State observers for mechanical systems with unilateral constraints: a discretization-based approach and experimental analysis

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

Vorgelegt von
Pascal Valentin Preiswerk
aus Basel (CH)

Hauptberichter:	Prof. Dr. ir. habil. Remco I. Leine
Mitberichter:	Prof. Dr. ir. Nathan van de Wouw
Tag der mündlichen Prüfung:	14.09.2022

Institut für Nichtlineare Mechanik
der Universität Stuttgart

2022

Preface

The work presented in this thesis is the result of my time at the Institute for Nonlinear Mechanics of the University of Stuttgart. During this time, I benefited from many great people through discussions, suggestions, collaborations and sharing ideas.

I am sincerely grateful to my doctoral advisor Prof. Dr. Remco I. Leine for his guidance and support during my research, as well as for giving me the opportunity to pursue my interests at his institute. His ability to present complex matters in a comprehensible way has always been a great inspiration for me. I enjoyed our many discussions in front of the whiteboard from which I learned a lot.

I would also like to thank my co-referee Prof. Dr. Nathan van de Wouw for reviewing my thesis and for his helpful comments on control-related questions.

Moreover, I thank my colleagues at the Institute for Nonlinear Mechanics for many discussions, for always having a sympathetic ear and for sharing good times. I am especially grateful to all my former and present colleagues who participated in our seminars, taught me the principles of mathematics and studied new subjects together. Special thanks go to Simon Sailer for working on various side-projects with me and for reviewing a good part of this manuscript.

Further thanks go to all the students who did a thesis under my supervision.

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

Stuttgart, June 2022 *Pascal Preiswerk*

Contents

Abstract

This monograph deals with the state observer design for mechanical systems with unilateral constraints. After the mathematical modeling is discussed in detail, necessary tools, which extend Lyapunov stability theory, are provided. A state observer design approach is investigated, which, in contrast to most existing observer designs, does not assume that closed contacts are instantaneously detected through measurements. In particular, a time discretization-based method is analyzed, which allows to circumvent some of the main difficulties caused by discontinuous time evolutions. Moreover, various state observer designs are implemented and tested on an experimental setup consisting of an impact oscillator.

In mechanical systems with unilateral constraints, collisions between different bodies can lead to abrupt velocity changes, which are conveniently modeled as instantaneous velocity jumps. The resulting time evolution is therefore discontinuous and, in contrast to smooth systems, cannot be described by ordinary differential equations. Among various available formalisms for such systems, measure differential inclusions are particularly suitable for modeling mechanical systems that include both unilateral constraints and friction. However, due to the discontinuous behavior, it is difficult to generalize analysis and design methods for smooth nonlinear systems to such systems.

The state estimation problem consists of reconstructing current state variables of an observed system from measurements. A common approach to achieve this goal is to design a state observer, which is an auxiliary system that is unilaterally coupled to the observed system through measurements and is designed such that the observer's state converges to the true, searched-for state. An important tool in the design process are Lyapunov-type stability theorems, which allow to assess qualitative stability properties of solutions solely based on the dynamical model. Specific theorems that address partial stability are provided, which fit into the existing Lyapunov stability theory for non-smooth dynamical systems.

It is an important aspect for the observer design whether or not contact information, such as collision time instants, can be extracted from measurements. Most available state observer designs for impulsive systems assume explicit knowledge of this contact information, which allows to let the impacts of the observer and the observed system coincide and to exploit the maximal monotonicity property of the contact force laws. In this thesis, particular attention is paid to the case where this contact information is unknown. For a class of mechanical systems, it is proposed to attack the observer problem by transforming and approximating the original continuous-time system by a discrete linear complementarity system through the use of the Paoli-Schatzman time discretization scheme. From there, a deadbeat observer and sufficient observability conditions are derived. It is shown that the discrete adaptation of an existing passivity-based observer design for linear complementarity systems can be applied. A key point in using a time discretization is that the discretization acts as a regularization, i.e. impacts take place over multiple time steps. This makes it possible to render the estimation error dynamics asymptotically stable. Furthermore, the so-called peaking phenomenon appears as singularity within the time discretization approach.

Finally, selected state observer designs are implemented and tested on an experimental setup, which includes a single unilateral constraint. The advantages and disadvantages of the approaches are discussed.

Zusammenfassung

Diese Monographie behandelt den Entwurf von Zustandsbeobachtern für mechanische Systeme mit einseitigen Bindungen. Nachdem die mathematische Modellierung im Detail diskutiert wurde, werden notwendige Werkzeuge vorgestellt, welche die bestehende Lyapunov Stabilitätstheorie erweitern. Es wird eine Entwurfsmethode untersucht, welche im Gegensatz zu den meisten bestehenden Beobachtern nicht erfordert, dass geschlossene Kontakte instantan über Messungen detektiert werden können. Die analysierte Methode basiert auf einer Zeitdiskretisierung, welche es erlaubt, einige durch unstetige Bewegungen verursachte Schwierigkeiten zu umgehen. Schließlich werden verschiedene Zustandsbeobachter an einem Stoß-Oszillator als Versuchsaufbau implementiert und getestet.

In mechanischen Systemen mit einseitigen Bindungen können Kollisionen zwischen verschiedenen Körpern auftreten. Diese führen zu abrupten Geschwindigkeitsänderungen, welche als instantane Geschwindigkeitssprünge modelliert werden. Die zeitliche Entwicklung der Zustände ist deshalb unstetig und kann, im Gegensatz zu glatten Systemen, nicht durch Differentialgleichungen beschrieben werden. Aus verschiedenen Formalismen für solche Systeme haben sich Maßdifferentialinklusionen als besonders praktisch zur Beschreibung von mechanischen Systemen mit einseitigen Bindungen und Reibung erwiesen. Viele Analyse- und Auslegungsmethoden für glatte nichtlineare Systeme lassen sich jedoch nur schwer für Systeme mit unstetigem Verhalten generalisieren.

Das Problem der Zustandsschätzung bezieht sich darauf, aus Messdaten aktuelle Zustände eines beobachteten Systems zu rekonstruieren. Ein verbreiteter Ansatz um dies zu erreichen ist die Verwendung eines Zustandsbeobachters. Dieser besteht aus einem virtuellen System, das über Messungen einseitig an das beobachtete System gekoppelt ist und so entworfen wird, dass sein Zustand zum unbekannten, gesuchten Zustand konvergiert. Ein wichtiges Werkzeug beim Beobachterentwurf sind Lyapunov Stabilitätstheoreme, welche einen Rückschluss auf die qualitativen Stabilitätseigenschaften zulassen, ohne explizite Lösungen des Systems zu kennen. Spezifische Theoreme zur partiellen Stabilität werden vorgestellt, welche sich in die

bestehende Lyapunov Stabilitätstheorie für nicht-glatte dynamische Systeme einfügen lassen.

Ein wichtiger Aspekt des Beobachterentwurfs ist es, ob Kontaktinformationen, wie z.B. die Zeitpunkte von Kollisionen, instantan aus Messungen extrahiert werden können. Die meisten bestehenden Zustandsbeobachter für impulsive Systeme erfordern die explizite Kenntis von Kontaktinformationen, so dass man Stöße im Beobachter und im beobachteten System zu den gleichen Zeitpunkten auftreten lassen kann. Dies erlaubt es in einem zweiten Schritt die Maximalmonotonie der Kontaktkraftgesetze auszunutzen. In dieser Arbeit wird speziell auf den Fall eingegangen, dass solche Kontaktinformationen nicht verfügbar sind. Für eine Klasse von mechanischen Systemen wird vorgeschlagen, das Beobachterproblem durch die Überführung des zeitkontinuierlichen Systems auf ein diskretes lineares Komplementaritätssystem mit Hilfe der Paoli-Schatzman Diskretisierung anzugehen. Dadurch lassen sich ein Totschlag-Beobachter und hinreichende Beobachtbarkeitskriterien herleiten. Es wird gezeigt, dass die diskrete Anpassung eines existierenden passivitätsbasierten Beobachters für lineare Komplementaritätssysteme angewendet werden kann. Ein wichtiger Vorteil der Zeitdiskretisierung ist, dass die Diskretisierung als Regularisierung wirkt, d.h. Stöße werden auf mehrere diskrete Zeitpunkte verteilt. Dies ermöglicht es, die Dynamik des Schätzfehlers asymptotisch zu stabilisieren. Weiter ist das sogenannte Peaking Phänomen als Singularität in der Zeitdiskretisierung zu sehen.

Schließlich werden ausgewählte Zustandsbeobachter an einem experimentellen Versuchsaufbau, mit einer einzelnen einseitigen Bindung, implementiert und getestet. Die Vor- und Nachteile der gewählten Ansätze werden diskutiert.

1 Introduction

This monograph is concerned with the state observer design for mechanical systems with unilateral constraints. In particular, the case where it cannot be instantaneously concluded from measurements whether or not contacts between multiple bodies are open or closed, is addressed. Furthermore, selected state observers are implemented on an experimental setup, tested and compared. This introductory chapter gives a motivation and outline for the subsequent chapters. An overview of the relevant literature is given to put the present work into a greater context. Furthermore, the aim and scope are presented.

1.1 Motivation

Mechanical systems with unilateral constraints are found in many engineering applications. In some systems involving multiple bodies, engineers aim to avoid the consequences of contacts, such as collisions or friction. As a classical example, rail squealing is an unwanted effect in railway vehicles passing curves, which is caused by friction induced oscillations and shall be avoided as much as possible. Appropriate models are required to analyze and reduce such effects. In other applications however, unilateral constraints are essential in the sense that they form a key component of the design. For example, the particular tasks of walking for legged robots or grasping objects using robotic manipulators rely on contacts that can open and close. The control of such systems requires accurate models of unilateral constraints and a good understanding of their influence on the behavior of mechanical systems.

Modeling attempts for impacts and friction phenomena can be found throughout the history of mechanics. Impacts between different bodies have for example been modeled using rigid bodies together with spring and damper elements representing the contacts. Such an approach has the disadvantage that it is difficult to identify the corresponding parameter values in applications. Furthermore, this modeling approach leads to stiff ordinary differential equations, which are problematic for numerical integration. In contrast, non-smooth models allow for discontinuities in the time evolution, which are an idealization of macroscopic observations. Think about dropping a ball from a certain height and letting it bounce on the ground. Then, macroscopically, the ball seems to behave as a rigid body which instantaneously changes its velocity when contact with the ground occurs. Similar observations can be made for slip-stick transitions in frictional contacts. Until now, a variety of mathematical formalisms has been introduced and studied by different scientific communities, all of which are suitable to describe the dynamics of certain non-smooth systems which exhibit such discontinuities in their time evolution. A formalism that allows to incorporate set-valued force laws (which are convenient for modeling impulsive and non-impulsive contact forces) and proved to be particularly fruitful for the derivation of numerical integration schemes is given by measure differential inclusions. They will take a central role in this thesis, but it is important to keep the connections to other frameworks in mind, as specific analysis or design problems can more easily be solved in one or another framework.

The design of complex systems featuring unilateral constraints, does not only require suitable models, but also systematic analysis and design methods. There has been a great effort to generalize existing methods for smooth non-linear systems to be applicable to non-smooth systems, including stability theory and various control design strategies. However, in some cases such a generalization proves to be a difficult task, for example for the problem of how to obtain estimates of the state of a system for which not all states can directly be measured, which is a classical problem in control theory. The need for such internal information is due to many purposes such as state-feedback control, system identification, monitoring or decision making. As in most applications only a limited number of sensors with a limited accuracy can be used due to cost and physical constraints, it is desirable to reconstruct the required information from only a few measurements. There are two main approaches to achieve this goal in combination with an accurate system model. One is optimization-based, i.e. the initial state at the beginning of a moving time window is estimated such that a certain functional of the error between a model-based predicted output and the measurements is minimized over that window. Since in most cases an online state estimation is required, the computation time is a limiting factor in this approach.

More commonly, a dynamical system approach is taken, in which a *state observer* is used, which is an auxiliary system that is unilaterally coupled to the observed system through measurements and is designed such that the observer's state converges to the true searched-for state. A typical core tool in the design process of state observers are Lyapunov-type stability theorems, which are used to assess the stability of the estimation error dynamics (i.e. the dynamics of the point-wise error between the trajectories of the observer and the observed system). While stability theory has been successfully generalized to non-smooth systems, it contains inherent difficulties when it comes to the analysis of such an error dynamics. In fact, the very definition of stability has to be altered in such a case. The closely related problems of tracking control and controlled synchronization also rely on the stability analysis of an error dynamics. These problems have been successfully solved for some classes of non-smooth systems by introducing alternative stability definitions and related theorems. However, these tools did not lead to solutions for the state observer problem, which motivates the need for fundamental research in this area.

1.2 Literature overview

In this section, an overview of some relevant literature in the fields of non-smooth dynamical systems and state observer design is given. As the number of works published in both of these areas is very large, this review is by no means complete, but helps to put the present work into context.

Non-smooth dynamical systems

Non-smooth systems refer to dynamical systems with a non-differentiable or discontinuous time evolution. They can be found in many research areas in science and engineering. It is therefore no surprise that various formalisms and modeling approaches have been developed in different scientific communities.

Particularly in non-smooth mechanics, on which the focus lies in this thesis, the works of Moreau [78, 79] and Jean [55, 56] build an important foundation. Therein, the dynamics is described by a measure differential inclusion (MDI), allowing for non-differentiable and discontinuous time evolutions (see also [7, 68, 77]). The framework makes extensive use of convex and non-smooth analysis (as treated for example in [6, 93]) and permits the derivation of numerical integration schemes. Set-valued force and impact laws are treated by Glocker and Pfeiffer [41, 87]. A comprehensive overview of numerical methods and modeling approaches for non-smooth dynamical systems is given by Acary and Brogliato [1]. Furthermore, Brogliato gives a broad overview of many additional topics in non-smooth mechanics in [22].

MDIs, which allow for non-differentiable and discontinuous trajectories, can in some sense be seen as a generalization of ordinary differential equations with discontinuous right-hand sides [38] and differential inclusions [5, 31, 38, 67], which both are capable of describing non-differentiable but continuous time evolutions. This will become clear in Chapter 2 of this monograph.

Systems that contain a continuous dynamics (described by continuous differential equations) and a discrete dynamics are also collectively referred to as hybrid systems. In hybrid system theory, which is often used in control theory literature, the two dynamics are typically formulated separately and come with a switching law that describes which one (or to be more precise, which hybrid mode) is active at a given point in time. Refer for example to the books of Goebel et al. [45] and Haddad et al. [48] for a broad treatment of hybrid systems. A disadvantage of these hybrid systems is however, that *accumulation points* (meaning an infinite number of impacts / state jumps or other types of switching events occurring within a finite time interval) cannot be described. Note that MDIs also contain a continuous and a discrete (or impulsive) part, but they are both combined in one compact equality and are capable to describe such accumulation points.

In addition to the system descriptions above, many more formalisms and various sub-classes of systems exist, which will not be discussed here.

State observer design

For linear systems, the state observer design problem can be considered solved, as general state observer designs are available, together with verifiable observability conditions (which assure that the searched-for information can in fact be reconstructed from the available measurements). The well-known works of Luenberger [70, 71] for deterministic linear systems and Kalman [58, 59] for stochastic linear systems build a foundation for an observer theory for linear dynamical systems [83]. In contrast, for non-linear dynamical systems, no such general observer designs are available. There is however a large number of results that apply to non-linear systems with specific structures. In particular, so-called *normal forms* are known, which are sub-classes of non-linear systems, for which a general observer design is available. Examples are systems with additive output nonlinearities (which only depend on the known input and output), for which a Luenberger-type observer can be designed [64] or state-affine systems [14, 15], for which a Kalman-like observer exists. Furthermore, triangular forms allow for high gain observers, in which nonlinearities are dominated through correction terms with sufficiently large gains [49]. It is natural to search for invertible state transformations that can transform a given non-linear systems to one of the known normal forms. An observer can then be designed for the transformed system and the state estimation is obtained via an inversion of the used transformation.

For various classes of non-smooth systems with a non-differentiable but continuous time evolution, observer designs have been analyzed using different formalisms. For bi-modal piecewise linear systems, a Luenberger-type observer was proposed by Juloski et al. [57] and experimentally tested by Doris et al. [36]. Therein, sufficient conditions for the stability of the estimation error dynamics were given based on piecewise Lyapunov functions and linear matrix inequalities. For piecewise affine systems, related results can be found from van de Wouw and Pavlov [99]. For Lur'e-type systems, Brogliato and Heemels [23] as well as Arcak and Kokotovic [4] investigate observer designs based on passivity theory. For bi-modal state-dependent switched systems, Fiore et al. [39] investigated an observer design based on contraction theory.

For non-smooth systems with impulsive motion, i.e. with a discontinuous time evolution, it is an important aspect whether or not the time instants, at which state jumps occur, can instantaneously be extracted from measurements. If these time instants are known, one can let the observer exhibit state jumps at the exact same time instants as the observed system, which greatly facilitates the design process. For example, well-known stability concepts can then directly be used to analyze the estimation error dynamics. Most available observer designs for such systems assume that these jump time instants are known. Heemels et al. [51] provided a passivity-based observer design for linear complementarity systems (LCSs) in which state jumps are induced externally (through impulsive inputs). For mechanical systems with unilateral constraints, Menini and Tornambè [75] investigate velocity observers, i.e. a reduced-order state observer for the case that all positions are measured, within the hybrid systems framework. Therein, accumulation points are explicitly excluded from the analysis. Within the same framework, Martinelli et al. [73] investigate observer design for the case that the observed system is only observable through impacts. Tanwani et al. [97], presented observers for various system classes, including certain measure differential inclusions, for which it is assumed that the time instants of state jumps are known. Also as measure differential inclusions, Tanwani et al. [98] deal with state observers for Lagrangian systems, specifically including mechanical systems with unilateral constraints. It is assumed therein, that all positions are measured and observers are discussed, that provide estimates for the full state or only for the corresponding velocities (similar to Menini [75]). Under the assumption that it can only be detected whether contacts are open or closed (for example through tactile sensors), Baumann and Leine [11] proposed synchronization-based state observers for certain linear mechanical systems (linear except for the unilateral constraints).

For the case that the time instants of state jumps are *not* available through measurements, only few results are available in literature. Furthermore, the applicability of the results is very limited. Kim et al. [63] and Menini and Tornambè [76] analyzed invertible state transformations, which allow to transform impulsive hybrid systems to a system representation without any state jumps. If one succeeds in finding such a transformation, a conventional observer might be applied to the transformed system and a state estimate is obtained by using the inverse transformation. However, finding a suitable state transformation is a very difficult task (possibly impossible), which was only demonstrated to be successful for mechanical one degree-of-freedom systems with fully elastic impacts. Also within the hybrid systems framework, Bernard and Sanfelice [13] proposed a local observer for linear systems (linear except for the state jumps) which relies on a sufficiently fast linear observer in between state jumps and correction terms that are not active in the vicinity of state jump time instants.

Related subjects

A number of related subjects are involved with similar difficulties as the state observer design for impulsive systems, as they also require to analyze the stability of an error dynamics. A major complication in such an analysis is the so-called peaking phenomenon, which refers to the fact that a slight mismatch in the jump time instants of two compared trajectories leads to a temporarily large Euclidean point-wise error, even if the trajectories are nearly matching [17,19,68,94]. Due to this peaking, standard stability concepts are not applicable. For tracking control, Biemond et al. [16,17,19] proposed to use a more general distance function to define the error, and provided corresponding stability definitions and theorems [18]. A similar approach has been used by Baumann et al. [8,9] to achieve controlled synchronization of impulsive systems. Rijnen et al. [91,91,92] propose an alternative approach to tracking control and a velocity estimation method.

1.3 Aim and scope

From the literature review, two main gaps can be identified regarding the observer design for impulsive mechanical systems. First, even for low dimensional, linear systems with one single unilateral constraint, no state observer design is available if the impact time instants (where state jumps occur) cannot instantaneously be detected through measurements. For example, if one wants to estimate the state of a two-mass impact oscillator with one unilateral constraint, where only the position of the non-impacting mass can be measured, it is not at all clear how to solve this problem.

Second, only few experimental studies on observer design for impulsive mechanical systems have been conducted, even for the case that the impact time instants are known. Furthermore, no direct comparison can be found, which is in part due to the fact, that different observer designs might use different measurements and are therefore not directly comparable (at least their performance cannot be compared in a fair way).

Based on these observations, the main objectives of this thesis can be summarized as follows. It is a first objective to provide the reader with the necessary mathematical background for modeling impulsive mechanical systems. Moreover, an overview of the most relevant formalisms will be given and some interconnections between them will be pointed out.

A second aim is to discuss stability theory as a fundamental tool for designing state observers. In particular, tools for evaluating partial stability will be provided, which directly extend existing Lyapunov-type theorems for measure differential inclusions. Based on these results, the main difficulties stemming from the discontinuous time evolution will be highlighted.

The main objective is to investigate if a new approach can be taken towards state observer designs which do not require explicit knowledge of the impact time instants of the observed system. As pointed out earlier, this problem stands as a major challenge and has been treated by only a minor number of publications.

Finally, it is an objective to design and construct an experimental setup consisting of an impact oscillator. Using this setup, selected state observers will be designed, experimentally tested and compared as far as possible.

In accordance with these objectives, the subsequent chapters are structured as outlined in the next section.

1.4 Outline

This first and introductory chapter has given a motivation, an overview of relevant literature and defined the objectives and the scope treated in this thesis.

Chapter 2 gives a short introduction to various formalisms that are used to describe non-smooth dynamical systems. Therein, measure differential inclusions take a central role. Discontinuous differential equations and inclusions, as well as their solution concepts, are discussed first as a stepping stone. Then the meaning and use of MDIs are explained. Finally, linear complementarity systems are taken up as a selected system class, which will reoccur in its discrete form in Chapter 6.

