)$, because Z is time-normalized. Hence, by (5.66), the virtual work (5.65) reads as

$$\delta W_R = -Y^c(T_R) + \mathcal{L}_Z\big(Y^v(T_R)\big)\,,$$

which evaluated along a motion $\dot{\gamma}$ gives

$$\delta W_R \circ \dot{\gamma} = -\delta T_R + (\underline{\delta} T_R)^{\cdot}. \tag{5.67}$$

Hereby, (5.52) and (5.55) as well as the fact that $\dot{\gamma}$ is an integral curve of Z have been used.

Equation (5.36) is used to express the virtual work of the inertia in terms of the particle motions. To do so, δT_R and $\underline{\delta} T_R$ are first computed separately. Using (5.36) in (5.52) yields

$$\delta T_R = \frac{1}{2} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \int_S \bar{\mathbf{v}}_x \cdot \bar{\mathbf{v}}_x \, \mathrm{d}m = \int_S \delta \mathbf{v}_x \cdot \mathbf{v}_x \, \mathrm{d}m \,,$$

where $\bar{\mathbf{v}}_x$ denotes the family of r-velocity vectors induced by $\dot{\bar{\gamma}}_\varepsilon$ and the variation $\delta \mathbf{v}_x$ of the r-velocity vector is

$$\delta \mathbf{v}_x = \frac{\partial \bar{\mathbf{v}}_x}{\partial \varepsilon}\Big|_{\varepsilon=0}. \tag{5.68}$$

Similarly, using (5.36) in (5.55) leads to

$$\underline{\delta} T_R = \frac{1}{2} \frac{\partial}{\partial \varepsilon}\Big|_{\varepsilon=0} \int_S (\mathbf{v}_x + \varepsilon \, \delta \mathbf{r}_x) \cdot (\mathbf{v}_x + \varepsilon \, \delta \mathbf{r}_x) \, \mathrm{d}m = \int_S \delta \mathbf{r}_x \cdot \mathbf{v}_x \, \mathrm{d}m \,,$$

where also (5.48) has been used. Hence, the virtual work of the inertia forces (5.67) can locally be written as

$$\delta W_R \circ \dot{\gamma} = \int_S (\delta \mathbf{r}_x \cdot \mathbf{v}_x)^{\cdot} \mathrm{d}m - \int_S \delta \mathbf{v}_x \cdot \mathbf{v}_x \, \mathrm{d}m \,. \tag{5.69}$$

Identifying the reference frames $\mathbf{e}_i^I$ leading to resting Cartesian charts ${}_I\phi_O$ with respect to r with resting frames of $\mathbb{E}^3$ allows to use (5.22), which translates to $\mathbf{v}_x = \dot{\mathbf{r}}_{Ox}$ and $\mathbf{a}_x = \ddot{\mathbf{r}}_{Ox}$. Moreover, it holds that $\bar{\mathbf{v}}_x = \frac{\partial}{\partial t} \bar{\mathbf{r}}_{Ox}$ from which by (5.68) follows that $\delta \mathbf{v}_x = \delta(\dot{\mathbf{r}}_{Ox}) = (\delta \mathbf{r}_{Ox})^{\cdot}$ due to the symmetry of second derivatives. Consequently, the virtual work (5.69) can be written equivalently as

$$\begin{aligned} \delta W_R \circ \dot{\gamma} &= \int_S (\delta \mathbf{r}_x \cdot \dot{\mathbf{r}}_{Ox})^{\cdot} \mathrm{d}m - \int_S \delta \dot{\mathbf{r}}_x \cdot \dot{\mathbf{r}}_{Ox} \, \mathrm{d}m \\ &= \int_S \delta \mathbf{r}_x \cdot \ddot{\mathbf{r}}_{Ox} \, \mathrm{d}m = \int_S \delta \mathbf{r}_x \cdot \mathbf{a}_x \, \mathrm{d}m \,. \end{aligned} \tag{5.70}$$

5.6. Forces

A fundamental concept in mechanics is the concept of forces acting on a mechanical system. Forces model the interaction of the particles of the system with each

other or with the system's environment. The concept of force adopted in this thesis is based on three postulates which in substance correspond to the laws of motion given by Newton and can be seen as generalizations thereof.

Postulate 1. *In the absence of forces with respect to a reference field* $R : \mathcal{M} \supseteq U_R \to A^1\mathcal{M}$*, the motions of the mechanical system are inertial motions.*

This postulate, which corresponds to Newton's first law, by Definition 5.9 implies that the second-order field Z having motions as integral curves corresponds to Z_R in the absence of forces. Or in other words, it implies that the *acceleration w.r.t. R of the system*

$$A_R := Z - Z_R \tag{5.71}$$

is caused by a force. The acceleration A_R, as the difference of second-order fields, is a vertical vector field. Similar to Newton's second law, the relation between the acceleration A_R and the force causing it is established by the following postulate.

Postulate 2. *The force causing the relative acceleration* $A_R = Z - Z_R$ *is given by*

$$f_R = \hat{M} \cdot A_R \,, \tag{5.72}$$

where $\hat{M} \cdot A_R$ *is the one-form* $\hat{M}(\,\cdot\,, A_R)$ *and* $\hat{M}$ *is the bundle metric on* $\mathrm{Ver}(A^1\mathcal{M})$ *induced by the Galilean metric* M*, which models the mass of the mechanical system.*

Let $\dot{\gamma}$ be a motion of the system, and as such an integral curve of Z. Evaluating (5.72) along a motion then reads as

$$f_R \circ \dot{\gamma} = \hat{M} \cdot (\ddot{\gamma} - Z_R \circ \dot{\gamma}) \,, \tag{5.73}$$

where $\hat{M} = \hat{M} \circ \dot{\gamma}$ has been used. By (4.61), equation (5.73) takes the form

$$f_R \circ \dot{\gamma} = \hat{M} \cdot \left(\nabla^R_{\dot{\gamma}} \dot{\gamma}\right)^v \,,$$

where $\left(\nabla^R_{\dot{\gamma}} \dot{\gamma}\right)^v = A_R \circ \dot{\gamma}$ is the acceleration of the system along a motion.

Since the bundle metric $\hat{M}$ induces the bijection

$$\hat{M} \cdot : \Gamma\big(\mathrm{Ver}(A^1\mathcal{M})\big) \to \Gamma\big(\mathrm{Ver}^*(A^1\mathcal{M})\big), \; V \mapsto f = \hat{M} \cdot V$$

between sections of the vertical bundle $\mathrm{Ver}(A^1\mathcal{M})$ and sections of the dual of the vertical bundle $\mathrm{Ver}^*(A^1\mathcal{M})$, Postulate 2 motivates the following.

Definition 5.11. A *force* acting on a mechanical system is a smooth section of the dual of the vertical bundle $\mathrm{Ver}^*(A^1\mathcal{M})$, i.e., a force is a $\mathrm{C}^\infty(A^1\mathcal{M})$-linear map

$$f \colon \Gamma(\mathrm{Ver}(A^1\mathcal{M})) \to \mathrm{C}^\infty(A^1\mathcal{M})$$

on the space of vertical vector fields.

It is crucial to see that the definition of forces is independent of the Galilean metric M, that is, it depends only on the configuration space-time $\mathcal{M}$ and its time structure ϑ. Hence, a force acting on a system is left unchanged if the system's mass distribution and therefore its Galilean metric are changed. By Postulate 2, however, the relative acceleration A_R caused by the same force is clearly affected by a change in mass distribution. More about this nuance in the definition of force and an epistemological discussion of it can be found in Hamel [56].

Let Ω be the action form of the mechanical system and Z be its second-order field. The differential two-form

$$\Phi_R = \Omega - \Omega_R \tag{5.74}$$

describes how the action form Ω differs from the action form Ω_R related to Z_R. The property $\partial\Phi_R = 0$ is easily asserted, because $\partial\Omega = \partial\Omega_R = 0$ by Theorem 4.5.

It is possible to associate the two-form Φ_R from (5.74) to the force f_R which by (5.72) is related to the vertical vector field A_R of (5.71). Indeed, in terms of the coordinate fields induced by a natural chart, the coefficients of the fields

$$f_R = f_i \,\mathrm{d}u^i \quad \text{and} \quad A_R = \; Z - Z_R = \; (Z^i - Z^i_R)\frac{\partial}{\partial u^i}$$

by (5.72) fulfill the relation

$$f_i = M_{ij}(Z^j - Z^j_R)\,, \tag{5.75}$$

where Z^j and Z^j_R are the coefficients of Z and Z_R, respectively. Using the representation (4.49) of the action form, the two-form $\Phi_R = \Omega - \Omega_R$ is given as

$$\begin{aligned}\Phi_R &= M_{ij}(Z^i - Z^i_R)\,\mathrm{d}q^j \wedge \mathrm{d}t + \frac{1}{2}M_{ij}\Big(\frac{\partial Z^i}{\partial u^k} - \frac{\partial Z^i_R}{\partial u^k}\Big)(\mathrm{d}q^j - u^j\mathrm{d}t)\wedge(\mathrm{d}q^k - u^k\mathrm{d}t)\\ &= f_j\,\mathrm{d}q^j\wedge\mathrm{d}t + \frac{1}{2}\frac{\partial f_j}{\partial u^k}(\mathrm{d}q^j - u^j\mathrm{d}t)\wedge(\mathrm{d}q^k - u^k\mathrm{d}t),\end{aligned} \tag{5.76}$$

where the last equality uses (5.75) and that the coefficients $M_{ij} = M_{ij}\circ\pi$ are independent of $u^1, \ldots, u^n$ and form a symmetric matrix. The local expression (5.76) shows that the two-form Φ_R is semi-basic. Moreover, since Φ_R is associated to the force f_R and additionally $\partial\Phi_R = 0$ holds, the following definition is appropriate.

Definition 5.12. A *force two-form* is a differential two-form $\Phi \in \Omega^2(A^1\mathcal{M})$ that is semi-basic and satisfies $\partial\Phi = 0$.

Equation (5.76) is the local representation of a map $f_R \mapsto \Phi_R$ relating forces to force two-forms. The following theorem gives a coordinate-free definition of this map and shows that it is a bijection.

Theorem 5.1 (Eugster et al. [40], Theorem 2)**.** *The formulae*

$$\varphi = f \circ \mu, \quad \varphi = -Z \lrcorner \Phi, \quad \Phi = -\tfrac{1}{2}\big(\partial\varphi + \hat{\vartheta} \wedge \varphi\big)$$

define bijections between

(i) the forces, i.e., smooth sections $f \in \Gamma(\mathrm{Ver}^*(A^1\mathcal{M}))$*,*

(ii) the semi-basic one-forms φ *with* $Z \lrcorner \varphi = 0$*,*

(iii) the force two-forms, i.e., the semi-basic two-forms Φ *with* $\partial\Phi = 0$*,*

where Z *is an arbitrary second-order field and* μ *is the vertical homomorphism on* $A^1\mathcal{M}$*. In local coordinates, it holds that*

$$\begin{aligned} f &= f_i \, \mathrm{d}u^i \,, \\ \varphi &= f_i (\mathrm{d}q^i - u^i \mathrm{d}t) \,, \\ \Phi &= f_i \, \mathrm{d}q^i \wedge \mathrm{d}t + \frac{1}{2} \frac{\partial f_i}{\partial u^j} (\mathrm{d}q^i - u^i \mathrm{d}t) \wedge (\mathrm{d}q^j - u^j \mathrm{d}t) \,. \end{aligned} \tag{5.77}$$

Definition 5.13. For a virtual displacement $Y \in \mathrm{Virt}(\mathcal{M})$, the *virtual work of the force* f is the scalar function

$$\delta W_f = \Phi(Z, Y^c) \,, \tag{5.78}$$

where Φ is the unique force two-form related to f by Theorem 5.1 and Z is the second-order field having motions $\dot{\gamma}$ of the system as integral curves.

By Theorem 5.1, it holds that $Z \lrcorner \Phi = -f \circ \mu$, such that the virtual work (5.78) of f can be written as

$$\delta W_f = \ \Phi(Z, Y^c) = Z \lrcorner \Phi(Y^c) = \ -f \circ \mu(Y^c) = -f(Y^v) \,, \tag{5.79}$$

where (5.57) has been used for the last equality. Evaluated along a motion of the system, it follows that

$$\delta W_f \circ \dot{\gamma} = -f(Y^v) \circ \dot{\gamma} = -f(\underline{\delta}\dot{\gamma}) = -f(\delta\gamma^v) \,. \tag{5.80}$$

Using the local representations (5.56) and (5.77) of Y^v and f, respectively, the virtual work (5.79) and (5.80) locally read as

$$\delta W_f = \ -(\delta q^i \circ \pi) \, f_i \quad \text{and} \quad \delta W_f \circ \dot{\gamma} = \ -\delta q^i \, f_i = -\delta \mathbf{q}^{\mathrm{T}} \mathbf{f} \,, \tag{5.81}$$

where the usual overload of notation is introduced by defining $f_i(t) = f_i \circ \dot{\gamma}(t)$.

It is assumed that a force acting on a mechanical system stems from a force distribution $F : S \to \mathrm{Ver}^*(A^1\mathcal{E})$ such that

$$f(\delta\gamma^v) = \int_S F_x\big(\delta\xi_x^v\big) \mathrm{d}x \tag{5.82}$$

where $F_x = F(x)$ denotes the force acting on the particle x and $\mathrm{d}x$ is the volume element of the system S.

The Galilean metric g of the space-time $\mathcal{E}$ induces the bundle metric $\hat{g}$ on the bundle of vertical vectors $\mathrm{Ver}(A^1\mathcal{E})$. As such, $\hat{g}$ establishes a pointwise isomorphism between $\mathrm{Ver}(A^1\mathcal{E})$ and its dual bundle. Hence, to any force $F_x \in \mathrm{Ver}^*(A^1\mathcal{E})$ acting on the particle x, the vertical vector field $F_x^\sharp$ implicitly defined by $F_x(\cdot) = \hat{g}(F_x^\sharp, \cdot)$ is bijectively associated. For every $a \in A^1\mathcal{E}$, the vertical lift $v_a : A^0_{\pi(a)}\mathcal{E} \to \mathrm{Ver}_a(A^1\mathcal{E})$ is a vector space isomorphism and hence allows to introduce the "vector field" $\check{F}_x^\sharp : A^1\mathcal{E} \to A^0\mathcal{E}$ by

$$\big(\check{F}_x^\sharp\big)_{\pi(a)} = v_a^{-1}\big((F_x^\sharp)_a\big) \in A^0_{\pi(a)}\mathcal{E}$$

for all $a \in A^1\mathcal{E}$. Finally, the choice of a reference point O allows to uniquely relate $\check{F}_x^\sharp$ to the *force vector* $\mathbf{F}_x$ defined by

$$(0, \mathbf{F}_x) = \mathrm{D}\phi_O \check{F}_x^\sharp\,,$$

because $\check{F}_x^\sharp$ is spacelike, see (5.4). Consequently, the integrand of (5.82) can be written as

$$F_x(\delta\xi_x^v) = \hat{g}(F_x^\sharp, \delta\xi_x^v) = g(\check{F}_x^\sharp, \delta\xi_x) = \mathbf{F}_x \cdot \delta\mathbf{r}_x\,,$$

where we have used the definitions (5.13) and (4.48) of g and $\hat{g}$, respectively, and the fact that $F_x^\sharp$ is the pointwise vertical lift of $\check{F}_x^\sharp$. With that, the virtual work of the force takes the form

$$\delta W_f \circ \dot{\gamma} = -\int_S \mathbf{F}_x \cdot \delta\mathbf{r}_x \,\mathrm{d}x = -\int_S \mathbf{F}_x \cdot \frac{\partial \mathbf{r}_{Ox}}{\partial q^i}\bigg|_\gamma \delta q^i \,\mathrm{d}x\,, \tag{5.83}$$

where (5.49) has been used. Finally, a comparison of (5.83) with (5.81) reveals that the components f_i of the force f are given by

$$\mathbf{f} = \int_S \mathbf{F}_x \cdot \frac{\partial \mathbf{r}_{Ox}}{\partial \mathbf{q}}\bigg|_\gamma \mathrm{d}x\,, \tag{5.84}$$

where as usual $\mathbf{f} = (f_1, \ldots, f_n)$.

Example 5.2 (Two-point interaction)**.** Often, a force interaction takes place only between two particles P and Q of a mechanical system S. The associated force distribution $\mathbf{F}_x$ vanishes everywhere except for the particles P and Q. Assume that $\mathbf{F}_Q = -\mathbf{F}_P = \lambda\mathbf{d}$, where the unit vector $\mathbf{d}$ is the force direction and λ the magnitude of the force interaction. The components of the force f acting on the system follow from (5.84) as

$$\mathbf{f} = \lambda\mathbf{d} \cdot \bigg(\frac{\partial \mathbf{r}_{OQ}}{\partial \mathbf{q}}\bigg|_\gamma - \frac{\partial \mathbf{r}_{OP}}{\partial \mathbf{q}}\bigg|_\gamma\bigg) = \lambda\mathbf{d} \cdot \frac{\partial \mathbf{r}_{PQ}}{\partial \mathbf{q}}\bigg|_\gamma = \lambda\mathbf{w}\,,$$

where we have implicitly defined the n-tuple $\mathbf{w}$, which is called *generalized force direction*. It is useful to relate the generalized force direction to a scalar kinematic quantity. Consider for example the one-form α on $\mathcal{M}$, which evaluated along a motion yields $\alpha(\dot\gamma) = \mathbf{d}\cdot(\mathbf{v}_Q - \mathbf{v}_P) = \mathbf{d}\cdot\dot{\mathbf{r}}_{PQ}$, where (5.22) has been used. Then by the chain rule it is found that

$$\alpha(\dot\gamma) = \mathbf{d}\cdot\left(\frac{\partial \mathbf{r}_{PQ}}{\partial \mathbf{q}}\Big|_\gamma \dot{\mathbf{q}} + \frac{\partial \mathbf{r}_{PQ}}{\partial t}\Big|_\gamma\right)$$

and one concludes that the one-form seen as a function on the state space is locally represented as $\alpha = \mathbf{w}^{\mathrm{T}}\mathbf{u} + \chi$ with respect to a natural chart of $A^1\mathcal{M}$. Consequently,

$$\mathbf{w} = \frac{\partial \alpha}{\partial \mathbf{u}}\,.$$

In the special case, where $\mathbf{d} = \mathbf{r}_{PQ}/||\mathbf{r}_{PQ}||$, one similarly finds that the generalized force direction can be computed by

$$\mathbf{w} = \frac{\partial g}{\partial \mathbf{q}}\,, \tag{5.85}$$

where g is the smooth function on $\mathcal{M}$, which evaluated along γ yields $g\circ\gamma = \mathbf{d}\cdot\mathbf{r}_{PQ}$. In fact, it holds that $(g\circ\gamma)\dot{} = \mathrm{d}g(\dot\gamma) = \mathbf{d}\cdot\dot{\mathbf{r}}_{PQ}$.

The force distribution F models the interactions of the particles of the system with each other and with the environment, i.e., particles which are not part of the system. The following distinction is made.

Definition 5.14. Let $K \subseteq S$ denote a *subsystem* of the mechanical system S. Forces are called *internal forces*, if they model the interaction between particles $x \in K$. Otherwise, if the forces model interactions between particles $x \in K$ and the environment, i.e., particles $\bar{x} \notin K$, they are called *external forces*. Denoting with $F^{K,\mathrm{i}}$ the distribution of internal forces and with $F^{K,\mathrm{e}}$ the distribution of external forces, the forces acting on the subsystem K decompose as $F^K = F^{K,\mathrm{i}} + F^{K,\mathrm{e}}$, which naturally induces the splittings

$$F_x^K = F_x^{K,\mathrm{i}} + F_x^{K,\mathrm{e}} \quad \text{and} \quad \mathbf{F}_x^K = \mathbf{F}_x^{K,\mathrm{i}} + \mathbf{F}_x^{K,\mathrm{e}}$$

for the force acting on a particle $x \in K$.

This brings us to the variational law of interaction, which is postulated as a generalization of Newton's third law.

Postulate 3. *Let $F^{K,i}$ denote the distribution of internal forces of an arbitrary subsystem $K \subseteq S$ of a mechanical system, then*

$$\int_K F_x^{K,i}((y^{rig})^v)\mathrm{d}x = 0 \quad \forall y^{rig}\,, \tag{5.86}$$

where the rigid virtual displacement field y^{rig} *are defined as the virtual displacement fields* $y^{rig} \in \mathrm{Virt}(\mathcal{E})$ *locally given by*

$$\mathrm{D}\phi_O y^{rig} = \big(0, \delta\mathbf{c}(t) + \delta\boldsymbol{\varphi}(t) \times \mathbf{r}_{Ox}\big)$$

for an arbitrary reference point O *and* $\phi_O(e) = (t, \mathbf{x} = \mathbf{r}_{Ox})$. *Condition* (5.86) *can equivalently be written as*

$$\int_K \mathbf{F}_x^{K,i} \cdot (\delta\mathbf{c} + \delta\boldsymbol{\varphi} \times \mathbf{r}_{Ox}) \mathrm{d}x = 0 \quad \forall \delta\mathbf{c}, \forall \delta\boldsymbol{\varphi}\,. \tag{5.87}$$

Example 5.3 (Newton's third law)**.** In the case where $K = \{P, Q\}$ consists of two particles only, the integral in (5.87) is just a sum and the variational law of interaction reads as

$$\mathbf{F}_P \cdot (\delta\mathbf{c} + \delta\boldsymbol{\varphi} \times \mathbf{r}_{OP}) + \mathbf{F}_Q \cdot (\delta\mathbf{c} + \delta\boldsymbol{\varphi} \times \mathbf{r}_{OQ}) = 0 \quad \forall \delta\mathbf{c}, \forall \delta\boldsymbol{\varphi}\,. \tag{5.88}$$

Setting $\delta\boldsymbol{\varphi} = 0$ shows that the forces necessarily satisfy

$$\mathbf{F}_P = -\mathbf{F}_Q\,,$$

which can be recognized as Newton's third law. Let $\mathbf{F}_Q = -\mathbf{F}_P = \lambda\mathbf{d}$, where the unit vector $\mathbf{d}$ is the force direction and λ the magnitude of the force interaction. Using this when setting $\delta\mathbf{c} = 0$ in (5.88), one obtains that the force has to satisfy

$$\lambda(\mathbf{r}_{PQ} \times \mathbf{d}) = 0\,.$$

Hence, the magnitude of the force can only be nonzero if the force direction $\mathbf{d}$ is parallel to $\mathbf{r}_{PQ}$ whenever P and Q are separated, i.e., if $\mathbf{r}_{PQ} \neq 0$. Conversely, any other $\mathbf{d}$ is only possible if $\mathbf{r}_{PQ} = 0$.

A force f^{p} is a *potential force* if the related *potential force two-form* Φ^{p} is closed, i.e., if

$$\mathrm{d}\Phi^{\mathrm{p}} = 0. \tag{5.89}$$

With this definition every force f and every related force two-form Φ can be split into respective sums

$$f = f^{\mathrm{p}} + f^{\mathrm{np}} \quad \text{and} \quad \Phi = \Phi^{\mathrm{p}} + \Phi^{\mathrm{np}}\,, \tag{5.90}$$

where f^{p} is a potential force and its related force two-form Φ^{p} is a potential force two-form. The remaining parts f^{np} and Φ^{np} are referred to as *nonpotential force* and *nonpotential force two-form*, respectively.

Evaluating condition (5.89) in local coordinates shows that $f^{\mathrm{p}} = f_i^{\mathrm{p}} \mathrm{d}u^i$ is a potential force if and only if its coefficients f_i^{p} can be written as

$$f_i^{\mathrm{p}} = E_i + B_{ij} u^j, \tag{5.91}$$

with functions E_i and B_{ij} which do not depend on $u^1, \dots, u^n$. Moreover, these functions have to fulfill

$$B_{ij} = -B_{ji}$$

together with

$$\frac{\partial B_{ij}}{\partial q^k} + \frac{\partial B_{ki}}{\partial q^j} + \frac{\partial B_{jk}}{\partial q^i} = 0 \quad \text{and} \quad \frac{\partial B_{ij}}{\partial t} = \frac{\partial E_i}{\partial q^j} - \frac{\partial E_j}{\partial q^i}\,,$$

which thanks to the suggestive use of the letters B and E can be identified as a generalized version of Maxwell's equations. Consequently, the potential force two-form related to (5.91) has the local form

$$\Phi^{\mathrm{P}} = E_i \, \mathrm{d}q^i \wedge \mathrm{d}t + \frac{1}{2} B_{ij} \, \mathrm{d}q^i \wedge \mathrm{d}q^j\,,$$

implying that a potential force two-form is basic.

According to the Poincaré lemma,[6] a closed differential form is locally exact, that is, there exists a neighborhood $W \subseteq A^1\mathcal{M}$ and a one-form ϕ defined on W such that

$$\Phi^{\mathrm{P}}\big|_W = \mathrm{d}\phi. \tag{5.92}$$

As a one-form ϕ satisfying (5.92), consider the locally defined basic one-form

$$\phi = -(V \circ \pi)\, \mathrm{d}t + (A_i \circ \pi)\, \mathrm{d}q^i\,, \tag{5.93}$$

with functions V and A_i defined on $\mathcal{M}$ and $\pi : A^1\mathcal{M} \to \mathcal{M}$ denoting the natural projection of the state space. For reasons of brevity, π is often dropped in the notation, e.g., $V = V \circ \pi$. With the one-form ϕ from (5.93), it holds that

$$E_i = -\Big(\frac{\partial V}{\partial q^i} + \frac{\partial A_i}{\partial t}\Big) \quad \text{and} \quad B_{ij} = 2\frac{\partial A_j}{\partial q^i}. \tag{5.94}$$

The potential force in the decomposition (5.90) of an arbitrary force can be chosen freely. In classical mechanics one often chooses potential forces which are locally described by a scalar potential V only, i.e. $A_i = 0$. In this case it follows from (5.94), (5.91) that

$$f_i^{\mathrm{P}} = -\frac{\partial V}{\partial q^i}$$

and therefore

$$f^{\mathrm{P}} = -\frac{\partial V}{\partial q^i}\, \mathrm{d}u^i \quad \text{and} \quad \Phi^{\mathrm{P}} = -\frac{\partial V}{\partial q^i}\, \mathrm{d}q^i \wedge \mathrm{d}t\,.$$

With (5.81), the virtual work of the force f^{P} is

$$\delta W_{f^{\mathrm{P}}} = \frac{\partial V}{\partial q^i}(\delta q^i \circ \pi) \quad \text{and} \quad \delta W_{f^{\mathrm{P}}} \circ \dot{\gamma} = \frac{\partial V}{\partial q^i} \delta q^i = \delta V\,, \tag{5.95}$$

[6] See Theorem 3.3.

where the variation δV can either be computed by (5.45) or (5.52), depending on whether V is considered as a function on $\mathcal{M}$ or as $V = V \circ \pi$, which is a function on $A^1\mathcal{M}$.

5.7. Principle of Virtual Work

The Postulates 1 and 2 establish that with respect to a reference field R, the action form of a mechanical system can be decomposed as

$$\Omega = \Omega_R + \Phi_R \,, \tag{5.96}$$

where Ω_R is the action form defining inertial motions and Φ_R is a force two-form.

Definition 5.15. For a virtual displacement $Y \in \mathrm{Virt}(\mathcal{M})$, the *virtual work of the mechanical system* is the scalar function

$$\delta W = \Omega(Z, Y^c) \,,$$

where Z is the second-order field of the mechanical system.

It follows from (5.96) that the virtual work of a system with respect to a reference field can be written as the sum of the virtual work of the inertia of the system and the forces acting on the system, that is,

$$\delta W = \ \Omega_R(Z, Y^c) + \Phi_R(Z, Y^c) = \delta W_R + \delta W_{f_R} \,, \tag{5.97}$$

where (5.62) and (5.78) have been used.

The second-order field Z of the system, by Definition 5.5, is the vector field fulfilling the conditions

$$\hat{\vartheta}(Z) = 1 \quad \text{and} \quad Z \lrcorner \Omega(\hat{Y}) = \Omega(Z, \hat{Y}) = 0 \quad \forall \hat{Y} \in \mathrm{Vect}(A^1\mathcal{M}) \,. \tag{5.98}$$

However, since it is a priori known that Z is a second-order field, $\hat{\vartheta}(Z) = 1$ is trivially satisfied. Moreover, by Theorem 4.5 (i), all vector fields which are either second-order or vertical lie in the kernel of $Z \lrcorner \Omega$. Consequently, the conditions (5.98) can be simplified to

$$Z \lrcorner \Omega(Y^c) = \Omega(Z, Y^c) = 0 \quad \forall Y \in \mathrm{Virt}(\mathcal{M}) \,,$$

because it is clear from (5.53) that the complete lift Y^c of Y is neither second-order nor vertical. Thus, the second-order field Z is characterized by the vanishing of the virtual work of the system for all virtual displacements on M. This characterization together with the decomposition (5.97) is the culmination of this theory, as it incorporates Postulates 1 and 2 by means of the definitions of force and force two-form. This is why the principle of virtual work can be formulated as the fundamental principle for the description of the dynamics of time-dependent finite-dimensional mechanical systems.

Principle of Virtual Work. *With respect to a reference field R, let δW_R be the virtual work of the inertia related to the kinetic energy T_R and let δW_{f_R} be the virtual work of the forces f_R acting on the system.*
The second-order field Z of the system is characterized by the condition

$$\delta W = \delta W_R + \delta W_{f_R} = 0 \quad \forall Y \in \mathrm{Virt}(\mathcal{M})\,. \tag{5.99}$$

Consequently, a motion $\dot\gamma$ is a solution of the equations of motion

$$\ddot\gamma = Z \circ \dot\gamma\,. \tag{5.100}$$

Combining (5.99) *and* (5.100)*, a motion $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ of the system is characterized by*

$$\delta W \circ \dot\gamma = (\underline{\delta} T_R)^{\cdot} - \delta T_R - f_R(\delta\gamma^v) = 0 \quad \forall \delta\gamma \in \mathrm{Virt}(\gamma(\mathcal{I}))\,, \tag{5.101}$$

see (5.67) *and* (5.80)*.*

The principle of virtual work (5.101) is readily recognized to have the form of *Lagrange's central equation.*[7] In fact, since $\underline{\delta} T_R = Y^v(T_R) \circ \dot\gamma$, the chart representations (5.56) and (5.81) can be used to rewrite the principle of virtual work locally as

$$\delta W \circ \dot\gamma = \frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{\partial T_R}{\partial \mathbf{u}}\delta\mathbf{q}\Big) - \delta T_R - \delta\mathbf{q}^{\top}\mathbf{f} = 0 \quad \forall \delta\mathbf{q}\,, \tag{5.102}$$

where $\delta\mathbf{q} = (\delta q^1, \ldots, \delta q^n)$ collects the coefficients of the virtual displacements $\delta\gamma$. By (5.69) and (5.83) the central equation of Lagrange can be expressed in terms of the particle's motions as

$$\delta W \circ \dot\gamma = \frac{\mathrm{d}}{\mathrm{d}t}\int_S \delta\mathbf{r}_x \cdot \mathbf{v}_x \,\mathrm{d}m - \int_S \delta\mathbf{v}_x \cdot \mathbf{v}_x \,\mathrm{d}m - \int_S \delta\mathbf{r}_x \cdot \mathbf{F}_x \,\mathrm{d}x = 0 \quad \forall \delta\mathbf{q}\,. \tag{5.103}$$

However, this is not the only representation of (5.101) in terms of the motions of the particles, as with the expressions (5.70) and (5.83) for the virtual work of the inertia and the forces it holds that

$$\delta W \circ \dot\gamma = \int_S \delta\mathbf{r}_x \cdot (\mathbf{a}_x \,\mathrm{d}m - \mathbf{F}_x \,\mathrm{d}x) = 0 \quad \forall \delta\mathbf{q}\,. \tag{5.104}$$

In classical mechanics literature this representation is often referred to as the principle of virtual work.[8] In our context however, (5.104) as well as Lagrange's central equation in the form (5.102) and (5.103) are mere representation of the

[7] See Hamel [55] and Section 2.4 in Bremer [15].
[8] See Glocker [48] and Bremer [15].

principle of virtual work (5.101). The same holds for the local representations of the principle of virtual work

$$\delta W \circ \dot{\gamma} = \delta \mathbf{q}^{\mathrm{T}} \Big(\frac{\mathrm{d}}{\mathrm{d}t} \Big(\frac{\partial T_R}{\partial \mathbf{u}} \Big) - \frac{\partial T_R}{\partial \mathbf{q}} - \mathbf{f} \Big) = 0 \quad \forall \delta \mathbf{q} \tag{5.105}$$

called fourth form of the fundamental equation in [101, p.74] and gained from using the representations (5.64) and (5.81) of the virtual work contributions with respect to the natural chart of the state space. After carrying out the total time derivative in (5.105), the local representation (5.37) of the kinetic energy can be used to find that the principle of virtual work (5.105) has the form

$$\delta W \circ \dot{\gamma} = \delta \mathbf{q}^{\mathrm{T}} \big(\mathbf{M} \ddot{\mathbf{q}} - \mathbf{h} \big) = 0 \quad \forall \delta \mathbf{q} \,, \tag{5.106}$$

where clearly all quantities are evaluated along the motion and where the local functions h_i on the state space $A^1\mathcal{M}$ have been introduced as

$$h_i = -\frac{\partial^2 T_R}{\partial t \partial u^i} - u^j \frac{\partial^2 T_R}{\partial q^j \partial u^i} + \frac{\partial T_R}{\partial q^i} + f_i \,. \tag{5.107}$$

The terms involving partial derivatives of the kinetic energy are sometimes referred to as *gyroscopic terms* or *gyroscopic forces*. It can be verified that these terms correspond to $(M_{ij} \circ \pi) Z_R^j$, where Z_R^j are the component functions of the second-order field associated to the kinetic energy by Ω_R, see (4.54). Equation (4.62) is used to restate (5.107) as

$$h_i = -\big((M_{ij} \Gamma^j_{\mu\nu}) \circ \pi \big) u^\mu u^\nu + f_i \,,$$

where $\Gamma^j_{\mu\nu}$ are the Christoffel symbols defined by the kinetic energy, the Greek indices range form 0 to n and $u^0 = 1$ is assumed.

It can be shown, that also the Hamel–Boltzmann equation[9] is a representation of the principle of virtual work (5.101) and is obtained by the choice of a particular chart of the state space. The details can be found in Section 4.10.2. of Winandy [127].

Every representation of the principle of virtual work can be used to derive the equations of motion characterizing the motions of the system. The modeling process always starts by determining the mass distribution m of the system and making kinematic assumptions about the system. The latter means that one has to model the kinematic map κ and that a reference field r on space-time $\mathcal{E}$ as well as a compatible reference field R on the configuration space-time $\mathcal{M}$ have to be chosen. These kinematic assumptions suffice to determine the virtual work of the inertia of the system. After that, the interaction of the system's particles

[9] See Hamel [57], p. 480f.

with each other and with the environment of the system has to be modeled by appropriately introducing a force distribution F leading to the virtual work of the forces acting on the system.

In the modeling process, the choice of a reference field corresponds to the agreement about which particles are regarded to be at rest. This choice is completely arbitrary. However, it affects the modeling process of the forces. To see that, let R and $\bar{R}$ be two reference fields on $\mathcal{M}$. By (5.96), the action form of the mechanical system can then be decomposed in two manners as

$$\Omega = \Omega_R + \Phi_R = \Omega_{\bar{R}} + \Phi_{\bar{R}} \,. \tag{5.108}$$

Obviously, the two different choices of the reference field might result in different inertial motions of the system if $\Omega_R \neq \Omega_{\bar{R}}$. In that case different forces need to be considered when modeling the action form of the system and therefore the motions of the system. The force needed to compensate the difference in the inertial motions resulting from the choice of $\bar{R}$ rather then R is the force two form

$$\Psi_{\bar{R},R} = \Omega_R - \Omega_{\bar{R}} \,,$$

which is called *inertia force two-form*. This is readily seen by comparing the telescopic expansion

$$\Omega_R + \Phi_R = \Omega_{\bar{R}} + (\Omega_R - \Omega_{\bar{R}}) + \Phi_R$$

with (5.108). In fact, it can be concluded that

$$\Phi_{\bar{R}} = \Psi_{\bar{R},R} + \Phi_R \,.$$

Two examples of inertia forces are the Coriolis forces and the centrifugal forces, see Sect. 39 in Landau and Lifshitz [70] or Sects. IV.4–5 in Lanczos [69]. It is shown by Eugster et al. [40] that inertia forces are potential forces. Thus, the statement that a mechanical system is only subjected to potential forces is independent of the reference field.

5.8. Lagrange's Equations of the Second Kind

Any motion $\dot{\gamma} : \mathbb{R} \supseteq \mathcal{I} \to A^1\mathcal{M}$ of a mechanical system fulfills the principle of virtual work. With respect to a natural chart of the state space, a motion is represented as $\Phi \circ \dot{\gamma}(t) = (t, \mathbf{q}(t), \dot{\mathbf{q}}(t))$ and is characterized by the equations of motion

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{\partial T_R}{\partial \mathbf{u}}\Big) - \frac{\partial T_R}{\partial \mathbf{q}} - \mathbf{f} = \mathbf{M}\ddot{\mathbf{q}} - \mathbf{h} = 0 \,, \tag{5.109}$$

which can be seen from the local representation (5.105) and (5.106) of the principle of virtual work. This form of the equations of motion is known as *Lagrange's equations of the second kind*, see Lagrange [67].

A splitting of the forces acting on the system into a sum of potential and nonpotential forces, see (5.90), can be used to reformulate Lagrange's equations of the second kind. In fact, using the splitting in (5.96), the action form of the system takes the form

$$\Omega = \Omega_R + \Phi_R^{\mathrm{p}} + \Phi_R^{\mathrm{np}} . \tag{5.110}$$

At least locally, the potential force two-form Φ_R^{p} is the exterior derivative of the one-form ϕ_R given in (5.93). Using the rules (4.47), it can be seen that the chart representation (5.93) of ϕ_R is equivalent to

$$\phi_R = (-V_R + A_i^R u^i)\mathrm{d}t + \partial(-V_R + A_i^R u^i), \tag{5.111}$$

because V_R and A_i^R only depend on $t, x^1, ..., x^n$. By comparing (5.111) to the definition (5.59) of the action form Ω_R induced by the kinetic energy T_R of the mechanical system, it is clear that the sum $\Omega_R + \Phi_R^{\mathrm{p}}$ can be locally written as

$$\Omega_R + \Phi_R^{\mathrm{p}} = \Omega_R + \mathrm{d}\phi_R = \mathrm{d}[(T_R - V_R + A_i^R u^i)\mathrm{d}t + \partial(T_R - V_R + A_i^R u^i)] , \tag{5.112}$$

which longs for the following definition.

Definition 5.16. Let T_R be the kinetic energy of the mechanical system with respect to a reference field R. Let Φ_R^{p} be the (locally) exact potential force two-form given in the natural chart (5.29) by

$$\Phi_R^{\mathrm{p}} = \mathrm{d}\phi_R \quad \text{with} \quad \phi_R = -V_R\mathrm{d}t + A_i^R\mathrm{d}x^i ,$$

where the component functions V_R and A_R^i of ϕ_R only depend on $t, x^1, \ldots, x^n$. The *Lagrangian* of the mechanical system with respect to the reference field R is the function

$$L_R \coloneqq T_R - V_R + A_i^R u^i \tag{5.113}$$

on the state space and it defines the *Cartan one-form*

$$\omega_R \coloneqq L_R\hat{\vartheta} + \partial L_R .$$

With this definition, (5.112) can be rewritten as

$$\Omega_R + \Phi_R^{\mathrm{p}} = \mathrm{d}(L_R\hat{\vartheta} + \partial L_R) = \mathrm{d}\omega_R . \tag{5.114}$$

Consequently, the action form of the mechanical system by (5.110) takes the form

$$\Omega = \Omega_R + \Phi_R = \ \mathrm{d}\omega_R + \Phi_R^{\mathrm{np}} , \tag{5.115}$$

where $\Phi_R = \Phi_R^{\mathrm{p}} + \Phi_R^{\mathrm{np}}$. Using (5.114) together with the definition (5.59) of the action form Ω_R in (5.115) leads to

$$\Omega = \mathrm{d}(T_R\hat{\vartheta} + \partial T_R) + \Phi_R = \mathrm{d}(L_R\hat{\vartheta} + \partial L_R) + \Phi_R^{\mathrm{np}} \tag{5.116}$$

revealing that in the local representation (5.105), and consequently in (5.109), the potential forces can be accounted for by replacing the kinetic energy T_R with the Lagrangian L_R. Hence, Lagrange's equations of the second kind can be expressed as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L_R}{\partial \mathbf{u}}\right) - \frac{\partial L_R}{\partial \mathbf{q}} - \mathbf{f}^{\mathrm{np}} = 0\,.$$

Similar to Lagrange's equations, also Hamilton's equations can be derived from (5.116) as a particular chart representation of the equations of motion. Confer Capobianco et al. [22] for the details.

5.9. Principle of Virtual Action and Hamilton's Principle

All objects introduced so far have been defined as fields on the space-time or the state space of the system, e.g., the kinetic energy, the virtual displacements, forces or the virtual work. This allowed to formulate the principle of virtual work (5.99), which as a field equation defines a second-order field Z. The single geometric object Z then by (5.100) defines all motions of the system as its integral curves. Combining (5.99) and (5.100), i.e., evaluating the principle of virtual work (5.99) along a motion of the system has led to (5.101), which directly characterizes the motion of the system and bypasses the second-order field Z. These two views on the principle of virtual work actually represent two schools of thought within mechanics.

While the approach using Z is purely (differential) geometric, the approach characterizing the curves directly can be used to establish a link to the calculus of variations and functional analysis. This link amplifies the range of mathematical tools that can be used to study the motions of a mechanical system. As a first step in that direction, the following is defined.

Definition 5.17. With respect to a reference field R, let T_R be the kinetic energy of the mechanical system and f_R denote the forces acting on the system. For a motion $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$, the *virtual action of the mechanical system* is defined as

$$\delta A \coloneqq \delta A_R + \delta A_{f_R} = \int_{\mathcal{I}} \big(\delta T_R + f_R(\delta\gamma^v)\big)\mathrm{d}t\,, \tag{5.117}$$

where the *virtual action of the inertia* and the *virtual action of* f_R are respectively given by

$$\delta A_R \coloneqq \int_{\mathcal{I}} \delta T_R\,\mathrm{d}t \quad \text{and} \quad \delta A_{f_R} \coloneqq \int_{\mathcal{I}} f_R(\delta\gamma^v)\,\mathrm{d}t\,. \tag{5.118}$$

Suppose, the interval $\mathcal{I}$ is given as $\mathcal{I} = (t_0, t_1)$, then it follows from the fundamental theorem of calculus, that the virtual action (5.117) of the system is linked to its virtual work (5.101) by

$$\delta A = -\int_{\mathcal{I}} \delta W \circ \dot{\gamma} \, \mathrm{d}t - \underline{\delta} T_R(t_1) + \underline{\delta} T_R(t_0) \,. \tag{5.119}$$

The boundary terms in (5.119) vanish, if virtual displacements $\delta\gamma \in \mathrm{Virt}(\gamma(\mathcal{I}))$ with compact support in $\mathcal{I}$ are chosen, because then $\delta\gamma(t_0) = \delta\gamma(t_1) = 0$. Hence, denoting the set of virtual displacements with compact support on $\mathcal{I}$ as $\mathrm{Virt}_0(\gamma(\mathcal{I}))$, it holds that

$$\delta A = -\int_{\mathcal{I}} \delta W \circ \dot{\gamma} \, \mathrm{d}t \quad \forall \delta\gamma \in \mathrm{Virt}_0(\gamma(\mathcal{I})) \,. \tag{5.120}$$

It is immediately clear, that if γ is a motion of the mechanical system, that is, if γ fulfills the principle of virtual work (5.101), then the virtual action of the system vanishes for all virtual displacements with compact support in $\mathcal{I}$. Using the chart representation (5.105) of the virtual work in (5.120), it is a consequence of the fundamental lemma of the calculus of variations, that the vanishing of δA for all virtual displacements with compact support in $\mathcal{I}$ implies that γ is a motion of the mechanical system. Thus, the following characterization of the motion is equivalent to the principle of virtual work.

Principle of Virtual Action. *With respect to a reference field R, let T_R be the kinetic energy of a mechanical system and let f_R be the forces acting on it. A time-parametrized curve $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ is a motion of the system if and only if its virtual action vanishes for all virtual displacements with compact support in $\mathcal{I}$, i.e.,*

$$\delta A = 0 \quad \forall \delta\gamma \in \mathrm{Virt}_0(\gamma(\mathcal{I})) \,. \tag{5.121}$$

Since the virtual action is linear in its dependency on $\delta\gamma$, the virtual action can be interpreted as a linear functional on the space $\mathrm{Virt}_0(\gamma(\mathcal{I}))$. With this view, the toolbox of functional analysis can be used to study the motion. Moreover, this is the right framework, if one wants to loosen the differentiability assumptions on the motion.[10] This will be important for the development of nonsmooth mechanics in Section 6.1.

If one considers mechanical systems which are only subjected to potential forces, the link to the calculus of variations can be established. In that case, the

[10] The theory of Schwartz' distributions, which are continuous linear functionals on the space of smooth functions with compact support, is for example used to define generalized solutions of differential equations. These generalized solutions do not satisfy the differentiability requirements of the differential equation but still solve it in a precisely defined sense, see Schwartz [113] and [114].

action form and therefore the motions of a mechanical system can be characterized by a Lagrangian L_R with respect to a reference field, see (5.116).

Definition 5.18. Let L_R be a Lagrangian of a mechanical system with respect to a reference field R. Introducing the set of time-parametrized curves $\gamma : \mathcal{I} = [t_0, t_1] \to \mathcal{M}$ with fixed endpoints $\gamma_0, \gamma_1 \in \mathcal{M}$ as

$$\mathcal{C}(\gamma_0, \gamma_1) = \left\{ \gamma \text{ time-param.} \,\middle|\, \gamma(t_0) = \gamma_0,\ \gamma(t_1) = \gamma_1 \right\},$$

the *action* of the mechanical system is the functional

$$A : \mathcal{C}(\gamma_0, \gamma_1) \to \mathbb{R},\ \gamma \mapsto A(\gamma) = \int_{\mathcal{I}} L_R \circ \dot{\gamma}\, \mathrm{d}t\,. \tag{5.122}$$

A motion γ is said to *stationarize* the action if for every one-parameter family of motions $\bar{\gamma}_\varepsilon$ lying in $\mathcal{C}(\gamma_0, \gamma_1)$, the derivative of $A(\bar{\gamma}_\varepsilon)$ with respect to ε vanishes for $\varepsilon = 0$. One easily convinces oneself that such families of motion imply $\delta\gamma(t_0) = \delta\gamma(t_1) = 0$, such that the action is stationary in γ if

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\bigg|_{\varepsilon=0} A(\bar{\gamma}_\varepsilon) = \int_{\mathcal{I}} \delta L_R\, \mathrm{d}t = 0 \quad \forall \delta\gamma \in \mathrm{Virt}_0(\gamma(\mathcal{I}))\,, \tag{5.123}$$

where δL_R is the variation (5.52) of the Lagrangian L_R. Since the potential forces can be derived from a scalar potential $V_R : \mathcal{M} \to \mathbb{R}$, we have $f_R(\delta\gamma^v) = -\delta V_R$ by (5.95) and $L_R = T_R - V_R$ by (5.113). Using this when comparing the principle of virtual action (5.121) with the stationarity condition (5.123), shows that both are identical, which proves the following.

Principle of Hamilton. *Consider a mechanical system which is only subjected to potential forces and let L_R be a Lagrangian of the system with respect to a reference field R. A time-parametrized curve $\gamma : \mathcal{I} = [t_0, t_1] \to \mathcal{M}$ with fixed endpoints $\gamma_0, \gamma_1 \in \mathcal{M}$ is a motion of the system if and only if γ stationarizes the action* (5.122) *of the system.*

In the principle of Hamilton, the restriction to Lagrangians of the form $L_R = T_R - V_R$ has not been mentioned, because the principle still holds in general. A proof can be found in Capobianco et al. [22]. It is also shown in [22] that two more equivalent versions of the Hamilton's Principle can be stated. These principles use action functionals which are similar to (5.122) but are defined on two different sets of curves $\beta : \mathbb{R} \supseteq \mathcal{I} \to A^1\mathcal{M}$.

5.10. Bilateral Constraints

In the modeling process it is sometimes advantageous to not ab initio contemplate kinematic restrictions on the particles of the system, that is, in κ, but to impose

bilateral constraints at a later stage. For instance, a rigid body pendulum can be modeled by starting with a rigid body that can freely move in space. In a second step, the pivot can be added by constraining the possible configurations the rigid body can attain.

In general, the mechanical system before the constraints are applied is said to be *free* and, as opposed to this, the system is referred as being *constrained* if the constraints are applied. Clearly, the notions of free and constrained are relative to the constraints in consideration.

Let $(\mathcal{M}, \vartheta, M)$ denote the Galilean configuration space-time of the free system. Then, there are two possibilities to formulate a condition characterizing the constrained system. For example, a constraint function $g : \mathcal{M} \supseteq U \to \mathbb{R}$ can be used to demand that space-time configurations $p \in \mathcal{M}$ of the constrained system have to fulfill

$$g(p) = 0\,.$$

Such a constraint is said to be formulated at *position level*. A second way to formulate a constraint is to use a one-form $\alpha \in \Omega^1(\mathcal{M})$ restricting the states $v \in A^1\mathcal{M}$ of the constrained system such that

$$\alpha(v) = 0\,.$$

Such a constraint is called a *velocity level constraint*. Every position level constraint can be formulated at velocity level as $\mathrm{d}g(v) = 0$, since for a motion γ of the constrained system it holds that $g \circ \gamma = 0$, which is equivalent to $(g \circ \gamma)^{\cdot} = \mathrm{d}g(\dot{\gamma}) = 0$ for appropriate initial conditions. Clearly, a motion of the free system will not voluntarily fulfill an imposed constraint, but must be forced to do so. The force needed is called *constraint force*. In classical mechanics, it is assumed that every constraint can be formulated at velocity level. Hence, a set of constraints defines a distribution Δ on $\mathcal{M}$. A distribution $\Delta \subseteq T\mathcal{M}$ is *compatible with the time structure ϑ on $\mathcal{M}$* if in the neighborhood of each point of $\mathcal{M}$ there exists a time-normalized vector field lying in Δ.

Definition 5.19. Let $(\mathcal{M}, \vartheta, M)$ denote the Galilean configuration space-time of the free system. A *set of bilateral constraints* is a distribution Δ of rank $f + 1$ on $\mathcal{M}$ that is compatible with the time structure ϑ on $\mathcal{M}$ together with a constraint force f_R^C, which is such that the motions $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ of the constrained system satisfy the constraints in the sense that

$$\dot{\gamma}_{\gamma(t)} \in \Delta_{\gamma(t)} \quad \forall t \in \mathcal{I}\,. \tag{5.124}$$

By (5.124), the constraint distribution describes the portion of the tangent bundle to the configuration space-time $\mathcal{M}$ in which the constrained system is allowed to move. Hence, the bundle $A^1\mathcal{M} \cap \Delta$ is interpreted as the set of admissible states for the constrained system.

Locally, $n-f$ linearly independent one-forms $\alpha^\nu \in \Omega^1(\mathcal{M})$ defined on $U \subseteq \mathcal{M}$ can be found, such that the fibers of the distribution Δ are characterized as

$$\Delta_p = \left\{ X_p \in T_p\mathcal{M} \;\middle|\; \alpha^\nu_p(X_p) = 0, \quad \nu = 1, \dots, n-f \right\} \tag{5.125}$$

for all $p \in U$. Let the defining one-forms be locally represented as

$$\alpha^\nu = W_i{}^\nu \mathrm{d}q^i + \chi^\nu \mathrm{d}t \tag{5.126}$$

with respect to the dual basis field induced by an adapted chart of $\mathcal{M}$. Hence, a motion with $\phi \circ \gamma(t) = (t, \mathbf{q}(t))$ satisfies the constraints, i.e., fulfills (5.124), if $\alpha^\nu(\dot\gamma) = 0$ for all ν, which in coordinates reads as

$$\mathbf{W}^{\mathrm{T}}(\gamma)\, \dot{\mathbf{q}} + \boldsymbol{\chi}(\gamma) = 0\,.$$

Since the constraint distribution Δ is compatible with the time structure ϑ, there is a local vector field B_0 with $\alpha^\nu(B_0) = 0$ and $\vartheta(B_0) = 1$ implying that the differential one-forms $\alpha^1, \dots, \alpha^{n-f}$ and ϑ are linearly independent. Since the one-forms (5.126) and $\vartheta = \mathrm{d}t$ are linearly independent, it follows that the one-forms $W_i{}^\nu \mathrm{d}q^i$ need to be mutually linearly independent. This is the case if the coefficient matrix $W_i{}^\nu(p)$ has full column rank $n-f$ for every $p \in U$. Consequently, because the kernel of the coefficient matrix is of dimension f, one can locally find f linearly independent vector fields

$$B_\mu = B^i{}_\mu \frac{\partial}{\partial q^i} \quad \text{with} \quad W_i{}^\nu \mathrm{d}q^i(B_\mu) = W_i{}^\nu B^i{}_\mu = 0 \tag{5.127}$$

for $\nu = 1, \dots, n-f$ and $\mu = 1, \dots, f$. In matrix notation, the coefficients of the vector fields B_μ satisfy $\mathbf{W}^{\mathrm{T}}\mathbf{B} = 0$. By (5.126), it holds that $\alpha^\nu(B_\mu) = 0$, which shows that the vector fields B_μ lie in the constraint distribution. Moreover, since B_0 is time-normalized, it can be locally written as

$$B_0 = \frac{\partial}{\partial t} + \beta^i \frac{\partial}{\partial q^i}$$

showing that the vector fields $B_0, B_1, \dots, B_f$ are linearly independent. Therefore, they can be used to span the constraint distribution, that is, for all $p \in U$

$$\Delta_p = \mathrm{span}\left\{ (B_0)_p, \dots, (B_f)_p \right\}.$$

Definition 5.20. For a given set of bilateral constraints characterized by the constraint distribution Δ, a virtual displacement $Y_p \in A^0_p\mathcal{M}$ is *compatible with the constraints* if it satisfies

$$Y_p \in \Delta_p \tag{5.128}$$

Therefore, the distribution $A^0\mathcal{M} \cap \Delta$ denotes the set of compatible virtual displacements. The set of compatible virtual displacement fields is defined as the set $\mathrm{Virt}^C(\mathcal{M}) \subset \mathrm{Virt}(\mathcal{M})$ of virtual displacement fields on $\mathcal{M}$ lying in the distribution $A^0\mathcal{M} \cap \Delta$.

It follows from (5.128) that a virtual displacement which is compatible with the constraints fulfills $\alpha_p^\nu(Y_p) = 0$ by (5.125). Moreover, as a virtual displacement, it is spacelike, i.e., $\mathrm{d}t_p(Y_p) = 0$. Therefore, it can be retrieved from the local description (5.126) that the mutually linearly independent one-forms ${W_i}^\nu \mathrm{d}q^i$ together with $\mathrm{d}t$ are defining one-forms for the set of compatible virtual displacements. Hence, $A^0\mathcal{M} \cap \Delta$ is a distribution of rank f and locally

$$\begin{aligned} A_p^0\mathcal{M} \cap \Delta_p = & \left\{ Y_p \in A_p^0\mathcal{M} \;\middle|\; {W_i}^\nu(p)\, \mathrm{d}q_p^i(Y_p) = 0, \quad \nu = 1, \dots, n-f \right\} \\ = & \ \mathrm{span}\left\{ (B_1)_p, \dots, (B_f)_p \right\}, \end{aligned} \tag{5.129}$$

where the second expression immediately follows from the definition (5.127) of the vectors B_μ. Consequently, a virtual displacement field $Y = \delta q^i \frac{\partial}{\partial q^i}$ is compatible with the constraints if its coefficients satisfy

$$\mathbf{W}^{\mathrm{T}} \delta\mathbf{q} = 0\,. \tag{5.130}$$

In classical mechanics, the *principle of d'Alembert–Lagrange* plays a central role for the treatment of constrained mechanical systems and is formulated as follows.

Definition 5.21. A set of bilateral constraints with constraint force f_R^C is *ideal* if the virtual work of the constraint force vanishes for all virtual displacement fields which are compatible with the constraint, that is,

$$\delta W_{f_R^C} = \Phi_R^C\big(Z^C, Y^c\big) = -f_R^C\big(Y^v\big) = 0 \quad \forall Y \in \mathrm{Virt}^C(\mathcal{M})\,.$$

In that case the constraint force is called *ideal* as well.

Proposition 5.2. *A set of bilateral constraints is ideal if and only if the constraint forces can locally be written as*

$$f_R^C = \ ({W_i}^\nu \circ \pi)\lambda_\nu \, \mathrm{d}u^i\,, \tag{5.131}$$

where ${W_i}^\nu$ are the component functions of the defining one-forms of the set of compatible virtual displacements (5.129) *and λ_ν are real functions on $A^1\mathcal{M}$.*

Proof. Using that a general constraint force has the form $f_R^C = f_i^C \, \mathrm{d}u^i$ and a virtual displacement field locally reads as $Y = \delta q^i \frac{\partial}{\partial q^i}$, Proposition 5.2 asserts that

$$\delta\mathbf{q}^{\mathrm{T}}\mathbf{f}^C = 0 \ \forall \delta\mathbf{q} \text{ with } \mathbf{W}^{\mathrm{T}}\delta\mathbf{q} = 0 \quad \Leftrightarrow \quad \exists \boldsymbol{\lambda} \text{ such that } \mathbf{f}^C = \mathbf{W}\boldsymbol{\lambda}$$

where we have used (5.130), Definition 5.21 and have dropped the π in the last equation. The existence of $\boldsymbol{\lambda}$ such that the equivalence holds is guaranteed by the Fredholm alternative, see [116, p. 337]. □

Denoting the virtual work of the free system with δW, it is a consequence of Definition 5.19 that the principle of virtual work describing the constrained motions reads as

$$\delta W^C \circ \dot{\gamma} = \delta W \circ \dot{\gamma} + \delta W_{f_R^C} \circ \dot{\gamma} = 0 \quad \forall \delta\gamma \in \mathrm{Virt}(\gamma(\mathcal{I}))\,. \tag{5.132}$$

By (5.131) and (5.105) the principle of virtual work of the constrained system locally reads as

$$\delta W^C \circ \dot{\gamma} = \delta \mathbf{q}^{\mathrm{T}} \Big(\frac{\mathrm{d}}{\mathrm{d}t} \Big(\frac{\partial T_R}{\partial \mathbf{u}} \Big) - \frac{\partial T_R}{\partial \mathbf{q}} - \mathbf{f} - \mathbf{W}\boldsymbol{\lambda} \Big) = 0 \quad \forall \delta \mathbf{q}\,, \tag{5.133}$$

where $\lambda_\nu = \lambda_\nu \circ \dot{\gamma}$ is such that $\dot{\gamma}_{\gamma(t)} \in \Delta_{\gamma(t)}$ for every $t \in \mathcal{I}$. Let us assume, as discussed in the beginning of this chapter, that the constraint distribution is locally characterized by position and velocity level constraints respectively defined by functions g^σ on $\mathcal{M}$ and one-forms $\alpha^\kappa \in \Omega^1(\mathcal{M})$. Consequently, it follows from (5.133) that the motion γ of the constrained system satisfies the equations of motion

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \Big(\frac{\partial T_R}{\partial \mathbf{u}} \Big) - \frac{\partial T_R}{\partial \mathbf{q}} - \mathbf{f} - \mathbf{W}\boldsymbol{\lambda} = \mathbf{M}\ddot{\mathbf{q}} - \mathbf{h} - \mathbf{W}\boldsymbol{\lambda} = 0 \\ g^\sigma \circ \gamma = 0\,, \quad \sigma = 1, \ldots, n_g \\ \alpha^\kappa(\dot{\gamma}) = 0\,, \quad \kappa = n_g + 1\,, \ldots, n - f\,, \end{aligned} \tag{5.134}$$

where $n_g \leq n - f$ denotes the number of position level constraints. This form of the equations of motion of an ideally constrained system are called *Lagrange's equations of the first kind*.[11]

Since to every position level constraint g^σ in (5.134) the defining one-form $\alpha^\sigma = \mathrm{d}g^\sigma$ of the constraint distribution is associated, it follows from the comparison of

$$\alpha^\sigma = \mathrm{d}g^\sigma = \frac{\partial g^\sigma}{\partial q^i} \mathrm{d}q^i + \frac{\partial g^\sigma}{\partial t} \mathrm{d}t$$

with (5.126) that

$$W_i{}^\sigma = \frac{\partial g^\sigma}{\partial q^i}\,. \tag{5.135}$$

The defining one-forms α^ν of the constraint distribution can be interpreted as real-valued functions on the state space $A^1\mathcal{M}$. With that interpretation, which is fairly common in classical mechanics textbooks, the defining one-forms (5.126) read as $\alpha^\nu = W_i{}^\nu u^i + \chi^\nu$, implying

$$W_i{}^\nu = \frac{\partial \alpha^\nu}{\partial u^i}\,. \tag{5.136}$$

[11] Actually, Lagrange [68] only treated the case where all constraints are on position level.

Moreover, every constraint can be formulated as an *acceleration level constraint*

$$(\alpha^\nu \circ \dot{\gamma})\dot{} = \mathrm{d}\alpha^\nu(\ddot{\gamma}) = 0\,, \tag{5.137}$$

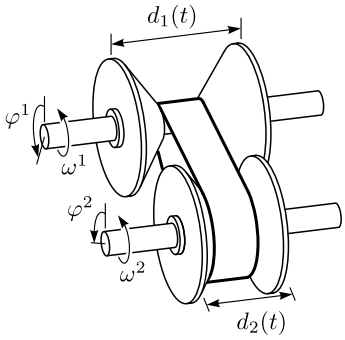
which for suitable initial conditions is obviously equivalent to the velocity level constraint $\alpha^\nu \circ \dot{\gamma} = \alpha^\nu(\dot{\gamma}) = 0$. This shows, that $\ddot{\gamma}$ of the constrained system has to lie in the distribution $\hat{\Delta} \subseteq T(A^1\mathcal{M})$ defined by the one-forms $\mathrm{d}\alpha^\nu$. It is clear from (5.137) that $\hat{\Delta} = T(\Delta \cap A^1\mathcal{M})$, which implies that $\hat{\Delta}$ is integrable. This is not surprising, because it has been assumed that every constraint can be formulated on velocity level, which is respected in Definition 5.19. It would be an interesting task to incorporate "nonclassical" constraints, that is, constraints which can only be formulated on acceleration level, into the presented theory. This would be done by replacing the distribution $\Delta \subseteq T\mathcal{M}$ with a distribution $\hat{\Delta} \subseteq T(A^1\mathcal{M})$ in the definition of a set of bilateral constraints. However, elaborating this goes beyond the scope of the present thesis. More details on this subject can be found in Weber [126].[12]

Definition 5.22. A set of constraints is *holonomic*, if the constraint distribution Δ is integrable. Otherwise, the set of constraints is said to be *nonholonomic*.

Frobenius' theorem, that is, Theorem 4.1, gives necessary and sufficient conditions for the integrability of a distribution and can therefore be used to check whether a set of constraints is holonomic or not.

Example 5.4. *Continuously variable transmission (CVT)*

A CVT is an automatic transmission that can change seamlessly through a continuous range of gear ratios. A common realization of a CVT consists of two pulleys, which both have two cone-shaped opposed parts. A belt mounted between the two pulleys is responsible for the power transmission between the two shafts. The rotation angle and the angular velocity of the i-th shaft are denoted as φ^i and ω^i, respectively. The system's configuration space-time $\mathcal{M}$ is three-dimensional and a natural chart of the state space $A^1\mathcal{M}$ is given by a time function t as well as $\mathbf{q} = (\varphi^1, \varphi^2)$ and $\mathbf{u} = (\omega^1, \omega^2)$. The distances d_1 and d_2 between the conic parts of the pulleys can be adjusted, such



[12] In [126], the notion of holonomic refers to the integrability of $\hat{\Delta}$. Therefore, classical mechanics constraints are always holonomic in their terminology.

that the effective radius of the belt around the shaft is changed, which allows to set the variable gear ratio a between the shaft's angular velocities, i.e., establish a relation $\omega^2 = a(t)\,\omega^1$. This relation constitutes a velocity level constraint $\alpha := \omega^2 - a(t)\,\omega^1 = 0$, where α is interpreted as a real function of the state space. Being linear in the velocities, α can also be interpreted as the one-form $\alpha = \mathrm{d}\varphi^2 - a(t)\,\mathrm{d}\varphi^1$ on $\mathcal{M}$, which is locally defining the constraint distribution Δ. It is readily seen, that Δ is precisely the distribution studied in Example 4.1.[13] Hence, the constraint distribution is not integrable, implying that the constraint is nonholonomic.

Definition 5.22 gives an important classification of a set of constraints. To see this, recall that by (5.124) the velocity $\dot{\gamma}_\gamma(t)$ of the constrained system lies in the constraint distribution Δ for any time. Now, if the set of constraints is holonomic, it follows from Frobenius' theorem that the space-time manifold $\mathcal{M}$ is foliated by the integral manifolds of Δ. Hence, any motion γ of the constrained system lies in one and only one integral manifold $\mathcal{N} \subseteq \mathcal{M}$. Moreover, in the holonomic case, the space-time configurations that can be attained along a motion of the constrained system are constrained to a leaf $\mathcal{N}$ of the foliation induced by Δ. The leaf is obviously determined by the initial conditions of the motion. Otherwise, if the set of constraints is nonholonomic, the space-time configurations that can be attained are unconstrained.

For a set of holonomic constraints, the integral manifold $\mathcal{N}$ of the foliation on $\mathcal{M}$ induced by Δ satisfies

$$\mathcal{N} \cap U = \left\{ p \in U \mid g^\nu(p) = c^\nu, \quad \nu = 1, \ldots, n-f \right\}$$

for some $\mathbf{c} = (c^1, \ldots, c^{n-f}) \in \mathbb{R}^{n-f}$, see (4.10). This implies that if a set of constraints is holonomic, it can be represented by a set of position constraints, because the functions g^ν can always be redefined such that $\mathcal{N}$ is characterized by $\mathbf{c} = \mathbf{0}$. Hence, a set of position level constraints, or more generally a set of holonomic constraints, aims at restricting the motion of the free system to a submanifold $\mathcal{N} \subseteq \mathcal{M}$. Since for a point $p \in \mathcal{N}$ it holds that $T_p\mathcal{N} = \Delta_p$ by the definition of an integral manifold, the tangent bundle $T\mathcal{N}$ is the restriction of the constraint distribution to the submanifold, i.e.,

$$T\mathcal{N} = \Delta|_{\mathcal{N}}\,. \tag{5.138}$$

Consequently, condition (5.124) takes the form

$$\dot{\gamma}_{\gamma(t)} \in T_{\gamma(t)}\mathcal{N} = \Delta_{\gamma(t)} \quad \forall t \in \mathcal{I}\,. \tag{5.139}$$

[13] It suffices to identify $(q^1, q^2, q^3) = (t, \varphi^1, \varphi^2)$ in Example 4.1.

The characterization of the motion by the principle of virtual work (5.132) together with (5.124) can consequently be restated as

$$\begin{aligned} \delta W^C \circ \dot{\gamma} = \ \delta W \circ \dot{\gamma} + \delta W_{f_R^C} \circ \dot{\gamma} = 0 \quad \forall \delta\gamma \in \mathrm{Virt}(\gamma(\mathcal{I})) \\ \gamma(t) \in \mathcal{N} \quad \forall t \in \mathcal{I}\,, \end{aligned} \tag{5.140}$$

where obviously the second equation implies (5.139). Since the constraint distribution is compatible with the time structure ϑ on $\mathcal{M}$, the time structure ϑ is nonvanishing on $\mathcal{N}$. Hence, $\bar{\vartheta} = \vartheta|_{\mathcal{N}}$ defines a time structure on $\mathcal{N}$. Moreover, $\bar{M} = M|_{\mathcal{N}}$ is a Galilean metric on $\mathcal{N}$.