In Chapter 3, by employing the principle of virtual action, a derivation of the equality of measures describing non-smooth mechanical systems is presented. Next, constitutive contact laws, stemming from unilateral constraints, are reviewed. Subsequently, two relevant time discretization schemes are discussed. Finally, state transformations that eliminate state jumps are touched upon.

Chapter 4 treats partial stability for MDIs. First, the needed stability concepts are defined and relevant Lyapunov-type theorems are given. These are a direct extension of the well-known Lyapunov stability theorems for ODEs and their generalization to MDIs. It will become clear, that these theorems are a fundamental tool for various closely related tasks, such as tracking control, synchronization and state observer design. Therefore, all of these tasks are strongly complicated by the peaking phenomenon that is explained in the same chapter. Finally, as it gives a feeling for necessary refinements and subtleties in the use of Lyapunov-type tools, the stability of switched systems is discussed particularly.

In Chapter 5, some existing state observer designs that require knowledge on state jump time instants are reviewed, with a focus on mechanical systems. Furthermore, a slight extension is given to account for Coulomb friction.

Chapter 6 treats state observer design for a class of mechanical systems in the case of unknown state jump time instants. The analysis builds upon a new discretization-based approach, that makes use of the scheme of Paoli and Schatzman. It is shown that this specific discretization leads to a discrete LCS. From there, a dead-beat observer and sufficient observability conditions are derived. It is demonstrated that an existing observer design for discrete LCSs can be applied after a necessary extension. Finally, the usefulness of the approach is discussed by analyzing the results in a numerical example.

In Chapter 7, an experimental setup consisting of an impact oscillator is built. Selected observer designs are then implemented and their performance is analyzed in experiments. The advantages and disadvantages of the different designs are explained.

Finally, a conclusion on the outcome of this work is presented in Chapter 8 and recommendations for future work are given.

2

Formalisms in Non-smooth Dynamics

In this chapter, some mathematical formalisms are reviewed, that in contrast to ordinary differential equations (ODEs) allow to accommodate non-smooth solutions (i.e. solutions that are non-differentiable or discontinuous). The discussion is limited to a selection of formalisms that are most relevant for the subsequent treatment of mechanical systems with unilateral constraints.

Absolutely continuous solutions, which may be non-differentiable, can be described by ordinary differential equations with discontinuous right-hand sides, together with a generalized solution concept that allows non-differentiable functions to be considered as solutions. Various such solution concepts exist in literature, two of which are due to Carathéodory [32] and to Filippov [38]. In particular, the solution concept of Filippov to discontinuous ODEs leads to the formalism of differential inclusions (DIs). Moreover, these solution concepts are equivalent to replacing the original systems with more general dynamical systems, which can be understood as 'integral versions' of ODEs and DIs, namely measure differential equations (MDEs) and measure differential inclusions (MDIs). It will be shown that measure differential inclusions, as treated by Moreau [78], are capable of describing solutions that are both discontinuous and non-differentiable (more precisely, solutions are allowed to be functions of bounded variation). In particular, MDIs are sufficient to fully describe mechanical systems with unilateral constraints and friction, which will be discussed in the subsequent chapter. As a less general system class, linear complementarity systems (LCSs) are discussed as well, as they build a foundation for the state observer design treated in the following chapters. Also, links between some system classes are established, since the analysis of a specific problem might be easier in one or another framework. Moreover, if two system classes are equivalent (in general or under some conditions), results obtained for one system class might be transported to the other. For discrete-time systems, various such equivalences are established in [52].

2.1 Discontinuous ODEs and differential inclusions

A first system class that allows for non-smooth solutions are ordinary differential equations (ODEs) with discontinuous right-hand sides. Besides their many applications, they also serve as a stepping stone towards the more general measure differential inclusions that will be discussed subsequently. Here, we restrict ourselves to discontinuous ODEs[1] of the form

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}) = \mathbf{f}_i(t, \mathbf{x}) \quad \text{for } \mathbf{x} \in \mathcal{X}_i, \tag{2.1}$$

where the right-hand side $\mathbf{f} : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ is a vector-valued function which is discontinuous in $\mathbf{x}$. This is expressed by stating multiple right-hand sides $\mathbf{f}_i(t, \mathbf{x})$, each of which is continuous in their respective domain $\mathcal{X}_i$. The variable t is referred to as time, $\mathbf{x}$ is called state and the dot symbol $(\dot{\ })$ denotes the derivative with respect to time, i.e. $\dot{\mathbf{x}} = \mathrm{d}\mathbf{x}/\mathrm{d}t$. The state-space $\mathbb{R}^n$ is divided into a finite number of disjunct, non-empty subsets $\mathcal{X}_i$ with $i \in \{1, 2, \cdots, N\}$ and $\cup_i \mathcal{X}_i = \mathbb{R}^n$. These subsets are separated by hyper-surfaces $\mathcal{S}_{ij}$, referred to as switching surfaces, which constitute sets of discontinuity points of $\mathbf{f}$. Typically, the switching surfaces, say between $\mathcal{X}_i$ and $\mathcal{X}_j$, are characterized by functions $h_{ij} : \mathbb{R}^n \to \mathbb{R}$ through $\mathcal{S}_{ij} = \{\mathbf{x} \in \mathbb{R}^n \mid h_{ij}(\mathbf{x}) = 0\}$. The set $\mathcal{S} = \cup_{i,j} \mathcal{S}_{ij}$ contains all discontinuity points of $\mathbf{f}$.

A continuously differentiable function $\mathbf{x} : [t_0, t_1] \to \mathbb{R}^n$, with the given initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$, is a **classical solution** of (2.1) if it satisfies (2.1) for all $t \in [t_0, t_1]$. If the right-hand side is continuous, i.e. $\mathbf{f}_i(t, \mathbf{x}) = \mathbf{f}_j(t, \mathbf{x})$ for all i, j and $\mathbf{x}$ such that $h_{ij}(\mathbf{x}) = 0$, then all solutions are classical solutions. If (2.1) is discontinuous at the switching surfaces however, other definitions of solutions are required, that can accommodate non-differentiable functions as solutions. Clearly, non-differentiable functions do not satisfy (2.1), since their time derivative does not exist at the points where they are not differentiable. Therefore, one has to specify in what sense such functions have to fulfill (2.1) in order to be considered a solution. This can be done in various ways. Two frequently used such solution concepts are due to Carathéodory [32] and Filippov [38], which both require solutions to be absolutely

[1] also referred to as state-dependent *switched systems* in literature [69]. In contrast, if the condition $\mathbf{x} \in \mathcal{X}_i$ in (2.1) is replaced by $i = \sigma(t)$, with a piecewise constant function $\sigma(t)$ (called the switching function), then the system is referred to as a time-dependent switched system.

continuous.[2] As such, solutions are in particular continuous, but not necessarily differentiable.

A function $\mathbf{x} : [t_0, t_1] \to \mathbb{R}^n$, with the initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$ is called a **Carathéodory solution** [32] of (2.1), if it is absolutely continuous and it fulfills (2.1) *almost everywhere* (a.e.) on $[t_0, t_1]$, i.e. for all $t \in [t_0, t_1]$ except for a set of time instants of Lebesgue measure zero. Roughly speaking, this means that (2.1) has to hold everywhere except for some isolated time instants. This solution concept is equivalent to requiring that solutions satisfy

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{f}(\tau, \mathbf{x}(\tau))\, \mathrm{d}\tau \tag{2.2}$$

for all $t \in [t_0, t_1]$, where the integral can be understood in the sense of either Lebesgue or Riemann-Stieltjes[3]. Therefore, Carathéodory solutions fulfill (2.1) in an 'integral sense'.

Example 2.1. As a simple example [96], consider the discontinuous ODE

$$\dot{x} = \begin{cases} 1 & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases} \tag{2.3}$$

and take an initial condition $x(0) = -1$. As long as $x(t) < 0$, the dynamics is given by $\dot{x} = 1$ and $x(t)$ increases until it reaches $x(1) = 0$. From there, the dynamics is given by $\dot{x} = 0$ such that the Carathéodory solution will stay at a value of $x(t) = 0 \;\forall\; t > 1$ and therefore has a 'kink' at $t = 1$.

Equation (2.2) gives rise to a more general system class. To see this, note that using the properties of the Riemann-Stieltjes integral, (2.2) can equivalently be written as

$$\int_{\mathcal{I}} \mathrm{d}\mathbf{x} = \int_{\mathcal{I}} \mathbf{f}(t, \mathbf{x}(t))\, \mathrm{d}t, \tag{2.4}$$

where $\mathcal{I} \subseteq [t_0, t_1]$ is a compact time interval. Equation (2.4) constitutes a specific **measure differential equation** (MDE). The name stems from the fact that the integrals on both sides of (2.4) can be seen as (signed) vector measures,[3] assigning a tuple of real numbers to every $\mathcal{I}$. For brevity, the notation $\mathrm{d}\mathbf{x} : \mathcal{I} \mapsto \int_{\mathcal{I}} \mathrm{d}\mathbf{x}$ and $\mathbf{f}(t, \mathbf{x})\, \mathrm{d}t : \mathcal{I} \mapsto \int_{\mathcal{I}} \mathbf{f}(t, \mathbf{x})\, \mathrm{d}t$ is used such that the MDE (2.4) can be written in short as

$$\mathrm{d}\mathbf{x} = \mathbf{f}(t, \mathbf{x})\, \mathrm{d}t. \tag{2.5}$$

[2] as defined in Definition A.1.

[3] An introduction to integration and measure theory can be found in Appendix A. Therein, differential measures are defined as well and the related notation is clarified.

For relations of the form (2.5), it is said that $\mathbf{f}(t, \mathbf{x})$ is the **density** of $\mathrm{d}\mathbf{x}$ with respect to $\mathrm{d}t$. By explicitly stating the arguments, (2.5) would read $\mathrm{d}\mathbf{x}(\mathcal{I}) = \mathbf{f}(t, \mathbf{x})\, \mathrm{d}t(\mathcal{I})$.

Remark 2.2. The MDE (2.5) allows for an alternative interpretation in terms of functionals. Let $\varphi : \mathcal{I} \to \mathbb{R}$ be a continuous real-valued function on the interval $\mathcal{I}$ with compact support in $\mathcal{I}$ (i.e. its support is contained in $\mathcal{I}$), written as $\varphi \in \mathrm{C}_0(\mathcal{I}, \mathbb{R})$. Then, the integral equation (2.4) is equivalent to

$$\int_{\mathcal{I}} \varphi \, \mathrm{d}\mathbf{x} = \int_{\mathcal{I}} \varphi \, \mathbf{f}(t, \mathbf{x}(t))\, \mathrm{d}t \quad \forall\, \varphi \in \mathrm{C}_0(\mathcal{I}, \mathbb{R}). \tag{2.6}$$

Therefore, by using the alternative short notation $\mathrm{d}\mathbf{x} : \varphi \mapsto \int_{\mathcal{I}} \varphi \, \mathrm{d}\mathbf{x}$ and likewise $\mathbf{f}(t, \mathbf{x})\, \mathrm{d}t : \varphi \mapsto \int_{\mathcal{I}} \varphi \, \mathbf{f}(t, \mathbf{x})\, \mathrm{d}t$, equation (2.5) holds in terms of functionals, which by stating the arguments then reads $\mathrm{d}\mathbf{x}[\varphi] = \mathbf{f}(t, \mathbf{x})\, \mathrm{d}t[\varphi]$. The functional $\mathrm{d}\mathbf{x}[\varphi]$ is referred to as **differential measure** of $\mathbf{x}$.

Going back to the original discontinuous ODE (2.1), Carathéodory solutions might not exist, while other solution concepts, such as the subsequent Filippov concept, do provide a solution. This is typically the case if solutions approach a switching surface from both sides, such that solutions starting in the vicinity of the switching surface are constrained to 'slide' on this switching surface.

Example 2.3. As a standard example [1, p. 62], consider the simple discontinuous ODE

$$\dot{x} = \begin{cases} 1 & \text{if } x < 0 \\ -1 & \text{if } x \geq 0 \end{cases} \tag{2.7}$$

and take an initial condition $x(0) = -1$. As long as $x(t) < 0$, the dynamics is given by $\dot{x} = 1$ and $x(t)$ increases until it reaches $x(1) = 0$. From there, the solution can neither increase nor decrease, because that would violate (2.7). However, $x(t) = 0 \;\forall\, t > 1$ violates (2.7) as well, such that no Carathéodory solution exists.

A function $\mathbf{x} : [t_0, t_1] \to \mathbb{R}^n$, with the initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$ is called a **Filippov solution** of (2.1), if it is a solution of

$$\dot{\mathbf{x}} \in \boldsymbol{\mathcal{F}}(t, \mathbf{x}(t)) \tag{2.8}$$

where $\boldsymbol{\mathcal{F}} : \mathbb{R} \times \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ is a set-valued extension constructed from $\mathbf{f}$. Here, $\mathcal{P}(\mathbb{R}^n)$ denotes the collection of all subsets of $\mathbb{R}^n$. Roughly speaking, $\boldsymbol{\mathcal{F}}(t, \mathbf{x})$ is taken as the smallest convex set that contains all the values of $\mathbf{f}(t, \mathbf{x})$ in a neighborhood of $\mathbf{x}$. In other words, it is the closed convex hull of all limit values of $\mathbf{f}(t, \hat{\mathbf{x}})$ for $\hat{\mathbf{x}} \to \mathbf{x}$ and $\hat{\mathbf{x}} \notin \mathcal{S}$.

Systems of the form (2.8), with an arbitrary set-valued $\boldsymbol{\mathcal{F}}$, are referred to as **differential inclusions** (DIs). A function $\mathbf{x} : [t_0, t_1] \to \mathbb{R}^n$, with the initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$ is called a solution of (2.8), if it is absolutely continuous and it fulfills (2.8) almost everywhere on $[t_0, t_1]$. This solution concept is equivalent to specifying that a solution has to fulfill

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^{t} \hat{\mathbf{f}}(\tau, \mathbf{x}(\tau))\,\mathrm{d}\tau \tag{2.9}$$

with $\hat{\mathbf{f}}(t, \mathbf{x}(t)) \in \boldsymbol{\mathcal{F}}(t, \mathbf{x}(t))$ for all $t \in [t_0, t_1]$.

Example 2.4. Consider again the discontinuous ODE (2.7) in Example 2.3, for which no Carathéodory solution exists. The set-valued Filippov extension of the right-hand side at the switching surface $x = 0$ is given by $\mathcal{F}(t, x = 0) = [-1, 1]$. Outside the switching function, the set-valued extension equals the right-hand side of the ODE, i.e. $\mathcal{F}(t, x > 0) = -1$ and $\mathcal{F}(t, x < 0) = 1$. Now take the initial condition $x(0) = -1$. Then the dynamics is given by $\dot{x} = 1$ as long as $x(t) < 0$, until the solution reaches $x(1) = 0$. From there, the solution can neither increase nor decrease, since that would violate the DI $\dot{x} \in \mathcal{F}(t, x(t))$. However, $x(t) = 0 \;\forall\; t > 1$ is a valid solution, because $0 \in [-1, 1] = \mathcal{F}(t, 0)$. Therefore, the Filippov solution exists and 'slides' along the switching surface $x = 0$ for $t > 1$.

Just as the integral version of a discontinuous ODEs gave rise to measure differential equations, the integral equation (2.9) also motivates a more general system class. Again, by using Riemann-Stieltjes integrals, (2.9) can equivalently be written as

$$\int_{\mathcal{I}} \mathrm{d}\mathbf{x} = \int_{\mathcal{I}} \hat{\mathbf{f}}(t, \mathbf{x}(t))\,\mathrm{d}t \quad \text{with} \quad \hat{\mathbf{f}}(t, \mathbf{x}) \in \boldsymbol{\mathcal{F}}(t, \mathbf{x}) \,\forall\, t \tag{2.10}$$

where $\mathcal{I} \subseteq [t_0, t_1]$ is a compact interval. Using the short notation, (2.10) can be written as

$$\mathrm{d}\mathbf{x} = \hat{\mathbf{f}}(t, \mathbf{x})\,\mathrm{d}t \quad \text{with} \quad \hat{\mathbf{f}}(t, \mathbf{x}) \in \boldsymbol{\mathcal{F}}(t, \mathbf{x}) \,\forall\, t. \tag{2.11}$$

System (2.11) is a specific **measure differential inclusion** (MDI). Similar to Remark 2.2 for MDEs, the MDI (2.11) can be interpreted in a functional sense as well.

Such MDIs can also be used to describe discontinuous solutions, after a natural generalization, which is discussed in the next section. Note that MDIs are the most general system class discussed so far and they specifically include MDEs, DIs as well as discontinuous ODEs.

Remark 2.5. Even if both a Filippov and a Carathéodory solution exist, they are not necessarily equal. This can be seen by constructing a Filippov solution to the

ODE (2.3) in Example 2.1. The Filippov DI reads

$$\dot{x} \in \mathcal{F}(t,x) = 1, \tag{2.12}$$

i.e. the discontinuous ODE is replaced by a continuous ODE. Starting from the initial condition $x(0) = -1$, the Filippov solution is therefore given by $x(t) = t - 1 \;\forall\, t > 0$, which strongly differs from the Carathéodory solution given in Example 2.1.

2.2 Measure differential inclusions

So far, it was shown that discontinuous ODEs and DIs are both system classes that allow for absolutely continuous solutions (which are possibly non-differentiable but continuous), if a suitable solution concept is provided. Equivalently, this can be achieved by replacing the original system with specific MDEs or MDIs, which specify the density of the differential measure $\mathrm{d}\mathbf{x}$ of the solution with respect to $\mathrm{d}t$. More general MDIs allow for solutions from more general function classes. Of particular interest are special functions of bounded variation, which can be decomposed as $\mathbf{x} = \mathbf{x}_{\mathrm{ac}} + \mathbf{x}_{\mathrm{s}}$ into an absolutely continuous part $\mathbf{x}_{\mathrm{ac}}$ and a step function $\mathbf{x}_{\mathrm{s}}$.

A step function $\mathbf{x}_{\mathrm{s}}$ is a piecewise constant function with discontinuities at a countable set of time instants $\{t_1, t_2, ...\}$. If the step heights (or jumps) of each component at the discontinuity points are given by a function $\mathbf{a}(t)$ through

$$\mathbf{x}_{\mathrm{s}}^{+}(t) - \mathbf{x}_{\mathrm{s}}^{-}(t) = \mathbf{a}(t) \quad \forall\, t \in \{t_1, t_2, ...\}, \tag{2.13}$$

then the step function can be expressed as

$$\mathbf{x}_{\mathrm{s}}^{+}(t) = \mathbf{x}_{\mathrm{s}}^{-}(t_0) + \int_{[t_0,t]} \mathbf{a}(\tau)\,\mathrm{d}\eta. \tag{2.14}$$

Therein, $\mathrm{d}\eta$ denotes the sum $\mathrm{d}\eta = \sum_k \mathrm{d}\delta_{t_k}$ of Dirac measures, which are such that for any open, closed or half-open interval $\mathcal{I}$ it holds that

$$\int_{\mathcal{I}} \mathbf{a}\,\mathrm{d}\delta_{t_k} = \begin{cases} \mathbf{a}(t_k) & \text{if } t_k \in \mathcal{I} \\ 0 & \text{if } t_k \notin \mathcal{I}. \end{cases} \tag{2.15}$$

As before, equation (2.14) is equivalently expressed as

$$\mathrm{d}\mathbf{x}_{\mathrm{s}} = \mathbf{a}(t)\,\mathrm{d}\eta \tag{2.16}$$

using the short notation, i.e. the differential measure $\mathrm{d}\mathbf{x}_{\mathrm{s}}$ of a step function has a density with respect to $\mathrm{d}\eta$, and the density corresponds to the step heights. Now, let $\mathbf{a}(t) = \mathbf{g}(t, \mathbf{x}(t)) \in \mathcal{G}(t, \mathbf{x}(t))$ with a given set-valued function $\mathcal{G}$, where $\mathbf{g}(t, \mathbf{x})$ is written to mean a dependence on $t, \mathbf{x}^+(t)$ and $\mathbf{x}^-(t)$ (following [78] and [68]). Then, using $\mathrm{d}\mathbf{x} = \mathrm{d}\mathbf{x}_{\mathrm{ac}} + \mathrm{d}\mathbf{x}_{\mathrm{s}}$ results in

$$\int_{\mathcal{I}} \mathrm{d}\mathbf{x} = \int_{\mathcal{I}} \mathbf{f}(t, \mathbf{x})\,\mathrm{d}t + \mathbf{g}(t, \mathbf{x})\,\mathrm{d}\eta \quad \text{with} \quad \begin{array}{l} \mathbf{f}(t, \mathbf{x}) \in \mathcal{F}(t, \mathbf{x}), \\ \mathbf{g}(t, \mathbf{x}) \in \mathcal{G}(t, \mathbf{x}), \end{array} \tag{2.17}$$

which is a measure differential inclusion that allows for special functions of bounded variation as solutions. In short, the MDI (2.17) reads

$$\mathrm{d}\mathbf{x} = \mathbf{f}(t, \mathbf{x})\,\mathrm{d}t + \mathbf{g}(t, \mathbf{x})\,\mathrm{d}\eta \quad \text{with} \quad \begin{array}{l} \mathbf{f}(t, \mathbf{x}) \in \mathcal{F}(t, \mathbf{x}), \\ \mathbf{g}(t, \mathbf{x}) \in \mathcal{G}(t, \mathbf{x}). \end{array} \tag{2.18}$$

Remark 2.6. The MDI (2.17) or (2.18) was motivated by combining a differential inclusion (2.8) and the description (2.13) of a step function. Conversely, a non-impulsive and an impulsive dynamics can be extracted from the MDI. Indeed, evaluating (2.17) over an isolated discontinuity point t_k, i.e.

$$\mathbf{x}^+(t_k) - \mathbf{x}^-(t_k) = \int_{\{t_k\}} \mathrm{d}\mathbf{x} = \int_{\{t_k\}} \mathbf{f}(t, \mathbf{x})\,\mathrm{d}t + \int_{\{t_k\}} \mathbf{g}(t, \mathbf{x})\,\mathrm{d}\eta \tag{2.19}$$

and using the fact that $\int_{\{t_k\}} \mathbf{f}(t, \mathbf{x})\,\mathrm{d}t = \mathbf{0}$ and $\int_{\{t_k\}} \mathbf{g}(t, \mathbf{x})\,\mathrm{d}\eta = \mathbf{g}(t_k, \mathbf{x}(t_k))$ yields the impulsive dynamics

$$\mathbf{x}^+(t_k) - \mathbf{x}^-(t_k) = \mathbf{g}(t_k, \mathbf{x}(t_k)), \tag{2.20}$$

with $\mathbf{g}(t_k, \mathbf{x}(t_k)) \in \mathcal{G}(t_k, \mathbf{x}(t_k))$. Moreover, evaluating (2.17) over a time interval $[t_0, t]$, which does not contain any discontinuity points, i.e.

$$\int_{[t_0, t]} \mathrm{d}\mathbf{x} = \int_{[t_0, t]} \mathbf{f}(t, \mathbf{x})\,\mathrm{d}t + \int_{[t_0, t]} \mathbf{g}(t, \mathbf{x})\,\mathrm{d}\eta, \tag{2.21}$$

using the fact that $\int_{[t_0, t]} \mathbf{g}(t, \mathbf{x})\,\mathrm{d}\eta = \mathbf{0}$ and taking the time derivative on both sides of (2.21) gives

$$\dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t)) \tag{2.22}$$

for almost every $t \in [t_a, t_b]$, with $\mathbf{f}(t, \mathbf{x})) \in \mathcal{F}(t, \mathbf{x})$.

For a more compact notation, the MDI (2.18) is also written as

$$\mathrm{d}\mathbf{x} \in \mathrm{d}\mathbf{\Gamma}(t, \mathbf{x}), \tag{2.23}$$

and $\mathrm{d}\mathbf{\Gamma}(t, \mathbf{x})$ is referred to as a set-valued measure function. A function $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ of locally bounded variation, with the given initial condition $\mathbf{x}^-(t_0) = \mathbf{x}_0$, is a **solution** of the MDI (2.23), if it fulfills (2.23) for all $t \geq t_0$ [68]. Even though at points of discontinuity t_k, the solution $\mathbf{x}(t_k)$ might not be defined, (2.23) is still fulfilled. It is common to write $\mathbf{x}(t, t_0, \mathbf{x}_0) = \mathbf{x}(t)$ to explicitly state the dependence on the initial conditions.

Compared to systems of differential equations, there are a number of properties of MDIs that make the analysis more difficult. A discussion of these properties can for example be found in [68] for general MDIs and in [85] for MDIs modeling mechanical systems with unilateral constraints. First, as for ordinary differential equations, the existence and uniqueness of solutions is not given in general. While for ODEs the Lipschitz continuity of the right-hand side (together with linear boundedness w.r.t. t in the case of non-autonomous ODEs) is sufficient for existence and uniqueness of solutions [60], no such general condition is available for MDIs. Second, the solutions of MDIs do not depend continuously on the initial conditions in some cases, meaning that two solutions can diverge strongly within a short time horizon, even if their initial conditions are close to each other. A typical example is the motion of a point mass whose position is restricted by two unilateral constraints. In the vicinity of the intersection of these unilateral constraints, the future motion of the point mass may strongly depend on which constraint causes an impact first [77, p. 129]. It is a third difficulty that, as for the example of the restricted point mass, solutions may be confined to an admissible subset $\mathcal{A}$ of the state space. In this chapter, it is assumed that at least one solution to the investigated MDI exists. Furthermore, it is assumed that the MDI is consistent in the sense that for initial conditions $\mathbf{x}_0 \in \mathcal{A}$ in the admissible set, all solutions remain in $\mathcal{A}$.

2.3 Linear complementarity systems

A system class of particular interest are **linear complementarity systems** (LCSs), which are linear dynamical systems that contain a variable which is determined by a linear complementarity problem (LCP).

Continuous formulation

In its continuous-time formulation, an LCS can be written as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{w}, \tag{2.24a}$$

$$\mathbf{z} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{w}, \tag{2.24b}$$

$$0 \leq \mathbf{z} \perp \mathbf{w} \geq 0, \tag{2.24c}$$

with the state $\mathsf{x} \in \mathbb{R}^n$ and the variables $\mathsf{z} \in \mathbb{R}^p$ and $\mathsf{w} \in \mathbb{R}^p$ (referred to as complementarity variables). Therein, the inequalities have to be understood component-wise, i.e. $\mathsf{z} = (z_1, \cdots, z_p)^\mathsf{T} \geq 0$ is written to mean that every component $z_i \geq 0$, $i \in \{1, \cdots, p\}$. Furthermore, the notation $\mathsf{z} \perp \mathsf{w}$ is used to express the orthogonality $\mathsf{z}^\mathsf{T}\mathsf{w} = 0$. The last line (2.24c) is referred to as an **inequality complementarity**. Here it is assumed that the matrices $\mathsf{A}, \mathsf{B}, \mathsf{C}$ and D are constant, but in general they can also be time dependent. The LCS consists of a linear ODE augmented by an algebraic equation and an inequality complementarity. The meaning of the LCS (2.24) only becomes clear after providing a suitable solution concept. With the goal of allowing non-differentiable and discontinuous solutions, such a solution concept was formalized by Heemels [53]. It does not rely on differential measures but on so-called Bohl distributions, which only allow for accumulation points in forward time. This allowed the authors to give sufficient conditions for the existence and uniqueness of solutions [51, 53].

It is sometimes convenient to reformulate the inequality complementarity using normal cones, which leads to alternative system descriptions for the LCS (2.24). The **normal cone** $\mathcal{N}_C$ to a non-empty, closed and convex set $C \subset \mathbb{R}^n$, at a point $\mathbf{x} \in C$, is defined as

$$\mathcal{N}_C(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^n \,|\, \mathbf{y}^\mathsf{T}(\mathbf{x}^* - \mathbf{x}) \leq 0, \, \forall \, \mathbf{x}^* \in C\} \tag{2.25}$$

while $\mathcal{N}_C(\mathbf{x})$ is empty for $\mathbf{x} \notin C$. In other words, the normal cone consists of the set of normal vectors to C in $\mathbf{x}$. If $\mathbf{x}$ is in the interior of C, the normal cone is $\mathcal{N}_C(\mathbf{x}) = \{0\}$.

Note that $\mathbf{x}$ in (2.25) is any element of $\mathbb{R}^n$, not necessarily the state of a dynamical system, which is why it has been written with serifs. As a *general notation* in this thesis, variables contained in the description of a dynamical system in first order form are written without serifs, whereas other variables (especially in the description of mechanics) are written with serifs. This will allow to some extent to use standard notations from systems and control theory and mechanics in parallel.

If C is taken as the set $\mathbb{R}_0^{n+}$, which is defined as the set of all elements of $\mathbb{R}^n$ with non-negative components, then a useful relationship between normal cones and inequality complementarities of the form $0 \leq \mathbf{x} \perp \mathbf{y} \geq 0$ can be established. Specifically, the equivalence

$$-\mathbf{y} \in \mathcal{N}_{\mathbb{R}_0^{n+}}(\mathbf{x}) \quad \Leftrightarrow \quad 0 \leq \mathbf{x} \perp \mathbf{y} \geq 0 \tag{2.26}$$

holds. Indeed, if the i-th component x_i of $\mathbf{x}$ is positive, then by the definition of the normal cone, the i-th component y_i of $\mathbf{y}$ has to be zero. Otherwise, if $x_i = 0$, it follows from (2.25) that $y_i \geq 0$, such that the inequality complementarity holds for each component i.

Using the equivalence (2.26), the inequality complementarity $0 \leq \mathbf{z} \perp \mathbf{w} \geq 0$ in the LCS can be replaced by $-\mathbf{w} \in \mathcal{N}_{\mathbb{R}_0^{p+}}(\mathbf{z})$, which for the case $\mathbf{D} = 0$ results in

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{w} \quad \text{with} \; -\mathbf{w} \in \mathcal{N}_{\mathbb{R}_0^{p+}}(\mathbf{C}\mathbf{x}), \tag{2.27}$$

which forms a differential inclusion. Similarly, for $\mathbf{D} \neq 0$, one obtains a differential algebraic inclusion.

LCSs are connected to a number of other system classes. For example, due to the complementarity $z_i w_i = 0 \, \forall \, i \in \{1, \cdots, p\}$, (2.24) contains 2^p 'modes' and each mode is described by a differential algebraic equation. As pointed out in [51], the LCS can also be reformulated as more general hybrid systems, but such a reformulation is generally cumbersome and leads to large system descriptions.

Discrete formulation

In Chapter 6, an LCS will be used in its discrete-time formulation, which is written as

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{w}_k, \\ \mathbf{z}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{w}_k, \\ 0 \leq \mathbf{z}_k &\perp \mathbf{w}_k \geq 0, \end{aligned} \tag{2.28}$$

with the discrete state $\mathbf{x}_k \in \mathbb{R}^n$ and the complementary variables $\mathbf{z}_k \in \mathbb{R}^p$ and $\mathbf{w}_k \in \mathbb{R}^p$. For discrete LCSs, various links to other system classes, such as mixed logical dynamical systems [12], extended linear complementarity systems [34], piece-wise affine systems and others have been established in [52]. Without discussing the details, it should be noted that these links and equivalences are much broader than in the continuous-time case. As an example, let the complementarity variables z_k, w_k and be scalar. Then, for a given $\mathbf{x}_k$, the variable w_k can be calculated as

$$w_k = \begin{cases} -D^{-1}\mathbf{C}\mathbf{x}_k & \text{if } D^{-1}\mathbf{C}\mathbf{x}_k \leq 0 \\ 0 & \text{if } D^{-1}\mathbf{C}\mathbf{x}_k > 0. \end{cases} \tag{2.29}$$

Using (2.29), w_k can be eliminated in (2.28), which yields the piece-wise linear system

$$\mathbf{x}_k = \begin{cases} (\mathbf{A} - D^{-1}\mathbf{B}\mathbf{C})\mathbf{x}_k & \text{if } D^{-1}\mathbf{C}\mathbf{x}_k \leq 0 \\ \mathbf{A}\mathbf{x}_k & \text{if } D^{-1}\mathbf{C}\mathbf{x}_k > 0. \end{cases} \tag{2.30}$$

3

Non-smooth multibody systems

In this chapter, the description of a non-smooth mechanical multibody system as a measure differential inclusion (MDI) is derived. For smooth mechanical systems, whose state trajectories are continuously differentiable, the principle of virtual work leads to the equations of motion as a set of differential equations. It is shown, that the equivalent principle of virtual action allows for a generalization to non-smooth systems, whose state trajectories exhibit discontinuities in time, leading to an equality of measures. In addition, set-valued force laws lead to a measure differential inclusion. Readers who are not familiar with differential measures (or Stieltjes measures and integrals) are referred to Appendix A prior to reading this chapter. Subsequently, two time discretization schemes are shown and their main difference is discussed. Finally, state transformations are discussed, which lead to continuous solutions in the transformed state space.

3.1 Equality of measures

Principle of virtual work

Classically, the equations of motion, which describe the non-impulsive dynamics of a mechanical system, can be derived from a fundamental axiom: the principle of virtual work. A mechanical system $\mathcal{S}$ consists of particles (or material points) whose positions in Euclidean space at a given time t are described by $\boldsymbol{\xi}(\mathbf{X}, t) \in \mathbb{R}^3$. Therein, $\mathbf{X} \in \mathbb{R}^3$ are particle coordinates in a reference position. The map $\boldsymbol{\xi} : (\mathbf{X}, t) \mapsto \boldsymbol{\xi}(\mathbf{X}, t)$

is called a motion of the system $\mathcal{S}$. The **virtual work** of a mechanical system $\mathcal{S}$ is defined as the functional

$$\delta W[\delta\boldsymbol{\xi}] := -\int_{\mathcal{S}} \delta\boldsymbol{\xi}^{\mathsf{T}}(\ddot{\boldsymbol{\xi}}\mathrm{d}m - \mathrm{d}\mathbf{F}). \tag{3.1}$$

Recall that the dot symbol $(\dot{\ })$ over variables refers to the time derivative. Furthermore, if $\mathcal{S}$ is a collection of bodies, the integrals have to be understood in the sense of $\int_{\mathcal{S}}(\cdot)\mathrm{d}m := \int_{\mathcal{S}}(\cdot)\rho\mathrm{d}V$ (i.e. $\mathrm{d}m$ is the mass distribution) and $\int_{\mathcal{S}}(\cdot)\mathrm{d}\mathbf{F} := \int_{\mathcal{S}}(\cdot)\mathbf{b}\mathrm{d}V + \int_{\partial\mathcal{S}}(\cdot)\mathbf{s}\mathrm{d}A$ with force density $\mathbf{b}$ and surface force density $\mathbf{s}$ (i.e. $\mathrm{d}\mathbf{F}$ is the force distribution). A function family $\hat{\boldsymbol{\xi}}(\mathbf{X}, t, \varepsilon)$, depending on a variation parameter ε, can be constructed such that $\boldsymbol{\xi}(\mathbf{X}, t) = \hat{\boldsymbol{\xi}}(\mathbf{X}, t, \varepsilon_0)$ for a given ε_0. The **virtual displacements** $\delta\boldsymbol{\xi}$ are then defined by

$$\delta\boldsymbol{\xi}(\mathbf{X}, t) := \frac{\partial\hat{\boldsymbol{\xi}}}{\partial\varepsilon}(\mathbf{X}, t, \varepsilon_0)\delta\varepsilon, \tag{3.2}$$

with the variation $\delta\varepsilon = \varepsilon - \varepsilon_0$ and under the assumption that $\hat{\boldsymbol{\xi}}(\mathbf{X}, t, \varepsilon)$ is differentiable with respect to ε.
The **principle of virtual work**, as a postulate, states that at each time instant t the virtual work vanishes for all virtual displacements $\delta\boldsymbol{\xi}$, i.e.

$$\delta W[\delta\boldsymbol{\xi}] = 0 \ \ \forall\ \delta\boldsymbol{\xi}. \tag{3.3}$$

Clearly, to fulfill the principle of virtual work at all times, it is required that the particle velocities $\dot{\boldsymbol{\xi}}$ are continuous and differentiable. With the definition (3.1), the principle of virtual work reads as

$$\delta W[\delta\boldsymbol{\xi}] = -\int_{\mathcal{S}} \delta\boldsymbol{\xi}^{\mathsf{T}}\ddot{\boldsymbol{\xi}}\mathrm{d}m + \int_{\mathcal{S}} \delta\boldsymbol{\xi}^{\mathsf{T}}\mathrm{d}\mathbf{F} = 0 \ \ \forall\ \delta\boldsymbol{\xi}. \tag{3.4}$$

Integration by parts of the first term on the right hand side of (3.4) yields

$$\int_{\mathcal{S}} \delta\boldsymbol{\xi}^{\mathsf{T}}\ddot{\boldsymbol{\xi}}\mathrm{d}m = -\int_{\mathcal{S}} \delta\dot{\boldsymbol{\xi}}^{\mathsf{T}}\dot{\boldsymbol{\xi}}\mathrm{d}m + \frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathcal{S}} \delta\boldsymbol{\xi}^{\mathsf{T}}\dot{\boldsymbol{\xi}}\mathrm{d}m\right), \tag{3.5}$$

which eliminates the particle accelerations $\ddot{\boldsymbol{\xi}}$. Furthermore, a more compact form can be obtained by using the definition of the **kinetic energy** T, which is given by

$$T := \frac{1}{2}\int_{\mathcal{S}} \dot{\boldsymbol{\xi}}^{\mathsf{T}}\dot{\boldsymbol{\xi}}\mathrm{d}m. \tag{3.6}$$

Indeed, from the definition (3.6) it follows that the first term on the right hand side of (3.5) is equal to $-\delta T$. Using this fact and (3.5) in (3.4) results in

$$\delta W[\delta\boldsymbol{\xi}] = -\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathcal{S}} \delta\boldsymbol{\xi}^{\mathsf{T}}\dot{\boldsymbol{\xi}}\mathrm{d}m\right) + \delta T + \int_{\mathcal{S}} \delta\boldsymbol{\xi}^{\mathsf{T}}\mathrm{d}\mathbf{F} = 0 \ \ \forall\ \delta\boldsymbol{\xi}, \tag{3.7}$$

which is known as **Lagrange's central equation** [21].

Generalized coordinates

In the following, the degrees of freedom are restricted to a finite number f, such that all particle coordinates can be determined from a finite set of (minimal) generalized coordinates $\mathbf{q} \in \mathbb{R}^f$. This restriction requires holonomic constraints (of the form $g(\boldsymbol{\xi}) = 0$) to be imposed on the possibly infinite dimensional model in particle coordinates. For example, setting the distance between any two particles in the Euclidean space to a constant value yields a rigid body model. In that case, the particle positions can be expressed as $\mathbf{r}(\mathbf{X}, \mathbf{q}(t), t)$, such that $\boldsymbol{\xi}(\mathbf{X}, t) = \mathbf{r}(\mathbf{X}, \mathbf{q}(t), t)$ describes the constrained motion. Equivalently, a set of (non-minimal) generalized coordinates $\mathbf{z} \in \mathbb{R}^h$ with $h \geq f$ might be used. The space $\mathbb{R}^h$ is then referred to as the **configuration space**. However, if additional constraints restrict the motion to a subspace or manifold of the configuration space, then that subspace is referred to as configuration manifold.

Using minimal coordinates $\mathbf{q} \in \mathbb{R}^f$, it follows from

$$\dot{\mathbf{r}} = \frac{\partial \mathbf{r}}{\partial \mathbf{q}}\dot{\mathbf{q}} + \frac{\partial \mathbf{r}}{\partial t}, \tag{3.8}$$

by taking the derivative with respect to $\dot{\mathbf{q}}$ on both sides, that

$$\frac{\partial \dot{\mathbf{r}}}{\partial \dot{\mathbf{q}}} = \frac{\partial \mathbf{r}}{\partial \mathbf{q}}. \tag{3.9}$$

Hence, the kinetic energy $T = \frac{1}{2}\int_{\mathcal{S}} \dot{\mathbf{r}}^{\mathsf{T}}\dot{\mathbf{r}}\mathrm{d}m$ may be expressed in terms of minimal coordinates and velocities, which takes the form

$$T(\mathbf{q}, \dot{\mathbf{q}}, t) = T_2(\mathbf{q}, \dot{\mathbf{q}}, t) + T_1(\mathbf{q}, \dot{\mathbf{q}}, t) + T_0(t), \tag{3.10}$$

with the terms $T_2 = \frac{1}{2}\dot{\mathbf{q}}^{\mathsf{T}}\mathbf{M}(\mathbf{q}, t)\dot{\mathbf{q}}$, $T_1 = \mathbf{a}(\mathbf{q}, t)^{\mathsf{T}}\dot{\mathbf{q}}$ and $T_0(t)$. Therein, $\mathbf{M}(\mathbf{q}, t)$ is referred to as **mass matrix**.

Admissible virtual displacements $\delta\mathbf{r}$ are in accordance with all (holonomic) constraints and are induced by the virtual displacements $\delta\mathbf{q}$ through $\delta\mathbf{r} = (\partial\mathbf{r}/\partial\mathbf{q})\delta\mathbf{q} = (\partial\dot{\mathbf{r}}/\partial\dot{\mathbf{q}})\delta\mathbf{q}$. As a result it follows that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathcal{S}} \delta\mathbf{r}^{\mathsf{T}}\dot{\mathbf{r}}\mathrm{d}m\right) &= \frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{S}} \delta\mathbf{q}^{\mathsf{T}}\left(\frac{\partial \mathbf{r}}{\partial \mathbf{q}}\right)^{\mathsf{T}}\dot{\mathbf{r}}\mathrm{d}m\right] \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left[\int_{\mathcal{S}} \delta\mathbf{q}^{\mathsf{T}}\left(\frac{\partial \dot{\mathbf{r}}}{\partial \dot{\mathbf{q}}}\right)^{\mathsf{T}}\left(\frac{\partial(\frac{1}{2}\dot{\mathbf{r}}^{\mathsf{T}}\dot{\mathbf{r}})}{\partial \dot{\mathbf{r}}}\right)^{\mathsf{T}}\mathrm{d}m\right] \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left[\delta\mathbf{q}^{\mathsf{T}}\left(\frac{\partial T}{\partial \dot{\mathbf{q}}}\right)^{\mathsf{T}}\right],
\end{aligned} \tag{3.11}$$

where the chain rule of differentiation has been used to obtain the second line and $T = \frac{1}{2}\int_{\mathcal{S}} \dot{\mathbf{r}}^\mathsf{T}\dot{\mathbf{r}}\mathrm{d}m$ to arrive at the last line. Evaluating (3.7) for admissible virtual displacement and using (3.11) results in

$$\delta W[\delta\mathbf{q}] = -\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{\mathbf{q}}}\delta\mathbf{q}\right)^\mathsf{T} + \delta T + \int_{\mathcal{S}} \delta\mathbf{q}^\mathsf{T}\left(\frac{\partial \mathbf{r}}{\partial \mathbf{q}}\right)^\mathsf{T}\mathrm{d}\mathbf{F} = 0 \ \ \forall\, \delta\mathbf{q}, \tag{3.12}$$

which is a necessary condition for (3.7) to hold.