Proposition 5.3 (Winandy [127], p. 196f)**.** *Let Z^C denote the second-order field of the constrained system on $A^1\mathcal{M}$. If the set of constraints is holonomic and ideal, then $\bar{Z} = Z^C|_{A^1\mathcal{N}}$ is the second-order field which by Theorem 4.5 uniquely defines the action form $\bar{\Omega} = \Omega|_{A^1\mathcal{N}}$ on $A^1\mathcal{N}$, where $\mathcal{N}$ is an integral manifold of the constraint distribution.*

Consequently, $\mathcal{N}$ is the configuration space-time and $A^1\mathcal{N}$ the state space of the constrained system. It is crucial to see that $\bar{\Omega}$ is the restriction of the action form Ω of the free system, implying that in the transition from $\mathcal{M}$ to $\mathcal{N}$ no constraint forces have to be considered.

— 6 —

Generalization to Nonsmooth Mechanics

Natura non facit saltus.

—*Gottfried W. Leibniz*

In this chapter, the theory developed in Chapter 5 is generalized to include nonsmooth mechanical systems. In Section 6.1, the continuity assumption of the motion, which so far has been assumed to be smooth, is relaxed to a class of functions allowing velocity jumps. Moreover, the principle of virtual action is generalized such that it also characterizes nonsmooth motions of the mechanical system. The concept of conic distribution is used to define unilateral constraints in Section 6.2. Finally, the description of mechanical systems with frictional contact is addressed in Section 6.3.

6.1. Nonsmooth Motions and the Principle of Virtual Action

In this section, the continuity assumption on the motions of a mechanical system is relaxed to a class of functions allowing velocity jumps. Following Moreau [91], it is assumed that the velocities of a mechanical system are functions of bounded variation. This implies that the motion $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ is absolutely continuous. Moreover, it implies that for any time t, the left limit $\dot{\gamma}^-(t)$ and the right limit $\dot{\gamma}^+(t)$ exist and are finite. Furthermore, γ and $\dot{\gamma}$ are differentiable almost everywhere, except for a countable set of time instants $\{\tau_1, \tau_2, \dots\}$, where the velocity is regarded as right-continuous, that is, $\dot{\gamma}(\tau_i) = \dot{\gamma}^+(\tau_i)$. Last but not least, for any time function t on $\mathcal{M}$, $(t \circ \gamma)^{\cdot}$ is assumed to be continuous, such that a motion can still be defined as a time-parametrized curve, implying that time passes uniformly along motions of the mechanical system.

It is evident, that a motion admitting velocity jumps cannot be a solution of the equations of motion, that is, the solution of a second-order differential equation, because the second derivative of the motion with respect to time, does not exist for all time instants. Hence, the principles of mechanics need to be adapted in order to cope with nonsmooth motions. The principle of virtual action is used as the entry point to generalize the principles of mechanics.

Since the inertial motion of a mechanical system is smooth, by Postulate 1, the velocity jumps must be caused by a force, which as a function of time has the same continuity properties as the motion. Clearly, a force in the sense of Definition 5.11 cannot cause a velocity jump. Thus, the concept of force must be generalized.

Definition 6.1. A *force* is a covector field

$$f\colon \mathcal{I} \to \mathrm{Ver}^*(A^1\mathcal{M}),\ t \mapsto \big(\dot\gamma(t), f_{\dot\gamma(t)}\big)\,,$$

along the motion $\dot\gamma : \mathbb{R} \supseteq \mathcal{I} \to A^1\mathcal{M}$ of the mechanical system.

It is clear, that a force in the sense of Definition 5.11 always induces a force along a motion of the system when evaluated along $\dot\gamma$. Since the velocity jumps of the motion occur only at a countable set $\{\tau_1, \tau_2, \dots\}$ of time instants, the force which is causing the jumps is zero except for these time instants. Such a force is called an *impulsive force*. While the virtual action contribution of a nonimpulsive force is still given by (5.118), the virtual action of an impulsive force F is defined as

$$\delta A_F = \sum_i F(\delta\gamma^v)(\tau_i)\,,$$

where $\tau_1, \tau_2, \dots$ are the time instants for which the impulsive force is nonzero. With this definition in mind, the following postulate can be stated, which characterizes the nonsmooth motions of a mechanical system.

Postulate 4. *With respect to a reference field R, let T_R be the kinetic energy of a mechanical system and let f_R and F_R be the nonimpulsive and impulsive forces acting on it, respectively. The motions of the mechanical system are the time-parametrized curves $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ for which the virtual action*

$$\delta A = \int_{\mathcal{I}} \big(\delta T_R + f_R(\delta\gamma^v)\big)\mathrm{d}t + \sum_i F_R(\delta\gamma^v)(\tau_i) \tag{6.1}$$

vanishes for all virtual displacements with compact support in $\mathcal{I}$, i.e.,

$$\delta A = 0 \quad \forall \delta\gamma \in \mathrm{Virt}_0(\gamma(\mathcal{I}))\,, \tag{6.2}$$

where $\tau_1, \tau_2, \dots$ are the time instants for which the impulsive force is nonzero.

The above postulate generalizes the principle of virtual action to nonsmooth mechanics, but it is nevertheless referred to as principle of virtual action. In order to derive the equations of motion of a nonsmooth mechanical system, assume that $\mathcal{I} = [t_0, t_1)$, such that the virtual action in (6.2) can be rewritten as

$$\delta A = \sum_{i=0}^{N} \int_{[\tau_i, \tau_{i+1})} \big(\delta T_R + f_R(\delta\gamma^v)\big)\mathrm{d}t + \sum_{i=1}^{N} F_R(\delta\gamma^v)(\tau_i)\,, \tag{6.3}$$

where $\tau_0 = t_0$ and $\tau_{N+1} = t_1$ have been used. Moreover, N denotes the possibly infinite number of velocity jumps, which occur at the time instants $\tau_1, \tau_2, \dots$ and are caused by the impulsive force.

Since $\dot{\gamma}$ is differentiable on every interval (τ_i, τ_{i+1}), it follows from the fundamental theorem of calculus that

$$0 = -\sum_{i=0}^{N} \int_{[\tau_i, \tau_{i+1})} (\underline{\delta} T_R)^{\cdot} \, \mathrm{d}t + \sum_{i=0}^{N} \left(\underline{\delta} T_R^-(\tau_{i+1}) - \underline{\delta} T_R^+(\tau_i) \right). \tag{6.4}$$

Adding (6.4) to (6.3) gives

$$\delta A = -\int_{\mathcal{I}} \delta W \circ \dot{\gamma} \, \mathrm{d}t + \sum_{i=1}^{N} \left(\underline{\delta} T_R^-(\tau_i) - \underline{\delta} T_R^+(\tau_i) + F_R(\delta\gamma^v)(\tau_i) \right) \tag{6.5}$$

for all $\delta\gamma \in \mathrm{Virt}_0(\gamma(\mathcal{I}))$, where we have used the virtual work in (5.101) and that necessarily $\delta\gamma(t_0) = \delta\gamma(t_1) = 0$ for virtual displacements with compact support in $\mathcal{I}$.

It is immediately apparent from Postulate 4, that a time-parametrized curve γ is a motion of the mechanical system if (and only if)[1] for any $\delta\gamma \in \mathrm{Virt}_0(\gamma(\mathcal{I}))$, the curve γ satisfies

$$\delta W \circ \dot{\gamma} = 0 \quad \text{a.e.} \quad \text{and} \quad \underline{\delta} T_R^+ - \underline{\delta} T_R^- - F_R(\delta\gamma^v) = 0 \tag{6.6}$$

for $t \in \{\tau_1, \tau_2, \dots\}$. The abbreviation *a.e.* in (6.6) stands for *almost everywhere* and expresses that the equation is fulfilled everywhere except on a set of Lebesgue measure zero. In this special case, this means that $\delta W \circ \dot{\gamma}(t) = 0$ for all $t \in \mathcal{I} \backslash \{\tau_1, \tau_2, \dots\}$. Sometimes a more descriptive terminology is used and it is said that $\delta W \circ \dot{\gamma}(t) = 0$ holds *for almost all* t. Hence, the principle of virtual work (5.101) holds for almost all time instants t. The characterization of the motion using (6.6) can be seen as a generalization of the principle of virtual work to nonsmooth mechanics, which is why it will be referred to as principle of virtual work as well.

By Section 5.8, the first equation in (6.6) implies, that a motion of the system fulfills Lagrange's equations of the second kind almost everywhere, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T_R}{\partial \mathbf{u}}\right) - \frac{\partial T_R}{\partial \mathbf{q}} - \mathbf{f} = \mathbf{M}\ddot{\mathbf{q}} - \mathbf{h} = 0 \quad \text{a.e.},$$

where the representation of the motion with respect to a natural chart of the state space is $\Phi \circ \dot{\gamma}(t) = (t, \mathbf{q}(t), \dot{\mathbf{q}}(t))$. The second equation in (6.6) describes

[1] The "only if" direction is a consequence of the fundamental lemma of the calculus of variations, see Lemma 1 in Gelfand and Fomin [45].

the motion at the time instants at which an impulsive force acts. To derive the equations of motion for these times, remember that $\underline{\delta} T_R = Y^v(T_R) \circ \dot{\gamma}$ and use the chart representations (5.56) and $F_R = F_i \, \mathrm{d}u^i$ in (6.6), which leads to

$$\frac{\partial T_R^+}{\partial \mathbf{u}} - \frac{\partial T_R^-}{\partial \mathbf{u}} = \mathbf{M}\,(\dot{\mathbf{q}}^+ - \dot{\mathbf{q}}^-) = \mathbf{F}$$

for $t \in \{\tau_1, \tau_2, \dots\}$ and is referred to as *impact equation*.

It is customary to introduce the atomic measure η, defined such that for a function of bounded variation $h : \mathcal{I} \to \mathbb{R}$ we have

$$\int_{\mathcal{I}} h(t) \, \mathrm{d}\eta \coloneqq \sum_{i=1}^{N} h(\tau_i) \,, \tag{6.7}$$

where $\tau_1, \tau_2, \dots$ are the discontinuity points of h. Hence, the atomic measure η can be regarded as a sum of Dirac point measures δ_{τ_i}, because (6.7) can be stated as

$$\int_{\mathcal{I}} h(t) \, \mathrm{d}\eta = \sum_{i=1}^{N} \int_{\mathcal{I}} h(t) \, \mathrm{d}\delta_{\tau_i} \,. \tag{6.8}$$

Using (6.7), the virtual action (6.1) and (6.5) can be rewritten as

$$\begin{aligned} \delta A = & \int_{\mathcal{I}} \big(\delta T_R + f_R(\delta\gamma^v)\big)\mathrm{d}t + \int_{\mathcal{I}} F_R(\delta\gamma^v)\mathrm{d}\eta \\ = & -\int_{\mathcal{I}} \delta W \circ \dot{\gamma} \, \mathrm{d}t + \int_{\mathcal{I}} \big(\underline{\delta} T_R^- - \underline{\delta} T_R^+ + F_R(\delta\gamma^v)\big)\mathrm{d}\eta \end{aligned} \tag{6.9}$$

for all $\delta\gamma \in \mathrm{Virt}_0(\gamma(\mathcal{I}))$. With respect to a natural chart of the state space, the motion locally reads as $\Phi \circ \dot{\gamma}(t) = (t, \mathbf{q}(t), \dot{\mathbf{q}}(t))$ such that the virtual action in (6.9) is

$$\delta A = -\int_{\mathcal{I}} \delta\mathbf{q}^{\mathrm{T}}(\mathbf{M}\,\ddot{\mathbf{q}} - \mathbf{h})\mathrm{d}t + \int_{\mathcal{I}} \delta\mathbf{q}^{\mathrm{T}}\big(\mathbf{M}\,(\dot{\mathbf{q}}^- - \dot{\mathbf{q}}^+) + \mathbf{F}\big)\mathrm{d}\eta \,, \tag{6.10}$$

for all $\delta\mathbf{q}$ with compact support in $\mathcal{I}$, where (5.106), (5.56) and (5.37) have been used. To stress that the time derivatives in (6.10) do not exist for all time instants, often new variables $\mathbf{u}$ and $\mathbf{a}$ are introduced such that $\dot{\mathbf{q}} = \mathbf{u}$ and $\ddot{\mathbf{q}} = \dot{\mathbf{u}} = \mathbf{a}$ whenever the derivatives exist. This can be expressed by defining the corresponding differential measures by

$$\mathrm{d}\mathbf{q} = \mathbf{u} \, \mathrm{d}t \quad \text{and} \quad \mathrm{d}\mathbf{u} = \mathbf{a} \, \mathrm{d}t + (\mathbf{u}^+ - \mathbf{u}^-) \, \mathrm{d}\eta \,. \tag{6.11}$$

Using this newly introduced variables in (6.10), the principle of virtual action (6.2) characterizing nonsmooth motions locally reads as

$$\begin{aligned} \delta A = & -\int_{\mathcal{I}} \delta\mathbf{q}^{\mathrm{T}}(\mathbf{M}\mathbf{a}-\mathbf{h})\mathrm{d}t + \int_{\mathcal{I}} \delta\mathbf{q}^{\mathrm{T}}\big(\mathbf{M}\,(\mathbf{u}^{-}-\mathbf{u}^{+})+\mathbf{F}\big)\mathrm{d}\eta \\ = & -\int_{\mathcal{I}} \delta\mathbf{q}^{\mathrm{T}}\big(\mathbf{M}\,\mathrm{d}\mathbf{u}-\mathbf{h}\,\mathrm{d}t-\mathbf{F}\mathrm{d}\eta\big) = 0 \end{aligned} \tag{6.12}$$

for all $\delta\mathbf{q}$ with compact support in $\mathcal{I}$. Equivalently, but more compactly, the principle of virtual action can locally be stated in the form of the equality of measures

$$\mathbf{M}\,\mathrm{d}\mathbf{u}-\mathbf{h}\,\mathrm{d}t-\mathbf{F}\,\mathrm{d}\eta = \mathbf{M}\,\mathrm{d}\mathbf{u}-\mathrm{d}\mathbf{F} = 0\,, \tag{6.13}$$

where all impulsive and nonimpulsive forces as well as the gyroscopic terms present in $\mathbf{h}$, see (5.107), are gathered in one measure $\mathrm{d}\mathbf{F} = \mathbf{h}\,\mathrm{d}t + \mathbf{F}\,\mathrm{d}\eta$.

6.2. Unilateral Constraints

The example par excellence, where unilateral constraints are used, is to model contact between parts of a mechanical system with each other or with the system's environment. The fact that one part of the system cannot penetrate another part of the system can be expressed as a one-sided restriction of the space-time configurations $p \in \mathcal{M}$ of the system. Using a local function $g : \mathcal{M} \supseteq U \to \mathbb{R}$, which is called *gap function* and might be the distance between the two parts in consideration, such a restriction is of the form

$$g(p) \geq 0 \tag{6.14}$$

and constitutes a unilateral constraint at *position level*. A motion γ of the system then has to satisfy $g \circ \gamma \geq 0$.

Since the description of contact phenomena has also been the motivating example for the development of a nonsmooth theory of mechanics, it is standing to reason that unilateral constraints are treated within nonsmooth mechanics, see Section 6.1.

Suppose, for some time t_0, the motion of the unilaterally constrained system satisfies $g \circ \gamma(t_0) > 0$, then the motion is locally unconstrained, because for a short amount of time the motion can evolve in any direction without violating the constraint (6.14). At some later time $t > t_0$, the motion might activate the constraint in the sense that $g \circ \gamma(t) = 0$. At this point it is necessary that

$$\lim_{\bar{t}\uparrow t}(g\circ\gamma)'(\bar{t}) = \mathrm{d}g(\dot{\gamma}^{-}(t)) \leq 0 \quad \text{and} \quad \lim_{\bar{t}\downarrow t}(g\circ\gamma)'(\bar{t}) = \mathrm{d}g(\dot{\gamma}^{+}(t)) \geq 0 \tag{6.15}$$

hold, where the first condition expresses that the motion has not violated the constraint just before activating the constraint and the second condition guarantees that the motion does not violate the constraint in the near future of t. This shows, that if the left inequality is strict, i.e., the motion approaches the unilateral constraint with non-zero velocity, then a velocity jump must occur because of the second inequality. Hence, a velocity jump must be allowed whenever the motion activates the constraint.

Multiplying the first condition in (6.15) by minus one, suggests that one can formulate (6.14) at velocity level as a restriction of the states $(p, v_p) \in A^1\mathcal{M}$ given by

$$\mathrm{d}g_p(v_p) \geq 0 \quad \text{if} \quad g(p) \leq 0\,. \tag{6.16}$$

A unilateral position level constraint is said to be *active at* p if it holds that $g(p) \leq 0$. Admittedly, the inequality $g(p) \leq 0$ in (6.16) is not obvious from the considerations leading to (6.15). In fact one might be tempted to use $g(p) = 0$ instead. However, the inequality is needed in the proof of Moreau's viability lemma,[2] which guarantees that the constraint (6.16) is equivalent to (6.14).

On velocity level, the constraint (6.16) represents a restriction of the velocities v_p to a cone in $T_p\mathcal{M}$ for all $p \in \mathcal{M}$ for which the constraint is active. A *velocity level constraint* must not necessarily stem from a position level constraint, but can be formulated using a local one-form $\alpha \in \Omega^1(\mathcal{M})$ to impose a restriction

$$\alpha_p(v_p) \geq 0\,,$$

on states $(p, v_p) \in A^1\mathcal{M} \subset T\mathcal{M}$, where, similar to (6.16), it must be specified for which points $p \in \mathcal{M}$ the restriction holds. As an example of a pure velocity level constraint, think of a freewheel, which admits only relative velocities in one direction.

In classical mechanics, it is assumed that every constraint can be formulated at velocity level. Hence, a set of unilateral constraints defines a conic distribution $\mathfrak{D}$ on $\mathcal{M}$. A conic distribution $\mathfrak{D} \subseteq T\mathcal{M}$ is *compatible with the time structure* ϑ *on* $\mathcal{M}$ if in the neighborhood of each point of $\mathcal{M}$ there exists a time-normalized vector field lying in $\mathfrak{D}$.

Definition 6.2. Let $(\mathcal{M}, \vartheta, M)$ denote the Galilean configuration space-time of the free system. A *set of unilateral constraints* is a conic distribution $\mathfrak{D}$ on $\mathcal{M}$ that is compatible with the time structure ϑ on $\mathcal{M}$ together with nonimpulsive and impulsive constraint forces f_R^C and F_R^C. The constraint forces are such that the motions $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ of the constrained system satisfy the constraints in the sense that

$$-\dot{\gamma}^-_{\gamma(t)} \in \mathfrak{D}_{\gamma(t)} \quad \text{and} \quad \dot{\gamma}^+_{\gamma(t)} \in \mathfrak{D}_{\gamma(t)} \quad \forall t \in \mathcal{I}\,. \tag{6.17}$$

[2] See Proposition 6.2, on page 121.

The constraint distribution describes the portion of the tangent bundle to the configuration space-time $\mathcal{M}$ in which the constrained system is allowed to move, see (6.17). Hence, the bundle $A^1\mathcal{M} \cap \mathfrak{D}$ can be seen as the set of admissible right limits of states for the constrained system. Often, the conditions (6.17) are referred to as kinematic consistency of the motion on velocity level in backward and forward time, respectively. It will be shown later, that if the system is subjected to position level constraints only, the constraint distribution is induced by these constraints and the conditions (6.17) correspond to the formulation of the constraints on velocity level.

Let the virtual action of the free system be denoted as δA, then it follows from Definition 6.2, that the principle of virtual action describing the motions of the constrained system is

$$\delta A^C \coloneqq \delta A + \int_{\mathcal{I}} \left(f_R^C(\delta\gamma^v)\mathrm{d}t + F_R^C(\delta\gamma^v)\mathrm{d}\eta \right) = 0 \quad \forall \delta\gamma \in \mathrm{Virt}_0(\gamma(\mathcal{I}))\,, \tag{6.18}$$

where the constraint forces are such that the motions satisfy the unilateral constraints, i.e., such that the motions fulfill (6.17).

Definition 6.3. For a given set of unilateral constraints characterized by the conic constraint distribution $\mathfrak{D}$, a virtual displacement $Y_p \in A_p^0\mathcal{M}$ is *compatible with the constraints* if it satisfies

$$Y_p \in \mathfrak{D}_p\,.$$

Therefore, the distribution $A^0\mathcal{M} \cap \mathfrak{D}$ denotes the set of compatible virtual displacements. The set of compatible virtual displacement fields is defined as the set $\mathrm{Virt}^C(\mathcal{M}) \subset \mathrm{Virt}(\mathcal{M})$ of virtual displacement fields on $\mathcal{M}$ lying in the distribution $A^0\mathcal{M} \cap \mathfrak{D}$.

Using k linearly independent one-forms $\alpha^\nu \in \Omega^1(\mathcal{M})$ defined on $U \subseteq \mathcal{M}$ the conic distribution is locally characterized by

$$\mathfrak{D}_p = \left\{ X_p \in T_p\mathcal{M} \;\middle|\; \alpha_p^\nu(X_p) \geq 0,\ \forall \nu \in I(p) \right\} \tag{6.19}$$

for every $p \in U$, see (4.14). Herein, I is a map assigning a subset of $\{1, \ldots, k\}$ to every p, i.e., $I(p)$ is an index set for every $p \in \mathcal{M}$. As will be discussed later, in the case of position level constraints, it is the index set of active position level constraints at p that plays the role of $I(p)$. Let the defining one-forms α^ν in (6.19) be locally represented as

$$\alpha^\nu = \; W_i{}^\nu \mathrm{d}q^i + \chi^\nu \mathrm{d}t \tag{6.20}$$

with respect to the dual basis fields induced by an adapted chart of $\mathcal{M}$. Since the defining one-form of the spacelike distribution $A^0\mathcal{M}$ is $\vartheta = \mathrm{d}t$, the set of

compatible virtual displacements is locally characterized by

$$A_p^0\mathcal{M} \cap \mathfrak{D}_p = \left\{ X_p \in A_p^0\mathcal{M} \mid \ {W_i}^{\nu}(p)\, \mathrm{d}q_p^i(X_p) \geq 0,\ \forall \nu \in I(p) \right\}, \tag{6.21}$$

for all p in the domain of α^{ν}. As in the case of bilateral constraints, the one-forms ${W_i}^{\nu}\mathrm{d}q^i$ are locally defining one-forms for the distribution of compatible virtual displacements $A^0\mathcal{M} \cap \mathfrak{D}$. By (6.21), a virtual displacement field $Y = \delta q^i \frac{\partial}{\partial q^i}$ is compatible with the unilateral constraints if its coefficients satisfy

$$\mathbf{W}_I^{\mathrm{T}} \delta\mathbf{q} \geq 0\,, \tag{6.22}$$

where $\mathbf{W}_I(p)$ is the matrix resulting from $\mathbf{W}(p)$ by discarding all columns whose index is not contained in $I(p)$.

Similar to the case of bilateral constraints, the *principle of d'Alembert–Lagrange*[3] is formulated as follows.

Definition 6.4. Consider a set of unilateral constraints with constraint forces f_R^C and F_R^C, which are nonimpulsive and impulsive, respectively. The set of constraints is *ideal*, if

$$f_R^C(\delta\gamma^v) \geq 0 \quad \text{and} \quad F_R^C(\delta\gamma^v) \geq 0 \quad \forall \delta\gamma \in \mathrm{Virt}_0^C(\gamma(\mathcal{I}))\,. \tag{6.23}$$

The constraint forces of a set of ideal constraints are called *ideal constraint forces.*

Proposition 6.1. *A set of unilateral constraints is ideal if and only if there exist real-valued functions of time $\lambda_\nu, \Lambda_\nu \geq 0$ such that the constraint forces can locally be written as*

$$\begin{aligned} &f_R^C = {W_i}^{\nu}\lambda_\nu\, \mathrm{d}u^i \quad \textit{and} \quad F_R^C = {W_i}^{\nu}\Lambda_\nu\, \mathrm{d}u^i \\ &\textit{with } \lambda_\nu(t) = \Lambda_\nu(t) = 0 \textit{ for all } \nu \notin I(\gamma(t))\,, \end{aligned} \tag{6.24}$$

where ${W_i}^{\nu} = {W_i}^{\nu} \circ \gamma$ are the component functions of the defining one-forms of the set of compatible virtual displacements (6.21) *evaluated along the motion γ of the system.*

Proof. With respect to the basis induced by a natural chart, a general constraint force and a virtual displacement field have the form $f_R^C = f_i^C \mathrm{d}u^i$ and $Y = \delta q^i \frac{\partial}{\partial q^i}$, respectively. Hence, Proposition 6.1 asserts that

$$\delta\mathbf{q}^{\mathrm{T}}\mathbf{f}^C \geq 0\ \forall \delta\mathbf{q} \text{ with } \mathbf{W}_I^{\mathrm{T}}\delta\mathbf{q} \geq 0 \quad \Leftrightarrow \quad \exists \boldsymbol{\lambda}_I \geq 0 :\ \mathbf{f}^C = \mathbf{W}_I\boldsymbol{\lambda}_I\,, \tag{6.25}$$

where (6.22) as well as (6.23) have been used. Hereby, the tuple $\boldsymbol{\lambda}_I(t)$ is introduced, which collects the $\lambda_\nu(t)$ with $\nu \in I(\gamma(t))$ such that $\mathbf{W}_I\boldsymbol{\lambda}_I = \mathbf{W}\boldsymbol{\lambda}$ because $\lambda_\nu(t) = 0$ if $\nu \notin I(\gamma(t))$. The equivalence (6.25) is guaranteed by Farka's Lemma, see [13, p. 165]. The proof for the impulsive force F_R^C is identical. □

[3] See Hamel [57], p. 91 or Leine et al. [79].

As a consequence of (6.24), the virtual action of the ideal constraint forces locally reads as

$$\int_{\mathcal{I}} \left(f_R^C(\delta\gamma^v)\mathrm{d}t + F_R^C(\delta\gamma^v)\mathrm{d}\eta \right) = \int_{\mathcal{I}} \delta\mathbf{q}^{\mathrm{T}}\mathbf{W}\left(\boldsymbol{\lambda}\mathrm{d}t + \boldsymbol{\Lambda}\mathrm{d}\eta\right) = \int_{\mathcal{I}} \delta\mathbf{q}^{\mathrm{T}}\mathbf{W}\mathrm{d}\mathbf{P}\,, \quad (6.26)$$

where the differential measure of the percussion $\mathbf{P}$ is introduced as

$$\mathrm{d}\mathbf{P} \coloneqq \boldsymbol{\lambda}\mathrm{d}t + \boldsymbol{\Lambda}\mathrm{d}\eta\,.$$

Using (6.26) and that the virtual action δA of the free system is given by (6.12), the virtual action (6.18) of an ideally constrained system can locally be written as

$$\delta A^C = -\int_{\mathcal{I}} \delta\mathbf{q}^{\mathrm{T}}\left(\mathbf{M}\,\mathrm{d}\mathbf{u} - \mathbf{h}\,\mathrm{d}t - \mathbf{F}\mathrm{d}\eta - \mathbf{W}\mathrm{d}\mathbf{P}\right). \quad (6.27)$$

In the case where all constraints are formulated at position level, that is, if the set of constraints is defined by k gap functions $g^\nu : \mathcal{M} \supseteq U \to \mathbb{R}$, the conic constraint distribution $\mathfrak{D}$ is locally given by

$$\mathfrak{D}_p = \left\{ X_p \in T_p\mathcal{M} \;\middle|\; \mathrm{d}g_p^\nu(X_p) \geq 0,\ \forall \nu \in A(p) \right\}, \quad (6.28)$$

for all $p \in U$, where $A(p) = \{\nu \,|\, g^\nu(p) \leq 0\}$ is the set of active constraints on position level, see (6.16). Since to every position level constraint g^ν the defining one-form $\alpha^\nu = \mathrm{d}g^\nu$ of the conic constraint distribution is associated, it follows from the comparison of

$$\alpha^\nu = \mathrm{d}g^\nu = \frac{\partial g^\nu}{\partial q^i}\mathrm{d}q^i + \frac{\partial g^\nu}{\partial t}\mathrm{d}t$$

with (6.20) that

$$W_i{}^\nu = \frac{\partial g^\nu}{\partial q^i}\,. \quad (6.29)$$

Moreover, the position level constraints define the subset $\mathcal{C} \subset \mathcal{M}$ which is locally given by

$$\mathcal{C} \cap U = \left\{ p \in U \,|\, g^\nu(p) \geq 0, \quad \nu = 1, \ldots, k \right\}. \quad (6.30)$$

This set can be interpreted as the admissible set of space-time configurations of the constrained system. The question that immediately arises is, whether the constrained motion, which by Definition 6.2 fulfills (6.17), lies within the admissible set $\mathcal{C}$. This question is answered by the following proposition, which sometimes is called the *viability lemma of Moreau*.

Proposition 6.2 (Moreau [91], Propositions 2.2 and 2.4)**.** *Let $\gamma : \mathbb{R} \supseteq \mathcal{I} \to \mathcal{M}$ be a motion of a mechanical system. The motion γ fulfills*

$$\gamma(t) \in \mathcal{C} \quad \forall t \in \mathcal{I}\,, \quad (6.31)$$

if and only if there is a $t_0 \in \mathcal{I}$ for which $\gamma(t_0) \in \mathcal{C}$ and the motion satisfies the unilateral constraints, that is, for all $t \in \mathcal{I}$ it holds that

$$-\dot{\gamma}^-_{\gamma(t)} \in \mathfrak{D}_{\gamma(t)} \quad \textit{and} \quad \dot{\gamma}^+_{\gamma(t)} \in \mathfrak{D}_{\gamma(t)}\,.$$

It is a consequence of Proposition 6.2 that for a set of constraints which is formulated at position level, condition (6.17) can be replaced by (6.31), such that by (6.18) the motion is described by

$$\delta A^C = 0 \quad \forall \delta\gamma \in \mathrm{Virt}_0(\gamma(\mathcal{I})) \quad \text{such that} \quad \gamma(t) \in \mathcal{C} \quad \forall t \in \mathcal{I}\,. \tag{6.32}$$

This is very similar to what has been done in the case of bilateral constraints, see (5.140). Another similarity can be revealed by comparing (6.28) with the definition (4.16) of the tangent cone $T_{\mathcal{C}}(p)$ to the set $\mathcal{C}$ at p. One finds, that $T_{\mathcal{C}} = \mathfrak{D}|_{\mathcal{C}}$, which reminds of (5.138).

Finally, since the constraint distribution (6.28) defined by position level constraints comes with the particular index set $A(p) = \{\nu \mid g^\nu(p) \leq 0\}$, Proposition 6.1 characterizing ideal constraint forces can be reformulated.

Corollary 6.1. *A set of unilateral constraints, all formulated on position level, is ideal if the constraint forces can locally be written as*

$$f^C_R = W_i{}^\nu \lambda_\nu \,\mathrm{d}u^i \quad \textit{and} \quad F^C_R = W_i{}^\nu \Lambda_\nu \,\mathrm{d}u^i\,,$$

where λ_ν and Λ_ν are real-valued functions of time which satisfy the Signorini conditions (Signorini's law)

$$g^\nu \in \mathcal{N}_{\mathbb{R}^-_0}(-\lambda_\nu) \quad \textit{and} \quad g^\nu \in \mathcal{N}_{\mathbb{R}^-_0}(-\Lambda_\nu)\,. \tag{6.33}$$

Herein, the abbreviation $g^\nu = g^\nu \circ \gamma$ has been used.

Proof. The Signorini condition for λ_ν is equivalent to

$$g^\nu \geq 0,\ \lambda_\nu \geq 0,\ g^\nu \lambda_\nu = 0 \text{ (no summation)}\,.$$

Indeed, if $\lambda_\nu > 0$, the argument of the normal cone is an interior point of the set $\mathbb{R}^-_0$ and the normal cone is zero, implying $g^\nu = 0$. On the other hand, if $\lambda_\nu = 0$, it follows from the definition of the normal cone (2.21) that $\mathcal{N}_{\mathbb{R}^-_0}(0)$ is the set of positive numbers including zero and therefore $g^\nu \geq 0$. Especially, if $\nu \notin A(\gamma(t))$, i.e., $g^\nu(t) > 0$, we have that $\lambda_\nu(t) = 0$. Moreover, $\lambda_\nu \geq 0$ is such that the motion satisfies the constraints (6.30), i.e., $g^\nu \geq 0$. The same reasoning hold for Λ_ν, such that the restrictions on λ_ν and Λ_ν found in Proposition 6.1 are retrieved. □

As a consequence, (6.27) can be used in the principle of virtual action (6.32) to see that the motions of an ideally constrained system satisfy

$$\mathbf{M}\,\mathrm{d}\mathbf{u} - \mathbf{h}\,\mathrm{d}t - \mathbf{F}\mathrm{d}\eta - \mathbf{W}\mathrm{d}\mathbf{P} = 0\,,$$

where the components of the percussion measure $\mathrm{d}\mathbf{P} = \boldsymbol{\lambda}\mathrm{d}t + \boldsymbol{\Lambda}\mathrm{d}\eta$ are such that the Signorini conditions (6.33) are fulfilled.

This section is closed with a remark which is overdue. Recall that the (vector) distributions are special cases of conic distributions, see (4.15), and that the intersection of conic distributions is again a conic distribution. Hence, the unilateral constraints treated in this section comprise also bilateral constraints. An important implication is that in the context of nonsmooth mechanics, bilateral constraints come with nonimpulsive and impulsive constraint forces, which is in contrast to smooth mechanics, cf. Definition 5.19. Furthermore, this opens the possibility to define the conic constraint distribution by a combination of unilateral and bilateral constraints. This circumstance is for example extensively used in contact mechanics.

6.3. Mechanical Systems with Frictional Contact

This section bridges the gap between the geometry of a finite-dimensional mechanical system and the actual modeling of such a system, which takes place in a chart. As an archetype of mechanical system one may think of a flexible multibody system. To make the flexible parts fit in this framework, it must be assumed that also the flexible parts can be described by finitely many coordinates. This is typically achieved by modeling the flexible parts as continua, which are then spatially discretized by an appropriate method, e.g., the finite element method. Moreover, a hard frictional contact model is introduced to describe the interaction of contacting bodies.

Since the mechanical model aims at quantitatively describing the motion of a specific system, e.g., a robot, a set of *generalized position coordinates* q^i, such as angles and displacements, have to be chosen. This corresponds to the choice of an adapted chart ϕ with coordinates $(t, \mathbf{q}) = (t, q^1, \dots, q^n)$ of the configuration space-time $\mathcal{M}$ of the system. The coordinates of the induced natural chart Φ of the state space $A^1\mathcal{M}$ are $(t, \mathbf{q}, \mathbf{u}) = (t, q^1, \dots, q^n, u^1, \dots, u^n)$, where $\mathbf{u}$ are called *generalized velocities* of the system. Since for the remainder of this thesis only these two charts are used, the notation can be lightened. For quantities f depending on the point of $\mathcal{M}$, the notation $f(t, \mathbf{q}) = f \circ \phi^{-1}(t, \mathbf{q})$ is introduced, and similarly $g(t, \mathbf{q}, \mathbf{u}) = g \circ \Phi^{-1}(t, \mathbf{q}, \mathbf{u})$ for quantities g defined on $A^1\mathcal{M}$.

With respect to the natural chart Φ, a motion of the mechanical system has the chart representation $\Phi \circ \dot{\gamma}(t) = (t, \mathbf{q}(t), \mathbf{u}(t))$, where $\dot{\mathbf{q}} = \mathbf{u}$ for almost all

times t, which is expressed as

$$\mathrm{d}\mathbf{q} = \mathbf{u}\,\mathrm{d}t\,, \tag{6.34}$$

see (6.11). Moreover, recall that the velocity measure in (6.11) is defined as

$$\mathrm{d}\mathbf{u} = \mathbf{a}\,\mathrm{d}t + (\mathbf{u}^+ - \mathbf{u}^-)\,\mathrm{d}\eta\,, \tag{6.35}$$

where $\mathbf{a}$ is called the *generalized acceleration* of the mechanical system. By (6.35) it holds that $\mathbf{a} = \dot{\mathbf{u}}$ for almost all t. Considering that all forces acting on the mechanical system are represented by $\mathrm{d}\mathbf{F}$, the motions of the mechanical system must satisfy the equality of measures

$$\mathbf{M}(t,\mathbf{q})\,\mathrm{d}\mathbf{u} = \mathrm{d}\mathbf{F}\,, \tag{6.36}$$

where $\mathbf{M} = \mathbf{M}^{\mathrm{T}}$ denotes the mass matrix of the system, see (6.13). Note that in (6.36) the mass matrix is evaluated along the motion.

The mechanical system is assumed to be subjected to $n_g + n_\alpha$ ideal bilateral constraints formulated at position and velocity level, respectively, as

$$\boldsymbol{g}(t,\mathbf{q}) = 0 \in \mathbb{R}^{n_g} \quad \text{and} \quad \boldsymbol{\alpha}(t,\mathbf{q},\mathbf{u}) = 0 \in \mathbb{R}^{n_\alpha}\,. \tag{6.37}$$

These constraints are for example used to model joints, guidances or other kinematic restrictions present in the mechanical system. It is a consequence of the principle of d'Alembert-Lagrange that the constraint forces of an ideal constraint have the particular form

$$\mathrm{d}\mathbf{F}_{g\alpha} = \mathbf{W}_g \mathrm{d}\mathbf{P}_g + \mathbf{W}_\alpha \mathrm{d}\mathbf{P}_\alpha\,,$$

see (6.26), where by (5.135) and (5.136), the generalized force directions are

$$\mathbf{W}_g = \left(\frac{\partial \boldsymbol{g}}{\partial \mathbf{q}}\right)^{\mathrm{T}} \quad \text{and} \quad \mathbf{W}_\alpha = \left(\frac{\partial \boldsymbol{\alpha}}{\partial \mathbf{u}}\right)^{\mathrm{T}},$$

respectively.

Consider that parts of the mechanical system can come into contact with each other and with the environment. It is assumed that the contact in the system can be described by $n_N < \infty$ pairs of contact points, i.e., pairs of particles which eventually come into contact. Let P and Q be the contact points of the k-th contact pair and assume without loss of generality that P lies on part I and Q on part II, see Figure 6.1. Assuming that the contours of the contacting bodies are smooth, the contact points are defined as the surface points that lie on a connecting line orthogonal to both of the bodies contours. Defining $\mathbf{n}$ to be the outward normal of part I at P, the *gap function* of the k-th contact pair is

$$g_N^k \coloneqq \mathbf{n} \cdot \mathbf{r}_{PQ}\,,$$

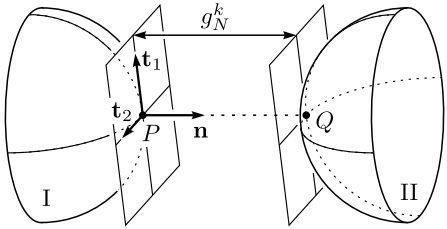


Figure 6.1: Contact kinematics of k-th contact pair.

where $\mathbf{r}_{PQ}$ is the vector pointing from P to Q. Strictly speaking, $\mathbf{r}_{PQ}$ is the vector pointing from ξ_P to ξ_Q and $\mathbf{n}$ is the outward normal of the image $\kappa(\mathrm{I}, \gamma)$ of part I at ξ_P. However, to avoid a cumbersome language, it is convenient to identify the particle with the event in space-time it occupies at a certain time instant. As is custom in the field of multibody dynamics, a contact pair k is simply referred to as contact k. The k-th contact is called

- *open* if $g_N^k > 0$, i.e., the contacting bodies are separated
- *closed* if $g_N^k = 0$, i.e., the contacting bodies are in contact
- *penetrating* if $g_N^k < 0$, i.e., the contacting bodies penetrate each other.

To hinder the contacts from penetrating, the ideal unilateral constraints at position level

$$\boldsymbol{g}_N(t, \mathbf{q}) \geq 0\,,$$

are imposed. The corresponding ideal constraint forces by (6.26) and (6.29) have the form

$$\mathrm{d}\mathbf{F}_N = \mathbf{W}_N \mathrm{d}\mathbf{P}_N \quad \text{with} \quad \mathbf{W}_N = \left(\frac{\partial \boldsymbol{g}_N}{\partial \mathbf{q}}\right)^{\mathrm{T}}.$$

The normal contact percussion measure $\mathrm{d}\mathbf{P}_N = \boldsymbol{\lambda}_N \mathrm{d}t + \boldsymbol{\Lambda}_N \mathrm{d}\eta$ is composed of the nonimpulsive forces $\boldsymbol{\lambda}_N$ and the impulsive forces $\boldsymbol{\Lambda}_N$, which are assumed to satisfy the Signorini condition guaranteeing the ideality of the unilateral constraints, see Corollary 6.1.

For the nonimpulsive contact force λ_{Nk} of the k-th contact, the Signorini condition (6.33) reads as

$$g_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\lambda_{Nk})\,. \tag{6.38}$$

The physical interpretation of the Signorini condition is easiest understood by looking at (6.38) as an inequality complementarity condition

$$\lambda_{Nk} \geq 0, \quad g_N^k \geq 0, \quad \lambda_{Nk}\, g_N^k = 0 \quad \text{(no summation)}. \tag{6.39}$$

Indeed, if $\lambda_{Nk} > 0$, the argument of the normal cone is an interior point of the set $\mathbb{R}_0^-$ and the normal cone is zero, implying $g_N^k = 0$. On the other hand, if $\lambda_{Nk} = 0$, it follows from the definition of the normal cone (2.21) that $\mathcal{N}_{\mathbb{R}_0^-}(0)$ is the set of positive numbers including zero and therefore $g_N^k \geq 0$. It is easy to see from Signorini's condition in the form (6.39) that for open contacts, the normal contact force is zero and if the contact is closed, only forces are allowed which push the contact surfaces apart. In fact, by Example 5.2, or more precisely equation (5.85), the contact force λ_{Nk} (Λ_{Nk}) can be interpreted as the magnitude of the two-point force interaction between the contacting particles P and Q. Moreover, the contact surfaces are subjected to compressive forces if $\lambda_{Nk} > 0$ ($\Lambda_{Nk} > 0$).

The set of active contacts is defined as the index set

$$A(t, \mathbf{q}) = \left\{ k = 1, \dots, n_N \;\middle|\; g_N^k(t, \mathbf{q}) \leq 0 \right\},$$

which is the set of active constraints involved in the local representation of the conic distribution, see (6.28). It is clear from this definition, that the set of inactive contacts, that is, the complement $\bar{A} = \{1, \dots, n_N\} \setminus A$ of the set A, is the set of open contacts.

Due to the ideality of the unilateral constraint, also the impulsive contact force Λ_{Nk} must satisfy the Signorini condition $g_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\Lambda_{Nk})$, which however does not fully determine the impulsive force. Hence, one is free to choose a suitable impact law. The *gap velocity* defined by

$$\dot{\boldsymbol{g}}_N(t, \mathbf{q}, \mathbf{u}) = \mathbf{W}_N^{\mathrm{T}}(t, \mathbf{q})\, \mathbf{u} + \frac{\partial \boldsymbol{g}_N}{\partial t}(t, \mathbf{q}) \tag{6.40}$$

is used to formulate the Newton-type impact law as

$$\begin{aligned} k \in A: &\quad \xi_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\Lambda_{Nk}) \\ k \in \bar{A}: &\quad \Lambda_{Nk} = 0\,. \end{aligned} \tag{6.41}$$

Herein, the kinematic quantity

$$\xi_N^k(t, \mathbf{q}, \mathbf{u}^-, \mathbf{u}^+) = \dot{g}_N^k(t, \mathbf{q}, \mathbf{u}^+) + e_N^k\, \dot{g}_N^k(t, \mathbf{q}, \mathbf{u}^-) \tag{6.42}$$

is introduced for the k-th contact with restitution coefficient $e_N^k \geq 0$. The impact law (6.41) implies that whenever a contact k participates actively in the impact process with $\Lambda_N^k > 0$, the post-impact velocity $\dot{g}_N^{k+} = \dot{g}_N^k(t, \mathbf{q}, \mathbf{u}^+)$ is related to the similarly defined pre-impact velocity by Newton's impact law

$$\dot{g}_N^{k+} = -e_N^k\, \dot{g}_N^{k-}\,.$$

More details on the intricacies of the generalized Newton's impact law (6.41) can be found in Glocker [51] and Winandy et al. [128]. It is readily checked that Λ_{Nk}

given by Newton's impact law (6.41) satisfies the Signorini condition. Indeed, for $g_N^k > 0$ it holds that $k \in \bar{A}$ and consequently $\Lambda_{Nk} = 0$. If $g_K^k = 0$, the normal cone inclusion in (6.41) assures that $\Lambda_{Nk} \geq 0$.

Dry friction between the surfaces of the k-th contact can be described by a Coulomb-type set-valued force law. Therefore, n_F^k velocity parameters $\boldsymbol{\alpha}_F^k(t, \mathbf{q}, \mathbf{u}) \in \mathbb{R}^{n_F^k}$ describing the relative motion of the surfaces are typically introduced, where $n_F^k = 1$ for planar friction, $n_F^k = 2$ for spatial friction and $n_F^k \geq 3$ for combined friction laws such as Coulomb–Contensou friction [75]. The friction forces have the form

$$\mathrm{d}\mathbf{F}_F = \mathbf{W}_F \mathrm{d}\mathbf{P}_F = \sum_{k=1}^{n_N} \mathbf{W}_F^k \mathrm{d}\mathbf{P}_{Fk} \quad \text{with} \quad \mathbf{W}_F^k = \left(\frac{\partial \boldsymbol{\alpha}_F^k}{\partial \mathbf{u}} \right)^{\mathrm{T}}, \tag{6.43}$$

where $\mathbf{W}_F = (\mathbf{W}_F^1 \dots \mathbf{W}_F^{n_N})$ and $\mathrm{d}\mathbf{P}_F = (\mathrm{d}\mathbf{P}_F^1, \dots, \mathrm{d}\mathbf{P}_F^{n_N})$. With C_F denoting the set of admissible (negative) friction forces, the constitutive laws for the nonimpulsive and impulsive friction forces of the k-th contact are

$$\boldsymbol{\alpha}_F^k \in \mathcal{N}_{C_F(\lambda_{Nk})}(-\boldsymbol{\lambda}_{Fk}) \quad \text{and} \quad \boldsymbol{\xi}_F^k \in \mathcal{N}_{C_F(\Lambda_{Nk})}(-\boldsymbol{\Lambda}_{Fk}) \tag{6.44}$$

whenever the contact k is active, i.e., $k \in A$. Moreover, the set C_F depends on the normal contact forces and

$$\boldsymbol{\xi}_F^k(t, \mathbf{q}, \mathbf{u}^-, \mathbf{u}^+) = \boldsymbol{\alpha}_F^k(t, \mathbf{q}, \mathbf{u}^+) + e_F^k \, \boldsymbol{\alpha}_F^k(t, \mathbf{q}, \mathbf{u}^-) \tag{6.45}$$

with restitution coefficient e_F^k. Otherwise, if the contact k is open, the friction forces are zero.

For the sake of simplicity, in this thesis we restrict ourselves to sets of admissible (negative) friction forces having the form

$$C_F(\lambda_{Nk}) = \left\{ \boldsymbol{\lambda}_{Fk} \in \mathbb{R}^{n_F^k} \;\middle|\; ||\boldsymbol{\lambda}_{Fk}|| \leq \mu_k \lambda_{Nk} \right\} \tag{6.46}$$

with friction coefficient μ_k and mention how the general case can be treated wherever it seems appropriate.

Example 6.1 (Isotropic spatial Coulomb friction)**.** The most prominent example of a friction law having a set of admissible friction forces of the form (6.46) is isotropic spatial Coulomb friction. As described in Leine and van de Wouw [77], in that case $n_F^k = 2$ since the velocity parameters correspond to two orthogonal relative velocities which are tangent to the contact surfaces. Specifically, using two mutually orthogonal vectors $\mathbf{t}_1$ and $\mathbf{t}_2$ spanning the tangent planes at the contact points, see Figure 6.1, the velocity parameters describing spatial Coulomb friction are

$$\boldsymbol{\alpha}_F^k = \begin{pmatrix} \mathbf{t}_1 \cdot (\mathbf{v}_Q - \mathbf{v}_P) \\ \mathbf{t}_2 \cdot (\mathbf{v}_Q - \mathbf{v}_P) \end{pmatrix}.$$

In view of Example 5.2, the two components of $\boldsymbol{\lambda}_{Fk} = (\lambda_{Fk,1}, \lambda_{Fk,2})$ are the magnitude of the friction force in the directions $\mathbf{t}_1$ and $\mathbf{t}_2$, respectively. This confirms that the friction force has indeed the form (6.43).

Under the premise that all forces which are not constraint or contact forces, such as spring forces, gyroscopic terms and dashpot forces, are nonimpulsive and can therefore be represented by a Lebesgue-density $\mathbf{h}(t, \mathbf{q}, \mathbf{u})$, the totality of forces acting on the mechanical system is represented by

$$\begin{aligned} \mathrm{d}\mathbf{F} &= \mathbf{h}\,\mathrm{d}t + \mathrm{d}\mathbf{F}_{g\alpha} + \mathrm{d}\mathbf{F}_N + \mathrm{d}\mathbf{F}_F \\ &= \mathbf{h}\,\mathrm{d}t + \mathbf{W}_g\mathrm{d}\mathbf{P}_g + \mathbf{W}_\alpha\mathrm{d}\mathbf{P}_\alpha + \mathbf{W}_N\mathrm{d}\mathbf{P}_N + \mathbf{W}_F\mathrm{d}\mathbf{P}_F\,. \end{aligned}$$

Consequently, the equality of measures (6.36) takes the form

$$\mathbf{M}\,\mathrm{d}\mathbf{u} = \mathbf{h}\,\mathrm{d}t + \mathbf{W}_g\mathrm{d}\mathbf{P}_g + \mathbf{W}_\alpha\mathrm{d}\mathbf{P}_\alpha + \mathbf{W}_N\mathrm{d}\mathbf{P}_N + \mathbf{W}_F\mathrm{d}\mathbf{P}_F\,. \tag{6.47}$$

We recall that the percussion measures $\mathrm{d}\mathbf{P}_X$ are composed by a nonimpulsive force $\boldsymbol{\lambda}_X$ and an impulsive force $\boldsymbol{\Lambda}_X$, that is,

$$\mathrm{d}\mathbf{P}_X = \boldsymbol{\lambda}_X\mathrm{d}t + \boldsymbol{\Lambda}_X\mathrm{d}\eta\,, \tag{6.48}$$

where X is used as a placeholder for any subscript. Using (6.35), one concludes from (6.47) that the acceleration of the system is characterized by the equations of motion

$$\mathbf{M}\,\mathbf{a} = \mathbf{h} + \mathbf{W}_g\boldsymbol{\lambda}_g + \mathbf{W}_\alpha\boldsymbol{\lambda}_\alpha + \mathbf{W}_N\,\boldsymbol{\lambda}_N + \mathbf{W}_F\,\boldsymbol{\lambda}_F \tag{6.49}$$

holding almost everywhere in time and the velocity jumps are given by the impact equations

$$\mathbf{M}\,(\mathbf{u}^+ - \mathbf{u}^-) = \mathbf{W}_g\boldsymbol{\Lambda}_g + \mathbf{W}_\alpha\boldsymbol{\Lambda}_\alpha + \mathbf{W}_N\,\boldsymbol{\Lambda}_N + \mathbf{W}_F\,\boldsymbol{\Lambda}_F\,, \tag{6.50}$$

where the forces are given by the corresponding constitutive laws, e.g., the contact laws, the impact laws and the law of gravity.

— 7 —

Time Finite Element based Moreau-type Integrators

In this chapter, the model for the description of mechanical systems with frictional contacts presented in Section 6.3 is discretized using a finite element method in time. Similar to Capobianco and Eugster [20], a continuous piecewise linear ansatz function for the generalized position coordinates $\mathbf{q}$ is used to arrive at two time-stepping schemes, namely Moreau's time-stepping scheme [20, 50, 91, 120] and the symmetric Moreau-type integrator [20], which are both extended in order to include bilateral constraints.

7.1. Time Finite Element Discretization

The finite element method aims at discretizing a variational equation by choosing ansatz functions for the involved quantities. In the case of mechanical systems with frictional contact, the variational equation at stake is the equality of measures (6.47), which is just a short form of the principle of virtual action stating that

$$\delta A = \int_{\mathcal{I}} \delta\mathbf{q}^{\mathrm{T}} \big(\mathbf{M}\,\mathrm{d}\mathbf{u} - \mathbf{h}\,\mathrm{d}t - \mathbf{W}\mathrm{d}\mathbf{P} \big) = 0 \tag{7.1}$$

for all $\delta\mathbf{q}$ with compact support in $\mathcal{I}$. Hereby, the percussion measures have been gathered as $\mathrm{d}\mathbf{P} = (\mathrm{d}\mathbf{P}_g, \mathrm{d}\mathbf{P}_\alpha, \mathrm{d}\mathbf{P}_N, \mathrm{d}\mathbf{P}_F)$ and $\mathbf{W} = (\mathbf{W}_g\ \mathbf{W}_\alpha\ \mathbf{W}_N\ \mathbf{W}_F)$ has been defined to keep notation short. The principle of virtual action (7.1) describes the motion of the system during the time interval $\mathcal{I}$.

Let the time interval be $\mathcal{I} = (t_0, t_N]$. Since the domain $\mathcal{I}$ of the motion is one-dimensional, choosing a finite element mesh corresponds to a partition of $\mathcal{I}$ into N temporal elements (intervals) $\mathcal{I}_i = (t_{i-1}, t_i]$ with $i = 1, \ldots, N$. It is assumed that all elements have the same temporal length Δt, that is, $t_{i+1} = t_i + \Delta t$ for all $i = 0, \ldots, N-1$. Introducing $\mathbf{q}_i$ to approximate the position coordinates of

the system at t_i, the piecewise linear ansatz

$$\mathbf{q}(t) \coloneqq \frac{t-t_i}{\Delta t}\mathbf{q}_{i+1} + \frac{t_{i+1}-t}{\Delta t}\mathbf{q}_i \quad \forall t \in \mathcal{I}_{i+1},\ i = 0, \dots, N-1 \tag{7.2}$$

is used to approximate the motion of the mechanical system. Herein, the $\mathbf{q}_i$ constitute the nodes of the finite element mesh. It follows from the fact that $\mathbf{u} = \dot{\mathbf{q}}$ for almost all t and the right-continuity of the velocity $\mathbf{u}$ that

$$\mathbf{u}(t) = \frac{1}{\Delta t}(\mathbf{q}_{i+1} - \mathbf{q}_i) \quad \forall t \in \mathcal{I}_{i+1},\ i = 0, \dots, N-1\,. \tag{7.3}$$

Hence, the motion described by (7.2) has a piecewise constant velocity. It is convenient to define

$$\mathbf{u}_i \coloneqq \frac{1}{\Delta t}(\mathbf{q}_{i+1} - \mathbf{q}_i) \tag{7.4}$$

for $i = 0, \dots, N-1$, such that $\mathbf{u}(t) = \mathbf{u}_i$ for all $t \in \mathcal{I}_{i+1}$, see (7.3). It readily follows that the discrete velocity $\mathbf{u}_i$ satisfies

$$\mathbf{u}_i = \mathbf{u}^+(t_i) = \mathbf{u}^-(t_{i+1})\,. \tag{7.5}$$

Since the acceleration $\mathbf{a}$ is the time derivative of the velocity $\mathbf{u}$ for almost all t and the velocity is piecewise constant, it holds that

$$\mathbf{a} = 0 \quad \text{a.e.}\,. \tag{7.6}$$

It is apparent from (7.3) that the velocity's discontinuity points are exactly the time instants $t_1, \dots, t_{N-1}$. Considering that the atomic measure η can be written as a sum of Dirac point measures, see (6.8), the velocity measure (6.35) is

$$\mathrm{d}\mathbf{u} = \sum_{i=0}^{N-2} (\mathbf{u}_{i+1} - \mathbf{u}_i)\, \mathrm{d}\delta_{t_{i+1}}\,, \tag{7.7}$$

which follows from (7.5) and (7.6). In the fashion of the Bubnov-Galerkin method, the ansatz for the virtual displacements is chosen of the same type as the ansatz for $\mathbf{q}$, that is, the virtual displacements are approximated by

$$\delta\mathbf{q}(t) \coloneqq \frac{t-t_i}{\Delta t}\delta\mathbf{q}_{i+1} + \frac{t_{i+1}-t}{\Delta t}\delta\mathbf{q}_i \quad \forall t \in \mathcal{I}_{i+1},\ i = 0, \dots, N-1 \tag{7.8}$$

with nodal values $\delta\mathbf{q}_i$. To assure that $\delta\mathbf{q}$ has compact support on $\mathcal{I}$, it must be imposed that $\delta\mathbf{q}_0 = \delta\mathbf{q}_N = 0$.

Using (7.7) as well as (7.2) and (7.8), the finite element discretization of the first part of the virtual action in (7.1) is

$$\int_{\mathcal{I}} \delta\mathbf{q}^{\mathrm{T}}\mathbf{M}\,\mathrm{d}\mathbf{u} = \sum_{i=0}^{N-2} \delta\mathbf{q}_{i+1}^{\mathrm{T}}\mathbf{M}_{i+1}(\mathbf{u}_{i+1} - \mathbf{u}_i)\,, \tag{7.9}$$

where $\mathbf{M}_{i+1} \coloneqq \mathbf{M}(t_{i+1}, \mathbf{q}_{i+1})$.

In contrast to (7.9), it is not sufficient to just use the ansatz functions to discretize the remaining parts of the virtual action in (7.1), but also a quadrature rule must be chosen. In order to do so, it is exploited that the temporal elements $\mathcal{I}_i$ form a partition of $\mathcal{I}$, such that

$$\int_{\mathcal{I}} \delta\mathbf{q}^{\mathrm{T}} \big(\mathbf{h}\,\mathrm{d}t + \mathbf{W}\mathrm{d}\mathbf{P}\big) = \sum_{i=0}^{N-1} \int_{\mathcal{I}_{i+1}} \delta\mathbf{q}^{\mathrm{T}} \big(\mathbf{h}\,\mathrm{d}t + \mathbf{W}\mathrm{d}\mathbf{P}\big)\,. \tag{7.10}$$

The integral over the element $\mathcal{I}_{i+1} = (t_i, t_{i+1}]$ can for example be approximated by the product of the evaluation of the integrand at the upper bound of the time interval with the length of the temporal element Δt. Specifically,

$$\int_{\mathcal{I}_{i+1}} \delta\mathbf{q}^{\mathrm{T}} \big(\mathbf{h}\,\mathrm{d}t + \mathbf{W}\mathrm{d}\mathbf{P}\big) \approx \delta\mathbf{q}_{i+1}^{\mathrm{T}} \big(\Delta t\,\mathbf{h}_{i+1}^{-} + \mathbf{W}_{i+1}\mathbf{P}_{i+1}\big)\,, \tag{7.11}$$

where $\mathbf{h}_{i+1}^{-} \coloneqq \mathbf{h}(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_i)$ and $\mathbf{W}_{i+1} \coloneqq \mathbf{W}(t_{i+1}, \mathbf{q}_{i+1})$. Moreover, the decomposition $\mathrm{d}\mathbf{P} = \boldsymbol{\lambda}\mathrm{d}t + \boldsymbol{\Lambda}\mathrm{d}\eta$ of the percussion measure has been used to introduce the discrete percussion

$$\mathbf{P}_{i+1} \coloneqq \Delta t\,\boldsymbol{\lambda}(t_{i+1}) + \boldsymbol{\Lambda}(t_{i+1})\,. \tag{7.12}$$

Using the quadrature rule (7.11) together with (7.9), the discretized version of the principle of virtual action (7.1) results as

$$\delta A \approx \sum_{i=0}^{N-2} \delta\mathbf{q}_{i+1}^{\mathrm{T}} \big(\mathbf{M}_{i+1}(\mathbf{u}_{i+1} - \mathbf{u}_i) - \Delta t\,\mathbf{h}_{i+1}^{-} - \mathbf{W}_{i+1}\mathbf{P}_{i+1}\big) = 0$$

for all $\delta\mathbf{q}_i$, $i = 1, \ldots, N-1$. Hence, it is necessary and sufficient that the discretized motion of the system satisfies

$$\mathbf{M}_{i+1}(\mathbf{u}_{i+1} - \mathbf{u}_i) - \Delta t\,\mathbf{h}_{i+1}^{-} - \mathbf{W}_{i+1}\mathbf{P}_{i+1} = 0\,, \tag{7.13}$$

which constitutes the velocity update of Moreau's time-stepping scheme.[1]

To obtain a different scheme, the integral over the element $\mathcal{I}_{i+1} = (t_i, t_{i+1}]$ in (7.10) is approximated using the trapezoidal rule, which yields

$$\int_{\mathcal{I}_{i+1}} \delta\mathbf{q}^{\mathrm{T}} \big(\mathbf{h}\,\mathrm{d}t + \mathbf{W}\mathrm{d}\mathbf{P}\big) \approx \frac{1}{2}\Big(\delta\mathbf{q}_i^{\mathrm{T}} \big(\Delta t\,\mathbf{h}_i^{+} + \mathbf{W}_i\mathbf{P}_i\big) + \delta\mathbf{q}_{i+1}^{\mathrm{T}} \big(\Delta t\,\mathbf{h}_{i+1}^{-} + \mathbf{W}_{i+1}\mathbf{P}_{i+1}\big)\Big)\,, \tag{7.14}$$

[1] Typically, Moreau's time-stepping scheme is assumed to have a possible impact, and therefore a velocity jump at the midpoint of the considered integration step, see for example [50, 91, 120]. The time instants t_i in the present formulation (7.13) are the possible impact times and correspond to the midpoints of the aforementioned common formulations.

where the notation $\mathbf{h}_i^+ \coloneqq \mathbf{h}(t_i, \mathbf{q}_i, \mathbf{u}_i)$ has been introduced. Using the quadrature rule (7.14) together with (7.9), the time finite element discretization of the principle of virtual action (7.1) is

$$\delta A = \sum_{i=0}^{N-2} \delta \mathbf{q}_{i+1}^{\mathrm{T}} \Big(\mathbf{M}_{i+1}(\mathbf{u}_{i+1} - \mathbf{u}_i) - \frac{\Delta t}{2} \big(\mathbf{h}_{i+1}^+ + \mathbf{h}_{i+1}^- \big) - \mathbf{W}_{i+1} \mathbf{P}_{i+1} \Big) = 0$$

for all $\delta \mathbf{q}_i$, $i = 1, \ldots, N-1$. This is satisfied if and only if the discretized motion fulfills

$$\mathbf{M}_{i+1}(\mathbf{u}_{i+1} - \mathbf{u}_i) - \frac{\Delta t}{2} \big(\mathbf{h}_{i+1}^+ + \mathbf{h}_{i+1}^- \big) - \mathbf{W}_{i+1} \mathbf{P}_{i+1} = 0 \,, \tag{7.15}$$

which is the velocity update formula of the symmetric Moreau-type integrator.

The position update is given by (7.4), that is, the position at the end of the time step

$$\mathbf{q}_{i+1} = \mathbf{q}_i + \Delta t \, \mathbf{u}_i \tag{7.16}$$

can be directly computed from the known state $(t_i, \mathbf{q}_i, \mathbf{u}_i)$ at the beginning of the time step. Hence, the position $\mathbf{q}_{i+1}$ is not influenced by any forces, which is why no constraints can be imposed on position level and the constraints (6.37) as well as the Signorini conditions (6.38) must be transferred to the velocity level before they can be imposed in the numerical schemes.

7.2. Velocity Level Constraints

In order to formulate the bilateral constraints (6.37) on velocity level, the constraint velocity

$$\dot{\boldsymbol{g}}(t, \mathbf{q}, \mathbf{u}) = \mathbf{W}_g^{\mathrm{T}}(t, \mathbf{q}) \mathbf{u} + \frac{\partial \boldsymbol{g}}{\partial t}(t, \mathbf{q})$$

is introduced. With that, the bilateral constraints (6.37) are equivalently formulated by demanding

$$\dot{\boldsymbol{g}}(t, \mathbf{q}, \mathbf{u}) = 0 \quad \text{and} \quad \boldsymbol{\alpha}(t, \mathbf{q}, \mathbf{u}) = 0 \,. \tag{7.17}$$

Clearly, these conditions are equivalent to the original constraints if and only if the initial conditions are chosen appropriately, that is, the initial conditions must fulfill the original constraints (6.37). Hence, for the stepping schemes of this chapter, the bilateral constraints can be imposed at velocity level by demanding

$$\dot{\boldsymbol{g}}(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_{i+1}) = 0 \quad \text{and} \quad \boldsymbol{\alpha}(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_{i+1}) = 0 \,. \tag{7.18}$$

As described in Glocker [48, p. 138], the gap velocity (6.40) can be used to formulate the Signorini condition (6.38) at velocity level as

$$\begin{array}{ll} k \in A: & \dot{g}_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\lambda_{Nk}) \\ k \in \bar{A}: & \lambda_{Nk} = 0 \end{array} \quad \text{a.e.} \,. \tag{7.19}$$

Similar to the case of bilateral constraints, the Signorini condition (7.19) is equivalent to (6.38) if the initial conditions of the motion satisfy $g_N^k \geq 0$.

7.3. Discrete Contact Laws

To arrive at discrete contact laws suitable for the stepping schemes presented in this chapter, i.e. for (7.13) and (7.15), the Signorini condition at velocity level (7.19) must be combined with the Newton-type impact law (6.41). Similarly, it is necessary to merge the friction laws described in (6.44).

Since the velocity of the system is continuous between velocity jumps, implying $\dot{g}_N^{k+} = \dot{g}_N^{k-} = \dot{g}_N^k$ for almost all t, it holds that

$$(1 + e_N^k)\dot{g}_N^k = \dot{g}_N^{k+} + e_N^k \dot{g}_N^{k-} = \xi_N^k \quad \text{a.e.}\,, \tag{7.20}$$

where the definition (6.42) of ξ_N^k has been used. Consequently, and in view of $e_N^k \geq 0$, the Signorini conditions at velocity level (7.19) can be stated using ξ_N^k instead of $\dot{g}_N^k$. Hence, for $k \in A$

$$\xi_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\Delta t\, \lambda_{Nk}) \tag{7.21}$$

where Proposition 2.3 has been invoked to scale the contact force by Δt, which denotes the temporal length of the element $\mathcal{I}_{i+1} = [t_i, t_{i+1})$. Recalling that the Newton-type impact law (6.41) is given by

$$\xi_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\Lambda_{Nk}) \tag{7.22}$$

in the case $k \in A$, it finally follows by Proposition 2.4 that (7.21) and (7.22) can be combined to

$$\xi_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\Delta t\, \lambda_{Nk} - \Lambda_{Nk})\,. \tag{7.23}$$

Evaluating (7.23) at the end of the time step yields the discrete normal contact law

$$\begin{aligned} k \in A_{i+1}: &\quad \xi_{N,i+1}^k \in \mathcal{N}_{\mathbb{R}_0^-}(-P_{Nk,i+1}) \\ k \in \bar{A}_{i+1}: &\quad P_{Nk,i+1} = 0\,, \end{aligned}$$

where it has been used that $P_{Nk,i+1} = \Delta t\, \lambda_{Nk}(t_{i+1}) + \Lambda_{Nk}(t_{i+1})$ by (7.12). Moreover, the notations

$$\xi_{N,i+1}^k \coloneqq \xi_N^k(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_i, \mathbf{u}_{i+1}) \quad \text{and} \quad A_{i+1} \coloneqq A(t_{i+1}, \mathbf{q}_{i+1})$$

have been introduced.

A similar reasoning leads to the discrete friction law. In fact, for the case where the contact is active, i.e., $k \in A$, the cone property of the normal cone allows to write $(1 + e_F^k)$ in front of $\boldsymbol{\alpha}_F^k$ in (6.44). Furthermore, for almost all t it

holds that $\mathbf{u}^+(t) = \mathbf{u}^-(t) = \mathbf{u}(t)$. Hence, it follows from (6.45) that the friction laws (6.44) can equivalently be written as

$$\boldsymbol{\xi}_F^k \in \mathcal{N}_{C_F(\Delta t\, \lambda_{Nk})}(-\Delta t\, \boldsymbol{\lambda}_{Fk}) \quad \text{and} \quad \boldsymbol{\xi}_F^k \in \mathcal{N}_{C_F(\Lambda_{Nk})}(-\boldsymbol{\Lambda}_{Fk})\,,$$

where the scaling of the friction force in the first normal cone inclusion immediately follows from the definition of the normal cone (2.21). Invoking Proposition 2.4 yields

$$\boldsymbol{\xi}_F^k \in \mathcal{N}_{C_F(\Delta t\, \lambda_{Nk}+\Lambda_{Nk})}\big(-\Delta t\, \boldsymbol{\lambda}_{Fk} - \boldsymbol{\Lambda}_{Fk}\big)\,. \tag{7.24}$$

Finally, evaluating (7.24) at the end of the time step leads to the discrete friction law

$$\begin{aligned} k \in A_{i+1}: &\quad \boldsymbol{\xi}_{F,i+1}^k \in \mathcal{N}_{C_F(P_{Nk,i+1})}(-\mathbf{P}_{Fk,i+1}) \\ k \in \bar{A}_{i+1}: &\quad \mathbf{P}_{Fk,i+1} = 0\,, \end{aligned}$$

where it has been used that $\mathbf{P}_{Fk,i+1} = \Delta t\, \boldsymbol{\lambda}_{Fk}(t_{i+1}) + \Lambda_{Nk}(t_{i+1})$ by (7.12) and the discrete kinematic quantity

$$\boldsymbol{\xi}_{F,i+1}^k = \boldsymbol{\xi}_F^k(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_i, \mathbf{u}_{i+1})$$

has been introduced.