Principle of virtual action

Up to this point, it was assumed that the generalized velocities are continuously differentiable functions, which is necessary in order for (3.12) to hold. However, on a macroscopic phenomenological level, it can be observed that colliding bodies experience discontinuities in their velocities. In order to incorporate such discontinuities in the description, the continuity requirement therefore has to be relaxed. This can be achieved by generalizing the principle of virtual action, which is a postulate equivalent to the principle of virtual work. More specifically, the **virtual action** δA [28] is defined as the time integral of the virtual work over an arbitrary closed time interval $\mathcal{I}$, i.e.

$$\delta A[\delta\mathbf{q}] := \int_{\mathcal{I}} \delta W[\delta\mathbf{q}]\mathrm{d}t. \tag{3.13}$$

The **principle of virtual action** states that the virtual action for a time interval $\mathcal{I}$ vanishes for arbitrary virtual displacements $\delta\mathbf{q}$,

$$\delta A[\delta\mathbf{q}] = 0 \ \ \forall\, \delta\mathbf{q}. \tag{3.14}$$

It immediately follows from the principle of virtual work, $\delta W[\delta\mathbf{q}] = 0 \ \forall\, \delta\mathbf{q}$, that the integral in (3.13) vanishes and therefore the principle of virtual action (3.14) is fulfilled. Conversely, the principle of virtual action implies the principle of virtual work and the two principles are therefore equivalent. Indeed, the first two terms on the right hand side of (3.12) can be rewritten as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{\mathbf{q}}}\delta\mathbf{q}\right)^\mathsf{T} - \delta T &= \left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{\mathbf{q}}}\right)\delta\mathbf{q} + \frac{\partial T}{\partial \dot{\mathbf{q}}}(\delta\mathbf{q})^{\cdot}\right]^\mathsf{T} - \left(\frac{\partial T}{\partial \mathbf{q}}\delta\mathbf{q} + \frac{\partial T}{\partial \dot{\mathbf{q}}}\delta\dot{\mathbf{q}}\right)^\mathsf{T} \\ &= \delta\mathbf{q}^\mathsf{T}\left[\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{\mathbf{q}}} - \frac{\partial T}{\partial \mathbf{q}}\right]^\mathsf{T},\end{aligned} \tag{3.15}$$

where the fact that $(\delta\mathbf{q})^{\cdot} = \delta\dot{\mathbf{q}}$ has been used to arrive at the last equality. The third term on the right hand side of (3.12) can be simplified by defining the generalized forces $\mathbf{f}$ as

$$\mathbf{f} := \int_{\mathcal{S}} \left(\frac{\partial \mathbf{r}}{\partial \mathbf{q}}\right)^{\mathsf{T}} \mathrm{d}\mathbf{F}. \tag{3.16}$$

Integrating (3.12), using (3.15) and (3.16) and applying the principle of virtual action leads to

$$\begin{aligned}
\delta A[\delta\mathbf{q}] &= -\int_{\mathcal{I}} \left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{\mathbf{q}}}\delta\mathbf{q}\right)^{\mathsf{T}} - \delta T - \int_{\mathcal{S}} \delta\mathbf{q}^{\mathsf{T}} \left(\frac{\partial \mathbf{r}}{\partial \mathbf{q}}\right)^{\mathsf{T}} \mathrm{d}\mathbf{F}\right] \mathrm{d}t && \text{(3.17a)}\\
&= -\int_{\mathcal{I}} \delta\mathbf{q}^{\mathsf{T}} \left[\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{\mathbf{q}}} - \frac{\partial T}{\partial \mathbf{q}}\right]^{\mathsf{T}} - \delta\mathbf{q}^{\mathsf{T}}\mathbf{f}\,\mathrm{d}t = 0 \;\; \forall\, \delta\mathbf{q} && \text{(3.17b)}\\
&= -\int_{\mathcal{I}} \delta\mathbf{q}^{\mathsf{T}} \left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{\mathbf{q}}}\right)^{\mathsf{T}} - \left(\frac{\partial T}{\partial \mathbf{q}}\right)^{\mathsf{T}} - \mathbf{f}\right] \mathrm{d}t = 0 \;\; \forall\, \delta\mathbf{q}. && \text{(3.17c)}
\end{aligned}$$

Finally, it follows from the fundamental lemma of the calculus of variations that the square bracket in (3.17c) vanishes for all t, which implies that the virtual work also vanishes for all t.

The reason why the principle of virtual action was introduced, is that it allows for a natural generalization to impulsive systems with discontinuous generalized velocities [28]. To see this, denote the generalized velocities as $\mathbf{u} := \dot{\mathbf{q}}$. Then (3.17c) can be reformulated as

$$\delta A[\delta\mathbf{q}] = -\int_{\mathcal{I}} \delta\mathbf{q}^{\mathsf{T}} \left[\mathrm{d}\left(\frac{\partial T}{\partial \mathbf{u}}\right)^{\mathsf{T}} - \left(\frac{\partial T}{\partial \mathbf{q}}\right)^{\mathsf{T}} \mathrm{d}t - \mathrm{d}\mathbf{R}\right] = 0 \;\; \forall\, \delta\mathbf{q}, \tag{3.18}$$

where $\mathrm{d}\mathbf{R} := \mathbf{f}\mathrm{d}t$ has been introduced and $\mathrm{d}(\partial T/\partial \mathbf{u}) = (\partial T/\partial \mathbf{u})^{\cdot}\mathrm{d}t$ has been used (refer to the discussion on densities in Appendix A.2 for details). Since (3.18) holds for all $\delta\mathbf{q}$, it particularly holds for virtual displacements with compact support on $\mathcal{I}$. In that case (3.18) can be written in short notation as

$$\mathrm{d}\left(\frac{\partial T}{\partial \mathbf{u}}\right)^{\mathsf{T}} - \left(\frac{\partial T}{\partial \mathbf{q}}\right)^{\mathsf{T}} \mathrm{d}t - \mathrm{d}\mathbf{R} = 0, \tag{3.19}$$

which is an **equality of measures**. The equality (3.19) remains meaningful for generalized velocities $\mathbf{u}$ that are special functions of locally bounded variation, written as $\mathbf{u} \in \mathrm{slbv}(\mathcal{I}, \mathbb{R}^f)$ (as defined in Appendix A.1). As such, $\mathbf{u}(t)$ can be discontinuous, with a countable set of discontinuity points. Furthermore, the left limit $\mathbf{u}^-(t)$ and the right limit $\mathbf{u}^+(t)$ of the generalized velocities exist for all t. Moreover, the generalized velocities $\mathbf{u}$ correspond to the time derivative of $\mathbf{q}$ wherever they are defined. As a consequence, the kinetic energy $T \in \mathrm{slbv}(\mathcal{I}, \mathbb{R})$ is also a special function of locally bounded variation. This implies $\mathrm{d}T = \dot{T}\mathrm{d}t + (T^+ - T^-)\mathrm{d}\eta$, i.e. T admits an additional density with respect to the atomic measure η which is defined as the finite sum of Dirac point measures at the continuity points $\{t_k\}$ (see Section 2.2), i.e.

$$\eta = \sum_k \delta_{t_k} \quad \text{with} \quad \delta_{t_k}(\mathcal{I}) = \begin{cases} 1 & \text{if } t_k \in \mathcal{I} \\ 0 & \text{if } t_k \notin \mathcal{I}. \end{cases} \tag{3.20}$$

Since $T \in \mathrm{slbv}(\mathcal{I}, \mathbb{R})$, its derivatives are also special functions of locally bounded variation, such that we have $(\partial T / \partial \mathbf{u}) \in \mathrm{slbv}(\mathcal{I}, \mathbb{R})$, which implies

$$\mathrm{d}\left(\frac{\partial T}{\partial \mathbf{u}}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \mathbf{u}}\right)\mathrm{d}t + \left[\left(\frac{\partial T}{\partial \mathbf{u}}\right)^+ - \left(\frac{\partial T}{\partial \mathbf{u}}\right)^-\right]\mathrm{d}\eta. \tag{3.21}$$

As a consequence, it is natural to generalize $\mathrm{d}\mathbf{R}$ to take the form $\mathrm{d}\mathbf{R} = \mathbf{f}\mathrm{d}t + \mathbf{f}_\eta \mathrm{d}\eta$ with some density $\mathbf{f}_\eta$, which is referred to as generalized impulsive force.

Therefore, the the generalization of (3.18) to nonsmooth systems can be written as

$$\begin{aligned} \delta A[\delta \mathbf{q}] = &\int_{\mathcal{I}} \delta \mathbf{q}^\mathsf{T} \left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \mathbf{u}}\right) - \frac{\partial T}{\partial \mathbf{q}} - \mathbf{f}\right]\mathrm{d}t \\ &+ \int_{\mathcal{I}} \delta \mathbf{q}^\mathsf{T} \left[\left(\frac{\partial T}{\partial \mathbf{u}}\right)^+ - \left(\frac{\partial T}{\partial \mathbf{u}}\right)^- - \mathbf{f}_\eta\right]\mathrm{d}\eta = 0 \ \ \forall\, \delta \mathbf{q}. \end{aligned} \tag{3.22}$$

From the fact that (3.22) holds for arbitrary $\delta\mathbf{q}$, it follows that the square brackets vanish at all times, leading to the **Lagrangian equations of the second kind**

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \mathbf{u}}\right) - \frac{\partial T}{\partial \mathbf{q}} = \mathbf{f}, \tag{3.23}$$

describing the non-impulsive motion, and their impulsive counterpart

$$\left(\frac{\partial T}{\partial \mathbf{u}}\right)^+ - \left(\frac{\partial T}{\partial \mathbf{u}}\right)^- = \mathbf{f}_\eta. \tag{3.24}$$

Note that the two equations (3.23) and (3.24) together are equivalent to the equality of measures (3.19).

After evaluating the derivatives and grouping terms, equation (3.23) then leads to the equations of motion of the form

$$\mathbf{M}(\mathbf{q}, t)\dot{\mathbf{u}} - \mathbf{h}_g(\mathbf{q}, \mathbf{u}, t) = \mathbf{f}. \tag{3.25}$$

It follows directly from (3.10), i.e. $T = T_2 + T_1 + T_0$, that the equations of motion include the term $\mathbf{M}(\mathbf{q}, t)\dot{\mathbf{u}}$, since $T_2 = \frac{1}{2}\mathbf{u}^\mathsf{T}\mathbf{M}(\mathbf{q}, t)\mathbf{u}$. In particular, if the particle coordinates are scleronomic, i.e. not explicitly depending on time, the kinetic energy is of the quadratic form $T(\mathbf{q},\mathbf{u}) = T_2(\mathbf{q},\mathbf{u}) = \frac{1}{2}\mathbf{u}^\mathsf{T}\mathbf{M}(\mathbf{q})\mathbf{u}$, with a mass matrix which does not explicitly depend on time. All other terms resulting form taking the derivatives in (3.23) are referred to as gyroscopic accelerations and are gathered in $\mathbf{h}_g(\mathbf{q},\mathbf{u},t)$. Similarly, equation (3.24) leads to the impact equations of the form

$$\mathbf{M}(\mathbf{q},t)(\mathbf{u}^+ - \mathbf{u}^-) = \mathbf{f}_\eta, \tag{3.26}$$

and the equality of measures (3.19) results in

$$\mathbf{M}(\mathbf{q},t)\,\mathrm{d}\mathbf{u} - \mathbf{h}_g(\mathbf{q},\mathbf{u},t)\,\mathrm{d}t = \mathrm{d}\mathbf{R}. \tag{3.27}$$

3.2 Constitutive contact laws

The equality of measures is to be complemented with constitutive laws, in particular contact laws. They include set-valued force laws, used to model contact forces such as friction and unilateral constraint forces, as well as normal and frictional impact laws, describing velocity changes due to collisions. It is convenient to formulate separate force laws for the normal and tangential components of the contact forces and to express them as normal cone inclusions.

Two bodies are in contact if two points, one on the surface of each body, are in contact. For two bodies $\mathcal{B}_1$ and $\mathcal{B}_2$ that can be in contact, let P be the body-fixed point on $\mathcal{B}_1$, which momentarily (i.e. at a fixed time instant) has the shortest distance to $\mathcal{B}_2$. Similarly, let Q be the body-fixed point on $\mathcal{B}_2$ that momentarily has the shortest distance to $\mathcal{B}_1$. The contact distance (or gap) g_N is the signed distance between P and Q. In other words, the two bodies are in contact if $g_N = 0$, they are separated if $g_N > 0$ and they penetrate each other if $g_N < 0$. Assuming that the surfaces of all bodies are smooth, the connecting line between P and Q is always normal to both surfaces.

Let $(\mathbf{n},\mathbf{t}_1,\mathbf{t}_2)$ be an orthonormal body-fixed frame attached to point P, such that $\mathbf{n}(\mathbf{q},t)$ is the unit outward normal in P and $\mathbf{t}_1(\mathbf{q},t)$ and $\mathbf{t}_2(\mathbf{q},t)$ span the tangential plane to the body surface in P, as shown in Figure 3.1. The relative position $\mathbf{r}_{PQ}$ of Q with respect to P can then be written as

$$\mathbf{r}_{PQ} = g_N\mathbf{n}, \tag{3.28}$$

which does not contain any tangential components, as P and Q always lie on a connecting line along $\mathbf{n}$. The *contact distance* in normal direction is therefore given by the projection

$$g_N := \mathbf{n}^\mathsf{T}\mathbf{r}_{PQ}. \tag{3.29}$$

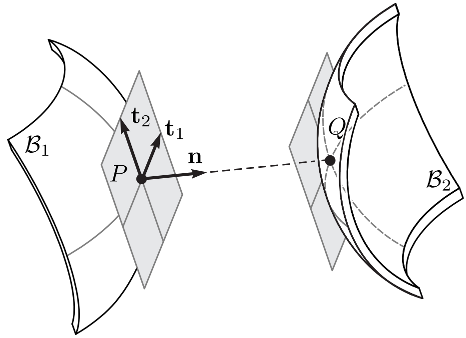

Figure 3.1: Contact geometry.

The relative velocity $\boldsymbol{\gamma} := \mathbf{v}_Q - \mathbf{v}_P$ between the contact points, referred to as *contact velocity*, is decomposed into the components

$$\mathbf{v}_Q - \mathbf{v}_P = \dot{\mathbf{r}}_{PQ} = \gamma_N \mathbf{n} + \gamma_{T_1} \mathbf{t}_1 + \gamma_{T_2} \mathbf{t}_2. \tag{3.30}$$

Therefore, the relative velocity in normal direction is the projection

$$\gamma_N = \mathbf{n}^\mathsf{T} (\mathbf{v}_Q - \mathbf{v}_P), \tag{3.31}$$

whereas the relative velocity in tangential direction, i.e. the projection onto the $\mathbf{t}_1$-$\mathbf{t}_2$-plane, is given by

$$\boldsymbol{\gamma}_T := (\gamma_{T_1}\ \gamma_{T_2})^\mathsf{T} = (\mathbf{t}_1\ \ \mathbf{t}_2)^\mathsf{T} (\mathbf{v}_Q - \mathbf{v}_P). \tag{3.32}$$

The contact forces $\mathbf{F}_P$, acting on $\mathcal{B}_1$ in P, and $\mathbf{F}_Q$, acting on $\mathcal{B}_2$ in Q, are decomposed in a similar fashion as

$$\mathbf{F}_Q = -\mathbf{F}_P = \lambda_N \mathbf{n} + \lambda_{T_1} \mathbf{t}_1 + \lambda_{T_2} \mathbf{t}_2, \tag{3.33}$$

where, due to the law of interaction, the contact forces are of equal magnitude but of opposite direction. In view of their incorporation into the equality of measures (3.27), it is convenient to also decompose the virtual work δW^C of the contact forces into a normal and a tangential part. To do this, first calculate δW^C as

$$\delta W^C = \delta \mathbf{r}_{OP}^\mathsf{T} \mathbf{F}_P + \delta \mathbf{r}_{OQ}^\mathsf{T} \mathbf{F}_Q = (\delta \mathbf{r}_{OQ} - \delta \mathbf{r}_{OP})^\mathsf{T} \mathbf{F}_Q = \left(\frac{\partial \mathbf{r}_{PQ}}{\partial \mathbf{q}} \delta \mathbf{q} \right)^\mathsf{T} \mathbf{F}_Q. \tag{3.34}$$

Because of $\dot{\mathbf{r}}_{PQ} = (\partial \mathbf{r}_{PQ}/\partial \mathbf{q})\dot{\mathbf{q}} + (\partial \mathbf{r}_{PQ}/\partial t)$ it follows that $\partial \mathbf{r}_{PQ}/\partial \mathbf{q} = \partial \dot{\mathbf{r}}_{PQ}/\partial \dot{\mathbf{q}}$. Therefore, with (3.30) and (3.33), the virtual work δW^C equates to

$$\begin{aligned} \delta W^C &= \delta \mathbf{q}^\mathsf{T} \left(\frac{\partial \dot{\mathbf{r}}_{PQ}}{\partial \dot{\mathbf{q}}} \right)^\mathsf{T} \mathbf{F}_Q \\ &= \delta \mathbf{q}^\mathsf{T} \left(\mathbf{n} \frac{\partial \gamma_N}{\partial \dot{\mathbf{q}}} + \mathbf{t}_1 \frac{\partial \gamma_{T_1}}{\partial \dot{\mathbf{q}}} + \mathbf{t}_2 \frac{\partial \gamma_{T_2}}{\partial \dot{\mathbf{q}}} \right)^\mathsf{T} (\lambda_N \mathbf{n} + \lambda_{T_1} \mathbf{t}_1 + \lambda_{T_2} \mathbf{t}_2) \\ &= \delta \mathbf{q}^\mathsf{T} \left[\left(\frac{\partial g_N}{\partial \mathbf{q}} \right)^\mathsf{T} \lambda_N + \left(\frac{\partial \gamma_{T_1}}{\partial \dot{\mathbf{q}}} \right)^\mathsf{T} \lambda_{T_1} + \left(\frac{\partial \gamma_{T_2}}{\partial \dot{\mathbf{q}}} \right)^\mathsf{T} \lambda_{T_2} \right], \end{aligned} \tag{3.35}$$

where the fact that $\gamma_N = \dot{g}_N$ was used to obtain the last equality. Finally, the virtual work $\delta W^N := \delta \mathbf{q}^\mathsf{T} (\partial g_N / \partial \mathbf{q})^\mathsf{T} \lambda_N$ of the normal components can be introduced and all the remaining terms are summarized in δW^T, such that

$$\delta W^C = \delta W^N + \delta W^T. \tag{3.36}$$

Multiple contacts, i.e. n pairs of contact points on the surfaces of two contacting bodies, lead to multiple contact distances, which are denoted g_N^i for the i-th contact and summarized in $\boldsymbol{g}_N := (g_N^1, \cdots, g_N^n)^\mathsf{T}$.

Unilateral constraints

Unilateral constraints are inequality constraints of the form $\boldsymbol{g}_N(\mathbf{q}, t) \geq \mathbf{0}$, where the notation $\boldsymbol{g}_N = (g_N^1, \cdots, g_N^n)^\mathsf{T} \geq \mathbf{0}$ is used to express that every component $g_N^i \geq 0,\ i \in \{1, \cdots, n\}$. These unilateral constraints correspond to not admitting penetration between contacting bodies, with negative values representing penetration, which shall be inadmissible. To prevent penetration of contacting bodies, contact forces λ_N^i are required in all contacts. The assumption that no adhesion and no distance effects are present between contacting bodies, leads to the constitutive law

$$\mathbf{0} \leq \boldsymbol{\lambda}_N \perp \boldsymbol{g}_N \geq \mathbf{0}, \tag{3.37}$$

where the notation $\boldsymbol{\lambda}_N \perp \boldsymbol{g}_N$ is used to express the orthogonality $\boldsymbol{\lambda}_N^\mathsf{T} \boldsymbol{g}_N = 0$. Hence, (3.37) is equivalent to $g_N^i \geq 0, \lambda_N^i \geq 0, g_N^i \lambda_N^i = 0$ for all i and is therefore referred to as an **inequality complementarity**. The law (3.37) is often called **Signorini condition** and by (2.26) can be written as a normal cone inclusion of the form

$$-\boldsymbol{\lambda}_N \in \mathcal{N}_{\mathbb{R}_0^{n+}}(\boldsymbol{g}_N), \tag{3.38}$$

where the set $\mathbb{R}_0^{n+} := \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{y} \geq \mathbf{0}\}$ was used. As before, (3.38) has to be understood component-wise, i.e. $-\lambda_N^i \in \mathcal{N}_{\mathbb{R}_0^+}(g_N^i)$ for all $i \in \{1, \cdots, n\}$.

As shown in [41, chap. 7], the force law (3.38) can be equivalently be formulated on velocity level, using *contact velocities* $\gamma_N^i := (\partial g_N^i/\partial \mathbf{q})^\mathsf{T}\mathbf{u} + (\partial g_N^i/\partial t)$. The reformulation is due to the continuity properties of $\mathbf{q}(t)$ and $\mathbf{u}^\pm(t)$ and reads

$$\begin{aligned} g_N^i = 0: & \quad -\lambda_N^i \in \mathcal{N}_{\mathbb{R}_0^+}(\gamma_N^i), \\ g_N^i > 0: & \quad \lambda_N^i = 0, \end{aligned} \tag{3.39}$$

which is convenient in view of the time discretization that will follow in Section 3.4.

The generalized normal contact force $\mathbf{f}_N$ can be obtained by a balance of the virtual work of the normal contact forces. By using the components in (3.36) one obtains

$$\delta W^N = \delta\mathbf{q}^\mathsf{T}\mathbf{f}_N = \delta\mathbf{q}^\mathsf{T}\left(\frac{\partial \boldsymbol{g}_N}{\partial \mathbf{q}}\right)^\mathsf{T}\boldsymbol{\lambda}_N \quad \forall\, \delta\mathbf{q}. \tag{3.40}$$

As a consequence, the generalized normal contact force $\mathbf{f}_N$ has the form

$$\mathbf{f}_N = \mathbf{W}_N\boldsymbol{\lambda}_N, \tag{3.41}$$

where $\mathbf{W}_N := (\partial \boldsymbol{g}_N/\partial\mathbf{q})^\mathsf{T}$ has been used to denote the so-called generalized force directions.