7.4. Numerical Implementation

This section sums up the two time-stepping schemes derived in the previous sections and describes how every time step of the scheme can be solved numerically. Since the two schemes, namely Moreau's time stepping scheme and the symmetric Moreau-type stepping scheme merely differ by the velocity update formula, they are treated simultaneously in what follows.

As initial conditions of the stepping scheme, the values $\mathbf{q}_0 = \mathbf{q}(t_0)$ and $\mathbf{u}_0 = \mathbf{u}(t_0)$ can be used.[2] Hence, at the beginning of a time step, the quantities $t_i, \mathbf{q}_i, \mathbf{u}_i$ are either known from the previous time step or from the initial conditions.

The step is initiated by computing $t_{i+1} = t_i + \Delta t$ and by performing the position update (7.16), i.e.,

$$\mathbf{q}_{i+1} = \mathbf{q}_i + \Delta t\, \mathbf{u}_i\,.$$

With that, the index set of active constraints $A_{i+1} = A(t_{i+1}, \mathbf{q}_{i+1})$ is computed.

[2] For the symmetric Moreau-type stepping scheme a more accurate initialization can be found in [20].

Next, the velocity update is due. Even though the velocity updates of the two schemes are different, they both have the same structure and can be formulated in the form

$$\begin{aligned} &\mathbf{R}_{\mathrm{s}}(\mathbf{u}_{i+1}, \mathbf{P}_{g,i+1}, \mathbf{P}_{\alpha,i+1}, \mathbf{P}_{N,i+1}, \mathbf{P}_{F,i+1}) = 0 \\ &k \in A_{i+1}: \quad \xi^k_{N,i+1} \in \mathcal{N}_{\mathbb{R}_0^-}(-P_{Nk,i+1}) \\ &\qquad\qquad \boldsymbol{\xi}^k_{F,i+1} \in \mathcal{N}_{C_F(P_{Nk,i+1})}(-\mathbf{P}_{Fk,i+1}) \\ &k \in \bar{A}_{i+1}: \quad P_{Nk,i+1} = 0,\ \mathbf{P}_{Fk,i+1} = 0\,. \end{aligned} \tag{7.25}$$

Since the index set A_{i+1} is fixed during the velocity update, (7.25) is readily recognized to be a normal cone inclusion problem, which can be numerically solved by any of the methods presented in Appendix A.

It remains to state the residuals $\mathbf{R}_{\mathrm{s}}$ for the two schemes. In the case of Moreau's time stepping scheme, it follows from (7.13) and (7.18) that the velocity update is described by

$$\mathbf{R}_{\mathrm{s}} = \begin{pmatrix} \mathbf{M}_{i+1}(\mathbf{u}_{i+1} - \mathbf{u}_i) - \Delta t\, \mathbf{h}^-_{i+1} - \mathbf{W}_{i+1}\mathbf{P}_{i+1} \\ \dot{\boldsymbol{g}}(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_{i+1}) \\ \boldsymbol{\alpha}(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_{i+1}) \end{pmatrix},$$

where $\mathbf{P}_{i+1} = (\mathbf{P}_{g,i+1}, \mathbf{P}_{\alpha,i+1}, \mathbf{P}_{N,i+1}, \mathbf{P}_{F,i+1})$. Similarly, it is a consequence of (7.15) that the velocity update of the symmetric Moreau-type integrator involves

$$\mathbf{R}_{\mathrm{s}} = \begin{pmatrix} \mathbf{M}_{i+1}(\mathbf{u}_{i+1} - \mathbf{u}_i) - \frac{\Delta t}{2}\left(\mathbf{h}^+_{i+1} + \mathbf{h}^-_{i+1}\right) - \mathbf{W}_{i+1}\mathbf{P}_{i+1} \\ \dot{\boldsymbol{g}}(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_{i+1}) \\ \boldsymbol{\alpha}(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_{i+1}) \end{pmatrix}.$$

— 8 —

Nonsmooth Generalized-α Scheme

In this chapter, a nonsmooth generalized-α method is derived, which extends the existing methods [17, 18, 25, 27] to account for frictional contact on velocity and acceleration level as well as a general kinematic equation. Moreover, a set of benchmark systems is devised and finally used to validate the performance of the presented scheme.[1]

8.1. Acceleration Level Constraints and Stabilization

In this section, the bilateral constraints as well as the Signorini condition are transferred to the acceleration level. Moreover, a stabilization on velocity and position level in the sense of Gear-Gupta-Leimkuhler is introduced to avoid constraint drift in the numerical scheme.

In order to formulate the bilateral constraints (6.37) on acceleration level, the constraint accelerations

$$\ddot{\boldsymbol{g}}(t,\mathbf{q},\mathbf{u},\mathbf{a}) = \mathbf{W}_g^{\mathrm{T}}(t,\mathbf{q})\,\mathbf{a} + \frac{\partial \dot{\boldsymbol{g}}}{\partial \mathbf{q}}(t,\mathbf{q},\mathbf{u})\,\mathbf{u} + \frac{\partial \dot{\boldsymbol{g}}}{\partial t}(t,\mathbf{q},\mathbf{u})$$
$$\dot{\boldsymbol{\alpha}}(t,\mathbf{q},\mathbf{u},\mathbf{a}) = \mathbf{W}_\gamma^{\mathrm{T}}(t,\mathbf{q})\,\mathbf{a} + \frac{\partial \boldsymbol{\alpha}}{\partial \mathbf{q}}(t,\mathbf{q},\mathbf{u})\,\mathbf{u} + \frac{\partial \boldsymbol{\alpha}}{\partial t}(t,\mathbf{q},\mathbf{u})$$

are introduced. With those, the bilateral constraints (6.37) are equivalently formulated by demanding

$$\ddot{\boldsymbol{g}}(t,\mathbf{q},\mathbf{u},\mathbf{a}) = 0 \quad \text{and} \quad \dot{\boldsymbol{\alpha}}(t,\mathbf{q},\mathbf{u},\mathbf{a}) = 0 \tag{8.1}$$

for almost all t as well as

$$\dot{\boldsymbol{g}}(t,\mathbf{q},\mathbf{u}^+) = 0 \quad \text{and} \quad \boldsymbol{\alpha}(t,\mathbf{q},\mathbf{u}^+) = 0 \tag{8.2}$$

whenever a velocity jump occurs. Clearly, these conditions are equivalent to the original constraints if the initial conditions are chosen appropriately, that

[1] The results of this chapter have been published in a similar form, see [21].

is, the initial conditions must fulfill the original constraints (6.37) as well as the respective constraints on velocity level (7.17).

As described in Glocker [48, p. 138], the gap acceleration

$$\ddot{\boldsymbol{g}}_N(t,\mathbf{q},\mathbf{u},\mathbf{a}) = \mathbf{W}_N^{\mathrm{T}}(t,\mathbf{q})\,\mathbf{a} + \frac{\partial \dot{\boldsymbol{g}}_N}{\partial \mathbf{q}}(t,\mathbf{q},\mathbf{u})\,\mathbf{u} + \frac{\partial \dot{\boldsymbol{g}}_N}{\partial t}(t,\mathbf{q},\mathbf{u})$$

allows to formulate the Signorini condition (6.38) at acceleration level as

$$\begin{aligned} k \in B: &\quad \ddot{g}_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\lambda_{Nk}) \\ k \in \bar{B}: &\quad \lambda_{Nk} = 0 \end{aligned} \qquad \text{a.e.}, \tag{8.3}$$

where the set

$$B(t,\mathbf{q},\mathbf{u}) = \left\{ k \in A \,\middle|\, \dot{g}_N^k(t,\mathbf{q},\mathbf{u}) \le 0 \right\} \tag{8.4}$$

characterizes the contacts that are active on position as well as on velocity level. The complement of B is again denoted as $\bar{B} = \{1,\dots,n_N\} \setminus B$ and by definition includes $\bar{A}$ as a subset, i.e., $\bar{A} \subseteq \bar{B}$. It is important to point out that if the motion fulfills the Signorini condition on one kinematic level, it does so also on all other kinematic levels provided that the initial conditions are compatible with the other kinematic levels.

It is well known that the constraints of a mechanical system can be formulated on acceleration level without changing its motions. Moreover, the acceleration level constraints come with favorable mathematical properties. During impact free time intervals for example, the presence of position level constraints leads to a DAE of index three. A formulation of the system with constraints on acceleration level reduces to a DAE of index one, which are often easier to solve numerically than higher index DAEs. However, the described index reduction by differentiation is prone to numerical drift, meaning that although the constraint is satisfied on acceleration level, the corresponding position and velocity level constraints are violated due to numerical integration errors. As a remedy, the position and velocity level constraints can be stabilized by introducing additional Lagrange multipliers. This stabilization, initially proposed by Gear et al. [44] and hence known by the name Gear–Gupta–Leimkuhler (GGL) method, can analogously be extended to unilateral constraints, see Brüls et al. [18].

To stabilize the constraints, the system's kinematics is formally extended to

$$\begin{aligned} \mathrm{d}\mathbf{q} &= (\mathbf{u} + \mathbf{u}_S)\,\mathrm{d}t \\ \mathrm{d}\mathbf{u} &= (\mathbf{a} + \mathbf{a}_S)\,\mathrm{d}t + (\mathbf{u}^+ - \mathbf{u}^-)\,\mathrm{d}\eta\,, \end{aligned} \tag{8.5}$$

where the added velocity $\mathbf{u}_S$ and the acceleration $\mathbf{a}_S$ are caused by the stabilization to (6.34) and (6.35), respectively. The multipliers $\boldsymbol{\nu}_g$, $\boldsymbol{\nu}_\alpha$ and $\boldsymbol{\nu}_N$ then take care

of the stabilization of (8.1) and (8.3) on velocity level by demanding

$$\begin{aligned}
&\mathbf{M}\,\mathbf{a}_S = \mathbf{W}_g\boldsymbol{\nu}_g + \mathbf{W}_\alpha\boldsymbol{\nu}_\alpha + \mathbf{W}_N\boldsymbol{\nu}_N \\
&\dot{\boldsymbol{g}}(t,\mathbf{q},\mathbf{u}) = 0 \quad \text{a.e.} \\
&\boldsymbol{\alpha}(t,\mathbf{q},\mathbf{u}) = 0 \quad \text{a.e.} \\
&\begin{aligned} k \in A: &\quad \dot{g}_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\nu_{Nk}) \\ k \in \bar{A}: &\quad \nu_{Nk} = 0\,. \end{aligned} \qquad \text{a.e.}
\end{aligned} \tag{8.6}$$

Note, that the stabilization (8.6) demands $\dot{\boldsymbol{g}} = 0$ for almost all time instants, which combined with (8.2) yields the condition $\dot{\boldsymbol{g}} = 0$ for all time instants. The same reasoning holds for $\boldsymbol{\alpha} = 0$.

The multipliers $\boldsymbol{\mu}_g$ and $\boldsymbol{\mu}_N$ are used to stabilize the constraints (8.1) and (8.3) on position level by

$$\begin{aligned}
&\mathbf{M}\,\mathbf{u}_S = \mathbf{W}_g\boldsymbol{\mu}_g + \mathbf{W}_N\boldsymbol{\mu}_N \\
&\boldsymbol{g}(t,\mathbf{q}) = 0 \\
&g_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\mu_{Nk})
\end{aligned} \tag{8.7}$$

for all contacts k.

It can be shown that, in absence of numerical errors, the solution of the equations of motion formulated with stabilized acceleration level constraints have vanishing Lagrange multipliers $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ and therefore $\mathbf{a}_S = \mathbf{u}_S = 0$. Moreover, it can be shown that the remaining quantities solve the equations of motion with constraints formulated on position and velocity level described in Section 6.3. This fact establishes the mechanical equivalence of the original equations of motion of Section 6.3 and the equations of motion with stabilized acceleration level constraints presented in this section.

8.2. Nonsmooth Generalized-α Discretization

To compute the motion of the mechanical system numerically, in this section a time-stepping scheme from the family of generalized-α methods is presented. The scheme is derived by integrating the equations of motion with stabilized acceleration level constraints over a time interval $\mathcal{I} = (t_i, t_{i+1}]$ and introducing appropriate discrete variables.

Considering the velocity $\mathbf{u}(t)$ as a right-continuous function, the velocity of the system at a time t can be written as

$$\mathbf{u}(t) = \mathbf{u}(t_i) + \int_{(t_i,t]} \mathrm{d}\mathbf{u} = \mathbf{u}(t_i) + \int_{t_i}^{t} \mathbf{a}\,\mathrm{d}t + \int_{(t_i,t]} \left(\mathbf{a}_S\,\mathrm{d}t + (\mathbf{u}^+ - \mathbf{u}^-)\,\mathrm{d}\eta\right), \tag{8.8}$$

where (8.5) has been used. Similarly, the position of the system at time t_{i+1} is

$$\mathbf{q}(t_{i+1}) = \mathbf{q}(t_i) + \int_{\mathcal{I}} \mathrm{d}\mathbf{q} = \mathbf{q}(t_i) + \int_{\mathcal{I}} \mathbf{u}\,\mathrm{d}t + \int_{\mathcal{I}} \mathbf{u}_S\,\mathrm{d}t\,,$$

which with the help of (8.8) can be reformulated to

$$\mathbf{q}(t_{i+1}) = \mathbf{q}(t_i) + \int_{\mathcal{I}} \mathbf{u}(t_i)\,\mathrm{d}t + \int_{\mathcal{I}} \int_{t_i}^{t} \mathbf{a}\,\mathrm{d}\tau\,\mathrm{d}t + \int_{\mathcal{I}} \mathbf{U}(t,t_i)\,\mathrm{d}t + \int_{\mathcal{I}} \mathbf{u}_S\,\mathrm{d}t\,, \quad (8.9)$$

where the last integral in (8.8) is denoted by

$$\mathbf{U}(t,t_i) \coloneqq \int_{(t_i,t]} (\mathbf{a}_S\,\mathrm{d}t + (\mathbf{u}^+ - \mathbf{u}^-)\,\mathrm{d}\eta)\,.$$

To derive the position and velocity updates of the scheme, numerical approximations of the integrals in (8.8) and (8.9) using quadratures have to be introduced. As approximants for the position, velocity and acceleration at some time instant t_i the quantities $\mathbf{q}_i$, $\mathbf{u}_i$ and $\mathbf{a}_i$ are introduced, respectively. Moreover, the discrete variables

$$\mathbf{U}_{i+1} \coloneqq \mathbf{U}(t_{i+1},t_i) \quad \text{and} \quad \mathbf{Q}_{i+1} \coloneqq \int_{\mathcal{I}} \mathbf{U}(t,t_i)\,\mathrm{d}t + \int_{\mathcal{I}} \mathbf{u}_S\,\mathrm{d}t \quad (8.10)$$

are introduced. The integrals of the acceleration $\mathbf{a}$ in (8.8) and (8.9) are discretized in the fashion of a generalized-α method [18] using the quadratures

$$\begin{aligned} \int_{\mathcal{I}} \mathbf{a}\,\mathrm{d}t &\approx \Delta t\big((1-\gamma)\bar{\mathbf{a}}_i + \gamma\bar{\mathbf{a}}_{i+1}\big) \\ \int_{\mathcal{I}} \int_{t_i}^{t} \mathbf{a}\,\mathrm{d}\tau\,\mathrm{d}t &\approx \frac{\Delta t^2}{2}\big((1-2\beta)\bar{\mathbf{a}}_i + 2\beta\bar{\mathbf{a}}_{i+1}\big)\,, \end{aligned} \quad (8.11)$$

where the auxiliary acceleration variables $\bar{\mathbf{a}}$ are linked to the approximants of the acceleration by

$$\alpha_m\bar{\mathbf{a}}_i + (1-\alpha_m)\bar{\mathbf{a}}_{i+1} = \alpha_f\mathbf{a}_i + (1-\alpha_f)\mathbf{a}_{i+1} \quad (8.12)$$

and the time step of the scheme is introduced as $\Delta t = t_{i+1} - t_i$. The coefficients α_f, α_m, β and γ can be chosen according to Newmark [94], Hilber-Hughes-Taylor [62], or Chung and Hulbert [26]. Here, the last option is chosen, which results in a second-order scheme with an adjustable level of numerical dissipation in the high-frequency range. The coefficients of the scheme are then given by

$$\alpha_m = \frac{2\rho_\infty - 1}{\rho_\infty + 1},\ \alpha_f = \frac{\rho_\infty}{\rho_\infty + 1},\ \gamma = \frac{1}{2} + \alpha_f - \alpha_m \text{ and } \beta = \frac{1}{4}\Big(\frac{1}{2} + \gamma\Big)^2\,,$$

where the spectral radius at infinite frequencies $\rho_\infty \in [0,1]$ controls the dissipation in the high-frequency range. For $\rho_\infty = 1$ the scheme shows the minimal and for $\rho_\infty = 0$ the maximal amount of dissipation in the high-frequency regime.

Finally, the position and velocity updates

$$\begin{aligned} \mathbf{u}_{i+1} &= \mathbf{u}_i + \Delta t\big((1-\gamma)\bar{\mathbf{a}}_i + \gamma\bar{\mathbf{a}}_{i+1}\big) + \mathbf{U}_{i+1} \\ \mathbf{q}_{i+1} &= \mathbf{q}_i + \Delta t\,\mathbf{u}_i + \frac{\Delta t^2}{2}\big((1-2\beta)\bar{\mathbf{a}}_i + 2\beta\bar{\mathbf{a}}_{i+1}\big) + \mathbf{Q}_{i+1} \end{aligned} \tag{8.13}$$

are obtained by using (8.11) and (8.10) in (8.8) and (8.9), respectively.

The equations of motion (6.49) are discretized as

$$\mathbf{M}_{i+1}\,\mathbf{a}_{i+1} = \mathbf{h}_{i+1} + \sum_{X\in\{g,\alpha,N,F\}} \mathbf{W}_{X,i+1}\,\boldsymbol{\lambda}_{X,i+1}\,,$$

where the subscript $i+1$ indicates that the quantity is evaluated at t_{i+1}, $\mathbf{q}_{i+1}$, for instance $\mathbf{M}_{i+1} = \mathbf{M}(t_{i+1}, \mathbf{q}_{i+1})$, and when applicable at $\mathbf{u}_{i+1}$, for example $\mathbf{h}_{i+1} = \mathbf{h}(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_{i+1})$. Moreover, $\boldsymbol{\lambda}_{X,i+1}$ approximates the forces $\boldsymbol{\lambda}_X$.

To find the discrete equations for $\mathbf{U}_{i+1}$, consider the approximation

$$\int_{\mathcal{I}} \mathbf{M}(\mathbf{a}_S\,\mathrm{d}t + (\mathbf{u}^+ - \mathbf{u}^-)\,\mathrm{d}\eta) \approx \mathbf{M}_{i+1}\mathbf{U}_{i+1}\,, \tag{8.14}$$

which is exact for a constant mass matrix. Furthermore, the approximation

$$\begin{aligned} \int_{\mathcal{I}} \mathbf{W}_K(\boldsymbol{\nu}_K\mathrm{d}t + \boldsymbol{\Lambda}_K\mathrm{d}\eta) &\approx \mathbf{W}_{K,i+1}\boldsymbol{\Lambda}_{K,i+1} \\ \int_{\mathcal{I}} \mathbf{W}_F\boldsymbol{\Lambda}_F\mathrm{d}\eta &\approx \mathbf{W}_{F,i+1}\boldsymbol{\Lambda}_{F,i+1}\,, \end{aligned} \tag{8.15}$$

is made, which are exact if $\mathbf{W}_{K/F}$ is constant and where the discrete variables

$$\boldsymbol{\Lambda}_{K,i+1} := \int_{\mathcal{I}} \boldsymbol{\nu}_K\mathrm{d}t + \boldsymbol{\Lambda}_K\mathrm{d}\eta \quad\text{and}\quad \boldsymbol{\Lambda}_{F,i+1} := \int_{\mathcal{I}} \boldsymbol{\Lambda}_F\mathrm{d}\eta \tag{8.16}$$

with $K \in \{g,\alpha,N\}$ have been introduced. In view of (6.50) and (8.6), the above discretizations (8.14) and (8.15) yield the discrete impact equations

$$\mathbf{M}_{i+1}\mathbf{U}_{i+1} = \sum_{K\in\{g,\alpha,N\}} \mathbf{W}_{K,i+1}\boldsymbol{\Lambda}_{K,i+1} + \mathbf{W}_{F,i+1}\boldsymbol{\Lambda}_{F,i+1}\,.$$

From the definition of the atomic measure (6.7), it can be seen that the discrete variables $\boldsymbol{\Lambda}_{K,i+1}$ and $\boldsymbol{\Lambda}_{F,i+1}$ consist of the sum of impulsive forces $\sum_k \boldsymbol{\Lambda}_K(t_k)$, respectively $\sum_k \boldsymbol{\Lambda}_F(t_k)$, corresponding to collisions in the time interval $\mathcal{I}$. In addition, $\boldsymbol{\Lambda}_{K,i+1}$ contains a contribution due to the stabilization of the constraints.

The discrete equation for $\mathbf{Q}_{i+1}$ is found from the approximation

$$\int_{\mathcal{I}} \int_{(t_i,t]} \mathbf{M}(\mathbf{a}_S \,\mathrm{d}\tau + (\mathbf{u}^+ - \mathbf{u}^-)\,\mathrm{d}\eta)\,\mathrm{d}t + \int_{\mathcal{I}} \mathbf{M}\mathbf{u}_S \,\mathrm{d}t \approx \mathbf{M}_{i+1}\mathbf{Q}_{i+1}\,.$$

Finally, it is by combining the impact equation (6.50) with the stabilizing conditions (8.6) and (8.7) that the discrete equation

$$\mathbf{M}_{i+1}\mathbf{Q}_{i+1} = \sum_{K\in\{g,N\}} \mathbf{W}_{K,i+1}\boldsymbol{\kappa}_{K,i+1} + \sum_{R\in\{\alpha,F\}} \frac{\Delta t}{2}\mathbf{W}_{R,i+1}\boldsymbol{\Lambda}_{R,i+1}$$

is motivated. Hereby, similar to (8.15) the approximation

$$\int_{\mathcal{I}} \int_{(t_i,t]} \mathbf{W}_K(\boldsymbol{\nu}_K \mathrm{d}t + \boldsymbol{\Lambda}_K \mathrm{d}\eta)\mathrm{d}t + \int_{\mathcal{I}} \mathbf{W}_K\boldsymbol{\mu}_K \mathrm{d}t \approx \mathbf{W}_{K,i+1}\boldsymbol{\kappa}_{K,i+1}$$

is used, where the discrete variables $\boldsymbol{\kappa}_{K,i+1}$ with $K \in \{g, N\}$ have been introduced as

$$\boldsymbol{\kappa}_{K,i+1} := \int_{\mathcal{I}} \int_{(t_i,t]} (\boldsymbol{\nu}_K \mathrm{d}\tau + \boldsymbol{\Lambda}_K \mathrm{d}\eta)\,\mathrm{d}t + \int_{\mathcal{I}} \boldsymbol{\mu}_K \mathrm{d}t\,. \tag{8.17}$$

Moreover, the remaining double integrals are approximated by

$$\int_{\mathcal{I}} \int_{(t_i,t]} (\boldsymbol{\nu}_\alpha \mathrm{d}t + \boldsymbol{\Lambda}_\alpha \,\mathrm{d}\eta)\,\mathrm{d}t \approx \frac{\Delta t}{2}\boldsymbol{\Lambda}_{\alpha,i+1} \quad \text{and} \quad \int_{\mathcal{I}} \int_{(t_i,t]} \boldsymbol{\Lambda}_F \,\mathrm{d}\eta\,\mathrm{d}t \approx \frac{\Delta t}{2}\boldsymbol{\Lambda}_{F,i+1}\,.$$

Similar to standard DAE solvers presented in Hairer and Wanner [54], the bilateral constraints on all kinematic levels are discretized by just evaluating them at the end of the time step. For the constraints originating from a position level constraint that is

$$\boldsymbol{g}(t_{i+1}, \mathbf{q}_{i+1}) = 0, \quad \dot{\boldsymbol{g}}(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_{i+1}) = 0 \quad \text{and} \quad \ddot{\boldsymbol{g}}(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_{i+1}, \mathbf{a}_{i+1}) = 0\,,$$

whereas for constraints originating from a velocity level constraint, the conditions

$$\boldsymbol{\alpha}(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_{i+1}) = 0 \quad \text{and} \quad \dot{\boldsymbol{\alpha}}(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_{i+1}, \mathbf{a}_{i+1}) = 0$$

are imposed by the scheme.

8.3. Discrete Normal Contact Laws

In order to derive an event-capturing time-stepping scheme, the contact dynamics during a time step $\mathcal{I} = (t_i, t_{i+1}]$ is not resolved in all detail and discrete contact laws are derived, which capture the contact dynamics occurring during the time step. More precisely, discrete normal contact laws are derived such that at the

end of the time step impenetrability is satisfied on all kinematic levels while capturing the effects of Newton's impact law.

The discretization of the normal contact laws is started from the acceleration level Signorini conditions (8.3). In view of $e_N^k \geq 0$ and (7.20), the condition $\dot{g}_N^k \leq 0$ can be replaced with $\xi_N^k \leq 0$ in (8.4). The discrete Signorini conditions then result by evaluating all quantities in (8.3) at the end of the time step. Thus, we have

$$\begin{aligned} k \in B_{i+1}: &\quad \ddot{g}_{N,i+1}^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\lambda_{Nk,i+1}) \\ k \in \bar{B}_{i+1}: &\quad \lambda_{Nk,i+1} = 0 \end{aligned}, \tag{8.18}$$

where the involved discrete variables are defined as

$$\begin{aligned} g_{N,i+1}^k &= g_N^k(t_{i+1}, \mathbf{q}_{i+1}) \\ \xi_{N,i+1}^k &= \xi_N^k(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_i, \mathbf{u}_{i+1}) \\ \ddot{g}_{N,i+1}^k &= \ddot{g}_N^k(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_{i+1}, \mathbf{a}_{i+1}) \end{aligned}$$

and the discrete version of (8.4) is introduced as

$$B_{i+1} = \left\{ k \in A_{i+1} \;\middle|\; \xi_{N,i+1}^k \leq 0 \right\}, \text{ where } \quad A_{i+1} = A(t_{i+1}, \mathbf{q}_{i+1}). \tag{8.19}$$

To formulate the normal contact law on velocity level, Signorini's law on velocity level, the impact law and the stabilization condition are combined. For that, consider the case of active normal contact, i.e., $k \in A$. Then, Signorini's law and the stabilization condition

$$\dot{g}_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\lambda_{Nk}) \quad \text{and} \quad \dot{g}_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\nu_{Nk}) \tag{8.20}$$

hold almost everywhere, respectively. Using the cone property of the normal cone, $(1 + e_N^k)$ may be written in front of $\dot{g}_N^k$ in (8.20), which in view of (7.20) takes the form

$$\xi_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\lambda_{Nk}) \quad \text{and} \quad \xi_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\nu_{Nk}). \tag{8.21}$$

The proposed reformulation of the combined Signorini and stabilization condition has the same form as the impact law (6.41), which reads as

$$\xi_N^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\Lambda_{Nk}). \tag{8.22}$$

With this preparatory work, we can finally proceed toward a discrete law. Since we are interested in the end of the time step, the set A_{i+1} introduced in (8.19) is used to determine whether the contact k is active on position level at the end of the time step. If this contact is active, that is $k \in A_{i+1}$, the contact is assumed to be active during the whole time step. Furthermore, the kinematic quantity ξ_N^k is assumed to be constant within a time step $\mathcal{I} = (t_i, t_{i+1}]$ and to

correspond to $\xi^k_{N,i+1}$, which allows to combine (8.21) and (8.22) in integral form as

$$k \in A_{i+1}: \quad \xi^k_{N,i+1} \in \mathcal{N}_{\mathbb{R}^-_0}\left(-\int_{\mathcal{I}} \left((\lambda_{Nk} + \nu_{Nk})\mathrm{d}t + \Lambda_{Nk}\mathrm{d}\eta \right) \right), \tag{8.23}$$

where Proposition 2.4 has been used. If $k \in \bar{A}_{i+1}$, the integral in (8.23) is zero.

At this point it is appropriate to qualitatively discuss the assumptions leading to (8.23). For that, consider two cases. Either there is no collision in the time interval $\mathcal{I}$ and only nonimpulsive motion takes place on $\mathcal{I}$ (case i) or there is impulsive motion during the time interval $\mathcal{I}$ (case ii).

(i) *Purely nonimpulsive motion on $\mathcal{I}$:* Since the contact is closed at the end of the time step and there is no collision, the contact must either have been closed during the whole time step, or it must have closed at some $t_c \in \mathcal{I} = (t_i, t_{i+1}]$ with $\dot{g}^k_N(t_c) = 0$. In the first case, the contact velocity $\dot{g}^k_N$ vanishes on $\mathcal{I}$ and by (7.20) we have $\xi^k_N = 0$ on $\mathcal{I}$ justifying the approximation. In the second case, for which the contact closes during the time step we have $\dot{g}^k_N(t_i) < 0$ and $\dot{g}^k_N(t_{i+1}) \geq 0$, which is at best approximated by $\xi^k_{N,i+1} = 0$. This allows the integral over the contact forces in (8.23) to be non-zero, which can capture the exact dynamics.

(ii) *Impulsive motion on $\mathcal{I}$:* There are one or more collisions or other events causing impulsive motion (a dynamic catastrophe). Then we have to admit that the impulsive part of the motion is dominating the dynamics on $\mathcal{I}$ and the nonimpulsive motion may be neglected. The error which is then made in the nonimpulsive description by falsely considering ξ^k_N to be constant is then small and of the order of the time step. If the collision takes place at $t_c \in \mathcal{I} = (t_i, t_{i+1}]$, then $\xi^k_N(t_c, \mathbf{q}(t_c), \mathbf{u}^-(t_c), \mathbf{u}^+(t_c))$ is approximated by $\xi^k_N(t_{i+1}, \mathbf{q}(t_{i+1}), \mathbf{u}(t_i), \mathbf{u}(t_{i+1}))$. This approximation becomes exact in the limit of $\Delta t \downarrow 0$.

Finally, equation (8.23) suggests a discrete normal contact law on velocity level given by

$$\begin{aligned} k \in A_{i+1}: &\quad \xi^k_{N,i+1} \in \mathcal{N}_{\mathbb{R}^-_0}(-P_{Nk,i+1}) \\ k \in \bar{A}_{i+1}: &\quad P_{Nk,i+1} = 0\,, \end{aligned} \tag{8.24}$$

where the discrete percussion is defined by

$$\mathbf{P}_{N,i+1} = \mathbf{\Lambda}_{N,i+1} + \Delta t\big((1-\gamma)\bar{\boldsymbol{\lambda}}_{N,i} + \gamma\bar{\boldsymbol{\lambda}}_{N,i+1}\big)\,. \tag{8.25}$$

Hereby, the newly introduced auxiliary force variables $\bar{\boldsymbol{\lambda}}_N$ are linked to the approximants of the contact forces by

$$\alpha_m\bar{\boldsymbol{\lambda}}_{N,i} + (1-\alpha_m)\bar{\boldsymbol{\lambda}}_{N,i+1} = \alpha_f\boldsymbol{\lambda}_{N,i} + (1-\alpha_f)\boldsymbol{\lambda}_{N,i+1}\,. \tag{8.26}$$

To motivate (8.24) as an approximation of (8.23), one recognizes that the second part of the discrete percussion (8.25) is a generalized-α discretization of the integral of $\boldsymbol{\lambda}_N$ over $\mathcal{I}$, see (8.11) and (8.12). This, in combination with (8.16), indeed shows that the discrete percussion approximates the integral appearing in (8.23), that is,

$$\mathbf{P}_{N,i+1} \approx \int_{\mathcal{I}} \left((\boldsymbol{\lambda}_N + \boldsymbol{\nu}_N) \mathrm{d}t + \boldsymbol{\Lambda}_N \mathrm{d}\eta \right) \overset{(6.48)}{=} \int_{\mathcal{I}} \left(\mathrm{d}\mathbf{P}_N + \boldsymbol{\nu}_N \mathrm{d}t \right). \tag{8.27}$$

In order to state the contact law on position level, it can be observed that (8.23) also holds if the integral is only taken over a time span $(t_i, t] \subseteq \mathcal{I}$. Moreover, the inclusion (8.23) implies that the integral is non-negative if the contact k is closed at the end of the time step. Since, in addition, the integral is zero if the contact k is open at the end of the time step, it follows that

$$g_{N,i+1}^k \in \mathcal{N}_{\mathbb{R}_0^-} \left(- \int_{(t_i,t]} \left((\lambda_{Nk} + \nu_{Nk}) \mathrm{d}\tau + \Lambda_{Nk} \mathrm{d}\eta \right) \right) \tag{8.28}$$

for all $t \in \mathcal{I}$. Moreover, the stabilization on position level (8.7) is approximated by

$$g_{N,i+1}^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\mu_{Nk}(t)) \tag{8.29}$$

for all times t in $\mathcal{I}$. Finally, Proposition 2.4 is used to combine (8.28) and (8.29) to

$$g_{N,i+1}^k \in \mathcal{N}_{\mathbb{R}_0^-} \left(- \left(\int_{\mathcal{I}} \int_{(t_i,t]} \left((\lambda_{Nk} + \nu_{Nk}) \mathrm{d}\tau + \Lambda_{Nk} \mathrm{d}\eta \right) \mathrm{d}t + \int_{\mathcal{I}} \mu_{Nk} \mathrm{d}t \right) \right). \tag{8.30}$$

This inclusion suggests a discrete normal contact law on position level given by

$$g_{N,i+1}^k \in \mathcal{N}_{\mathbb{R}_0^-}(-\hat{\kappa}_{Nk,i+1}) \quad \text{with} \quad k = 1, \ldots, n_N \,, \tag{8.31}$$

together with

$$\hat{\boldsymbol{\kappa}}_{N,i+1} = \boldsymbol{\kappa}_{N,i+1} + \frac{\Delta t^2}{2} \left((1 - 2\beta) \bar{\boldsymbol{\lambda}}_{N,i} + 2\beta \bar{\boldsymbol{\lambda}}_{N,i+1} \right). \tag{8.32}$$

Hereby, the second part in (8.32) approximates the double integral of $\boldsymbol{\lambda}_N$ using the generalized-α method, see (8.11) and (8.12), and it therefore follows from (8.17) that

$$\hat{\boldsymbol{\kappa}}_{N,i+1} \approx \int_{\mathcal{I}} \int_{(t_i,t]} \left((\boldsymbol{\lambda}_N + \boldsymbol{\nu}_N) \mathrm{d}\tau + \boldsymbol{\Lambda}_N \mathrm{d}\eta \right) \mathrm{d}t + \int_{\mathcal{I}} \boldsymbol{\mu}_N \mathrm{d}t \,,$$

which confirms that (8.31) is indeed an approximation of (8.30).

8.4. Discrete Friction Laws

In (6.44), the friction law of the k-th contact is stated as

$$\boldsymbol{\alpha}_F^k \in \mathcal{N}_{C_F(\lambda_{Nk})}(-\boldsymbol{\lambda}_{Fk}) \tag{8.33}$$

for the case where the contact is active, i.e., $k \in A$. Using the cone property of the normal cone, one may write $(1 + e_F^k)$ in front of $\boldsymbol{\alpha}_F^k$ in (8.33). Furthermore, for almost all t it holds that $\mathbf{u}^+(t) = \mathbf{u}^-(t) = \mathbf{u}(t)$. Hence, it follows from (6.45) that the friction law can be equivalently written as

$$\boldsymbol{\xi}_F^k \in \mathcal{N}_{C_F(\lambda_{Nk})}(-\boldsymbol{\lambda}_{Fk}).$$

This reformulation of the friction law brings it in a similar form as the frictional impact law (6.44)

$$\boldsymbol{\xi}_F^k \in \mathcal{N}_{C_F(\Lambda_{Nk})}(-\boldsymbol{\Lambda}_{Fk})\,.$$

The following discretization process exploits that the set of admissible (negative) friction forces has the homogeneity property $C_F(\alpha) = \alpha C_F(1)$. Moreover, similar as for the laws in normal direction, $\boldsymbol{\xi}_F^k$ is considered to be constant on the short time-interval $\mathcal{I}$ and it is assumed to take the value

$$\boldsymbol{\xi}_{F,i+1}^k = \boldsymbol{\xi}_F^k(t_{i+1}, \mathbf{q}_{i+1}, \mathbf{u}_i, \mathbf{u}_{i+1})\,.$$

Specifically, this translates to the friction law as

$$\boldsymbol{\xi}_{F,i+1}^k \in \mathcal{N}_{C_F(\lambda_{Nk}(t))}\big(-\boldsymbol{\lambda}_{Fk}(t)\big), \quad \forall t \in \mathcal{I}.$$

Because $\boldsymbol{\xi}_{F,i+1}^k$ is constant on $\mathcal{I}$, relation (2.22), that is, Proposition 2.4 in integral form, can be invoked, which directly gives

$$\boldsymbol{\xi}_{F,i+1}^k \in \mathcal{N}_{C_F\left(\int_{\mathcal{I}} \lambda_{Nk}\mathrm{d}t\right)}\left(-\int_{\mathcal{I}} \boldsymbol{\lambda}_{Fk}\,\mathrm{d}t\right). \tag{8.34}$$

For the impulsive part of the motion on $\mathcal{I}$, we similarly have the approximation

$$\boldsymbol{\xi}_{F,i+1}^k \in \mathcal{N}_{C_F(\Lambda_{Nk}(t))}\big(-\boldsymbol{\Lambda}_{Fk}(t)\big), \quad \forall t \in \mathcal{I}$$

and using (2.22), the impact law is casted in integral form as

$$\boldsymbol{\xi}_{F,i+1}^k \in \mathcal{N}_{C_F\left(\int_{\mathcal{I}} \Lambda_{Nk}\,\mathrm{d}\eta\right)}\left(-\int_{\mathcal{I}} \boldsymbol{\Lambda}_{Fk}\mathrm{d}\eta\right). \tag{8.35}$$

Finally, (8.34) and (8.35) are combined by using Proposition 2.4 to

$$\boldsymbol{\xi}_{F,i+1}^k \in \mathcal{N}_{C_F\left(\int_{\mathcal{I}} \lambda_N^k\mathrm{d}t+\Lambda_N^k\mathrm{d}\eta\right)}\left(-\int_{\mathcal{I}} (\boldsymbol{\lambda}_F^k\mathrm{d}t + \boldsymbol{\Lambda}_F^k\mathrm{d}\eta)\right),$$

which motivates as approximation the discrete friction law

$$\begin{aligned} k \in A_{i+1}: &\quad \boldsymbol{\xi}^k_{F,i+1} \in \mathcal{N}_{C_F(P_{Nk,i+1})}(-\mathbf{P}_{Fk,i+1}) \\ k \in \bar{A}_{i+1}: &\quad \mathbf{P}_{Fk,i+1} = 0\,, \end{aligned} \tag{8.36}$$

where

$$\mathbf{P}_{F,i+1} = \boldsymbol{\Lambda}_{F,i+1} + \Delta t\big((1-\gamma)\bar{\boldsymbol{\lambda}}_{F,i} + \gamma\bar{\boldsymbol{\lambda}}_{F,i+1}\big)\,. \tag{8.37}$$

Hereby, the introduced auxiliary discrete friction forces are linked to the actual forces by

$$\alpha_m\bar{\boldsymbol{\lambda}}_{F,i} + (1-\alpha_m)\bar{\boldsymbol{\lambda}}_{F,i+1} = \alpha_f\boldsymbol{\lambda}_{F,i} + (1-\alpha_f)\boldsymbol{\lambda}_{F,i+1}\,. \tag{8.38}$$

Clearly, because of (8.16) and interpreting the second part of (8.37) as the generalized-α approximation of the integral of the nonimpulsive friction forces, it holds that

$$\mathbf{P}_{F,i+1} \approx \int_{\mathcal{I}} (\boldsymbol{\Lambda}_F \mathrm{d}\eta + \boldsymbol{\lambda}_F \mathrm{d}t) \overset{(6.48)}{=} \int_{\mathcal{I}} \mathrm{d}\mathbf{P}_F\,.$$

Using the discrete percussion $P_{Nk,i+1}$ in (8.36) is finally justified by (8.27) since the stabilizing Lagrange multipliers are zero for the exact solution.

Since the discrete friction law (8.36) combines the effects of nonimpulsive and impulsive friction forces, an additional friction law is needed to distinguish these effects and compute values for both $\boldsymbol{\lambda}_{Fk,i+1}$ and $\boldsymbol{\Lambda}_{Fk,i+1}$. The friction law (8.36) basically consists of three cases. Either the k-th contact is open ($k \in \bar{A}_{i+1}$) and $\mathbf{P}_{Fk,i+1} = 0$ or the contact is active ($k \in A_{i+1}$) and one of the following two cases holds. In the first case, the negative discrete percussion lies in the interior of the set of admissible negative friction forces $C_F(P_{Nk,i+1})$, which by the normal cone inclusion in (8.36) implies $\boldsymbol{\xi}^k_{F,i+1} = 0$. Hence, this can be seen as the case of sticking contact, which motivates the definition of the set of sticking contacts as

$$D^{\mathrm{st}}_{i+1} = \big\{k \in A_{i+1} \;\big|\; \boldsymbol{\xi}^k_{F,i+1} = 0\big\}\,.$$

The second case is the case of slipping contact, where $\boldsymbol{\xi}^k_{F,i+1}$ is nonzero and the negative discrete percussion must lie on the boundary of the set $C_F(P_{Nk,i+1})$ by (8.36). Obviously the set of slipping contacts is just the complement of the set of sticking contacts, that is, $D^{\mathrm{sl}}_{i+1} = A_{i+1} \setminus D^{\mathrm{st}}_{i+1}$. Hence, in case of slipping $k \in D^{\mathrm{sl}}_{i+1}$.

We are now ready to state the remaining discrete friction law as

$$\begin{aligned} k \in D^{\mathrm{st}}_{i+1}: &\quad \dot{\boldsymbol{\alpha}}^k_{F,i+1} \in \mathcal{N}_{C_F(\lambda_{Nk,i+1})}(-\boldsymbol{\lambda}_{Fk,i+1}) \\ k \in D^{\mathrm{sl}}_{i+1}: &\quad \boldsymbol{\lambda}_{Fk,i+1} = \boldsymbol{\lambda}^{\mathrm{sl}}_{Fk,i+1} \\ k \in \bar{A}_{i+1}: &\quad \boldsymbol{\lambda}_{Fk,i+1} = 0\,, \end{aligned} \tag{8.39}$$

where $\boldsymbol{\lambda}^{\mathrm{sl}}_{Fk,i+1}$ denotes the element on the boundary of $C_F(\lambda_{Nk,i+1})$ satisfying

$$\boldsymbol{\alpha}^k_{F,i+1} \in \mathcal{N}_{C_F(\lambda_{Nk,i+1})}(-\boldsymbol{\lambda}^{\mathrm{sl}}_{Fk,i+1})\,.$$

Moreover, $\dot{\boldsymbol{\gamma}}^k_{F,i+1}$ denotes the evaluation at the end of the time step of the acceleration

$$\dot{\boldsymbol{\alpha}}^k_F(t,\mathbf{q},\mathbf{u},\mathbf{a}) = (\mathbf{W}^k_F)^{\mathrm{T}}(t,\mathbf{q})\,\mathbf{a} + \frac{\partial \boldsymbol{\alpha}^k_F}{\partial \mathbf{q}}(t,\mathbf{q},\mathbf{u})\,\mathbf{u} + \frac{\partial \boldsymbol{\gamma}^k_F}{\partial t}(t,\mathbf{q},\mathbf{u})\,.$$

In our case, the set of admissible friction forces is given by (6.46) and one can easily verify that

$$\boldsymbol{\lambda}^{\mathrm{sl}}_{Fk,i+1} = -\mu_k \lambda_{Nk,i+1} \frac{\boldsymbol{\alpha}^k_{F,i+1}}{||\boldsymbol{\alpha}^k_{F,i+1}||}\,.$$

In essence, the discrete friction law (8.39) corresponds to the evaluation of (8.33) at the end of the time step, with the difference that in the sticking case, the friction law is formulated on acceleration level. More details on this matter can be found in Section 10.4 of Glocker [48]. This section is concluded with the remark, that the formulation of the sticking case on acceleration level is strictly necessary, because otherwise the two discrete friction laws would lead to an ambiguity. Specifically, in the sticking case, (8.36) would imply $\boldsymbol{\xi}^k_{F,i+1} = 0$ and the evaluation of (8.33) at the end of the time step would imply $\boldsymbol{\alpha}^k_{F,i+1} = 0$, which are the same condition in the case of $e^k_F = 0$. Hence, in that case the two conditions collapse and create an ambiguity which does not allow to compute $\boldsymbol{\lambda}_{Fk,i+1}$ and $\boldsymbol{\Lambda}_{Fk,i+1}$ independently.

8.5. Extension to General Velocities

The kinematic equation (6.34) is generalized to

$$\mathrm{d}\mathbf{q} = \dot{\mathbf{q}}(t,\mathbf{q},\mathbf{u})\,\mathrm{d}t\,, \quad \text{where} \quad \dot{\mathbf{q}}(t,\mathbf{q},\mathbf{u}) = \mathbf{B}(t,\mathbf{q})\mathbf{u} + \boldsymbol{\beta}(t,\mathbf{q}) \tag{8.40}$$

with $\mathbf{u}(t) \in \mathbb{R}^m$ and $\mathbf{B}(t,\mathbf{q}(t)) \in \mathbb{R}^{n\times m}$. Such a generalization is for example needed in rigid body dynamics, when the orientations of the bodies are described by unit quaternions and the components of the angular velocities with respect to a body fixed frame are chosen as velocity parameters of the system. Another prominent example is the use of minimal coordinates and minimal velocities for a nonholonomic system, where typically $m < n$.

It follows immediately from (8.40) that

$$\mathrm{d}\dot{\mathbf{q}} = \mathbf{B}(t,\mathbf{q})\mathrm{d}\mathbf{u} + \frac{\partial \dot{\mathbf{q}}(t,\mathbf{q},\mathbf{u})}{\partial \mathbf{q}}\mathrm{d}\mathbf{q} + \frac{\partial \dot{\mathbf{q}}(t,\mathbf{q},\mathbf{u})}{\partial t}\mathrm{d}t\,.$$

Gathering the densities with respect to $\mathrm{d}t$ after using (6.35) and (8.40) allows to rewrite the differential measure of $\dot{\mathbf{q}}$ as

$$\mathrm{d}\dot{\mathbf{q}} = \ddot{\mathbf{q}}(t, \mathbf{q}, \mathbf{u}, \mathbf{a})\mathrm{d}t + \mathbf{B}(t, \mathbf{q})(\mathbf{u}^+ - \mathbf{u}^-)\mathrm{d}\eta\,,$$

where we have introduced the function

$$\ddot{\mathbf{q}}(t, \mathbf{q}, \mathbf{u}, \mathbf{a}) = \mathbf{B}(t, \mathbf{q})\mathbf{a} + \frac{\partial \dot{\mathbf{q}}(t, \mathbf{q}, \mathbf{u})}{\partial \mathbf{q}}\dot{\mathbf{q}}(t, \mathbf{q}, \mathbf{u}) + \frac{\partial \dot{\mathbf{q}}(t, \mathbf{q}, \mathbf{u})}{\partial t}\,.$$

It is straightforward to see, that after introducing the stabilization as in (8.5), the position and velocity measures result as

$$\begin{aligned}\mathrm{d}\mathbf{q} &= \dot{\mathbf{q}}(t, \mathbf{q}, \mathbf{u} + \mathbf{u}_S)\,\mathrm{d}t\\ \mathrm{d}\dot{\mathbf{q}} &= \ddot{\mathbf{q}}(t, \mathbf{q}, \mathbf{u}, \mathbf{a} + \mathbf{a}_S)\mathrm{d}t + \mathbf{B}(t, \mathbf{q})(\mathbf{u}^+ - \mathbf{u}^-)\mathrm{d}\eta\,.\end{aligned}$$

By a similar reasoning as in Section 8.2, the corresponding position update formula is retrieved as

$$\mathbf{q}_{i+1} = \mathbf{q}_i + \Delta t\,\dot{\mathbf{q}}(t_i, \mathbf{q}_i, \mathbf{u}_i) + \frac{\Delta t^2}{2}\ddot{\mathbf{q}}\big(t_i, \mathbf{q}_i, \mathbf{u}_i, (1-2\beta)\bar{\mathbf{a}}_i + 2\beta\bar{\mathbf{a}}_{i+1}\big) + \mathbf{B}(t_i, \mathbf{q}_i)\mathbf{Q}_{i+1}$$

and generalizes the update formula (8.13).

8.6. Numerical Implementation

This section sums up the nonsmooth generalized-α method derived in the previous sections and describes how every time step of the scheme can be solved numerically. To do so, it is used that the scheme in every time step constitutes a normal cone inclusion problem, which can be solved using the ideas presented in Appendix A.

As initial conditions of the stepping scheme, the values of the kinematic quantities $\mathbf{q}_0$, $\mathbf{u}_0$, $\mathbf{a}_0$, $\mathbf{U}_0$ and $\mathbf{Q}_0$ as well as all discrete forces at the initial time t_0 are assumed to be known. The initial conditions must be compatible in the sense that they solve the discrete equations of motion as well as satisfy all constraints at the initial time t_0. As initial value for the auxiliary variables the first order approximation[2] $\bar{\mathbf{a}}_0 = \mathbf{a}_0$ and $\bar{\boldsymbol{\lambda}}_{\square,0} = \boldsymbol{\lambda}_{\square,0}$ is chosen.

Assuming that all quantities are known at the beginning of the time step, the presented scheme is formulated as a system of nonlinear equations $\mathbf{R}(\mathbf{x}) = 0$, where

$$\begin{aligned}\mathbf{x} = \big(&\mathbf{a}_{i+1},\, \mathbf{U}_{i+1},\, \mathbf{Q}_{i+1},\, \boldsymbol{\kappa}_{g,i+1},\, \boldsymbol{\Lambda}_{g,i+1},\, \boldsymbol{\lambda}_{g,i+1},\, \boldsymbol{\Lambda}_{\alpha,i+1},\, \boldsymbol{\lambda}_{\alpha,i+1},\\ &\boldsymbol{\kappa}_{N,i+1},\, \boldsymbol{\Lambda}_{N,i+1},\, \boldsymbol{\lambda}_{N,i+1},\, \boldsymbol{\Lambda}_{F,i+1},\, \boldsymbol{\lambda}_{F,i+1}\big)\,.\end{aligned} \tag{8.41}$$

[2] A more accurate initialization can be found on page 351 in [18].

The nonlinear equations are then solved by a semi-smooth Newton method. The computed $\mathbf{x}$ can subsequently be used to find the remaining quantities at the end of the time step. Hence, the positions, velocities and percussions at t_{i+1} are regarded as dependent on $\mathbf{x}$. In fact, one can solve (8.12) for $\bar{\mathbf{a}}_{i+1}$ and use it to compute $\mathbf{q}_{i+1}$ and $\mathbf{u}_{i+1}$ from (8.13) or its generalized counterparts of Section 8.5. Similarly, the auxiliary contact forces can be computed from (8.26) and (8.38), respectively. In turn, these can be directly inserted into (8.32), (8.25) and (8.37) to retrieve the respective values of $\hat{\boldsymbol{\kappa}}_{N,i+1}$, $\mathbf{P}_{N,i+1}$ and $\mathbf{P}_{F,i+1}$.

Having in mind that the just mentioned quantities depend on $\mathbf{x}$, the first part of the residual $\mathbf{R}$ is stated as

$$\mathbf{R}_{\mathrm{s}} = \begin{pmatrix} \mathbf{M}_{i+1}\,\mathbf{a}_{i+1} - \mathbf{h}_{i+1} - \sum_{X\in\{g,\alpha,N,F\}} \mathbf{W}_{X,i+1}\,\boldsymbol{\lambda}_{X,i+1} \\ \mathbf{M}_{i+1}\mathbf{U}_{i+1} - \sum_{X\in\{g,\alpha,N,F\}} \mathbf{W}_{X,i+1}\,\boldsymbol{\Lambda}_{X,i+1} \\ \mathbf{M}_{i+1}\mathbf{Q}_{i+1} - \sum_{K\in\{g,N\}} \mathbf{W}_{K,i+1}\boldsymbol{\kappa}_{K,i+1} - \sum_{R\in\{\alpha,F\}} \frac{\Delta t}{2}\mathbf{W}_{R,i+1}\boldsymbol{\Lambda}_{R,i+1} \\ \boldsymbol{g}(t_{i+1},\mathbf{q}_{i+1}) \\ \dot{\boldsymbol{g}}(t_{i+1},\mathbf{q}_{i+1},\mathbf{u}_{i+1}) \\ \ddot{\boldsymbol{g}}(t_{i+1},\mathbf{q}_{i+1},\mathbf{u}_{i+1},\mathbf{a}_{i+1}) \\ \boldsymbol{\alpha}(t_{i+1},\mathbf{q}_{i+1},\mathbf{u}_{i+1}) \\ \dot{\boldsymbol{\alpha}}(t_{i+1},\mathbf{q}_{i+1},\mathbf{u}_{i+1},\mathbf{a}_{i+1}) \end{pmatrix},$$

where the residual $\mathbf{R} = (\mathbf{R}_{\mathrm{s}}, \mathbf{R}_{\mathrm{c}})$ is split into a part $\mathbf{R}_{\mathrm{s}}$ containing all equations except the discrete contact laws, which are contemplated in $\mathbf{R}_{\mathrm{c}}$. Hereby, the subscript "s" is chosen for $\mathbf{R}_{\mathrm{s}}$ to indicate the smooth part of the residual. In order to state the remaining part of the residual, the normal cone inclusions arising in the discrete contact laws have to be formulated as equations. This is done using the results from Appendix A and gives a piecewise smooth residual $\mathbf{R}_{\mathrm{c}} = (\mathbf{R}_{\kappa_N}, \mathbf{R}_{\Lambda_N}, \mathbf{R}_{\lambda_N}, \mathbf{R}_{\Lambda_F}, \mathbf{R}_{\lambda_F})$, which is subsequently set up part by part.

Starting with the normal contact law on position level (8.31), relation (A.5) is used to restate the law as

$$\hat{\kappa}_{Nk,i+1} = -\mathrm{prox}_{\mathbb{R}_0^-}\left(r g^k_{N,i+1} - \hat{\kappa}_{Nk,i+1}\right) \quad k = 1,\dots,n_N\,. \tag{8.42}$$

Equivalently, by (A.8) the discrete contact law (8.31) can be restated as the residual

$$\begin{aligned} k \in \mathcal{A}_{i+1}: &\quad R^k_{\kappa_N} := g^k_{N,i+1} = 0 \\ k \in \bar{\mathcal{A}}_{i+1}: &\quad R^k_{\kappa_N} := \hat{\kappa}_{Nk,i+1} = 0\,, \end{aligned} \tag{8.43}$$

where the k-th component $R^k_{\kappa_N}$ of $\mathbf{R}_{\kappa_N}$ is implicitly defined and where the set

$$\mathcal{A}_{i+1} = \left\{k = 1,\dots,n_N \;\middle|\; r g^k_{N,i+1} - \hat{\kappa}_{Nk,i+1} \le 0\right\}$$

together with its complement $\bar{\mathcal{A}}_{i+1} = \{1,\dots,n_N\} \setminus \mathcal{A}_{i+1}$ are introduced. Since for contacts in $\mathcal{A}_{i+1}$ the gap must be closed at the end of the time step, see

(8.43), $\mathcal{A}_{i+1}$ contains the same contacts as A_{i+1} because due to the stabilization no contact is penetrated at the end of the time step. We can use this fact to state the residual for the contact law on velocity level.

Again, (A.8) is used to formulate the normal contact law (8.24) as

$$\begin{aligned} k \in \mathcal{A}_{i+1}: &\quad P_{Nk,i+1} = -\mathrm{prox}_{\mathbb{R}_0^-}\left(r\xi_{N,i+1}^k - P_{Nk,i+1}\right) \\ k \in \bar{\mathcal{A}}_{i+1}: &\quad P_{Nk,i+1} = 0\,. \end{aligned}$$

Similar as before, the k-th component $R_{\Lambda_N}^k$ of $\mathbf{R}_{\Lambda_N}$ follows from (A.8) as

$$\begin{aligned} k \in \mathcal{B}_{i+1}: &\quad R_{\Lambda_N}^k := \xi_{N,i+1}^k = 0 \\ k \in \bar{\mathcal{B}}_{i+1}: &\quad R_{\Lambda_N}^k := P_{Nk,i+1} = 0\,, \end{aligned} \tag{8.44}$$

where we have introduced the set

$$\mathcal{B}_{i+1} = \left\{k \in \mathcal{A}_{i+1} \;\middle|\; r\xi_{N,i+1}^k - P_{Nk,i+1} \leq 0\right\}$$

as well as its complement $\bar{\mathcal{B}}_{i+1} = \{1,\dots,n_N\} \setminus \mathcal{B}_{i+1}$. Since, $\mathcal{B}_{i+1}$ is the subset of closed contacts $\mathcal{A}_{i+1}$ for which by (8.44) $\xi_{N,i+1}^k = 0$, the set $\mathcal{B}_{i+1}$ and B_{i+1} contain the same contacts. The reasoning here is similar as on position level.

Finally, it can proceeded in the same manner with the acceleration level constraint (8.18) and state it as

$$\begin{aligned} k \in \mathcal{B}_{i+1}: &\quad \lambda_{Nk,i+1} = -\mathrm{prox}_{\mathbb{R}_0^-}\left(r\ddot{g}_{N,i+1}^k - \lambda_{Nk,i+1}\right) \\ k \in \bar{\mathcal{B}}_{i+1}: &\quad \lambda_{Nk,i+1} = 0\,. \end{aligned}$$

With this intermediate step, the k-th component $R_{\lambda_N}^k$ of $\mathbf{R}_{\lambda_N}$ is then implicitly defined by

$$\begin{aligned} k \in \mathcal{C}_{i+1}: &\quad R_{\lambda_N}^k := \ddot{g}_{N,i+1}^k = 0 \\ k \in \bar{\mathcal{C}}_{i+1}: &\quad R_{\lambda_N}^k := \lambda_{Nk,i+1} = 0\,, \end{aligned} \tag{8.45}$$

where we have introduced the set

$$\mathcal{C}_{i+1} = \left\{k \in \mathcal{B}_{i+1} \;\middle|\; r\ddot{g}_{N,i+1}^k - \lambda_{Nk,i+1} \leq 0\right\}$$

as well as its complement $\bar{\mathcal{C}}_{i+1} = \{1,\dots,n_N\} \setminus \mathcal{C}_{i+1}$. Again, due to the stabilization and (8.45), the sets $\mathcal{C}_{i+1}$ and C_{i+1} contain the same contacts along the discrete motion.

For the discrete friction laws, the procedure is the same. The inclusion (8.36) is reformulated as

$$\begin{aligned} k \in \mathcal{A}_{i+1}: &\quad \mathbf{P}_{Fk,i+1} = -\mathrm{prox}_{C_F(P_{Nk,i+1})}\left(r\boldsymbol{\xi}_{F,i+1}^k - \mathbf{P}_{Fk,i+1}\right) \\ k \in \bar{\mathcal{A}}_{i+1}: &\quad \mathbf{P}_{Fk,i+1} = 0\,, \end{aligned} \tag{8.46}$$

where (A.5) has been used. Using (A.10) allows to implicitly define the k-th contribution of the residual $\mathbf{R}_{\Lambda_F} = \left(\mathbf{R}^1_{\Lambda_F}, \dots, \mathbf{R}^{n_N}_{\Lambda_F}\right)$ as

$$\begin{aligned}
k \in \mathcal{D}^{\mathrm{st}}_{i+1}: \quad & \mathbf{R}^k_{\Lambda_F} := \boldsymbol{\xi}^k_{F,i+1} = 0 \\
k \in \mathcal{D}^{\mathrm{sl}}_{i+1}: \quad & \mathbf{R}^k_{\Lambda_F} := \mathbf{P}_{Fk,i+1} + \mu_k P_{Nk,i+1} \frac{\boldsymbol{\xi}^k_{F,i+1}}{||\boldsymbol{\xi}^k_{F,i+1}||} = 0 \\
k \in \bar{\mathcal{A}}_{i+1}: \quad & \mathbf{R}^k_{\Lambda_F} := \mathbf{P}_{Fk,i+1} = 0\,,
\end{aligned} \tag{8.47}$$

where we have introduced the sets

$$\mathcal{D}^{\mathrm{st}}_{i+1} = \left\{ k \in \mathcal{A}_{i+1} \;\middle|\; r\boldsymbol{\xi}^k_{F,i+1} - \mathbf{P}_{Fk,i+1} \in C_F(P_{Nk,i+1}) \right\}$$

and $\mathcal{D}^{\mathrm{sl}}_{i+1} = \mathcal{A}_{i+1} \setminus \mathcal{D}^{\mathrm{st}}_{i+1}$. Again it is clear from (8.47) that the set $\mathcal{D}^{\mathrm{st}}_{i+1}$ corresponds the set of sticking contacts D^{st}_{i+1} and consequently $\mathcal{D}^{\mathrm{sl}}_{i+1}$ to the set of slipping contacts D^{sl}_{i+1}. This, together with (A.5), allows to rewrite (8.39) as

$$\begin{aligned}
k \in \mathcal{D}^{\mathrm{st}}_{i+1}: \quad & \boldsymbol{\lambda}_{Fk,i+1} = -\mathrm{prox}_{C_F(\lambda_{Nk,i+1})}\left(r\dot{\boldsymbol{\alpha}}^k_{F,i+1} - \boldsymbol{\lambda}_{Fk,i+1}\right) \\
k \in \mathcal{D}^{\mathrm{sl}}_{i+1}: \quad & \boldsymbol{\lambda}_{Fk,i+1} = -\mu_k \lambda_{Nk,i+1} \frac{\boldsymbol{\alpha}^k_{F,i+1}}{||\boldsymbol{\alpha}^k_{F,i+1}||} \\
k \in \bar{\mathcal{A}}_{i+1}: \quad & \boldsymbol{\lambda}_{Fk,i+1} = 0\,.
\end{aligned} \tag{8.48}$$

The sticking contacts which are also sticking on acceleration level are gathered by

$$\mathcal{E}^{\mathrm{st}}_{i+1} = \left\{ k \in \mathcal{D}^{\mathrm{st}}_{i+1} \;\middle|\; r\dot{\boldsymbol{\alpha}}^k_{F,i+1} - \boldsymbol{\lambda}_{Fk,i+1} \in C_F(\lambda_{Nk,i+1}) \right\}.$$

Moreover, let $\mathcal{E}^{\mathrm{sl}}_{i+1} = \mathcal{D}^{\mathrm{st}}_{i+1} \setminus \mathcal{E}^{\mathrm{st}}_{i+1}$. Consequently, (A.10) allows to implicitly define the k-th contribution of the residual $\mathbf{R}_{\lambda_F} = \left(\mathbf{R}^1_{\lambda_F}, \dots, \mathbf{R}^{n_N}_{\lambda_F}\right)$ by reformulating (8.48) as

$$\begin{aligned}
k \in \mathcal{E}^{\mathrm{st}}_{i+1}: \quad & \mathbf{R}^k_{\lambda_F} := \dot{\boldsymbol{\alpha}}^k_{F,i+1} = 0 \\
k \in \mathcal{E}^{\mathrm{sl}}_{i+1}: \quad & \mathbf{R}^k_{\lambda_F} := \boldsymbol{\lambda}_{Fk,i+1} + \mu_k \lambda_{Nk,i+1} \frac{\dot{\boldsymbol{\alpha}}^k_{F,i+1}}{||\dot{\boldsymbol{\alpha}}^k_{F,i+1}||} = 0 \\
k \in \mathcal{D}^{\mathrm{sl}}_{i+1}: \quad & \mathbf{R}^k_{\lambda_F} := \boldsymbol{\lambda}_{Fk,i+1} + \mu_k \lambda_{Nk,i+1} \frac{\boldsymbol{\alpha}^k_{F,i+1}}{||\boldsymbol{\alpha}^k_{F,i+1}||} = 0 \\
k \in \bar{\mathcal{A}}_{i+1}: \quad & \mathbf{R}^k_{\lambda_F} := \boldsymbol{\lambda}_{Fk,i+1} = 0\,.
\end{aligned}$$

Now that the residual $\mathbf{R}$ has been defined, it is evident that it is piecewise smooth, i.e., it is smooth (differentiable) in $\mathbf{x}$ if the index sets $\mathcal{A}_{i+1}, \mathcal{B}_{i+1}, \dots$ do not change in the vicinity of $\mathbf{x}$. This enables us to use a semi-smooth Newton method to solve $\mathbf{R}(\mathbf{x}) = 0$. Hence, a time step of the scheme can be summarized as follows:

Time step with semi-smooth Newton method:

(i) As starting value ($\nu = 0$) use the vector $\mathbf{x}^0$ constructed like (8.41) but by choosing the known values at the beginning of the time step.

(ii) While $||\mathbf{R}(\mathbf{x}^\nu)||_\infty \leq \mathrm{TOL_n}$ and $\nu \leq \mathrm{MAXITER_n}$ do the Newton update

$$\mathbf{x}^{\nu+1} = \mathbf{x}^\nu - \mathbf{J}_\mathbf{R}(\mathbf{x}^\nu)^{-1}\mathbf{R}(\mathbf{x}^\nu) \tag{8.49}$$

and increase ν by one. Hereby, $\mathbf{J}_\mathbf{R}(\mathbf{x}^\nu)$ denotes the Jacobian matrix of $\mathbf{R}$ evaluated at $\mathbf{x}^\nu$, where the index sets $\mathcal{A}_{i+1}, \mathcal{B}_{i+1}, \ldots$ are held constant (equal to the sets arising in the computation of $\mathbf{R}(\mathbf{x}^\nu)$) while taking the partial derivatives.

(iii) Solve (8.12) for $\bar{\mathbf{a}}_{i+1}$ and use it to compute $\mathbf{q}_{i+1}$ and $\mathbf{u}_{i+1}$ from (8.13) or its generalized counterparts of Section 8.5. Similarly, compute the auxiliary contact forces from (8.26) and (8.38), respectively and insert these into (8.32), (8.25) and (8.37) leading to $\hat{\boldsymbol{\kappa}}_{N,i+1}$, $\mathbf{P}_{N,i+1}$ and $\mathbf{P}_{F,i+1}$.

For the sake of completeness, remark that step 3) is only needed for output purposes and is basically the first part of the computation of $\mathbf{R}$ in the subsequent time step. Moreover, it is clear from (8.49) that the Jacobian matrix $\mathbf{J}_\mathbf{R}$ must have full rank.

For mechanical systems without friction, the here presented scheme is very similar to the generalized-α method presented in Brüls et al. [18]. The only difference lies in the residuals describing the normal contact law on position and velocity level. Specifically, in [18] the authors use $\kappa_{Nk,i+1} = 0$ instead of $\hat{\kappa}_{Nk,i+1} = 0$ in (8.43) as well as $\Lambda_{Nk,i+1} = 0$ instead of $P_{Nk,i+1} = 0$ in (8.44). Even though by our experience, this small difference produces if at all minimal differences in the motion and the forces of the simulated system, it is more than just a subtlety. In fact, replacing the contact force quantities in (8.43) and (8.44) destroys the equivalence of the contact law formulated using the sets $\mathcal{A}_{i+1}$ and $\mathcal{B}_{i+1}$ with the corresponding discrete contact laws formulated as normal cone inclusions.