With (3.41) a force law for the generalized force $\mathbf{f}_N$ can be deduced as a normal cone inclusion as well. To see this, the set

$$\mathcal{A} := \{\mathbf{q} \in \mathbb{R}^f \mid \boldsymbol{g}_N(\mathbf{q}) \geq \mathbf{0}\} \tag{3.42}$$

of admissible generalized coordinates is introduced. In general, $\mathcal{A}$ can be non-convex, which requires a more general definition of the normal cone than for convex sets. Here, the constraint functions g_N^i are assumed to be of class C^1 and to fulfill the additional regularity condition that all level curves $g_N^i(\mathbf{q}) = 0$ intersect transversally. Roughly speaking, one then obtains a convex set when zooming in on any point of the set $\mathcal{A}$. Since $\mathbb{R}_0^{n+}$ is a convex cone, it follows from the definition of the normal cone (2.25) that

$$\begin{aligned} \mathcal{N}_{\mathbb{R}_0^{n+}}(\boldsymbol{g}_N) &= \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{y}^\mathsf{T}(\boldsymbol{g}_N^* - \boldsymbol{g}_N) \leq 0 \;\; \forall\, \boldsymbol{g}_N^* \in \mathbb{R}_0^{n+}\} \\ &= \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{y}^\mathsf{T}(\boldsymbol{g}_N^* - \boldsymbol{g}_N) \leq 0 \;\; \forall\, \boldsymbol{g}_N^* \in \lim_{\varepsilon\to 0+} \mathcal{B}_\varepsilon(\boldsymbol{g}_N, \mathbb{R}_0^{n+})\}, \end{aligned} \tag{3.43}$$

where the notation $\mathcal{B}_\varepsilon(\mathbf{x}, \mathcal{K}) := \{\mathbf{y} \in \mathcal{K} \mid \|\mathbf{x} - \mathbf{y}\| \leq \varepsilon\}$ was used for the intersection of the ball with constant radius $\varepsilon > 0$, around a point $\mathbf{x} \in \mathbb{R}^n$, with a general subset $\mathcal{K} \subset \mathbb{R}^n$. Note that the restriction of $\boldsymbol{g}_N^*$ to $\mathcal{B}_\varepsilon(\boldsymbol{g}_N, \mathbb{R}_0^{n+})$ is sufficient since $\mathbb{R}_0^{n+}$ is a cone. Next, by the definition of continuity, for every $\varepsilon > 0$ there exists a $\delta(\varepsilon) > 0$ such that the implication $\|\mathbf{q}^* - \mathbf{q}\| \leq \delta \Rightarrow \|\boldsymbol{g}_N(\mathbf{q}^*) - \boldsymbol{g}_N(\mathbf{q})\| \leq \varepsilon$ holds. Hence, by expressing the constraints $\boldsymbol{g}_N(\mathbf{q})$ as functions of the generalized coordinates and using their continuity, the normal cone can equivalently be written as

$$\begin{aligned}\mathcal{N}_{\mathbb{R}_0^{n+}}(\boldsymbol{g}_N(\mathbf{q})) &= \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{y}^\mathsf{T}(\boldsymbol{g}_N(\mathbf{q}^*) - \boldsymbol{g}_N(\mathbf{q})) \leq 0 \;\; \forall\, \mathbf{q}^* \in \lim_{\delta \to 0+} \mathcal{B}_\delta(\mathbf{q}, \mathcal{A})\} \\ &= \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{y}^\mathsf{T} \frac{\partial \boldsymbol{g}_N(\mathbf{q})}{\partial \mathbf{q}}(\mathbf{q}^* - \mathbf{q}) \leq 0 \;\; \forall\, \mathbf{q}^* \in \lim_{\delta \to 0+} \mathcal{B}_\delta(\mathbf{q}, \mathcal{A})\},\end{aligned} \tag{3.44}$$

where a first order approximation for the difference $\boldsymbol{g}_N(\mathbf{q}^*) - \boldsymbol{g}_N(\mathbf{q})$ was used to obtain the last equation. With $\mathbf{W}_N = (\partial \boldsymbol{g}_N / \partial \mathbf{q})^\mathsf{T}$, it follows from (3.41) and (3.44) that the force law for $\mathbf{f}_N$ is given by the normal cone inclusion

$$-\mathbf{f}_N \in \mathcal{N}_\mathcal{A}(\mathbf{q}), \tag{3.45}$$

where the normal cone to the (possibly non-convex) set $\mathcal{A}$ is defined as

$$\mathcal{N}_\mathcal{A}(\mathbf{q}) := \{\mathbf{y} \in \mathbb{R}^f \mid \mathbf{y}^\mathsf{T}(\mathbf{q}^* - \mathbf{q}) \leq 0 \;\; \forall\, \mathbf{q}^* \in \lim_{\varepsilon \to 0+} \mathcal{B}_\varepsilon(\mathbf{q}, \mathcal{A})\}. \tag{3.46}$$

Friction

A widely used law for modeling friction forces is Coulomb's friction law. In its spatial formulation, it gives a force law for the components $\boldsymbol{\lambda}_T^i = (\lambda_{T_1} \quad \lambda_{T_2})^\mathsf{T}$ of the i-th frictional contact in the $\mathbf{t}_1$-$\mathbf{t}_2$-plane, and can be formulated as a normal cone inclusion for closed contacts

$$\begin{aligned} g_N^i = 0: \;\; & \boldsymbol{\gamma}_T^i \in \mathcal{N}_{\mathcal{C}_T(\lambda_N^i)}(-\boldsymbol{\lambda}_T^i), \\ g_N^i > 0: \;\; & \boldsymbol{\lambda}_T^i = \mathbf{0}. \end{aligned} \tag{3.47}$$

Therein, $\boldsymbol{\gamma}_T^i$ is the relative sliding velocity and the set $\mathcal{C}_T$ describes the set of admissible friction forces $-\boldsymbol{\lambda}_T$. For isotropic Coulomb friction, it describes a disc $\mathcal{C}_T(\boldsymbol{\lambda}_N^i) = \{-\boldsymbol{\lambda}_T \mid \|\boldsymbol{\lambda}_T\| \leq \mu^i \lambda_N^i\}$ with a radius depending on the normal contact force λ_N^i and a constant friction coefficient μ^i. However, other shapes of the set $\mathcal{C}_T$ might be required. For example, an elliptic set $\mathcal{C}_T$ can be used to obtain an orthotropic friction law.

Similar to the normal contact forces, the generalized friction force $\mathbf{f}_T^i$ of the i-th contact takes the form

$$\mathbf{f}_T^i = \mathbf{W}_T^i \boldsymbol{\lambda}_T^i, \tag{3.48}$$

with $\mathbf{W}_T^i := (\partial \boldsymbol{\gamma}_T^i / \partial \mathbf{u})^\mathsf{T}$ denoting the generalized friction force directions. Indeed, by using the tangential components in (3.36), the balance of the virtual work reads

$$\delta W^{T,i} = \delta \mathbf{q}^\mathsf{T} \mathbf{f}_T^i = \delta \mathbf{q}^\mathsf{T} \left(\frac{\partial \boldsymbol{\gamma}_T^i}{\partial \mathbf{u}} \right)^\mathsf{T} \boldsymbol{\lambda}_T^i \;\; \forall\, \delta \mathbf{q}, \tag{3.49}$$

from which (3.48) directly follows.

Impact law

On a macroscopic phenomenological level, it is observed that collisions between bodies in a mechanical multibody system lead to instantaneous velocity jumps. These velocity discontinuities require impulsive forces Λ_N^i, acting normal to the contact plane of the i-th contact. The simultaneous occurrence of a velocity jump with a corresponding impulse is referred to as an **impact**. In that case, an impact law relates the post-impact relative velocities $\boldsymbol{\gamma}_N^+$ to the pre-impact relative velocities $\boldsymbol{\gamma}_N^-$ in normal direction to the contact plane. A widely used impact law is the generalized Newtonian impact law [40], which accounts for superfluous contacts (i.e. inactive in the sense that $\Lambda_N^i = 0$ and is formulated with $\boldsymbol{\xi}_N := \boldsymbol{\gamma}_N^+ + \boldsymbol{\varepsilon}\boldsymbol{\gamma}_N^-$ component-wise as

$$\begin{aligned} g_N^i &= 0 : 0 \leq \Lambda_N^i \perp \xi_N^i \geq 0, \\ g_N^i &> 0 : \Lambda_N^i = 0, \end{aligned} \tag{3.50}$$

where $\boldsymbol{\varepsilon} = \operatorname{diag}\{\varepsilon_i\}$ with constant coefficients of restitution $\varepsilon_i \in [0\ 1]$. With this, whenever a contact is active in the sense that $\Lambda_N^i > 0$, it follows that the relative normal contact velocity is reversed and scaled according to $\gamma_N^{i+} = -\varepsilon_i \gamma_N^{i-}$. The impact law (3.50) can alternatively be written as a normal cone inclusion for closed contacts, as in

$$\begin{aligned} g_N^i &= 0 : \xi_N^i \in \mathcal{N}_{\mathbb{R}_0^-}(-\Lambda_N^i), \\ g_N^i &> 0 : \Lambda_N^i = 0. \end{aligned} \tag{3.51}$$

In the case of frictional contacts, impulsive tangential forces $\boldsymbol{\Lambda}_T^i$ lead to tangential velocity discontinuities accordingly. A corresponding impact law [42] can be formulated with $\boldsymbol{\xi}_T^i = \boldsymbol{\gamma}_T^{i+} + \varepsilon_T^i \boldsymbol{\gamma}_T^{i-}$ similar to the normal direction as

$$\begin{aligned} g_N^i &= 0 : \boldsymbol{\xi}_T^i \in \mathcal{N}_{\mathcal{C}_T(\Lambda_N^i)}(-\boldsymbol{\Lambda}_T^i), \\ g_N^i &> 0 : \Lambda_T^i = 0, \end{aligned} \tag{3.52}$$

where $\mathcal{C}_T(\Lambda_N^i) = \{-\boldsymbol{\Lambda}_T \mid \|\boldsymbol{\Lambda}_T\| \leq \mu^i \Lambda_N^i\}$.

3.3 Measure differential inclusion

Using the structure of the generalized contact forces $\mathbf{f}_C = \mathbf{f}_N + \sum_i \mathbf{f}_T^i = \mathbf{W}_N \boldsymbol{\lambda}_N + \sum_i \mathbf{W}_T^i \boldsymbol{\lambda}_T^i$ and impulses $\mathbf{f}_\eta = \mathbf{W}_N \boldsymbol{\Lambda}_N + \mathbf{W}_T \boldsymbol{\Lambda}_T$ in the equality of measures (3.27) results in

$$\begin{aligned} \mathrm{d}\mathbf{q} &= \mathbf{u}\,\mathrm{d}t, \\ \mathbf{M}\,\mathrm{d}\mathbf{u} - \mathbf{h}(t, \mathbf{q}, \mathbf{u})\,\mathrm{d}t &= \mathbf{W}_N\,\mathrm{d}\mathbf{P}_N + \mathbf{W}_T\,\mathrm{d}\mathbf{P}_T, \end{aligned} \tag{3.53}$$

where $\mathbf{h}$ contains the gyroscopic accelerations $\mathbf{h}_g$ as well as all generalized forces which are not due to the contact forces. The matrix $\mathbf{W}_N$ is composed of the generalized force directions of all unilateral constraints and the matrix $\mathbf{W}_T = (\mathbf{W}_T^1 \cdots \mathbf{W}_T^m)$ contains the generalized force directions of all m frictional contacts. The percussion measures $\mathrm{d}\mathbf{P}_N$ and $\mathrm{d}\mathbf{P}_T$ are composed of the non-impulsive constraint forces $\boldsymbol{\lambda}_N$ and $\boldsymbol{\lambda}_T = (\boldsymbol{\lambda}_T^{1\mathsf{T}} \cdots \boldsymbol{\lambda}_T^{m\mathsf{T}})^\mathsf{T}$ as well as the corresponding impulsive forces $\boldsymbol{\Lambda}_N$ and $\boldsymbol{\Lambda}_T$, according to

$$\begin{aligned} \mathrm{d}\mathbf{P}_N &= \boldsymbol{\lambda}_N \,\mathrm{d}t + \boldsymbol{\Lambda}_N \,\mathrm{d}\eta\,, \\ \mathrm{d}\mathbf{P}_T &= \boldsymbol{\lambda}_T \,\mathrm{d}t + \boldsymbol{\Lambda}_T \,\mathrm{d}\eta. \end{aligned} \tag{3.54}$$

The components of all constraint forces and impulses are determined by the force laws (3.39), (3.47), (3.51) and (3.52).

For an even more compact formulation, the percussion measures can be summarized in $\mathrm{d}\mathbf{P} := (\mathrm{d}\mathbf{P}_N^\mathsf{T}\ \mathrm{d}\mathbf{P}_T^\mathsf{T})^\mathsf{T}$. With $\mathbf{W} = (\mathbf{W}_N\ \mathbf{W}_T)$ it then follows that

$$\begin{aligned} \mathrm{d}\mathbf{q} &= \mathbf{u}\,\mathrm{d}t, \\ \mathbf{M}\,\mathrm{d}\mathbf{u} - \mathbf{h}(t,\mathbf{q},\mathbf{u})\,\mathrm{d}t &= \mathbf{W}\,\mathrm{d}\mathbf{P}. \end{aligned} \tag{3.55}$$

3.4 Time discretization

In the following, two specific time discretization schemes for measure differential inclusions of the form (3.55) are discussed. Moreau's time-stepping scheme is perhaps the most widely used scheme within the Nonsmooth Dynamics community as it can be applied for the simulation of systems with multiple unilateral constraints with Coulomb friction. It directly discretizes the equality of measures and a combined contact-impact law on position-switched velocity level. As a consequence, the switching nature of the generalized Newtonian impact law is inherited. Conversely, the less known time discretization scheme of Paoli and Schatzman involves an impact law directly formulated on position level, instead of on position-switched velocity level. As a consequence, velocity changes over impacts take place over two consecutive time steps in the discretization. Both the Moreau and the Paoli-Schatzman scheme are part of the category of event-capturing schemes, that spread the effects of impacts over one or multiple time steps and can therefore overcome accumulation points. Conversely, event-driven schemes use higher order ODE solvers between impacts, but require every impact to be detected such that the impact equations can be solved. Event-driven schemes are therefore not applicable if accumulation points can occur.

Moreau's time-stepping scheme

In order to discretize the equality of measures (3.55), first note that both sides can be understood as measures of a given time interval $\mathcal{I}$, and can therefore be written in the integral form

$$\begin{aligned} \int_{\mathcal{I}} \mathrm{d}\mathbf{q} &= \int_{\mathcal{I}} \mathbf{u}\,\mathrm{d}t, \\ \int_{\mathcal{I}} (\mathbf{M}\,\mathrm{d}\mathbf{u} - \mathbf{h}(t,\mathbf{q},\mathbf{u})\,\mathrm{d}t) &= \int_{\mathcal{I}} \mathbf{W}\,\mathrm{d}\mathbf{P}, \end{aligned} \tag{3.56}$$

for a given time interval $\mathcal{I}$. Furthermore, by introducing $\mathbf{P} := \int_{\mathcal{I}} \mathrm{d}\mathbf{P}$ with the normal and tangential components $\mathbf{P}_N = \int_{\mathcal{I}} \mathrm{d}\mathbf{P}_N$ and $\mathbf{P}_T = \int_{\mathcal{I}} \mathrm{d}\mathbf{P}_T$, the force and impact laws can be gathered in a combined contact-impact law. Under the assumption that $\xi_N^i = \gamma_N^{i+} + \varepsilon_i \gamma_N^{i-}$ is constant on $\mathcal{I}$, the combined contact-impact law for the normal component reads

$$\begin{aligned} g_N^i &= 0: \quad \xi_N^i \in \mathcal{N}_{\mathbb{R}_0^-}(-P_N^i), \\ g_N^i &> 0: \quad P_N^i = 0 \end{aligned} \tag{3.57}$$

and for the tangential component, with $\boldsymbol{\xi}_T^i = \boldsymbol{\gamma}_T^{i+} + \varepsilon_{Ti}\boldsymbol{\gamma}_T^{i-}$ constant on $\mathcal{I}$, it holds that

$$\begin{aligned} g_N^i &= 0: \quad \boldsymbol{\xi}_T^i \in \mathcal{N}_{\mathcal{C}_T(P_N^i)}(-\mathbf{P}_T^i), \\ g_N^i &> 0: \quad \mathbf{P}_T^i = \mathbf{0}. \end{aligned} \tag{3.58}$$

For a given time interval $\mathcal{I} = [t_k, t_{k+1}]$, with $t_{k+1} = t_k + \Delta t$, the first equality of measures in (3.56) is discretized by approximating the integral on the right-hand side using the trapezoidal rule, leading to

$$\mathbf{q}_{k+1} - \mathbf{q}_k \approx \frac{1}{2}(\mathbf{u}_k + \mathbf{u}_{k+1})\Delta t. \tag{3.59}$$

Therein, quantities with an index k refer to apporoximants at time $t = t_k = k\Delta t$, e.g. $\mathbf{q}_k \approx \mathbf{q}(t_k)$. In order to discretize the second equality in (3.56), Moreau's scheme makes use of a midpoint $t_M = t_k + \frac{\Delta t}{2}$ between two consecutive discrete time instants, with $\mathbf{q}_M = \mathbf{q}_k + \mathbf{u}_k \frac{1}{2}\Delta t$. The discretization is

$$\mathbf{M}(\mathbf{q}_M)(\mathbf{u}_{k+1} - \mathbf{u}_k) - \mathbf{h}(t_M, \mathbf{q}_M, \mathbf{u}_k)\Delta t = \mathbf{W}(\mathbf{q}_M, t_M)\mathbf{P}_k, \tag{3.60}$$

where $\mathbf{P}_k$ is an approximant of $\mathbf{P} = \mathbf{P}_N + \mathbf{P}_T$ for $\mathcal{I} = [t_k, t_{k+1}]$. Finally, the discretized contact laws are

$$\begin{aligned} g_N^i(\mathbf{q}_M, t_M) &= 0: \quad \xi_{N,k}^i \in \mathcal{N}_{\mathbb{R}_0^-}(-P_{N,k}^i), \\ g_N^i(\mathbf{q}_M, t_M) &> 0: \quad P_{N,k}^i = 0 \end{aligned} \tag{3.61}$$

and

$$\begin{aligned} g_N^i(\mathbf{q}_M, t_M) = 0: &\quad \boldsymbol{\xi}_{T,k}^i \in \mathcal{N}_{\mathcal{C}_T(P_{N,k}^i)}(-\mathbf{P}_{T,k}^i), \\ g_N^i(\mathbf{q}_M, t_M) > 0: &\quad \mathbf{P}_{T,k}^i = \mathbf{0}. \end{aligned} \tag{3.62}$$

It is important to note that due to the contact laws (3.61) and (3.62) formulated on position-switched velocity level, the calculation of an index set $\mathcal{J} = \{i \mid g_N^i = 0\}$ is required in every time step to distinguish closed contacts from open contacts with $g_N^i > 0$. Furthermore, even if the normal cone inclusions are maximal monotone, the contact laws (3.61) and (3.62) are not monotone due to their switching nature.

Paoli-Schatzman scheme

As before, the starting point for the Paoli-Schatzman discretization scheme [85, 86] is the equality of measures (3.55). However, it is restricted to mechanical systems with a single frictionless unilateral constraint, or more generally, to multiple frictionless unilateral constraints which are decoupled in the sense that $\mathbf{w}_i^\mathsf{T}\mathbf{M}^{-1}\mathbf{w}_j = 0$ for $i \neq j$ with $\mathbf{w}_i = (\partial g_N^i / \partial \mathbf{q})^\mathsf{T}$. This scheme was originally motivated by the fact that it allows for a rigorous convergence proof [86] (more rigorous than can be given for Moreau's scheme). The relations (3.56) and (3.59) hold as before and the scheme can be written as

$$\begin{aligned} &\mathbf{q}_{k+1} = \mathbf{q}_k + \Delta t \mathbf{u}_{k+1}, \\ &\mathbf{M}(\mathbf{q}_k)(\mathbf{u}_{k+1} - \mathbf{u}_k) - \mathbf{h}(t_k, \mathbf{q}_k, \mathbf{u}_k)\Delta t = \mathbf{W}\left(\frac{\mathbf{q}_{k+1} + \varepsilon \mathbf{q}_{k-1}}{1+\varepsilon}\right)\mathbf{P}_k. \end{aligned} \tag{3.63}$$

To keep a simple notation, it is assumed here that the coefficients of restitution $\varepsilon_i = \varepsilon$ are equal for all contacts. The crucial difference to other discretizations is the somewhat heuristic formulation of the combined contact-impact law directly on position level, which is written with the kinematic variable

$$\boldsymbol{\zeta}_k := \boldsymbol{g}\left(\frac{\mathbf{q}_{k+1} + \varepsilon \mathbf{q}_{k-1}}{1+\varepsilon}\right) \tag{3.64}$$

as an inequality complementarity

$$0 \leq \boldsymbol{\zeta}_k \perp \mathbf{P}_k \geq 0, \tag{3.65}$$

or, alternatively, as a normal cone inclusion

$$-\mathbf{P}_k \in \mathcal{N}_{\mathbb{R}_0^{n+}}(\boldsymbol{\zeta}_k). \tag{3.66}$$

Since the discrete law (3.65) is formulated on position level, the calculation of an index set is avoided. This discrete impact law may seem somewhat heuristic as it is not a direct discretization of the combined contact-impact law (3.57). However, its meaningfulness becomes clear when evaluated over multiple time steps. To see this, let $\mathbf{W}$ be constant and $\boldsymbol{g}(\mathbf{q}) = \mathbf{W}^\top \mathbf{q}$ be linear for simplicity. Then $\boldsymbol{\zeta}_k$ can be taken as $\boldsymbol{\zeta}_k = \boldsymbol{g}_{k+1} + \varepsilon \boldsymbol{g}_{k-1}$ with $\boldsymbol{g}_k = \mathbf{W}^\top \mathbf{q}_k$. Letting $\boldsymbol{\zeta}_k$ vanish over two consecutive time steps, i.e. $\boldsymbol{\zeta}_{k-1} = \boldsymbol{\zeta}_k = \mathbf{0}$, it follows that

$$\begin{aligned}\frac{\boldsymbol{\zeta}_k - \boldsymbol{\zeta}_{k-1}}{\Delta t} &= \frac{\boldsymbol{g}_{k+1} - \boldsymbol{g}_k}{\Delta t} + \varepsilon \frac{\boldsymbol{g}_{k-1} - \boldsymbol{g}_{k-2}}{\Delta t} = \mathbf{W}^\top \left(\frac{\mathbf{q}_{k+1} - \mathbf{q}_k}{\Delta t} + \varepsilon \frac{\mathbf{q}_{k-1} - \mathbf{q}_{k-2}}{\Delta t} \right) \\ &= \mathbf{W}^\top (\mathbf{u}_{k+1} + \varepsilon \mathbf{u}_{k-1}) = \boldsymbol{\gamma}_{k+1} + \varepsilon \boldsymbol{\gamma}_{k-1} = \mathbf{0}.\end{aligned} \tag{3.67}$$

The last equality, $\boldsymbol{\gamma}_{k+1} + \varepsilon \boldsymbol{\gamma}_{k-1} = \mathbf{0}$, shows that Newton's impact law is fulfilled in a discretized sense over two time steps. In case of an impact, the relative normal velocity of the contact point is reversed over two time steps.

Example 3.1. To illustrate the impact behavior of the Paoli-Schatzman scheme, consider a bouncing ball system (also discussed by Paoli [85]), with one degree of freedom, described by $\mathrm{d}q = u\,\mathrm{d}t$ and $m\,\mathrm{d}u - mg\,\mathrm{d}t = \mathrm{d}P_N$ with $\mathrm{d}P_N$ according to (3.54). The ball is restricted to $q \geq 0$. As shown in Figure 3.2, on the right, the velocity reversals due to impacts take place over two consecutive time steps. For a coefficient of restitution $\varepsilon = 0.5$, the discrete solution (black) penetrates the unilateral constraint at impacts, while for $\varepsilon = 0$ (gray), the constraint is satisfied at the impact. The parameters for this example are $m = 1$, $g = 9.81$ and the initial condition is $q(0) = 0.75$ and $u(0) = 0$.

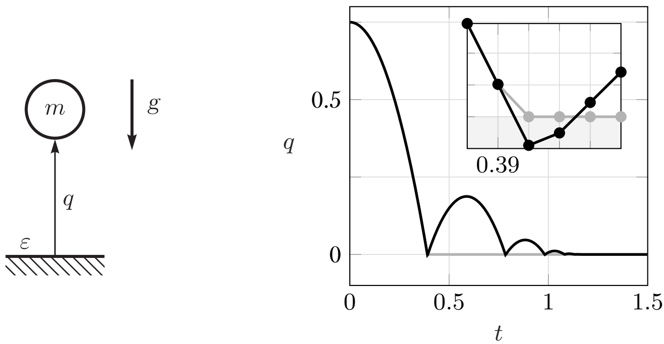

Figure 3.2: *Left:* The bouncing ball system. *Right:* Trajectories simulated with the Paoli-Schatzman scheme for coefficients of restitution $\varepsilon = 0.5$ (black) and $\varepsilon = 0$ (gray). Velocity reversals due to impacts occur over two consecutive time steps.

3.5 State transformations

The collection of generalized minimal coordinates $\mathbf{q} \in \mathbb{R}^f$ and generalized minimal velocities $\mathbf{u} \in \mathbb{R}^f$ in $\mathsf{x} = (\mathbf{q}^\mathsf{T}\ \mathbf{u}^\mathsf{T})^\mathsf{T} \in \mathbb{R}^n$ is referred to as the **state**[1]of the system. The space $\mathbb{R}^n$ (with $n = 2f$) of possible states is called the **state space** of the system. An integral curve in the state space, which is induced by a motion $\mathbf{q}(t)$, is referred to as a trajectory. A natural question that arises when analyzing or designing a general nonlinear system is, if there exists an alternative representation, which is of a special structure that simplifies the analysis or design. For smooth nonlinear systems of the form

$$\dot{\mathsf{x}} = \mathsf{f}(\mathsf{x}, t) \tag{3.68}$$

one typically seeks a state transformation, which is a mapping $\phi : \mathsf{x} \mapsto \mathsf{z}$ from the original state $\mathsf{x} \in \mathcal{X} \subset \mathbb{R}^n$ to a new state $\mathsf{z} = \phi(\mathsf{x}) \in \mathcal{Z} \subset \mathbb{R}^N$, such that the dynamics has a specific form when expressed in z. Typically such transformations are required to preserve the existence of solutions (if given for the original system). For systems of the form (3.68), state transformations are therefore usually required to be diffeomorphisms, i.e. ϕ has to be a bijection and both ϕ and its inverse ϕ^{-1} are required to be differentiable. In that case the Jacobian $(\partial\phi^{-1})/(\partial\mathsf{z})$, as well as its inverse, are continuous. The dynamics in the transformed state

$$\begin{aligned}\dot{\mathsf{z}} &= \frac{\partial\phi}{\partial\mathsf{x}}\mathsf{f}(\mathsf{x}, t) = \left.\frac{\partial\phi}{\partial\mathsf{x}}\right|_{\mathsf{x}=\phi^{-1}(\mathsf{z})} \mathsf{f}(\phi^{-1}(\mathsf{z}), t) \\ &= \left(\frac{\partial\phi^{-1}}{\partial\mathsf{z}}\right)^{-1} \mathsf{f}(\phi^{-1}(\mathsf{z}), t) =: \hat{\mathsf{f}}(\mathsf{z}, t)\end{aligned} \tag{3.69}$$

then has a continuous right-hand side if the original system has a continuous right-hand side.

When analyzing non-smooth dynamical systems, more general state transformations are applied. It is natural to seek coordinate transformations that render the trajectories of the transformed system continuous. In that case various difficulties that are caused by discontinuities in the state trajectories (such as the peaking phenomenon described in Section 4.3) could be circumvented. Two types of such transformations are reviewed in the following.

[1]Note that variables and functions related to the state space description are written without serifs, whereas in the original description of the mechanical system serifs are used (to avoid confusion if the same letter is assigned a different meaning in the state space description).

Zhuravlev-Ivanov transformation

With the goal of eliminating discontinuities in vibro-impact systems with one unilateral constraint, Ivanov [54] introduced non-smooth state transformations for systems of the form[2]

$$\begin{aligned} \mathrm{d}q &= u\,\mathrm{d}t, \\ \mathrm{d}u &= f(t,q,u)\,\mathrm{d}t + \mathrm{d}P_N, \end{aligned} \tag{3.70}$$

where $\mathrm{d}P_N = \lambda_N \mathrm{d}t + \Lambda_N \mathrm{d}\eta$, together with the contact force laws

$$-\lambda_N \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(u) & \text{if } q = 0 \\ 0 & \text{if } q > 0\,, \end{cases} \qquad -\Lambda_N \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(\xi) & \text{if } q = 0 \\ 0 & \text{if } q > 0 \end{cases} \tag{3.71}$$

with the kinematic variable $\xi = u^+ + \varepsilon u^-$. Note that system (3.70) and (3.71) includes a unilateral constraint $g(q) = q \geq 0$ and Newton's impact law. Ivanov's simplest transformation $\boldsymbol{\phi} : (q,u) \mapsto (s,v)$ that eliminates discontinuities is such that

$$\begin{aligned} q &= |s|, \\ u &= Rv\,\mathrm{sgn}(s) \quad \text{with} \quad R := 1 - \frac{1-\varepsilon}{1+\varepsilon}\,\mathrm{sgn}(sv), \end{aligned} \tag{3.72}$$

where the sign function is defined as

$$\mathrm{sgn}(t) := \begin{cases} 1 & \text{if } \ t > 0 \\ 0 & \text{if } \ t = 0 \\ -1 & \text{if } \ t < 0. \end{cases} \tag{3.73}$$

For a coefficient of restitution $\varepsilon = 1$ this state transformation is equal to what has previously been suggested by Zhuravlev [104]. In that case, the geometrical interpretation of the transformation $\boldsymbol{\phi}$ is that every point in the right half plane of the original state space (q,u), for which $q \geq 0$, is mapped to two equivalent points in the transformed state space (s,v). These two image points are equivalent by a point reflection at the origin, as shown in Figure 3.3 on the left. The resulting dynamics in the transformed state is described by

$$\begin{aligned} \dot{s} &= Rv, \\ \dot{v} &= R^{-1}\mathrm{sgn}(s) f(t,|s|,Rv\,\mathrm{sgn}(s)), \end{aligned} \tag{3.74}$$

[2] In [54], a hybrid system formulation is used, which is equivalent to the MDI given here. Also, the treatment can be extended to multiple degrees of freedom systems with a single unilateral constraint.

which is a system of differential equations with discontinuous right-hand sides. Due to the sign functions, discontinuities clearly occur for $s = 0$, but also for $v = 0$ as the variable R includes $\operatorname{sgn}(sv)$. As a consequence, solutions of (3.74) are generally non-differentiable at $s = 0$ and $v = 0$, but solutions are continuous functions of time.

Among other applications, the Zhuravlev-Ivanov transformation can be used for stability analysis [54] or controller design [84]. However, even though trajectories are continuous in the transformed representation, system (3.74) is not suitable for other purposes. For example, the design of state observers aims at determining the current state from measurements. Clearly, both s and $\operatorname{sgn}(s)$ cannot be measured, as s does not correspond to a physical quantity. All other measurements, such as the original states q and u, or functions thereof, are identical for trajectories in the transformed state space, which can be mapped onto each other by point wise mirroring at the origin. In that case it is said that system (3.74) is not observable for any physical measurement. However, one can argue that it is not necessary to distinguish between equivalent points, as they are mapped to the same points in the original state space. This leads to the ideas in the next section, which is concerned with transformations that can lead to an increase in the dimension of the state space.

Gluing and immersion transformations

Another way of achieving continuous trajectories is by finding a state transformation $\psi : \mathbf{x} \mapsto \mathbf{z}$ that maps the original state $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^n$ to the transformed state $\mathbf{z} \in \mathcal{Z} \subset \mathbb{R}^N$, which is such that the pre-impact states $\mathbf{x}^-(t)$ and the post-impact states $\mathbf{x}^+(t)$ are mapped to the same point $\mathbf{z}(t) = \mathbf{z}^+(t) = \mathbf{z}^-(t)$ in the transformed state space, for all possible solutions of the original non-smooth system. Hence for $\mathbf{x}^+ \neq \mathbf{x}^-$, the two points $\mathbf{x}^+$ and $\mathbf{x}^-$ are 'glued together' in a graphical sense in the transformed state space, as shown in Figure 3.3 on the right.

In geometric terms (without properly introducing the related objects) one intends to define an equivalence relation $\sim$ on $\mathcal{X}$, which is such that $\mathbf{x}^+(t) \sim \mathbf{x}^-(t)$ for all possible trajectories. Subsequently, the transformation is defined by a projection from $\mathcal{X}$ to the quotient set $\mathcal{X}/\sim$, followed by a one-to-one mapping from $\mathcal{X}/\sim$ to $\mathcal{Z}$.

In some cases, an increase in the state space dimension $N > n$ leads to a great simplification of the description. The state transformation is then referred to as an **immersion transformation**.

Kim et al. [62] introduced **gluing functions**, which are immersion transformations $\psi : \mathbf{x} \mapsto \mathbf{z}$ mapping $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^n$ to $\mathbf{z} \in \mathcal{Z} \subset \mathbb{R}^N$, $N \geq n$ which additionally fulfill a number of conditions ensuring the smoothness of trajectories in the transformed state space. Menini et al. [76] investigated immersion transformations for mechanical

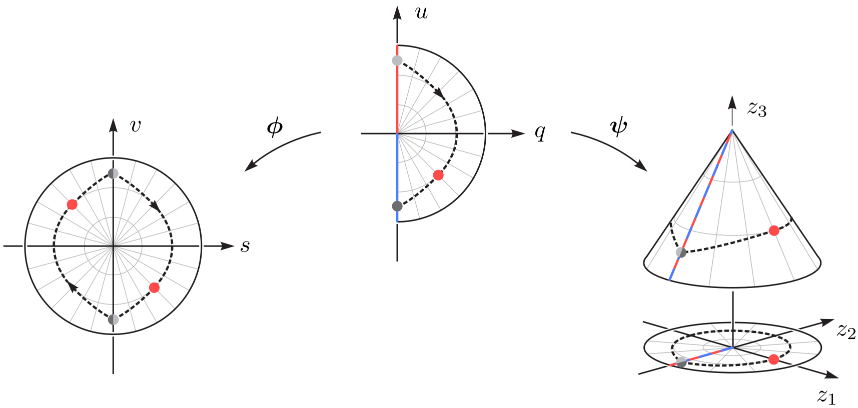

Figure 3.3: Coordinate transformations mapping discontinuous two dimensional state trajectories to continuous state trajectories, for the example of a mechanical one degree-of-freedom system with elastic impacts (state (q, u)). The Zhuralvev-Ivanov transformation ϕ leads to two equivalent points in the transformed state space (s, v). Immersion transformations ψ can map to a transformed state space of a higher dimension (z_1, z_2, z_3).

systems with a single unilateral constraint and perfectly elastic impacts (i.e. with a coefficient of restitution equal to one). As a convenient procedure, the authors suggest to define an auxiliary output, which is a chosen function of the measured output, and to define the transformed state as a number of time derivatives of this auxiliary output. This approach was shown to be useful for the analysis of simple one degree-of-freedom impact oscillators, for which the transformation also constitutes a gluing function. The core idea is illustrated by a simple example in the following.

Example 3.2 (Menini et al. [76]). Consider a one degree of freedom impact oscillator with elastic impacts of the form[3]

$$
\begin{aligned}
\mathrm{d}q &= u\,\mathrm{d}t, \\
m\,\mathrm{d}u &= -(kq + du)\,\mathrm{d}t + \mathrm{d}P_N, \\
y &= q
\end{aligned}
\tag{3.75}
$$

[3]In [76], a hybrid system formulation is used, which is equivalent to the MDI given here.

with $\mathrm{d}P_N = \lambda_N \mathrm{d}t + \Lambda_N \mathrm{d}\eta$, together with the combined contact force laws

$$-\lambda_N \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(u) & \text{if } q = 0 \\ 0 & \text{if } q > 0 \end{cases}, \qquad - \Lambda_N \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(\xi) & \text{if } q = 0 \\ 0 & \text{if } q > 0 \end{cases} \tag{3.76}$$

with the kinematic variable $\xi = u^+ + u^-$ (which corresponds to a coefficient of restitution $\varepsilon = 1$). For brevity, introduce $\bar{k} := k/m$ and $\bar{d} := d/m$. Then, by taking the time derivatives of the auxiliary output $\bar{y} = q^2$, an immersion transformation is found leading to a linear system. Indeed, with $\mathbf{x} = (q \;\; u)^\mathsf{T}$ and

$$\mathbf{z} = \boldsymbol{\psi}(\mathbf{x}) := \begin{pmatrix} q^2 \\ 2qu \\ 2u^2 - 2\bar{k}q^2 - 2\bar{d}qu \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \dot{\bar{y}} \\ \ddot{\bar{y}} \end{pmatrix}, \tag{3.77}$$

the transformed dynamics is linear and reads

$$\dot{\mathbf{z}} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -4\bar{d}\bar{k} & -(2\bar{d}^2 + 4\bar{k}) & -3\bar{d} \end{pmatrix} \mathbf{z}, \tag{3.78}$$

which is readily verified by taking the time derivative of (3.77). An impact equation is not required for the transformed dynamics. Indeed, the transformed state $\mathbf{z}$ is continuous as it consists of a combination of continuous monomials: q^2 is continuous as q is continuous, qu is continuous as $q = 0$ at points of discontinuity of u and u^2 is continuous since $u^+ = -u^-$ at all points of discontinuity of u.

4

Partial stability

Stability theory is concerned with the qualitative behavior of solutions with respect to limit sets (such as equilibria). Roughly speaking, if any solution that starts near a given limit set stays near that limit set for all future times, then the limit set is called stable. Lyapunov stability theory has originally been developed for smooth differential equations and provided useful tools to examine a system's stability properties in closed form. Since then, it has been generalized to various other systems including differential inclusions and measure differential inclusions. Furthermore, since the first definitions of stability [66], many new notions of stability have been introduced.

In this chapter, the notion of stability of partial equilibria is introduced for measure differential equations and its links to other stability concepts are discussed. Theorems allowing to conclude stability properties solely based on the dynamics equations also serve as important tools for designing state observers, tracking controllers or for analyzing synchronization. However, for impulsive systems with state jumps, a phenomenon commonly referred to as 'peaking' requires new approaches. This peaking phenomenon is caused by discontinuities in the solutions, which is why the stability of switched systems (without such discontinuities) and partial stability of discrete-time systems are discussed as well.

4.1 Stability concepts

Consider a measure differential inclusion (MDI) of the form

$$\mathrm{d}\mathbf{x} \in \mathrm{d}\boldsymbol{\Gamma}(t, \mathbf{x}), \tag{4.1}$$

where $\mathrm{d}\boldsymbol{\Gamma}(t, \mathbf{x})$ is a set-valued measure function and $\mathbf{x}(t)$ represents the state which is depending on time t. A function $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ of locally bounded variation, with the given initial condition $\mathbf{x}^-(t_0) = \mathbf{x}_0$, is a **solution** of the MDI (4.1), if it fulfills (4.1) for all $t \geq t_0$ [68]. At points of discontinuity t_i, the solution $\mathbf{x}(t_i)$ is not defined. However, (4.1) is still fulfilled as it has to be understood in an integral sense. It is common to write $\mathbf{x}(t, t_0, \mathbf{x}_0) = \mathbf{x}(t)$ to explicitly state the dependence on the initial conditions.

In the following, notions of stability are introduced, which are qualitative properties of solutions of the MDI (4.1) with respect to equilibria. A point $\mathbf{x}^*$ in the state space is an **equilibrium point** of (4.1) if there exists a solution of (4.1) for which it holds that $\mathbf{x}(t, t_0, \mathbf{x}^*) = \mathbf{x}^*$ for all $t \geq t_0$. As a consequence, for equilibrium points it must hold that $\mathrm{d}\mathbf{x} = \mathbf{0}$ which implies $\mathbf{0} \in \mathrm{d}\boldsymbol{\Gamma}(t, \mathbf{x}^*)$.

In many applications only a part of the state variables is of interest. To separate such state variables in the dynamics, let the state $\mathbf{x} \in \mathbb{R}^n$ be composed of two parts, $\mathbf{x} = (\mathbf{y}^\mathsf{T} \;\; \mathbf{z}^\mathsf{T})^\mathsf{T}$, with $\mathbf{y} \in \mathbb{R}^{n_1}$, $\mathbf{z} \in \mathbb{R}^{n_2}$ and $n_1 + n_2 = n$. An autonomous system

$$\mathrm{d}\mathbf{x} \in \mathrm{d}\boldsymbol{\Gamma}(\mathbf{x}), \tag{4.2}$$

can then equivalently be written as

$$\begin{aligned} \mathrm{d}\mathbf{y} &\in \mathrm{d}\boldsymbol{\Gamma}_1(\mathbf{y}, \mathbf{z}), \\ \mathrm{d}\mathbf{z} &\in \mathrm{d}\boldsymbol{\Gamma}_2(\mathbf{y}, \mathbf{z}). \end{aligned} \tag{4.3}$$

Solutions of (4.3) are written as $(\mathbf{y}(t, t_0, \mathbf{y}_0, \mathbf{z}_0), \mathbf{z}(t, t_0, \mathbf{y}_0, \mathbf{z}_0))$ for given initial conditions $\mathbf{y}^-(t_0) = \mathbf{y}_0$ and $\mathbf{z}^-(t_0) = \mathbf{z}_0$. A subspace of the state space, for which $\mathbf{y} = \mathbf{y}^*$, is a **partial equilibrium** of (4.3) if there exists a solution of (4.3) for which it holds that $\mathbf{y}(t, t_0, \mathbf{y}^*, \mathbf{z}_0) = \mathbf{y}^*$ for all $t \geq t_0$ and any initial condition $\mathbf{z}^-(t_0) = \mathbf{z}_0$. It follows that $\mathrm{d}\mathbf{y} = \mathbf{0}$ must hold for partial equilibria, implying $\mathbf{0} \in \mathrm{d}\boldsymbol{\Gamma}_1(\mathbf{y}^*, \mathbf{z})$ for all $\mathbf{z}$.

Note that non-autonomous systems can be seen as special cases of (4.3). Indeed, letting $\mathbf{z} = t$ be the time, (4.3) yields the nonautonomous MDI $\mathrm{d}\mathbf{y} \in \mathrm{d}\boldsymbol{\Gamma}_1(\mathbf{y}, t)$, together with $\dot{t} = 1$.