In cases where multiple contacts with linearly dependent generalized force directions are present in the system, the Jacobian matrix $\mathbf{J}_\mathbf{R}$ is singular. Hence, for this kind of system the equation $\mathbf{R}(\mathbf{x}) = 0$ must be solved by a method that does not invert this Jacobian matrix. A popular choice is to reformulate the system $\mathbf{R}(\mathbf{x}) = 0$ such that it can be solved by fixed point iterations, see Appendix A. For the aforementioned reformulation, the vector of unknowns is split as $\mathbf{x} = (\mathbf{y}, \mathbf{z})$, where

$$\begin{aligned} \mathbf{y} &= (\mathbf{a}_{i+1}, \mathbf{U}_{i+1}, \mathbf{Q}_{i+1}, \boldsymbol{\kappa}_{g,i+1}, \boldsymbol{\Lambda}_{g,i+1}, \boldsymbol{\lambda}_{g,i+1}, \boldsymbol{\Lambda}_{\alpha,i+1}, \boldsymbol{\lambda}_{\alpha,i+1}) \\ \mathbf{z} &= (\boldsymbol{\kappa}_{N,i+1}, \boldsymbol{\Lambda}_{N,i+1}, \boldsymbol{\lambda}_{N,i+1}, \boldsymbol{\Lambda}_{F,i+1}, \boldsymbol{\lambda}_{F,i+1}) \,. \end{aligned} \tag{8.50}$$

Since $\mathbf{y}$ does not contain any contact forces, $\frac{\partial \mathbf{R}_\mathrm{s}}{\partial \mathbf{y}}$ can be regular despite the presence of linearly dependent contact force directions. Assuming that this is the case, the implicit function theorem applied to $\mathbf{R}_\mathrm{s}(\mathbf{y}, \mathbf{z}) = 0$ assures the existence of a function $\mathbf{F}$ such that $\mathbf{y} = \mathbf{F}(\mathbf{z})$, i.e., the tuple $\mathbf{y}$ can be regarded as depending on the contact forces at the end of the time step. Moreover, it can be seen from the derivation of the residual $\mathbf{R}_\mathrm{c}$ above, that it can be equivalently formulated as $\mathbf{z} = \mathbf{p}(\mathbf{y}, \mathbf{z})$, where $\mathbf{p}$ makes use of the proximal point function. Indeed, recognize for example that (8.43) is equivalent to (8.42) or that (8.47) is equivalent to (8.46). This leaves us with an equation $\mathbf{z} = \mathbf{p}(\mathbf{F}(\mathbf{z}), \mathbf{z})$, which can be solved using fixed point iterations.

Numerically, the value $\mathbf{y} = \mathbf{F}(\mathbf{z})$ can be found by solving $\mathbf{R}_\mathrm{s}(\mathbf{y}, \mathbf{z}) = 0$ for a fixed value of $\mathbf{z}$ using Newton's method. The computed value $\mathbf{y}$ in turn is then used in the fixed point iteration. Therefore, a time step of the scheme can be implemented as follows:

Time step with fixed point iterations:

(i) As starting value ($\nu = \mu = 0$) use the vectors $\mathbf{y}^0$ and $\mathbf{z}^0$ constructed like (8.50) but by choosing the known values at the beginning of the time step.

(ii) While $||\mathbf{z}^\mu - \mathbf{p}(\mathbf{y}^0, \mathbf{z}^\mu)||_\infty \leq \mathrm{TOL}_\mathrm{fp}$ and $\mu \leq \mathrm{MAXITER}_\mathrm{fp}$ do the fixed point update:

  (i) While $||\mathbf{R}_\mathrm{s}(\mathbf{y}^\nu, \mathbf{z}^\mu)||_\infty \leq \mathrm{TOL}_\mathrm{n}$ and $\nu \leq \mathrm{MAXITER}_\mathrm{n}$, perform a Newton step

$$\mathbf{y}^{\nu+1} = \mathbf{y}^\nu - \mathbf{J}_{\mathbf{R}_\mathrm{s}}(\mathbf{y}^\nu, \mathbf{z}^\mu)^{-1} \mathbf{R}_\mathrm{s}(\mathbf{y}^\nu, \mathbf{z}^\mu)$$

  and increase ν by one. Hereby, $\mathbf{J}_{\mathbf{R}_\mathrm{s}}(\mathbf{y}^\nu, \mathbf{z}^\mu)$ denotes the Jacobian matrix of $\mathbf{R}_\mathrm{s}$ with respect to its $\mathbf{y}$ dependence.

  (ii) Use the converged solution $\mathbf{y}^\nu$ of step (i) to perform the fixed point update

$$\mathbf{z}^{\mu+1} = \mathbf{p}(\mathbf{y}^\nu, \mathbf{z}^\mu)\,.$$

  Subsequently increase μ by one and set $\nu = 0$ as well as $\mathbf{y}^0 = \mathbf{y}^\nu$.

(iii) Solve (8.12) for $\bar{\mathbf{a}}_{i+1}$ and use it to compute $\mathbf{q}_{i+1}$ and $\mathbf{u}_{i+1}$ from (8.13) or its generalized counterparts of Section 8.5. Similarly, compute the auxiliary contact forces from (8.26) and (8.38), respectively and insert these into (8.32), (8.25) and (8.37) leading to $\hat{\boldsymbol{\kappa}}_{N,i+1}$, $\mathbf{P}_{N,i+1}$ and $\mathbf{P}_{F,i+1}$.

Last but not least, it has to be mentioned that the parameter r used in the prox-equations as well as in the index sets $\mathcal{A}_{i+1}, \mathcal{B}_{i+1}, \ldots$ could be chosen differently for every contact and even differently for the normal contact law and the friction law of the same contact.

8.7. Examples

In this section, the presented nonsmooth generalized-α scheme is used to obtain the time evolution of some benchmark systems, which are all chosen such that particular features of the scheme can be validated separately.

rotating bouncing ball

Following [47, 102], consider a homogeneous rigid sphere of radius $R = 0.1$ and mass $m = 1$ which is constrained to move in the $(\mathbf{e}_x^I\text{-}\mathbf{e}_y^I)$-plane and which under the influence of gravity with gravitational acceleration $g = 9.81$ falls on a horizontal plane, see Figure 8.1. To parametrize the motion of the ball, the minimal coordinates $\mathbf{q} = (x, y, \varphi)$ are chosen, where the center of mass S of the sphere is addressed by the representation ${}_I\mathbf{r}_{OS} = (x, y, 0)$ of the position vector $\mathbf{r}_{OS}$ with respect to the basis I. Moreover, the angle φ describes the orientation of the sphere. The velocity parameters $\mathbf{u} = (u_x, u_y, u_\varphi)$ are chosen to correspond to $\dot{\mathbf{q}}$ whenever the time derivative of the coordinates exist. Consequently, the mass matrix and the force vector are

$$\mathbf{M} = \begin{pmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & \theta_S \end{pmatrix} \quad \text{and} \quad \mathbf{h} = \begin{pmatrix} 0 \\ -mg \\ 0 \end{pmatrix} \tag{8.51}$$

with the rotational inertia $\theta_S = \frac{2}{5}mR^2$ of the sphere. Finally, the contact with the plane is described by the gap function

$$g_N = y - R \quad \text{and} \quad \alpha_F = u_x + Ru_\varphi \,,$$

which is the relative horizontal velocity of the contact point with respect to the plane and with (6.46) describes planar Coulomb friction. The friction coefficient is $\mu = 0.2$ and the restitution coefficient $e_F = 0$.

To validate the presented scheme, three simulations[3] are instructive. For all of them $\mathbf{q}(0) = (0, 1, 0)$ is set as initial configuration and $\mathbf{u}(0) = (0, 0, \omega)$ is used, such that the ball has an initial rotational velocity ω.

The first case starts from rest, that is $\omega = 0$, and $e_N = 0.5$ is chosen. This results in the typical bouncing motion, which exhibits the Zeno phenomenon. The simulation result is shown in Figure 8.1 and asserts that the proposed scheme can overcome accumulation points.

The subsequent two cases are used to test the behavior of the scheme with respect to friction forces. For both cases, $e_N = 0$ implying that once the contact closes it remains closed, i.e., the post-impact velocity is $u_y^+ = 0$, which allows

[3] Simulation parameters: $r = 0.3$; $\Delta t = 2 \cdot 10^{-3}$; $\rho_\infty = 0.5$; $\mathrm{TOL}_\mathrm{n} = 10^{-6}$.

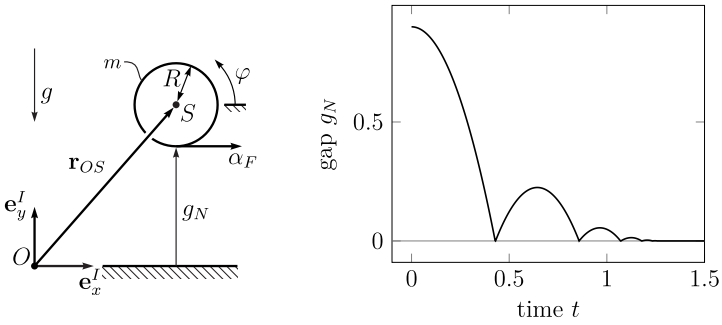


Figure 8.1: Sketch of the bouncing ball system (left) and simulated time evolution for the case where $\omega = 0$ and $e_N = 0.5$ (right).

us to validate friction. At the closing time instant a frictional impact occurs. Moreover, because the ball is constantly accelerated by gravitation, the pre-impact velocity is $u_y^- = -\sqrt{2g(y(0) - R)}$. It then follows from the impact equations (6.50), that the impulsive normal contact force is $\Lambda_N = m\sqrt{2g(y(0) - R)} \approx 4.2$. Furthermore, after the impact the nonimpulsive normal contact force compensates the gravitational force and therefore takes the value $\lambda_N = mg = 9.81$. Depending on the value of ω, two cases arise.

In the first case, the rotational velocity is high ($\omega = 50$) and the contact slides after the impact, implying that the friction forces attain the maximally allowed values $\Lambda_F = \mu\Lambda_N \approx 0.84$ and $\lambda_F = \mu\lambda_N \approx 1.96$. After a period of sliding contact, finally, the ball has slowed down enough such that a slip-stick transition takes place and the ball begins a pure rolling motion described by the kinematic condition $\gamma_F = 0$. Since the rolling motion is described by constant velocities, no net forces occur, implying $\lambda_F = 0$. Hence, at the slip-stick transition the nonimpulsive friction force instantly jumps to zero. Figure 8.2 shows that the described behavior of the friction forces is perfectly reproduced by the presented generalized-α scheme.

In the second case, when the rotational velocity is small enough, the contact sticks at impact and the ball exhibits a rolling motion directly after. As no friction force is needed for rolling, $\lambda_F = 0$ for all times. The pre-impact velocities are $u_x^- = 0$ as well as $u_\varphi^- = \omega$. Since the contact sticks directly after the impact, i.e., $\gamma_F^+ = 0$, from the impact equations (6.50) it follows that the impulsive friction force takes the value $\Lambda_F = -\frac{2}{7}mR\omega \approx -0.29$. Again, this behavior is in perfect accordance with the simulation, see Figure 8.2.

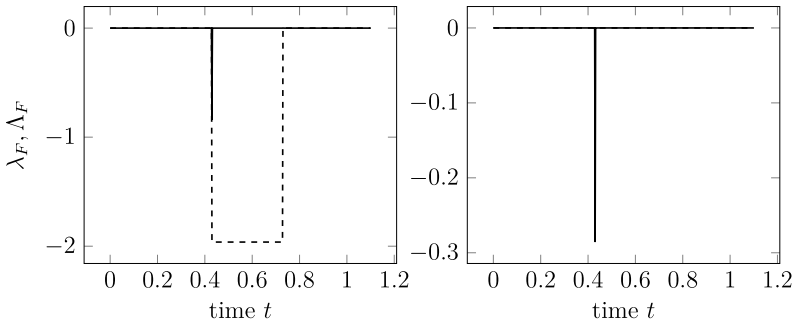


Figure 8.2: Simulated friction forces λ_F (dashed) and Λ_F (solid) for the cases $\omega = 50$ (left) and $\omega = 10$ (right).

BALL IN CORNER

A ball making frictional contact with a corner is a simple case where the Newton step (8.49) fails due to the singularity of the Jacobian matrix $\mathbf{J_R}$. Consider the ball described in Section 8.7 and assume it can get in contact with two inclined planes with inclination angles $\alpha = 45°$ and $\beta = 45°$, respectively, see Figure 8.3. The gap functions and the friction velocities are

$$\boldsymbol{g}_N = \begin{pmatrix} -x\sin\alpha + y\cos\alpha - R \\ x\sin\beta + y\cos\beta - R \end{pmatrix} \quad \text{and} \quad \boldsymbol{\alpha}_F = \begin{pmatrix} u_x\cos\alpha + u_y\sin\alpha + ru_\varphi \\ u_x\cos\beta - u_y\sin\beta + ru_\varphi \end{pmatrix}.$$

For the simulation[4] shown in Figure 8.3, the restitution coefficients were chosen

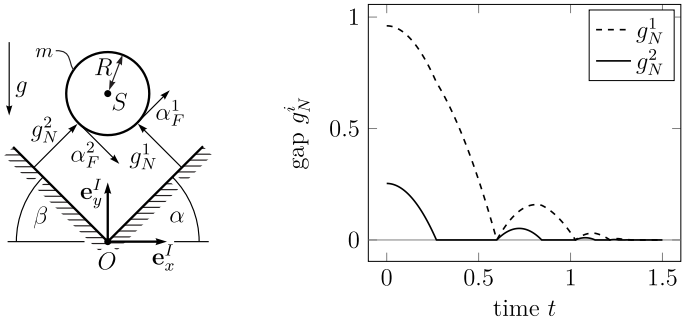


Figure 8.3: Sketch of the ball in corner system (left) and simulated time evolution of the gaps (right).

to be $e_N^1 = 0.5$, $e_N^2 = 0$ and $e_F^1 = e_F^2 = 0$ and the friction coefficients were set $\mu_1 = \mu_2 = 0.3$. Starting at rest with $\mathbf{q}(0) = (-0.5,\, 1,\, 0)$, the ball will eventually come to rest with both contacts closed. It is exactly in that situation that the Jacobian matrix $\mathbf{J_R}$ becomes singular and the simulation can only be continued with fixed point iterations as solution strategy. In Figure 8.3, this is the case for $t \approx 1.36$.

BALL IN CYLINDER

The importance of the stabilization of the unilateral constraint at position level can impressively be shown by simulating the ball of Section 8.7 rolling inside a cylinder of radius $R_\mathrm{c} = 1$. The centerline of the cylinder is assumed to be orthogonal to the $\mathbf{e}_x^I$-$\mathbf{e}_y^I$-plane, such that it can be identified with the point P, see Figure 8.4. The normal contact with the cylinder is described by the gap

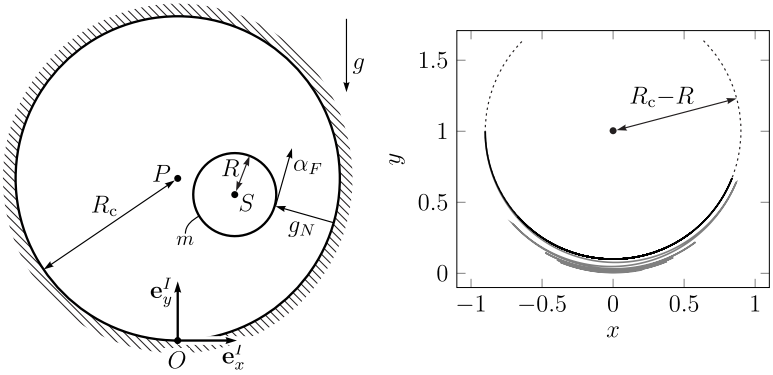


Figure 8.4: Sketch of the ball in cylinder system (left) and simulated time evolution of S (right). black: generalized-α scheme, gray: Moreau's time-stepping scheme.

function

$$g_N = R_\mathrm{c} - R - ||\mathbf{r}_{SP}||$$

with $e_N = 0$. Using ${}_I\mathbf{n} = (n_x,\, n_y,\, 0)$ to denote the components of the inward normal $\mathbf{n} = \mathbf{r}_{SP}/||\mathbf{r}_{SP}||$ of the cylinder, the tangent unit vector ${}_I\mathbf{t} = (n_y,\, -n_x,\, 0)$ is introduced. With that, the tangent velocity

$$\alpha_F = {}_I\mathbf{v}_S^\mathrm{T}\, {}_I\mathbf{t} + R u_\varphi$$

is defined, where the velocity of S is ${}_I\mathbf{v}_S = (u_x,\, u_y,\, 0)$. The friction coefficient $\mu = 0.1$ and the restitution coefficient $e_F = 0$ are chosen. Looking at the

[4] Simulation parameters: $r = 0.2$; $\Delta t = 10^{-4}$; $\rho_\infty = 0.5$; $\mathrm{TOL_n} = \mathrm{TOL_{fp}} = 10^{-6}$.

trajectory of S simulated[5] with the generalized-α scheme shows that the contact does not penetrate. To show that contact penetration is a big issue for this system, the trajectory of S resulting from the generalized-α scheme is compared to the trajectory gotten from a simulation[6] with Moreau's time-stepping scheme, which does not stabilize the unilateral constraint on position level, Chapter 7.[7] The comparison is shown in Figure 8.4. For the simulation the ball was assumed to start from rest with $\mathbf{q}(0) = (-0.9,\, 1,\, 0)$.

PAINLEVÉ ROD

It is well known that during the sliding motion of rigid bodies over a rough surface, a "frictional dynamic catastrophe" can occur, that is, impulsive motion which is not the result of a collision. This phenomenon, called Painlevé paradox, is for example studied in [46, 47, 78], where the following benchmark system is used. Consider a rigid homogeneous slender rod of mass $m = 1$, length $2l$

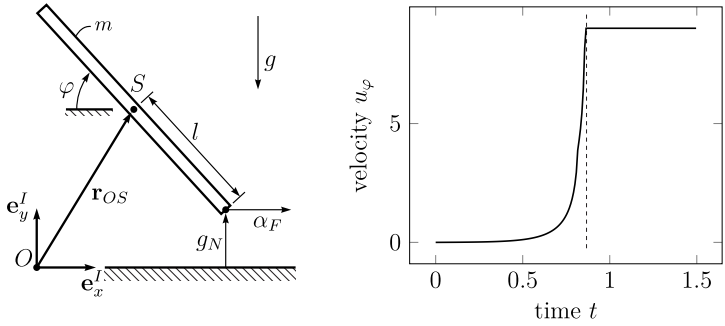


Figure 8.5: Sketch of the Painlevé rod (left) and simulated time evolution of the angular velocity of the rod (right).

and rotational inertia $\theta_S = \frac{1}{3}ml^2$, where $l = 1$. As shown in Figure 8.5, the rod moves in the $\mathbf{e}_x^I$-$\mathbf{e}_y^I$-plane and is under the influence of gravity with gravitational acceleration $g = 10$. Describing the orientation of the rod by the angle φ, the minimal coordinates $\mathbf{q} = (x,\, y,\, \varphi)$ are chosen, where the center of mass S of the sphere is addressed by ${}_I\mathbf{r}_{OS} = (x,\, y,\, 0)$. Using the natural velocity coordinates $\mathbf{u}$ corresponding to $\dot{\mathbf{q}}$ almost everywhere, the mass matrix and the force vector of the system have the form (8.51). The contact of the rod's tip with the ground is

[5] Simulation parameters: $r = 0.3$; $\Delta t = 10^{-2}$; $\rho_\infty = 0.5$; $\mathrm{TOL_n} = 10^{-6}$.

[6] Parameters for Moreau's time-stepping scheme: $r = 0.3$; $\Delta t = 10^{-2}$; $\mathrm{TOL_{fp}} = 10^{-6}$.

[7] Because the vector $\mathbf{h}$ is constant, Moreau's time-stepping scheme and the symmetric Moreau-type scheme of Chapter 7 coincide.

described by

$$g_N = y - l\sin\varphi \quad \text{and} \quad \alpha_F = u_x - l u_\varphi \sin\varphi\,.$$

The friction coefficient is set to $\mu = 5/3$ and $e_N = e_F = 0$ is used to model inelastic impact behavior. With the initial conditions $\mathbf{q}(0) = (0,\, l\sin(\varphi_0),\, \varphi_0)$ and $\mathbf{u}(0) = (v,\, 0,\, 0)$, the rod's tip is initially in contact and the rod slides with an initial inclination of $\varphi_0 = 31^\circ$ and an initial horizontal velocity $v = 30$. As analyzed in [47], this sliding motion eventually results in detaching of the contact. This detaching comes with a blowup of the accelerations and nonimpulsive contact forces, which is seen as an impact. The blowup of the accelerations leads to vertical asymptotes in the velocities, which constitute problems for integration schemes. However, since the asymptote is just an isolated singularity, event-capturing schemes with constant time step might overcome these blowups, as they do not try to fully resolve it. The simulation[8] result in Figure 8.5 shows that the presented generalize-α scheme can indeed deal with the Painlevé paradox.

guided hopper

The suitability of the presented nonsmooth generalized-α scheme for the simulation of flexible multibody systems is demonstrated by simulating a guided hopper. It consists of a vertically guided main body of mass $M = 3$, which is addressed by the coordinate y. At the hip H with ${}_I\mathbf{r}_{OH} = (x_0,\, y,\, 0)$ a rigid homogeneous rod of mass $m = 1.56$, length $L = 0.2$ and rotational inertia $\theta_S = 8.5 \cdot 10^{-3}$ around the rod's center of mass is attached to the main body. Herein, x_0 is the arbitrary horizontal position of the guidance. The orientation of the rod is prescribed by the angle $\alpha(t)$ which is a given function of time. A straight planar Euler–Bernoulli beam [39] with undeformed length L is connected to the knee K of the rod by an actuated rotational joint with prescribed actuation angle $\beta(t)$. For the linear elastic beam the axial stiffness $EA = 1.89 \cdot 10^7$, the bending stiffness $EI = 14.2$ and the mass line density $\varrho = 0.71$ are chosen. Moreover, following [58], the centerline of the beam is discretized with B-Spline shape functions. Using the parameter $\xi \in [0, 1]$ to parametrize the beam and denoting the generalized coordinates of the beam by $\mathbf{q}_\mathrm{b}$, a point C on the centerline of the beam is addressed by $\mathbf{r}_{OC}(\xi, \mathbf{q}_\mathrm{b})$. The generalized coordinates $\mathbf{q} = (y,\, \mathbf{q}_\mathrm{b})$ of the hopper as well as the auxiliary quantities $\mathbf{d} = \mathbf{r}_{OK} - \mathbf{r}_{OC}(0, \mathbf{q}_\mathrm{b})$ and $\varphi = \measuredangle(\mathbf{r}_{KH}, \mathbf{r}'_{OC}(0, \mathbf{q}_\mathrm{b}))$ are introduced, where $(\cdot)'$ denotes the derivative with respect to ξ. With these quantities, the bilateral constraints composing the knee joint are

$$\boldsymbol{g} = \begin{pmatrix} \mathbf{e}_x^I \cdot \mathbf{d} \\ \mathbf{e}_y^I \cdot \mathbf{d} \\ \varphi - \beta \end{pmatrix}\,.$$

[8] Simulation parameters: $r = 0.1$; $\Delta t = 8 \cdot 10^{-4}$; $\rho_\infty = 0.9$; $\mathrm{TOL}_\mathrm{n} = 10^{-8}$.

Finally, the contact of the endpoint P of the beam with the horizontal plane is described by

$$g_N = \mathbf{e}_y^I \cdot \mathbf{r}_{OC}(1, \mathbf{q}_\mathrm{b}) \quad \text{and} \quad \alpha_F = \mathbf{e}_x^I \cdot \mathbf{v}_C(1, \mathbf{q}_\mathrm{b}, \dot{\mathbf{q}}_\mathrm{b}) \,,$$

where $\mathbf{v}_C$ is the velocity of the centerline and corresponds to the time derivative of $\mathbf{r}_{OC}$ almost everywhere. The contact parameters are set to $e_N = e_F = 0$ and $\mu = 0.2$.

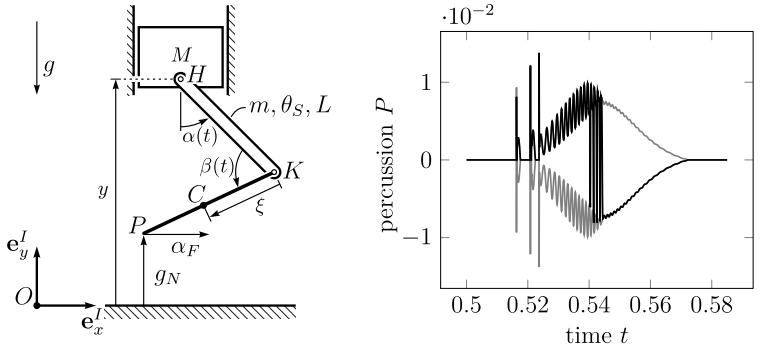


Figure 8.6: Sketch of the guided hopper (left) and simulated time evolution of the percussions (right). black: P_F, gray: $\pm\mu P_N$.

For the simulation,[9] the system is initially assumed to be at rest with $y(0) = 0.31$. Moreover, the beam is undeformed initially. For the actuation angles

$$\alpha(t) = \frac{\pi}{3} - \frac{\pi}{30}\big(1 - \cos(4\pi\, t)\big) \quad \text{and} \quad \beta(t) = \pi - 2\alpha(t)$$

have been chosen and gravity is contemplated by the gravitational acceleration $g = 9.81$. The percussions, plotted in Figure 8.6, show that the generalized-α scheme can cope with the complex contact dynamics arising in multibody systems containing flexible parts and time dependent bilateral constraints. This makes the presented scheme well suited for engineering applications.

TIPPETOP

The tippetop consists of a spherical main body at which a stick is attached. Starting form the standing position with the stick pointing upwards and with

[9] The generalized-α scheme with the following parameters was used: $r = 0.15$; $\Delta t = 5{\cdot}10^{-5}$; $\rho_\infty = 0$; $\mathrm{TOL_n} = 10^{-6}$. The beam was discretized by two elements of polynomial degree 2 and 5 Gauss quadrature points were used.

high spinning velocity around its symmetry axis, the top inverts and spins on the stick. Since the spinning velocity decreases due to dissipation, the top tumbles back to the standing position after a while.

Let $R_1 = 1.5 \cdot 10^{-2}$ be the radius of the sphere with center C_1 characterizing the main body of the top. The rounded end of the stick is described by a sphere of radius $R_2 = 5 \cdot 10^{-3}$ and midpoint C_2. The two midpoints as well as the top's center of mass S lie on the axis of symmetry of the top and the distances between S and the points C_1 and C_2 are $a_1 = 3 \cdot 10^{-3}$ and $a_2 = 1.6 \cdot 10^{-2}$, respectively, see Figure 8.7. The position of the top is described by the components ${}_I\mathbf{r}_{OS}$ of the position vector of S with respect to the resting basis I. To characterize the orientation of the top, we introduce the body fixed K-frame such that $\mathbf{e}_z^K$ lies on the symmetry axis of the top and points towards the stick. The transformation matrix $\mathbf{A}_{IK} = ({}_I\mathbf{e}_x^K \ {}_I\mathbf{e}_y^K \ {}_I\mathbf{e}_z^K)$ is then parametrized using a unit quaternion $\mathbf{p}$. Hence, the configuration of the top is described by $\mathbf{q} = ({}_I\mathbf{r}_{OS}, \mathbf{p})$. The generalized velocity $\mathbf{u} = ({}_I\mathbf{v}_S, {}_K\boldsymbol{\Omega})$ is composed by the representations of the velocity $\mathbf{v}_S$ of S and the angular velocity of the top $\boldsymbol{\Omega}$ with respect to the bases I and K, respectively. As mentioned in Section 8.5, this choice leads to a model with the generalized kinematic equation (8.40). The relevant quantities $\mathbf{B}$, $\boldsymbol{\beta}$, $\mathbf{M}$ and $\mathbf{h}$ describing a so parameterized rigid body under the influence of gravity is found in [109].

The mass of the top is $m = 6 \cdot 10^{-3}$ and the representation of the inertia tensor is the diagonal matrix ${}_K\boldsymbol{\Theta}_S = \mathrm{diag}(I_1, I_1, I_3)$ with $I_1 = 8 \cdot 10^{-7}$ and $I_3 = 7 \cdot 10^{-7}$.

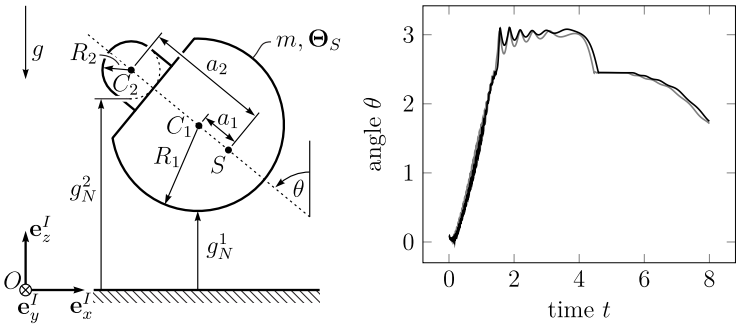


Figure 8.7: Sketch of the tippetop (left) and simulated time evolution of the angle θ (right). Black: presented model, gray: Möller et al. [88].

To describe the contact between the ground and either the main body ($i = 1$) or the stick ($i = 2$), the gap

$$g_N^i = \mathbf{e}_z^I \cdot \mathbf{r}_{OC_i} - R_i$$

is used. The friction between the top and the ground is modeled as Coulomb–Contensou friction [75], which uses the friction velocity

$$\boldsymbol{\gamma}_F^i = \begin{pmatrix} \mathbf{e}_x^I \cdot \mathbf{v}_{P_i} \\ \mathbf{e}_y^I \cdot \mathbf{v}_{P_i} \\ \frac{3\pi R}{16} \mathbf{e}_z^I \cdot \boldsymbol{\Omega} \end{pmatrix}, \tag{8.52}$$

where $\mathbf{v}_{P_i} = \mathbf{v}_{C_i} - R_i \boldsymbol{\Omega} \times \mathbf{e}_z^I$ denotes the velocity of the respective contact point and $R = 5 \cdot 10^{-4}$ denotes the assumed contact radius. For both contacts, $e_N = e_F = 0$ and $\mu = 0.3$ are used. It is shown by Möller et al. [88] that using the friction velocity (8.52) together with the set of admissible friction forces (6.46) results in a good approximation for the Coulomb–Contensou friction law used for the tippetop simulation in Leine and Glocker [75].

For the simulation,[10] the top is initially at the position ${}_I\mathbf{r}_{OS}(0) = (0,\ 0,\ z_0)$ with $z_0 = 1.2015 \cdot 10^{-2}$ and is inclined by the angle $\theta(0) = 0.1$, where θ denotes the angle between $\mathbf{e}_z^I$ and $\mathbf{e}_z^K$, that is, $\mathbf{e}_z^I \cdot \mathbf{e}_z^K = \cos\theta$. The center of mass S is assumed to be at rest while the top spins with initial angular velocity ${}_K\boldsymbol{\Omega} = (0,\ 0,\ 180)$. To assure that the quaternion $\mathbf{p}$ remains a unit-quaternion, $\mathbf{p}$ is normalized after every step.

It is apparent from Figure 8.7 that the simulation results using the presented scheme are in line with the results of Möller et al. [88][11]. This shows that the presented scheme is well suited for mechanical systems with spatial friction as well as models with a general kinematic equation (8.40).

[10] Simulation parameters: $r = 0.001$; $\Delta t = 1 \cdot 10^{-3}$; $\rho_\infty = 0.5$; $\mathrm{TOL_n} = 10^{-6}$.

[11] Parameters for Moreau's time-stepping scheme: $r = 0.001$; $\Delta t = 10^{-4}$; $\mathrm{TOL_{fp}} = 10^{-6}$.

— 9 —

Conclusions and Outlook

In this monograph, the mathematical description and the simulation of mechanical systems with frictional contact have been discussed. Thereby, the complete line of thought which underlies the mechanical modeling process has been addressed, i.e., the development of a geometric mechanical theory, its application for the description of a specific system and the temporal discretization of the dynamical equations needed for the simulation of the system. Furthermore, bridging the gap between theoretical and engineering mechanics has been a particular aim of this thesis. To that end, the vector space formalism of engineering mechanics has been retrieved as a local representation of the geometric theory. Changing the perspective, it can be stated that the presented mechanical theory provides a mathematical-theoretical superstructure to engineering mechanics.

The presented mechanical theory is founded on the concept of the Galilean manifold, which is used as a model for both the physical and the configuration space-time. Based on the ideas of Loos [82], the mathematical foundations of the theory have been elaborated in Chapter 4. In particular, the interrelation between nonlinear connections, action forms and the coordinate independent description of non-autonomous second-order ordinary differential equations on Galilean manifolds has been worked out in detail. Moreover, the similarities and differences between the geometry of Galilean and Riemannian manifolds, employed for the coordinate independent description of autonomous second-order ordinary differential equations, has been discussed.

With the aim of bringing the presented theory closer to the classical mechanics literature, the geometric theory is based on three postulates, which in essence correspond to the laws of motion given by Newton. This is in contrast to the approaches of Loos [82] and Winandy [127], who base their theory on the postulation of the action form of the mechanical system. Furthermore, the presented theory extends the work of Winandy [127] as it not only treats the dynamics of the system on the configuration space-time, but also links the system's dynamics to the motions of the particles constituting the system.

It has been a main goal of this thesis to bring engineering and theoretical mechanics closer together by embedding the framework of engineering mechanics in a geometric theory for finite-dimensional mechanical systems. The main benefit of having such a theoretical superstructure is that the geometric theory rigorously distinguishes between physical quantities by assigning distinct mathematical objects to them. In doing so, the physical properties of the quantities are encoded into the mathematics, which allows to precisely ascertain the role of different physical concepts in mechanics. In certain cases, this can help to relieve some obscurity which traditionally adheres to physical concepts such as for example inertia forces. In that spirit, some aspects of the presented theory have to be discussed in more detail, e.g.,

- The presented theory makes no use of the concept of inertial frame. A proper definition of inertial frame has to be given. Moreover, the interrelation between the inertial frames and the resting field have to be studied.
- In continuum mechanics, see for example [63], the velocity and acceleration are often defined with respect to an observer, such that with the change of observer also the notions of velocity and acceleration change. This has farreaching implications, particularly when it comes to the definition of objective (observer independent) rates of deformation. In this dissertation, the velocity and acceleration are linked to a reference field and are thus independent of the observer. It has to be inspected if the presented distinction between reference field and observer can be of benefit and how it relates to the other notions of observer found in mechanics.