Definition 4.1 (Stability of partial equilibria[1]). A partial equilibrium of system (4.3) with $\mathbf{y}(t) = \mathbf{y}^*$ is said to be

i. **stable** if for any number $\varepsilon > 0$ there exists a number $\delta(\varepsilon, \mathbf{z}_0) > 0$ such that for all solutions with initial conditions $\mathbf{y}^-(t_0) = \mathbf{y}_0$ and $\mathbf{z}^-(t_0) = \mathbf{z}_0$ it holds that

$$\|\mathbf{y}_0 - \mathbf{y}^*\| < \delta \quad \Rightarrow \quad \|\mathbf{y}(t, t_0, \mathbf{y}_0, \mathbf{z}_0) - \mathbf{y}^*\| < \varepsilon$$

for all $t \geq t_0$ where the solution is defined. It is called **uniformly stable** if $\delta = \delta(\varepsilon)$, independent of $\mathbf{z}_0$.

ii. locally **attractive** if there exists a number $\gamma(\mathbf{z}_0) > 0$ such that

$$\|\mathbf{y}_0 - \mathbf{y}^*\| < \gamma \quad \Rightarrow \quad \lim_{t\to\infty} \|\mathbf{y}(t, t_0, \mathbf{y}_0, \mathbf{z}_0) - \mathbf{y}^*\| = 0.$$

It is called globally attractive if $\gamma \to \infty$.

iii. **attractively stable** (locally or globally) if it is stable and attractive (locally or globally). It is further called **uniformly attractively stable** (locally or globally) if it is uniformly stable and attractive (locally or globally) with γ independent of $\mathbf{z}_0$.

As a special case, if $\mathbf{y}^* = \mathbf{0}$ is a stable partial equilibrium of system (4.1), then it is said that the **system is stable with respect to y** or that the $\mathbf{y}$-dynamics of system (4.3) is stable. According to Definition 4.1, stability of a partial equilibrium point expresses that all solutions starting nearby that partial equilibrium will stay near for all future times. If all solutions starting nearby are approaching the partial equilibrium as time tends to infinity, the partial equilibrium is attractive.

If $\mathbf{y} = \mathbf{x}$ is the full state, then Definition 4.1 coincides with the classical stability definition of an equilibrium point in the sense of Lyapunov.

4.2 Lyapunov-type methods

The stability notions introduced in Section 4.1 refer to qualitative properties of solutions of a general MDI (4.1). In many cases, these solutions cannot be found in closed form. Therefore, methods have been developed to prove stability of equilibria solely based on the MDI itself. Lyapunov originally developed such a method, which is now known as *Lyapunov's direct method*, for ordinary differential equations. The method is based on finding so-called Lyapunov functions which can be understood as a generalized measure for the deviation of the solution from an equilibrium. A generalization to autonomous MDIs is discussed in [68]. In a similar fashion (but more general), the subsequent theorem gives a tool for the assessment of partial stability for MDIs. It makes use of the following definition.

Definition 4.2 (Class $\mathcal{K}$ function). A continuous function $\alpha : \mathbb{R}_{0+} \to \mathbb{R}_{0+}$ is said to belong to **class $\mathcal{K}$** if $\alpha(0) = 0$ and it is strictly increasing.

[1] In literature, *partial stability* often has a different meaning and refers to the property that solutions starting near an equilibrium point $\mathbf{x}^*$ stay near that equilibrium point, but only in their $\mathbf{y}$-component, see for example [100]. A comparison of the different concepts can be found in [101] and a link to time-varying systems in [29].

Theorem 4.3. Let there be a partial equilibrium of system (4.3) with $\mathbf{y} = \mathbf{0}$. Let $\mathbf{x} = (\mathbf{y}^\mathsf{T}\ \mathbf{z}^\mathsf{T})^\mathsf{T}$ and define $\mathcal{D}_h := \{\mathbf{x} \in \mathcal{A} \mid \|\mathbf{y}\| < h\}$. Then the following statements hold.

i. If there exists a class $\mathcal{K}$ function α, a constant $h > 0$ and a scalar, continuously differentiable function $V(\mathbf{y}, \mathbf{z})$ with $V(\mathbf{0}, \mathbf{z}) = \mathbf{0}\,\forall\,\mathbf{z}$ such that

$$\alpha(\|\mathbf{y}\|) \leq V(\mathbf{y}, \mathbf{z}) \quad \forall\,\mathbf{x} \in \mathcal{A}, \tag{4.4a}$$

$$\mathrm{d}V(\mathbf{y}, \mathbf{z}) \leq 0 \qquad\qquad \forall\,\mathbf{x} \in \mathcal{D}_h, \tag{4.4b}$$

then the partial equilibrium with $\mathbf{y} = \mathbf{0}$ is stable (i.e. system (4.3) is stable with respect to $\mathbf{y}$).

ii. If there exist class $\mathcal{K}$ functions α and β and a constant $h > 0$ such that (4.4) holds and in addition

$$V(\mathbf{y}, \mathbf{z}) \leq \beta(\|\mathbf{y}\|) \quad \forall\,\mathbf{x} \in \mathcal{A}, \tag{4.5}$$

then the partial equilibrium with $\mathbf{y} = \mathbf{0}$ is uniformly stable (i.e. system (4.3) is uniformly stable w.r.t. $\mathbf{y}$).

iii. If there exist class $\mathcal{K}$ functions α, β and γ and a constant $h > 0$ such that (4.4), (4.5) hold and in addition

$$\dot{V}(\mathbf{y}, \mathbf{z}) \leq -\gamma(\|\mathbf{y}\|) \quad \forall\,\mathbf{x} \in \mathcal{D}_h, \tag{4.6}$$

whenever $\dot{V}$ is defined, then the partial equilibrium with $\mathbf{y} = \mathbf{0}$ is (locally) uniformly attractively stable (i.e. system (4.3) is (locally) uniformly attractively stable w.r.t. $\mathbf{y}$). The statement holds globally if $\mathcal{D}_h = \mathcal{A}$.

Proof. The proof of Theorem 4.3 is divided into three parts, one for each statement.

i. Let $\mathbf{x}(t_0) \in \mathcal{D}_h$. From $\mathrm{d}V \leq 0$ it directly follows that $V(\mathbf{y}(t), \mathbf{z}(t)) \leq V(\mathbf{y}(t_0), \mathbf{z}(t_0))$ for $t_0 \leq t$. Together with (4.4a) it therefore holds that

$$\alpha(\|\mathbf{y}\|) \leq V(\mathbf{y}(t), \mathbf{z}(t)) \leq V(\mathbf{y}_0, \mathbf{z}_0).$$

Since $V(\mathbf{y}, \mathbf{z}_0)$ is continuous in $\mathbf{y}$ and $V(\mathbf{0}, \mathbf{z}_0) = 0$, there exists a $\delta(\varepsilon, \mathbf{z}_0)$ for every $\varepsilon > 0$, such that $V(\mathbf{y}, \mathbf{z}_0) < \alpha(\min(\varepsilon, h)) < \alpha(\varepsilon)$ for all $\mathbf{y}$ with $\|\mathbf{y}\| < \delta$. Specifically, for $\|\mathbf{y}_0\| < \delta$ it follows that

$$\alpha(\|\mathbf{y}\|) \leq V(\mathbf{y}(t), \mathbf{z}(t)) \leq V(\mathbf{y}_0, \mathbf{z}_0) < \alpha(\varepsilon).$$

Finally, since α is monotonically increasing, it follows that $\|\mathbf{y}\| < \varepsilon$, which concludes the stability proof.

ii. Since β is continuous and $\beta(0) = 0$, there exists a number $\delta(\varepsilon)$ (independent of $\mathbf{z}_0$) for every number $\varepsilon > 0$, such that $\beta(\delta) < \alpha(\min(\varepsilon, h)) < \alpha(\varepsilon)$. Specifically, for $\|\mathbf{y}_0\| < \delta$ it follows with (4.5) that

$$\alpha(\|\mathbf{y}\|) \leq V(\mathbf{y}(t), \mathbf{z}(t)) \leq V(\mathbf{y}_0, \mathbf{z}_0) \leq \beta(\|\mathbf{y}_0\|) < \alpha(\varepsilon).$$

Since α is monotonically increasing, it follows that $\|\mathbf{y}\| < \varepsilon$, which concludes the uniform stability proof.

iii. Uniform stability follows from ii. Due to (4.4), $V(\mathbf{y}(t), \mathbf{z}(t))$ is non-increasing and bounded from below. Therefore the limit $L := \lim_{t\to\infty} V(\mathbf{y}(t), \mathbf{z}(t)) \geq 0$ exists. By reductio ad absurdum $L = 0$. To see this, assume $L > 0$. Since V is non-increasing, $V(\mathbf{y}, \mathbf{z}) \geq L$ for all $t \geq t_0$. With (4.5) it follows that $L \leq \beta(\|\mathbf{y}\|)$ and therefore $\rho := \beta^{-1}(L) \leq \|\mathbf{y}\|$. Therefore, the inequality

$$\begin{aligned} V(\mathbf{y}^+(t), \mathbf{z}^+(t)) &= V(\mathbf{y}_0, \mathbf{z}_0) + \int_{[t_0,t]} \mathrm{d}V \\ &\leq V(\mathbf{y}_0, \mathbf{z}_0) - \int_{[t_0,t]} \gamma(\|\mathbf{y}\|)\,\mathrm{d}t \\ &\leq V(\mathbf{y}_0, \mathbf{z}_0) - \gamma(\rho)(t - t_0) \end{aligned} \tag{4.7}$$

shows, that for $t^* = t_0 + V(\mathbf{y}_0, \mathbf{z}_0)/\gamma(\rho)$ it holds that $V(\mathbf{y}(t^*), \mathbf{z}(t^*)) = 0$, which contradicts the initial assumption.

Finally, as a consequence of $L = 0$, it must hold that $\lim_{t\to\infty} \alpha(\|\mathbf{y}(t)\|) = 0$, which implies $\lim_{t\to\infty} \mathbf{y}(t) = 0$ and therefore proves attractivity. □

The function V in Theorem 4.3 is referred to as **Lyapunov function**. The inequality $\mathrm{d}V(\mathbf{y}, \mathbf{z}) \leq 0$ in (4.4b) has to be understood in an integral sense, such that due to $\int_{[t_a,t_b]} \mathrm{d}V = V(\mathbf{y}^+(t_b), \mathbf{z}^+(t_b)) - V(\mathbf{y}^-(t_a), \mathbf{z}^-(t_a)) \leq 0$ for all $t_a \leq t_b$, the Lyapunov function is required to be non increasing along solutions of (4.3). The fact that V can depend on both $\mathbf{y}$ and $\mathbf{z}$ provides more freedom in the selection of a suitable Lyapunov function. In many cases however, it is sufficient to use a positive definite Lyapunov function that only depends on $\mathbf{y}$. In that case, the conditions (4.4a) and (4.5) are directly fulfilled (see for example Lemma 4.3 in [60]) and $\mathrm{d}V(\mathbf{y}) \leq 0$ for all $\mathbf{x} \in \mathcal{D}_h$ is sufficient for stability of the partial equilibrium. The fact that the inequalities (4.4a) and (4.5) only have to hold for $\mathbf{x} \in \mathcal{A}$ also provides more flexibility in the selection of Lyapunov function candidates.

It can be shown[2], that from $\mathbf{x} \in \mathrm{lbv}(\mathcal{I}, D)$, it follows that $(V \circ \mathbf{x}) \in \mathrm{lbv}(\mathcal{I}, \mathbb{R})$ if the function $V : D \to \mathbb{R}$ is Lipschitz continuous on $\mathcal{I}$. When checking the conditions of Theorem 4.3 for a such a Lyapunov function, it is useful to write the differential measure $\mathrm{d}V$ as

$$\mathrm{d}V = \dot{V}(\mathbf{x}(t))\,\mathrm{d}t + (V^+(\mathbf{x}(t)) - V^-(\mathbf{x}(t)))\,\mathrm{d}\eta, \tag{4.8}$$

where $V^+(\mathbf{x}(t)) = V(\mathbf{x}^+(t))$ due to continuity of V. The non-positivity of $\mathrm{d}V$ is then checked by evaluating if the densities $\dot{V}$ and $V^+ - V^-$ are non-positive.

Example 4.4. Consider again the impact oscillator (3.75) and (3.76) from Example 3.2 with a coefficient of restitution $\varepsilon < 1$ and the kinematic variable $\xi = u^+ + \varepsilon u^-$. Since the positions q are restricted to be non-negative, the set $\mathcal{A}$ of admissible states is $\mathcal{A} = \{\mathbf{x} = (q \;\; u)^\mathsf{T} \mid q \geq 0\}$. To verify stability of the equilibrium $\mathbf{x}^* = (q^* \;\; u^*)^\mathsf{T} = (0 \;\; 0)^\mathsf{T}$, take a quadratic Lyapunov function of the form $V = \mathbf{x}^\mathsf{T}\mathbf{P}\mathbf{x}$ with

$$\mathbf{P} = \frac{1}{2}\begin{pmatrix} k + \delta\frac{d}{m} & \delta \\ \delta & m \end{pmatrix}. \tag{4.9}$$

This Lyapunov function corresponds to the total mechanic energy, augmented by additional terms including $0 < \delta < d$. For the non-impulsive dynamics it follows

$$\begin{aligned}
\dot{V} &= 2\mathbf{x}^\mathsf{T}\mathbf{P}\dot{\mathbf{x}} \\
&= \left(k + \delta\frac{d}{m}\right) q\dot{q} + \delta\dot{q}u + \delta q\dot{u} + mu\dot{u} \\
&= \left(k + \delta\frac{d}{m}\right) qu + \delta u^2 + \left(\delta\frac{1}{m}q + u\right)(-kq - du + \lambda_N) \\
&= -(d-\delta)u^2 - \delta\frac{k}{m}q^2,
\end{aligned} \tag{4.10}$$

which is negative definite since m, d and k are strictly positive and $\delta < d$. In the last step, it has been used that $q\lambda_N = 0$ and $u\lambda_N = 0$ due to the force law (3.76). For the impulsive dynamics, it follows

$$\begin{aligned}
V^+ - V^- &= \mathbf{x}^{+\mathsf{T}}\mathbf{P}\mathbf{x}^+ - \mathbf{x}^{-\mathsf{T}}\mathbf{P}\mathbf{x}^- \\
&= (\mathbf{x}^+ + \mathbf{x}^-)^\mathsf{T}\mathbf{P}(\mathbf{x}^+ - \mathbf{x}^-) \\
&= \frac{1}{2}\begin{pmatrix} q^+ + q^- \\ u^+ + u^- \end{pmatrix}^\mathsf{T} \begin{pmatrix} k + \delta\frac{d}{m} & \delta \\ \delta & m \end{pmatrix} \begin{pmatrix} q^+ - q^- \\ u^+ - u^- \end{pmatrix} \\
&= -\frac{1}{2}m(1-\varepsilon^2)(u^-)^2,
\end{aligned} \tag{4.11}$$

[2] See for example [78] (Proposition 12.6) or [68] (Proposition 6.3).

where in the last step it has been used that $q^+ = q^- = 0$ whenever an impact occurs and $u^+ = -\varepsilon u^-$. Hence $V^+ - V^-$ is non-positive and the Lyapunov function V is decreasing during impulsive and non-impulsive motion, as visualized in Figure 4.1 for a selected initial condition and given parameters. Note that the shape of the Lyapunov function in the inadmissible set $\mathcal{A}^c$, i.e. for $q < 0$, is irrelevant.

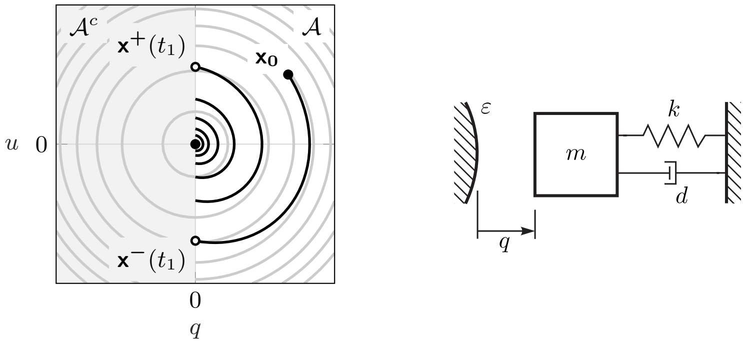

Figure 4.1: *Left:* Level curves of $V = \mathbf{x}^\mathsf{T}\mathbf{P}\mathbf{x}$ (gray) for Example 4.4 and solution (black) for a selected initial condition $\mathbf{x}_0 = (0.5 \quad 0.3)^\mathsf{T}$ and parameters $m = 1$, $k = 1$, $d = 0.2$, $\varepsilon = 0.8$ and $\delta = 1e - 3$. *Right:* The example system.

Remark 4.5. A **linear**, **time-invariant system** $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ is globally asymptotically stable (or, more precisely, its equilibrium $\mathbf{x}^* = \mathbf{0}$) if (and only if) for every matrix $\mathbf{Q} = \mathbf{Q}^\mathsf{T} \succ 0$ there exists a matrix $\mathbf{P} = \mathbf{P}^\mathsf{T} \succ 0$, such that the linear matrix equation

$$\mathbf{A}^\mathsf{T}\mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{Q} \tag{4.12}$$

is fulfilled ([30], Thm. 5.5), which is known as the **Lyapunov equation**. Furthermore, $V = \mathbf{x}^\mathsf{T}\mathbf{P}\mathbf{x}$ then serves as a Lyapunov function. Indeed, by calculating

$$\dot{V} = \dot{\mathbf{x}}^\mathsf{T}\mathbf{P}\mathbf{x} + \mathbf{x}^\mathsf{T}\mathbf{P}\dot{\mathbf{x}} = \mathbf{x}^\mathsf{T}(\mathbf{A}^\mathsf{T}\mathbf{P} + \mathbf{P}\mathbf{A})\mathbf{x} = -\mathbf{x}^\mathsf{T}\mathbf{Q}\mathbf{x} \prec 0, \tag{4.13}$$

it is seen that V is decreasing along solutions for $\mathbf{x} \neq \mathbf{0}$, since $\mathbf{Q}$ is positive definite. When searching a Lyapunov function, it can be useful to solve the linear matrix inequality (LMI)

$$\mathbf{A}^\mathsf{T}\mathbf{P} + \mathbf{P}\mathbf{A} \prec 0 \tag{4.14}$$

instead of (4.12), as one is usually not interested in any specific $\mathbf{Q}$, and efficient numerical solvers exist for LMIs.

Systems where only a part of the state variables are of interest are not the only application for Lyapunov-type partial stability methods. Various other stability problems can be recast as a partial stability problem, even though the entire state is of interest. To name just a few, the following stability problems are directly linked to partial stability.

- **Attractive incremental stability** refers to the qualitative system property that all solutions of a system $\mathrm{d}\mathbf{x} \in \mathrm{d}\boldsymbol{\Gamma}(\mathbf{x})$ converge to each other as time tends to infinity. To assess this property, one can view two solutions $\mathbf{x}_1$ and $\mathbf{x}_2$ of the original system as solutions of two identical systems

$$\begin{aligned} \mathrm{d}\mathbf{x}_1 &\in \mathrm{d}\boldsymbol{\Gamma}(\mathbf{x}_1), \\ \mathrm{d}\mathbf{x}_2 &\in \mathrm{d}\boldsymbol{\Gamma}(\mathbf{x}_2). \end{aligned} \tag{4.15}$$

  After introducing the error $\tilde{\mathbf{x}} := \mathbf{x}_1 - \mathbf{x}_2$ this is equivalent to

$$\begin{aligned} \mathrm{d}\tilde{\mathbf{x}} &\in \mathrm{d}\boldsymbol{\Gamma}(\tilde{\mathbf{x}} + \mathbf{x}_2) - \mathrm{d}\boldsymbol{\Gamma}(\mathbf{x}_2) =: \mathrm{d}\tilde{\boldsymbol{\Gamma}}(\tilde{\mathbf{x}}, \mathbf{x}_2), \\ \mathrm{d}\mathbf{x}_2 &\in \mathrm{d}\boldsymbol{\Gamma}(\mathbf{x}_2), \end{aligned} \tag{4.16}$$

  such that attractive incremental stability is equivalent to attractive stability of (4.16) with respect to $\tilde{\mathbf{x}}$.

- **Full state synchronization** refers to the qualitative property that solutions of two coupled, but possibly non-identical, systems

$$\begin{aligned} \mathrm{d}\mathbf{x} &\in \mathrm{d}\boldsymbol{\Gamma}(\mathbf{x}, \hat{\mathbf{x}}), \\ \mathrm{d}\hat{\mathbf{x}} &\in \mathrm{d}\hat{\boldsymbol{\Gamma}}(\mathbf{x}, \hat{\mathbf{x}}), \end{aligned} \tag{4.17}$$

  converge to each other. Once again, with the error $\tilde{\mathbf{x}} := \mathbf{x} - \hat{\mathbf{x}}$ this can be written as

$$\begin{aligned} \mathrm{d}\tilde{\mathbf{x}} &\in \mathrm{d}\boldsymbol{\Gamma}(\tilde{\mathbf{x}} + \hat{\mathbf{x}}, \hat{\mathbf{x}}) - \mathrm{d}\hat{\boldsymbol{\Gamma}}(\tilde{\mathbf{x}} + \hat{\mathbf{x}}, \hat{\mathbf{x}}) &&=: \mathrm{d}\tilde{\boldsymbol{\Gamma}}(\tilde{\mathbf{x}}, \hat{\mathbf{x}}), \\ \mathrm{d}\hat{\mathbf{x}} &\in \mathrm{d}\hat{\boldsymbol{\Gamma}}(\tilde{\mathbf{x}} + \hat{\mathbf{x}}, \hat{\mathbf{x}}) &&=: \mathrm{d}\check{\boldsymbol{\Gamma}}(\tilde{\mathbf{x}}, \hat{\mathbf{x}}), \end{aligned} \tag{4.18}$$

  and stability of (4.18) with respect to $\tilde{\mathbf{x}}$ is equivalent to full state synchronization of the two systems (4.17).