Another benefit of the geometric theory is that it allows to study the mechanical concepts with the tools of differential geometry. In Section 5.10 for example, the condition for bilateral constraints to be holonomic is given by Frobenius' theorem, which is a purely mathematical result. However, since the geometric theory comprises the formalism of engineering mechanics, the conditions given by Frobenius' theorem could be translated for the use in engineering mechanics. Specifically, the two equivalent conditions (4.6) and (4.9), which are formulated in vector-matrix notation, have been derived as local representations of the geometric conditions provided by Frobenius' theorem. In that spirit, it is important to pursue the work of retrieving classical results within the geometric framework such that they can be studied with the machinery of differential geometry and the results can be transferred to the engineering world. In particular, the following research tasks are identified:

- In classical mechanics, results such as the Noether's theorem, Liouville's theorem or the Hamilton–Jacobi equation play an important role for the

study of mechanical systems which are only subjected to potential forces. Hence, within the presented geometric mechanical theory on Galilean manifolds, coordinate-free versions of these results have to be found.

- It is known in classical mechanics, that for a mechanical system which is subjected to ideal holonomic bilateral constraints, one can find minimal coordinates such that the system's dynamics has the form of a free system, i.e., the constraint forces do not appear in the minimal description. The coordinate-free statement of this circumstance is given by Proposition 5.3. Similarly, if the constraints are ideal but nonholonomic, it is known that one can find minimal velocity coordinates, and hence a reduced state-space, such that the constraint forces do not appear in the minimal description. The coordinate-free version of this statement has to be established and studied in detail.
- Several technically important mechanical systems have a configuration manifold, that is, submanifolds of simultaneous events of the configuration space-time, which may be identified with a Lie group. In that case, it has to be studied how the Lie group structure can be exploited to reformulate the principles of mechanics. For example, it has to be investigated if the results of Chapter 13 in Marsden and Ratiu [84] can be retrieved in the context of Galilean manifolds.

The numerical schemes derived in Chapters 7 and 8 are event-capturing schemes, which have been developed to alleviate some of the deficits of the most widespread schemes such as Moreau's time-stepping scheme. Specifically, these weaknesses are the first-order accuracy, the numerical contact penetration and the high level of numerical dissipation. In general, a better performance comes along with more computational costs. For instance, the presented generalized-α scheme is of second-order accuracy during impact free time intervals and exhibits no numerical contact penetration. However, the dimension of the normal cone inclusion problem that has to be solved at every time step is three times larger than the one of Moreau's time-stepping scheme. These considerations lead to the following research tasks:

- In Chapter 7, piecewise linear ansatz functions have been used to derive numerical schemes for mechanical systems with frictional contact. To arrive at higher-order schemes, ansatz functions which are piecewise polynomials of higher-order can be used. However, in that case, it is not immediately clear how to discretize the contact laws accordingly, which has to be investigated.
- To make the presented nonsmooth generalized-α scheme computationally more efficient, different strategies for the solution of the normal cone

inclusion problem, which constitutes a time step of the scheme, have to be evaluated.

- A thorough comparison between the presented generalized-α scheme and the schemes of [28, 43] has to be done. Special attention should be given to the effects resulting from the different formulations of the friction law, which here is included on velocity and acceleration level whereas [28, 43] rely on a formulation on position and velocity level.
- Since flexible multibody systems with large rotations are often studied in the Lie group setting, see [9], the presented nonsmooth generalized-α scheme has to be adapted for the time-integration of the dynamics of systems with frictional contact described on a Lie group.

The development of mechanics always profited from a symbiotic relation between engineering and science. Engineering often provides the fundamental equations and the working principles of the different branches of mechanics, which are then precisely stated within a mathematically rigorous framework. This process is particularly fruitful if it is iterative and if the insights of mathematics find their way back to engineering. It has been the aim of this dissertation to contribute in both directions. On the one hand, a geometric mechanical theory has been developed, which allowed to retrieve new insights for engineering, e.g., the conditions in vector-matrix notation that can be used to assert whether a set of bilateral constraints is holonomic. On the other hand, a novel nonsmooth generalized-α scheme for mechanical systems with frictional contact has been proposed and numerically evaluated. However, a rigorous convergence proof for the scheme is missing, which is a typical task of numerical mathematics rather than engineering.

— *A* —

Normal Cone Inclusion Problems

The presented schemes for mechanical systems with frictional contacts in every time step results in a normal cone inclusion problem, that is, a set of nonlinear equations stemming from the discretization of the dynamics together with normal cone inclusions, which are invoked by the discrete contact laws. Hence, finding a numerical solution to a time step of the presented scheme in essence corresponds to finding a numerical solution to

$$\mathbf{y} \in \mathcal{N}_C(-\mathbf{z}) \quad \text{with} \quad \mathbf{R}_{\text{s}}(\mathbf{y}, \mathbf{z}) = 0\,, \tag{A.1}$$

where $\mathbf{R}_{\text{s}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is an implicit relationship between $\mathbf{y}$ and $\mathbf{z}$. An extensive collection of solutions strategies for (A.1) can be found in Acary et al. [4]. In what follows, we limit ourselves to two strategies based on the reformulation of the normal cone inclusion as an equation including the proximal point function.

For a closed convex non-empty set $C \subseteq \mathbb{R}^n$, the proximal point function is defined by

$$\text{prox}_C : \mathbb{R}^n \to \mathbb{R}^n,\ \mathbf{p} \mapsto \mathbf{q} = \arg\min_{\mathbf{p}^* \in C} \left(\frac{1}{2} ||\mathbf{p} - \mathbf{p}^*||^2 \right), \tag{A.2}$$

which maps a point $\mathbf{p}$ to the closest point $\mathbf{q} \in C$, where the distance between the points is measured by the Euclidean norm $||\,.\,||$. It is immediately clear from (A.2) that $\text{prox}_C(\mathbf{p}) = \mathbf{p}$ if and only if $\mathbf{p} \in C$. Using the just defined proximal point function, it can be shown that two points $\mathbf{x}$ and $\mathbf{y}$ fulfill the normal cone inclusion

$$\mathbf{y} \in \mathcal{N}_C(\mathbf{x})\,, \tag{A.3}$$

if and only if they satisfy the equation

$$\mathbf{x} = \text{prox}_C(\mathbf{x} + r\mathbf{y}) \tag{A.4}$$

for any $r > 0$, see [76].

Using the equivalence of (A.3) and (A.4) allows to reformulate (A.1) as

$$\mathbf{z} = -\text{prox}_C(r\mathbf{y} - \mathbf{z}) \quad \text{with} \quad \mathbf{R}_\text{s}(\mathbf{y}, \mathbf{z}) = 0\,, \tag{A.5}$$

which reduces the problem of finding the solution of (A.1) to numerically finding the solution of a nonlinear equation.

The equation (A.5) can for example be used to define a residual $\mathbf{R}$=($\mathbf{R}_\text{s}$, $\mathbf{R}_\text{c}$), where

$$\mathbf{R}_\text{c}(\mathbf{x}) := \mathbf{z} + \text{prox}_C(r\mathbf{y} - \mathbf{z}) = 0 \tag{A.6}$$

with $\mathbf{x} = (\mathbf{y}, \mathbf{z})$. Due to the presence of prox_C, the residual $\mathbf{R}$ is continuous but nonsmooth, hence (A.6) can be solved using the semi-smooth (nonsmooth) Newton method as presented by [4] and [7], where in the Newton update

$$\mathbf{x}^{\nu+1} = \mathbf{x}^\nu - \mathbf{J}_\mathbf{R}(\mathbf{x}^\nu)^{-1}\mathbf{R}(\mathbf{x}^\nu)$$

any regular element $\mathbf{J}_\mathbf{R}(\mathbf{x}^\nu)$ of the generalized Jacobian $\partial\mathbf{R}(\mathbf{x}^\nu)$ can be used. The semi-smooth Newton method reduces to the Newton method whenever $\mathbf{R}$ is differentiable. In that case, $\mathbf{J}_\mathbf{R}(\mathbf{x}^\nu)$ is the Jacobian matrix of $\mathbf{R}$ at $\mathbf{x}^\nu$.

Solving (A.6) with the semi-smooth Newton method has two drawbacks if $\mathbf{J}_\mathbf{R}(\mathbf{x}^\nu)$ is computed by finite differences. Firstly, the accuracy of the Jacobian strongly depends on the parameter r. Moreover, using finite differences to compute $\mathbf{J}_\mathbf{R}(\mathbf{x}^\nu)$ at a point where $\mathbf{R}$ is not differentiable generally leads to an arbitrary element in $\partial\mathbf{R}(\mathbf{x}^\nu)$. Hence, we have no control over which element of $\partial\mathbf{R}(\mathbf{x}^\nu)$ is ultimately used for the Newton update.

As a remedy, we can use an active set formulation to treat (A.6), which relies on the specific knowledge of the set C. In our case, the two sets of interest are $\mathbb{R}_0^-$ and $C_F(R)$.

For $\mathbb{R}_0^-$, the proximal point function is piecewise given as

$$\begin{aligned} p \in \mathbb{R}_0^- :&\quad \text{prox}_{\mathbb{R}_0^-}(p) = p \\ p \notin \mathbb{R}_0^- :&\quad \text{prox}_{\mathbb{R}_0^-}(p) = 0\,. \end{aligned} \tag{A.7}$$

It follows immediately from using (A.7) in (A.6) that the residual in (A.6) is equivalent to

$$\begin{aligned} ry - z \in \mathbb{R}_0^- :&\quad R_\text{c} := y = 0 \\ ry - z \notin \mathbb{R}_0^- :&\quad R_\text{c} := z = 0\,, \end{aligned} \tag{A.8}$$

where we abstain from the bold notation as all variables are scalar. Proceeding all the same for the set $C_F(R)$, the proximal point function is piecewise given by

$$\begin{aligned} \mathbf{p} \in C_F(R) :&\quad \text{prox}_{C_F(R)}(\mathbf{p}) = \mathbf{p} \\ \mathbf{p} \notin C_F(R) :&\quad \text{prox}_{C_F(R)}(\mathbf{p}) = R\frac{\mathbf{p}}{||\mathbf{p}||}\,. \end{aligned} \tag{A.9}$$

Moreover, using (A.9) in (A.6) allows to reformulate the residual equivalently to

$$\begin{aligned} r\mathbf{y} - \mathbf{z} \in C_F(R): &\quad \mathbf{R}_{\mathrm{c}} := \mathbf{y} = 0 \\ r\mathbf{y} - \mathbf{z} \notin C_F(R): &\quad \mathbf{R}_{\mathrm{c}} := \mathbf{z} + R\tfrac{\mathbf{y}}{||\mathbf{y}||} = 0\,, \end{aligned} \tag{A.10}$$

where the fact that $\mathbf{z} + R\frac{r\mathbf{y}-\mathbf{z}}{||r\mathbf{y}-\mathbf{z}||} = 0$ is solved by $\mathbf{z} = -R\frac{\mathbf{y}}{||\mathbf{y}||}$ has been used. Furthermore, $r\mathbf{y} - \mathbf{z} \notin C_F(R)$ implies with $\mathbf{z} \in C_F(R)$ that $||\mathbf{y}|| > 0$. In these active set formulations of (A.6) the parameter r is only present in the activation condition of the residuals and does not affect the accuracy of the computation of the Jacobian by finite differences. Moreover, since the active set formulation gives direct access to the non-differentiable points of the residual, we can choose to not switch between the two pieces of the residual during the numerical differentiation process, which leads to a well-behaved Jacobian and a more robust scheme.

Another popular strategy to solve (A.5) is the use of fixed point iterations, see [7] or [120]. For that, the implicit function theorem on $\mathbf{R}_{\mathrm{s}}(\mathbf{y}, \mathbf{z}) = 0$ is used, which guarantees the existence of a function $\mathbf{F}$ such that $\mathbf{y} = \mathbf{F}(\mathbf{z})$. Hence, the fixed point iterations follow as

$$\mathbf{z}^{\nu+1} = -\mathrm{prox}_C(r\mathbf{F}(\mathbf{z}^\nu) - \mathbf{z}^\nu)\,,$$

where generally $\mathbf{F}$ is not known analytically and $\mathbf{y}^\nu = \mathbf{F}(\mathbf{z}^\nu)$ must be found numerically by solving $\mathbf{R}_{\mathrm{s}}(\mathbf{y}^\nu, \mathbf{z}^\nu) = 0$ for $\mathbf{y}^\nu$ while treating $\mathbf{z}^\nu$ as a constant.

Bibliography

[1] R. Abraham and J. E. Marsden. *Foundations of Mechanics*. Addison-Wesley, 2nd edition, 1987.

[2] V. Acary. Higher order event capturing time-stepping schemes for nonsmooth multibody systems with unilateral constraints and impacts. *Applied Numerical Mathematics*, 62:1259–1275, 2012.

[3] V. Acary. Projected event-capturing time-stepping schemes for nonsmooth mechanical systems with unilateral contact and Coulomb's friction. *Computer Methods in Applied Mechanics and Engineering*, 256:224–250, 2013.

[4] V. Acary, M. Brémond, and O. Huber. On solving contact problems with Coulomb friction: formulations and numerical comparisons. In *Advanced Topics in Nonsmooth Dynamics*, pages 375–457. Springer, 2018.

[5] U. Aeberhard. *Geometrische Behandlung idealer Stösse*. PhD thesis, ETH Zurich, 2008.

[6] U. Aeberhard and C. Glocker. Energy considerations for excited perfect collisions. In *Proceedings of the Fifth EUROMECH Nonlinear Oscillations Conference*, pages 422–431, 2005.

[7] P. Alart and A. Curnier. A mixed formulation for frictional contact problems prone to Newton like solution methods. *Computer Methods in Applied Mechanics and Engineering*, 92:353–375, 1991.

[8] M. Arnold and O. Brüls. Convergence of the generalized-α scheme for constrained mechanical systems. *Multibody System Dynamics*, 18(2):185–202, 2007.

[9] M. Arnold, O. Brüls, and A. Cardona. Error analysis of generalized-α Lie group time integration methods for constrained mechanical systems. *Numerische Mathematik*, 129(1):149–179, 2015.

[10] V. I. Arnold. *Mathematical Methods of Classical Mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer, 2nd edition, 1989.

[11] M. Audin. *Geometry*. Springer, 2002.

[12] X. Bekaert and K. Morand. Connections and dynamical trajectories in generalised Newton-Cartan gravity I. An intrinsic view. *Journal of Mathematical Physics*, 57 (2), 2016.

[13] D. Bertsimas and J. N. Tsitsiklis. *Introduction to linear optimization*, volume 6. Athena Scientific Belmont, 1997.

[14] A. M. Bloch. *Nonholonomic Mechanics and Control*, volume 24 of *Interdisciplinary Applied Mathematics*. Springer Science & Business Media, 2nd edition, 2015.

[15] H. Bremer. *Elastic Multibody Dynamics: A Direct Ritz Approach*, volume 35 of *Intelligent Systems, Control, and Automation: Science and Engineering*. Springer, 2008.

[16] B. Brogliato. *Nonsmooth Mechanics: Models, Dynamics and Control*. Springer, 1999.

[17] O. Brüls, V. Acary, and A. Cardona. 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, 2014.

[18] O. Brüls, V. Acary, and A. Cardona. On the constraints formulation in the nonsmooth generalized-α method. In *Advanced Topics in Nonsmooth Dynamics*, pages 335–374. Springer, 2018.

[19] F. Bullo and A. D. Lewis. *Geometric control of mechanical systems: modeling, analysis, and design for simple mechanical control systems*, volume 49. Springer Science & Business Media, 2004.

[20] G. Capobianco and S. R. Eugster. Time finite element based Moreau-type integrators. *International Journal for Numerical Methods in Engineering*, 114(3): 215–345, 2018.

[21] G. Capobianco, J. Harsch, S. R. Eugster, and R. I. Leine. A nonsmooth generalized-alpha method for mechanical systems with frictional contact. *International Journal for Numerical Methods in Engineering*, 2021.

[22] G. Capobianco, T. Winandy, and S. R. Eugster. The principle of virtual work and Hamilton's principle on Galilean manifolds. *Journal of Geometric Mechanics*, 2021.

[23] E. Cartan. *Leçons sur les invariants intégraux*. A. Hermann & fils, 1922.

[24] P. Chadwick. *Continuum Mechanics: Concise Theory and Problems*. Dover Books on Physics. Dover Publications, 1999.

[25] Q.-Z. Chen, V. Acary, G. Virlez, and O. Brüls. A nonsmooth generalized-α scheme for flexible multibody systems with unilateral constraints. *International Journal for Numerical Methods in Engineering*, 96(8):487–511, 2013.

[26] J. Chung and G. Hulbert. A time integration algorithm for structural dynamics with improved numerical dissipation: the generalized-α method. *Journal of applied mechanics*, 60(2):371–375, 1993.

[27] A. Cosimo, J. Galvez, F. J. Cavalieri, A. Cardona, and O. Brüls. A robust nonsmooth generalized-α scheme for flexible systems with impacts. *Multibody System Dynamics*, 48(2):127–149, 2020.

[28] A. Cosimo, F. J. Cavalieri, J. Galvez, A. Cardona, and O. Brüls. A general purpose formulation for nonsmooth dynamics with finite rotations: Application to the woodpecker toy. *Journal of Computational and Nonlinear Dynamics*, 16 (3), 2021.

[29] M. Crampin. A geometrical version of the Helmholtz conditions in timedependent Lagrangian dynamics. *Journal of Physics A: Mathematical and Theoretical*, 17 (7):1437–1447, 1984.

[30] M. de León and P. R. Rodrigues. *Methods of differential geometry in analytical mechanics*, volume 158. Elsevier, 1989.

[31] H. Dombrowski and K. Horneffer. Die Differentialgeometrie des Galileischen Relativitätsprinzips. *Mathematische Zeitschrift*, 86:291–311, 1964.

[32] H. D. Dombrowski and K. Horneffer. Das verallgemeinerte Galileische Relativitätsprinzip. *Nachrichten der Akademie der Wissenschaften in Göttingen: Mathematisch-Physikalische Klasse*, pages 233–246, 1964.

[33] C. Duval and H. P. Künzle. Sur les connexions Newtoniennes et l'extension non tivial du groupe de Galilée. *C. R. Acad. Sci. Paris - Série A*, 285:813–816, 1977.

[34] J. Ehlers. Über den Newtonschen Grenzwert der Einsteinschen Gravitationstheorie. In *Grundlagenprobleme der modernen Physik*, pages 65–84. Bibliographisches Institut, 1981.

[35] J. Ehlers. Republication of: On the Newtonian limit of Einstein's theory of gravitation. *General Relativity and Gravitation*, 51(12):1–20, 2019.

[36] Élie Cartan. Sur les variétés à connexion affine et la théorie de la relativité généralisée (première partie). *Annales scientifiques de l'École Normale Supérieure*, 40:325–412, 1923.

[37] Élie Cartan. Sur les variétés à connexion affine, et la théorie de la relativité généralisée (première partie) (suite). *Annales scientifiques de l'Ecole Normale Supérieure*, 41:1–25, 1924.

[38] S. R. Eugster. *Geometric Continuum Mechanics and Induced Beam Theories*, volume 75 of *Lecture Notes in Applied and Computational Mechanics*. Springer, 2015.

[39] S. R. Eugster and J. Harsch. *A Variational Formulation of Classical Nonlinear Beam Theories*, pages 95–121. Springer International Publishing, 2020.

[40] S. R. Eugster, G. Capobianco, and T. Winandy. Geometric description of time-dependent finite-dimensional mechanical systems. *Mathematics and Mechanics of Solids*, 25(11):2050–2075, 2020.

[41] K. Friedrichs. Eine invariante Formulierung des Newtonschen Gravitationsgesetzes und des Grenzüberganges vom Einsteinschen zum Newtonschen Gesetz. *Mathematische Annalen*, 98:566–575, 1928.

[42] F. Gallissot. Les formes extérieures en mécanique. *Annales de l'institut Fourier*, 4:145–297, 1952.

[43] J. Galvez, F. J. Cavalieri, A. Cosimo, O. Brüls, and A. Cardona. A nonsmooth frictional contact formulation for multibody system dynamics. *International Journal for Numerical Methods in Engineering*, 121(16):3584–3609, 2020.

[44] C. W. Gear, B. Leimkuhler, and G. K. Gupta. Automatic integration of Euler-Lagrange equations with constraints. *Journal of Computational and Applied Mathematics*, 12:77–90, 1985.

[45] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Prentice-Hall, 1963.

[46] F. Génot and B. Brogliato. New results on Painlevé paradoxes. *European Journal of Mechanics-A/Solids*, 18(4):653–677, 1999.

[47] C. Glocker. *Dynamik von Starrkörpersystemen mit Reibung und Stößen*. PhD thesis, TU München, 1995.

[48] C. Glocker. *Set-Valued Force Laws: Dynamics of Non-Smooth Systems*, volume 1 of *Lecture Notes in Applied Mechanics*. Springer, 2001.

[49] C. Glocker. The geometry of Newtonian impacts with global dissipation index for moving sets. In *Proc. of the Int. Conf. on Nonsmooth/Nonconvex Mechanics*, pages 283–290. Citeseer, 2002.

[50] C. Glocker. Simulation of hard contacts with friction: An iterative projection method. In *Proceedings of a Conference in Honor of Jurgen Scheurle*. Springer, 2013.

[51] C. Glocker. Energetic consistency conditions for standard impacts. part I: Newton-type inequality impact laws and Kane's example. *Multibody System Dynamics*, 29 (1):77–117, 2013.

[52] C. Godbillon. *Géométrie différentielle et mécanique analytique*. Collection Méthodes. Hermann, 1969.

[53] H. Goldstein. *Classical mechanics*. Addison-Wesley, 7th edition, 1965.

[54] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II*. Springer, 2nd edition, 2002.

[55] G. Hamel. Die Lagrange-Eulerschen Gleichungen der Mechanik. *Zeitschr. für Math. u. Phys.*, 50(1-2):1–57, 1904.

[56] G. Hamel. Die Axiome der Mechanik. In R. Grammel, editor, *Grundlagen der Mechanik: Mechanik der Punkte und starren Körper*, volume 5 of *Handbuch der Physik*. Springer, 1927.

[57] G. Hamel. *Theoretische Mechanik: Eine einheitliche Einführung in die gesamte Mechanik*, volume 57 of *Grundlehren der mathematischen Wissenschaften*. Springer, 1949.

[58] J. Harsch and S. R. Eugster. *Finite Element Analysis of Planar Nonlinear Classical Beam Theories*, pages 123–157. Springer International Publishing, Cham, 2020.

[59] W. Hauger, D. Gross, and W. Schnell. *Technische Mechanik 3. Kinetik*. Springer-Verlag GmbH, 2012.

[60] P. Haupt. *Continuum Mechanics and Theory of Materials*. Advanced Texts in Physics. Springer, 2013.

[61] P. Havas. Four-dimensional formulations of Newtonian mechanics and their relation to the special and the general theory of relativity. *Rev. Mod. Phys.*, 36:938–965, 1964.

[62] H. M. Hilber, T. J. R. Hughes, and R. L. Taylor. Improved numerical dissipation for time integration algorithms in structural dynamics. *Earthquake Engineering & Structural Dynamics*, 5(3):283–292, 1977.

[63] G. A. Holzapfel. *Nonlinear Solid Mechanics, A Continuum Approach for Engineering*. Wiley, 2001.

[64] M. Jean and J. J. Moreau. Unilaterality and dry friction in the dynamics of rigid body collections. In *Proc. Contact Mechanics International Symposium, Lausanne, Switzerland, October 7-9*, pages 31–48, 1992.

[65] J. Kleinert, B. Simeon, and M. Obermayr. An inexact interior point method for the large-scale simulation of granular material. *Computer Methods in Applied Mechanics and Engineering*, 278:567–598, 2014.

[66] H. P. Künzle. Galilei and Lorentz structures on space-time : comparison of the corresponding geometry and physics. *Annales de l'I.H.P. Physique théorique*, 17 (4):337–362, 1972.

[67] J.-L. Lagrange. Théorie de la libration de la lune, et des autres phénomènes que dépendent de la figure non sphérique de cette planète. *Nouveaux Mémoires de l'Académie royale des Sciences et Belles-Lettres de Berlin*, 1780.

[68] J.-L. Lagrange. *Mécanique Analytique*. Éditions Jacques Gabay, Sceaux, 1788.

[69] C. Lanczos. *The variational principles of mechanics*, volume 4 of *Mathematical Expositions*. University of Toronto Press, 1952.

[70] L. D. Landau and E. M. Lifshitz. *Mechanics*, volume 1 of *Course of Theoretical Physics*. Pergamon Press, 1969.

[71] S. Lang. *Undergraduate Algebra*. Undergraduate Texts in Mathematics. Springer, 3rd edition, 2005.

[72] B. Langerock. Conic distributions and accessible sets. *Journal of Dynamical and Control Systems*, 14(1):43–70, 2008.

[73] J. M. Lee. *Manifolds and differential geometry*, volume 107. American Mathematical Society, 2009.

[74] J. M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, 2nd edition, 2013.

[75] R. I. Leine and C. Glocker. A set-valued force law for spatial Coulomb–Contensou friction. *European Journal of Mechanics – A/Solids*, 22:193–216, 2003.

[76] R. I. Leine and H. Nijmeijer. *Dynamics and Bifurcations of Non-Smooth Mechanical Systems*, volume 18 of *Lecture Notes in Applied and Computational Mechanics*. Springer, 2004.

[77] R. I. Leine and N. van de Wouw. *Stability and Convergence of Mechanical Systems with Unilateral Constraints*, volume 36 of *Lecture Notes in Applied and Computational Mechanics*. Springer-Verlag, 2008.

[78] R. I. Leine, B. Brogliato, and H. Nijmeijer. Periodic motion and bifurcations induced by the Painlevé paradox. *European Journal of Mechanics – A/Solids*, 21 (5):869–896, 2002.

[79] R. I. Leine, U. Aeberhard, and C. Glocker. Hamilton's principle as variational inequality for mechanical systems with impact. *Journal of Nonlinear Science*, 19 (6):633–664, 2009.

[80] R. I. Leine, A. Schweizer, M. Christen, J. Glover, P. Bartelt, and W. Gerber. Simulation of rockfall trajectories with consideration of rock shape. *Multibody System Dynamics*, 32(2):241–271, 2014.

[81] P. Libermann and C.-M. Marle. *Symplectic Geometry and Analytical Mechanics*, volume 35 of *Mathematics and Its Applications*. Springer, 1987.

[82] O. Loos. Analytische Mechanik. Seminarausarbeitung, Institut für Mathematik, Universität Innsbruck, 1982.

[83] O. Loos. Automorphism groups of classical mechanical systems. *Monatshefte für Mathematik*, 100(4):277–292, 1985.

[84] J. E. Marsden and T. Ratiu. *Introduction to Mechanics and Symmetry*, volume 17 of *Texts in Applied Mathematics*. Springer, 2nd edition, 1999.

[85] J. M. Maruskin. *Dynamical Systems and Geometric Mechanics*. De Gruyter, 2nd edition, 2018.

[86] P. W. Michor. *Topics in Differential Geometry*, volume 93 of *Graduate Studies in Mathematics*. American Mathematical Society, 2008.

[87] M. Möller. *Consistent Integrators for Non-Smooth Dynamical Systems*. PhD thesis, ETH Zurich, 2011.

[88] M. Möller, R. I. Leine, and C. Glocker. An efficient approximation of orthotropic set-valued force laws of normal cone type. In *7th Euromech Solid Mechanics Conference, Lisbon*, pages 7–11, 2009.

[89] G. Morandi, C. Ferrario, G. Lo Vecchio, G. Marmo, and C. Rubano. The inverse problem in the calculus of variations and the geometry of the tangent bundle. *Physics Reports*, 188(3-4):147–284, 1990.

[90] J. J. Moreau. Dynamique de systèmes à liasons unilatérales avec frottement sec éventuel; essais numériques. Technical report, Laboratoire de Mécanique Générale des Milieux Continus, Université des Sciences et Téchniques du Languedoc, 1986.

[91] J. J. Moreau. Unilateral contact and dry friction in finite freedom dynamics. In J. J. Moreau and P. D. Panagiotopoulos, editors, *Non-Smooth Mechanics and Applications. CISM Courses and Lectures*, pages 1–82. Springer, 1988.

[92] J. J. Moreau. *Friction, Arching and Contact Dynamics*, chapter Numerical investigation of shear zones in granular materials, pages 233–247. World Scientific, 1997.

[93] R. M. Murray, Z. Li, and S. S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. Taylor & Francis, 1994.

[94] N. M. Newmark. A method of computation for structural dynamics. *Journal of the Engineering Mechanics Division*, 85(3):67–94, 1959.

[95] C. Nouguier-Lehon, B. Cambou, and E. Vincens. Influence of particle shape and angularity on the behaviour of granular materials: a numerical analysis. *International Journal for Numerical and Analytical Methods in Geomechanics*, 27:1207–1226, 2003.

[96] G. E. Nützi. *Non-Smooth Granular Rigid Body Dynamics with Applications to Chute Flows*. PhD thesis, ETH Zurich, 2016.

[97] R. W. Ogden. *Non-linear Elastic Deformations*. Dover Publications, 1997.

[98] W. M. Oliva. *Geometric Mechanics*, volume 1798 of *Lecture Notes in Mathematics*. Springer, 2002.

[99] O. M. O'Reilly. *Engineering Dynamics - A Primer*. Springer, 2nd edition, 2010.

[100] J. G. Papastavridis. *Analytical Mechanics: A Comprehensive Treatise on the Dynamics of Constrained Systems*. World Scientific, 2014.

[101] L. A. Pars. *A Treatise on Analytical Dynamics*. Heinemann, 1965.

[102] F. Pfeiffer and C. Glocker. Contacts in multibody systems. *Journal of Applied Mathematics and Mechanics*, 64(5):773 – 782, 2000.

[103] F. Pfeiffer and T. Schindler. *Einführung in die Dynamik*. Springer, 1989.

[104] M. F. Rañada. Time-dependent Lagrangian systems: A geometric approach to the theory of systems with constraints. *Journal of Mathematical Physics*, 35(2): 748–758, 1994.

[105] F. Radjai and S. Roux. Turbulentlike fluctuations in quasistatic flow of granular media. *Physical Review Letters*, 89(6):064302, 2002.

[106] R. T. Rockafellar and R.-B. Wets. *Variational Analysis*. Springer, 2009.

[107] R. M. Rosenberg. *Analytical Dynamics of Discrete Systems*. Mathematical Concepts and Methods in Science and Engineering. Springer, 1st edition, 1977.

[108] C. Rüede and N. Straumann. On Newton-Cartan cosmology. *Helv. Phys. Acta*, 70:318–335, 1997.

[109] S. Sailer, S. R. Eugster, and R. I. Leine. The tippedisk: a tippetop without rotational symmetry. *Regular and Chaotic Dynamics*, 25(6):553–580, 2020.

[110] F. Scheck. *Mechanics*. Advanced Texts in Physics. Springer, 2005.

[111] T. Schindler and V. Acary. Timestepping schemes for nonsmooth dynamics based on discontinuous Galerkin methods: definition and outlook. *Mathematics and Computers in Simulation*, 95:180–199, 2014.

[112] S. Schoeder, H. Ulbrich, and T. Schindler. Discussion of the Gear–Gupta–Leimkuhler method for impacting mechanical systems. *Multibody System Dynamics*, 31(4):477–495, 2014.

[113] L. Schwartz. *Théorie des distributions*. Hermann, Paris, 1966.

[114] L. Schwartz. *Mathematics for the Physical Sciences*. Hermann, 1966.

[115] A. A. Shabana. *Dynamics of Multibody Systems*. EngineeringPro collection. Cambridge University Press, 2005.

[116] T. S. Shores. *Applied Linear Algebra and Matrix Analysis*. Springer, 2nd edition, 2007.

[117] J.-M. Souriau. *Structure des Systèmes Dynamiques*. Dunod, 1970.

[118] J.-M. Souriau. *Structure of Dynamical Systems*, volume 149 of *Progress in Mathematics*. Springer, 1997.

[119] M. Spivak. *A Comprehensive Introduction to Differential Geometry*, volume 2. Publish or Perish, Houston, Texas, 3rd edition, 1999.

[120] C. Studer. *Numerics of Unilateral Contacts and Friction: Modeling and Numerical Time Integration in Non-smooth Dynamics*, volume 47 of *Lecture Notes in Applied and Computational Mechanics*. Springer, 2009.

[121] C. Studer, R. Leine, and C. Glocker. Step size adjustment and extrapolation for time-stepping schemes in non-smooth dynamics. *International Journal for Numerical Methods in Engineering*, 76(11):1747–1781, 2008.

[122] R. A. Toupin. World invariant kinematics. *Archive for Rational Mechanics and Analysis*, 1(1):181–211, 1957. ISSN 1432-0673.

[123] A. A. Transeth, R. I. Leine, C. Glocker, and K. Y. Pettersen. 3-D snake robot motion: Nonsmooth modeling, simulations, and experiments. *IEEE Transactions on Robotics*, 24(2):361–376, 2008.

[124] A. Trautman. Sur la théorie Newtonienne de la gravitation. *C. R. Acad. Sci. Paris*, 257:617–620, 1963.

[125] M. Trümper. Lagrangian mechanics and the geometry of configuration spacetime. *Annals of Physics*, 149(1):203 – 233, 1983.

[126] R. W. Weber. Hamiltonian systems with constraints and their meaning in mechanics. *Archive for Rational Mechanics and Analysis*, 91(4):309–335, 1986.

[127] T. Winandy. *Dynamics of finite-dimensional mechanical systems*. PhD thesis, University of Stuttgart, 2019.

[128] T. Winandy, M. Baumann, and R. I. Leine. Variational analysis of inequality impact laws for perfect unilateral constraints. In *Advanced Topics in Nonsmooth Dynamics*, pages 47–92. Springer, 2018.

[129] C. Woernle. *Mehrkörpersysteme: Eine Einführung in die Kinematik und Dynamik von Systemen starrer Körper*. Springer, 2011.