- **State observers** are auxiliary systems, which are unilaterally coupled to another system and are designed such that full state synchronization between both systems is achieved. The state observer design for mechanical systems with unilateral constraints is discussed in Chapters 5 and 6.

4.3 Peaking

While the stability definitions in Section 4.1 are suitable for analyzing non-smooth trajectories (including state jumps), they can be insufficient when considering the error between two such trajectories. If one defines the error between two trajectories $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ as the point-wise difference $\tilde{\mathbf{x}} = \mathbf{x}_1 - \mathbf{x}_2$, then a small time mismatch between impact time instants (where the state jumps occur) leads to error peaks: even if two trajectories are close to each other, the Euclidean error norm $\|\tilde{\mathbf{x}}\|$ (that was used in Definition 4.1 and Theorem 4.3), becomes large in the time interval between the impact time instants. To illustrate this, Figure 4.2 shows two trajectories for the impact oscillator of Example 4.4 under a periodic external forcing (on the left), together with the time evolution of $\|\tilde{\mathbf{x}}\|$. While the impact oscillator of Example 4.4 cannot be expected to be globally incrementally stable, as it can generally exhibit chaotic behavior, the plots show that for selected parameters and excitation the solutions converge to each other (at least locally). However, due to the peaking phenomenon, the error dynamics is not stable by Definition 4.1 and therefore the introduced Lyapunov theorem cannot directly be used to investigate incremental stability, synchronization or state observers.

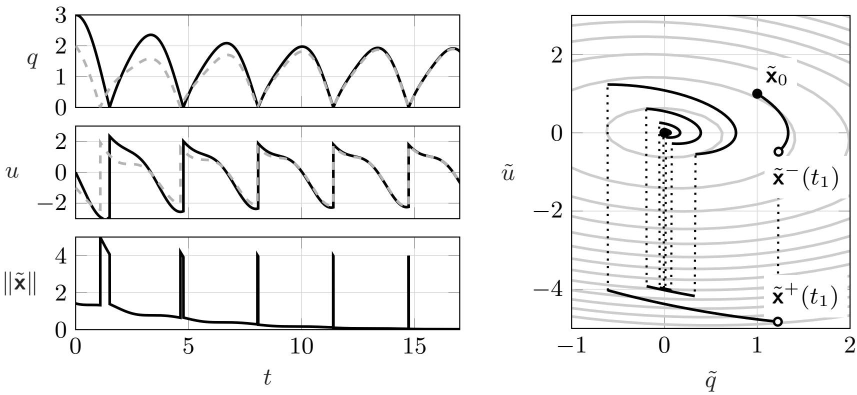

Figure 4.2: *Left:* Peaking of the error norm $\|\tilde{\mathbf{x}}\|$ between two trajectories $\mathbf{x}_1$ (solid black) and $\mathbf{x}_2$ (dashed gray) of the impact oscillator of Example 4.4. *Right:* State-space trajectory (black) of the error $\tilde{\mathbf{x}}$ and level curves of a quadratic function $V(\tilde{\mathbf{x}})$ (gray).

Approaches to circumvent the peaking problem can be grouped in two main categories. The first category is concerned with transforming or modifying the dynamics at hand, such that no more state jumps occur and the available Lyapunov-type theorems are applicable. For example, by applying the state transformations introduced in Section 3.5, it might be possible to obtain a transformed dynamics without state jumps, and therefore without peaking. Alternatively, one can regularize the impulsive dynamics by replacing the impact equations with a smooth dynamics approximating the non-smooth behavior. Such a regularization typically involves stiff differential equations, which makes regularization-based simulation and optimization of non-smooth systems difficult. However, it is a useful analysis approach, which can be taken for stability analysis.

The second category is concerned with alternative definitions of stability, which apply in the presence of peaking. For example, within the hybrid systems framework, Biemond et al. [18] propose alternative definitions of incremental stability and provide corresponding Lyapunov-type theorems. These definitions and theorems rely on a more general distance function $d(\mathbf{x}_1, \mathbf{x}_2)$ to be used in lieu of the Euclidean distance $\|\mathbf{x}_1 - \mathbf{x}_2\|$ that was used in Definition 4.1 and Theorem 4.3. The construction of specific distance functions, which do not exhibit peaking, was shown to be useful to solve the tracking problem for impulsive systems [17] and for controlled synchronization [9].

4.4 Stability of switched systems

In many cases, the regularization or discretization of a MDI leads to a switched system, which is characterized by differential or difference equations with multiple right-hand sides together with a switching law that describes which of these right-hand sides is active. Also, for the case of mechanical systems with unilateral constraints, the MDI itself can be seen as an approximation of a smooth, but switched, dynamics. In fact, the impact equations and the impact law describe discontinuous velocities, which can be observed on a macroscopic level, but are only an approximation of an unknown smooth dynamics on a microscopic level. In view of the peaking problem, it is therefore instructive to discuss the stability of switched systems as well. Here, only state dependent switching laws are considered, i.e. the state-space $\mathbb{R}^n$ can be divided into subsets $\mathcal{X}_i$, referred to as **cells**, and each cell is assigned with only one of the possible right-hand sides. Furthermore, the discussion is restricted to **piecewise affine (PWA) systems**, which are a well-studied subgroup of switched systems and can be written as

$$\dot{\mathbf{x}} = \mathbf{A}_i \mathbf{x} + \mathbf{b}_i \quad \text{for } \mathbf{x} \in \mathcal{X}_i \ (i \in \{1, \ldots, N\}), \tag{4.19}$$

with constant system matrices $\mathbf{A}_i \in \mathbb{R}^{n \times n}$, $\mathbf{b}_i \in \mathbb{R}^n$ and the state $\mathbf{x} \in \mathbb{R}^n$. The cells $\mathcal{X}_i$ are separated by a finite set of hyperplanes $\mathbf{H}_j \mathbf{x} + \mathbf{h}_j = 0$ $(j \in \{1, \ldots, \overline{N}\})$.

Here, it is assumed that the right-hand side of (4.19) is continuous. The PWA system (4.19) can be seen as a special case of the MDI (4.1), such that the (partial) stability Theorem 4.3 can directly applied. Corresponding Lyapunov functions are typically called **common Lyapunov functions**, as they are applied for all cells. To ensure that a common Lyapunov function V is non-increasing, it consequently has to hold that

$$\dot{V} = \frac{\partial V}{\partial \mathbf{x}}(\mathbf{A}_i\mathbf{x} + \mathbf{b}_i) \leq 0 \quad \forall\, \mathbf{x} \in \mathcal{X}_i \ (i \in \{1, ..., N\}). \tag{4.20}$$

However, for many PWA systems (and also for more general switched systems), it is not possible to find a single Lyapunov function that fulfills (4.20) for all cells. A much less conservative stability analysis can be based on multiple Lyapunov functions, e.g. one Lyapunov function V_i for each cell, such that

$$V(\mathbf{x}) = V_i(\mathbf{x}) \quad \text{for } \mathbf{x} \in \mathcal{X}_i. \tag{4.21}$$

Instead of (4.20), each V_i then only has to decrease inside its corresponding cell $\mathcal{X}_i$, i.e. for all $i \in \{1, ..., N\}$ it has to hold that

$$\dot{V}_i = \frac{\partial V_i}{\partial \mathbf{x}}(\mathbf{A}_i\mathbf{x} + \mathbf{b}_i) \leq 0 \quad \forall\, \mathbf{x} \in \mathcal{X}_i. \tag{4.22}$$

Furthermore, one has to ensure that $V_j(x^+) \leq V_i(x^-)$ whenever the solution transitions between neighboring cells, i.e. if $x^- \in \mathcal{X}_i$ and $x^+ \in \mathcal{X}_j$. The easiest way to ensure this is to use Lyapunov functions that are continuous at the hyperplanes which are separating the cells.

Example 4.6. Consider the bi-modal, piecewise linear (PWL) system

$$\dot{\mathbf{x}} = \begin{cases} \mathbf{A}_1\mathbf{x} & \text{if } \mathbf{H}\mathbf{x} > 0 \\ \mathbf{A}_2\mathbf{x} & \text{if } \mathbf{H}\mathbf{x} \leq 0, \end{cases} \tag{4.23}$$

with the matrices

$$\mathbf{A}_1 = \begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{d}{m} \end{pmatrix}, \ \mathbf{A}_2 = \begin{pmatrix} 0 & 1 \\ -\frac{k+\kappa}{m} & -\frac{d}{m} \end{pmatrix} \text{ and } \mathbf{H} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}^\mathsf{T}. \tag{4.24}$$

The system describes a one degree-of-freedom oscillator with a unilateral spring, as depicted in Figure 4.3 on the right, with mass m, damping d, two spring rates k and κ and the state $\mathbf{x} = (q \quad u)^\mathsf{T}$. Both springs are relaxed for $q = 0$ and $\mathbf{x}^* = \mathbf{0}$ is clearly an equilibrium. To prove asymptotic stability of $\mathbf{x}^*$, the piecewise quadratic Lyapunov function

$$V(\mathbf{x}) = \begin{cases} V_1(\mathbf{x}) = \mathbf{x}^\mathsf{T}\mathbf{P}_1\mathbf{x} & \text{if } \mathbf{H}\mathbf{x} > 0 \\ V_2(\mathbf{x}) = \mathbf{x}^\mathsf{T}\mathbf{P}_2\mathbf{x} & \text{if } \mathbf{H}\mathbf{x} \leq 0 \end{cases} \tag{4.25}$$

with the matrices

$$\mathbf{P}_1 = \frac{1}{2}\begin{pmatrix} k+\delta\frac{d}{m} & \delta \\ \delta & m \end{pmatrix}, \ \mathbf{P}_2 = \frac{1}{2}\begin{pmatrix} k+\kappa+\delta\frac{d}{m} & \delta \\ \delta & m \end{pmatrix} \tag{4.26}$$

with $\delta < d$ can be used. In fact, by a similar calculation as in (4.10), it is verified that $\dot{V}_1 = -(kd/m)q^2 - (\delta - d)u^2$ and $\dot{V}_2 = -((k+\kappa)d/m)q^2 - (\delta - d)u^2$, which are both negative definite. Figure 4.3, on the left, shows the solution of (4.23) for a given initial condition $\mathbf{x}_0 = (0.5 \quad 0.3)$, which is clearly attracted by the equilibrium $\mathbf{x}^* = \mathbf{0}$. Furthermore, as can be seen from the level curves, the piecewise quadratic Lyapunov function is decreasing along the solution.

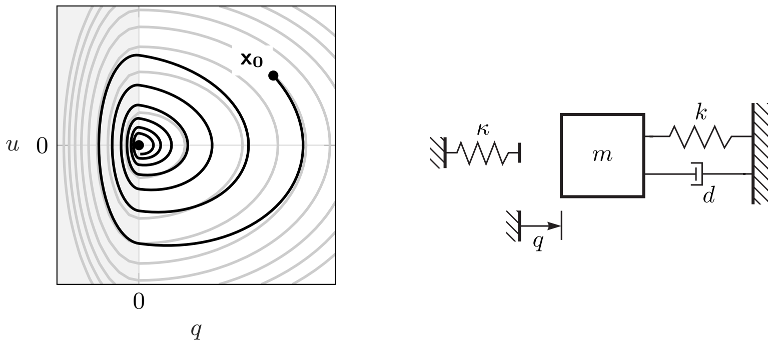

Figure 4.3: *Left:* Level curves of the piecewise quadratic Lyapunov function $V(\mathbf{x})$ (gray) for Example 4.6 and solution (black) for a selected initial condition $\mathbf{x}_0 = (0.5 \quad 0.3)^\top$ and parameters $m = 1$, $k = 1$ $\kappa = 10$, $d = 0.2$ and $\delta = 1e-3$. *Right:* The example system.

Remark 4.7. As for the non-switched case, LMI solvers are a useful tool for finding the Lyapunov functions of a time-invariant PWL system. The necessary condition (4.22) is fulfilled if the LMIs

$$\mathbf{A}_i^\top \mathbf{P}_i + \mathbf{P}_i \mathbf{A}_i \prec 0 \tag{4.27}$$

hold. However, the LMIs (4.27) only have to hold for $\mathbf{x} \in \mathcal{X}_i$ and they might only allow for a solution if this additional information is incorporated into the LMI.

4.5 Stability of discrete-time systems

For the (partial) stability of discrete-time systems, Lyapunov-type theorems exists which are very similar to the continuous-time case. In view of Chapter 6, where the time discretization of a mechanical system with unilateral constraints is analyzed, the main results are stated here for completeness. Consider a discrete-time dynamical system of the form

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k) \tag{4.28}$$

with the discrete state $\mathbf{x}_k \in \mathbb{R}^n$, a mapping $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ and an iteration parameter $k \in \mathbb{Z}$ defining the discrete time.[3] The index notation has to be understood in the sense that $\mathbf{x}_k = \mathbf{x}(k)$ and is used here for a more compact writing. As before, let the state $\mathbf{x}_k$ be composed of two parts, $\mathbf{x}_k = (\mathbf{y}_k^\mathsf{T} \;\; \mathbf{z}_k^\mathsf{T})^\mathsf{T}$, such that system (4.28) can equivalently be written as

$$\begin{aligned} \mathbf{y}_{k+1} &= \mathbf{F}_1(\mathbf{y}_k, \mathbf{z}_k), \\ \mathbf{z}_{k+1} &= \mathbf{F}_2(\mathbf{y}_k, \mathbf{z}_k). \end{aligned} \tag{4.29}$$

Solutions of (4.29) are written as $(\mathbf{y}(k, k_0, \mathbf{y}_0, \mathbf{z}_0), \mathbf{z}(k, k_0, \mathbf{y}_0, \mathbf{z}_0))$, indicating the dependence on given initial conditions $\mathbf{y}(k_0) = \mathbf{y}_0$ and $\mathbf{z}(k_0) = \mathbf{z}_0$, or in short as $(\mathbf{y}_k, \mathbf{z}_k)$, omitting the initial conditions. A subspace of the state space, for which $\mathbf{y} = \mathbf{y}^*$, is a **partial fixed point** of (4.29) if there exists a solution of (4.29) for which it holds that $\mathbf{y}(k, k_0, \mathbf{y}^*, \mathbf{z}_0) = \mathbf{y}^*$ for all $k \geq k_0$ and any initial condition $\mathbf{z}_0$. It follows that $\mathbf{F}_1(\mathbf{y}^*, \mathbf{z}_k) = \mathbf{y}^*$ for all $\mathbf{z}_k$ must hold for partial fixed points.

Definition 4.8 (Stability of partial fixed points). A partial fixed point with $\mathbf{y}_k = \mathbf{y}^*$ of system (4.29) is said to be

i. **stable** if for any number $\varepsilon > 0$ there exists a number $\delta(\varepsilon, \mathbf{z}_0) > 0$ such that for all solutions with initial conditions $\mathbf{y}_0$ and $\mathbf{z}_0$ it holds that

$$\|\mathbf{y}_0 - \mathbf{y}^*\| < \delta \quad \Rightarrow \quad \|\mathbf{y}(k, k_0, \mathbf{y}_0, \mathbf{z}_0) - \mathbf{y}^*\| < \varepsilon$$

for all $k \geq k_0$. It is called **uniformly stable** if $\delta = \delta(\varepsilon)$, independent of $\mathbf{z}_0$.

ii. locally **asymptotically stable** if there additionally exists a number $\gamma(\mathbf{z}_0) > 0$ such that

$$\|\mathbf{y}_0 - \mathbf{y}^*\| < \gamma \quad \Rightarrow \quad \lim_{k \to \infty} \|\mathbf{y}(k, k_0, \mathbf{y}_0, \mathbf{z}_0) - \mathbf{y}^*\| = 0,$$

and globally asymptotically stable if $\gamma \to \infty$.

---

[3] If the discrete-time system is obtained by uniformly sampling a continuous-time system with time t, then $t = k\Delta t$ with a fixed time step Δt.

Theorem 4.9. Let there be a partial fixed point of system (4.3) with $\mathbf{y}_k = \mathbf{0}$. Then the following statements hold.

i. If there exists a scalar, continuously differentiable function $V(\mathbf{y}, \mathbf{z})$ with $V(\mathbf{0}, \mathbf{z}) = \mathbf{0}\,\forall\,\mathbf{z}$ and a class $\mathcal{K}$ function α such that

$$\alpha(\|\mathbf{y}\|) \leq V(\mathbf{y}, \mathbf{z}), \tag{4.30a}$$

$$V(\mathbf{y}_{k+1}, \mathbf{z}_{k+1}) - V(\mathbf{y}_k, \mathbf{z}_k) \leq 0, \tag{4.30b}$$

then the partial fixed point with $\mathbf{y} = \mathbf{0}$ is stable (i.e. system (4.29) is stable with respect to $\mathbf{y}$).

ii. If there additionally exists a class $\mathcal{K}$ function β such that

$$V(\mathbf{y}, \mathbf{z}) \leq \beta(\|\mathbf{y}\|), \tag{4.31}$$

then the partial fixed point with $\mathbf{y} = \mathbf{0}$ is uniformly stable (i.e. system (4.3) is uniformly stable w.r.t. $\mathbf{y}$).

Proof. The proof is similar to the continuous time case and can be found in [82].

Remark 4.10. A **linear, time-invariant system** $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$ is globally asymptotically stable (i.e. its fixed point $\mathbf{x}^* = \mathbf{0}$) if (and only if) for every matrix $\mathbf{Q} = \mathbf{Q}^\top \succ 0$ there exists a matrix $\mathbf{P} = \mathbf{P}^\top \succ 0$, such that the linear matrix equation

$$\mathbf{A}^\top\mathbf{P}\mathbf{A} - \mathbf{P} = -\mathbf{Q} \tag{4.32}$$

is fulfilled, which the discrete version of the Lyapunov equation in Remark 4.5. Similar to the continuous case, $V_k = \mathbf{x}_k^\top\mathbf{P}\mathbf{x}_k$ then serves as a Lyapunov function, which can be seen by calculating

$$V_{k+1} - V_k = \mathbf{x}_{k+1}^\top\mathbf{P}\mathbf{x}_{k+1} - \mathbf{x}_k^\top\mathbf{P}\mathbf{x}_k = \mathbf{x}_k^\top(\mathbf{A}^\top\mathbf{P}\mathbf{A} - \mathbf{P}\mathbf{A})\mathbf{x}_k = -\mathbf{x}_k^\top\mathbf{Q}\mathbf{x}_k \prec 0. \tag{4.33}$$

The quadratic function V is decreasing along solutions for $\mathbf{x}_k \neq \mathbf{0}$ if $\mathbf{Q}$ is positive definite. If the specific choice of $\mathbf{Q}$ is not of interest, the LMI

$$\mathbf{A}^\top\mathbf{P}\mathbf{A} - \mathbf{P} \prec 0 \tag{4.34}$$

can be solved to find a suitable $\mathbf{P}$ which provides a Lyapunov function, instead of solving (4.32).

5

State observers with known impact time instants

The state estimation problem is concerned with obtaining non-measured state variables (or parameters) from a number of measurement signals. These unknown states are often needed for control purposes, monitoring or for parameter identification. The estimation typically relies on a model of the observed system. One of the two main approaches to solve this problem is optimization-based: one estimates the initial state at the beginning of a moving time window, such that a certain functional of the error between a model-based predicted output and the measurements is minimized over that window. More commonly however, a dynamical system approach is used. A state observer is set up, which is a (virtual) dynamical system, whose state (or a transformation thereof) asymptotically converges to the true state and therefore serves as a state estimation. Convergence is achieved through a unidirectional coupling between the state observer and the observed system through measurements. More precisely, for a system of the form

$$\begin{aligned} \mathrm{d}\mathbf{x} &\in \mathrm{d}\boldsymbol{\Gamma}(t, \mathbf{x}, \mathbf{v}), \\ \mathbf{y} &= \mathbf{h}(\mathbf{x}) \end{aligned} \tag{5.1}$$

with the state $\mathbf{x} \in \mathbb{R}^n$, input $\mathbf{v} \in \mathbb{R}^m$ and the (measured) output $\mathbf{y} \in \mathbb{R}^p$, a **state observer** is an auxiliary system

$$\begin{aligned} \mathrm{d}\mathbf{z} &\in \mathrm{d}\boldsymbol{\Delta}(t, \mathbf{z}, \mathbf{v}, \mathbf{y}), \\ \hat{\mathbf{x}} &= \psi(\mathbf{z}, \mathbf{v}, \mathbf{y}) \end{aligned} \tag{5.2}$$

with the observer state $\mathbf{z} \in \mathbb{R}^N$ and estimated state $\hat{\mathbf{x}} \in \mathbb{R}^n$, for which it holds that the dynamics of the estimation error $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$ is (globally) attractively stable (which particularly implies that $\lim_{t\to\infty} \|\mathbf{x} - \hat{\mathbf{x}}\| = \mathbf{0}$ and that small initial estimation errors result in small future estimation errors). The design process for such a state observer consists of the selection of a suitable measure function $\boldsymbol{\Delta}$ and a function ψ. In general $N \geq n$, and observers are referred to as embedding observers if $N > 0$. A good starting point is usually an identity observer, for which $N = n$ and $\hat{\mathbf{x}} = \mathbf{z}$.

For linear systems, general observability conditions, under which a state observer exists, can be given [30]. Furthermore, general state observer designs such as the Luenberger observer for time-invariant systems [70, 71] or the Kalman observer for time-varying systems [59] are known. For nonlinear systems, the design of a state observer is more difficult for several reasons. First, observability conditions are generally not only depending on the system but also on the input. Furthermore, if a state observer exists, it cannot be assumed that its state is of the same dimension as of the observed system. In some cases, a higher state dimension is necessary, which makes it more difficult to find its dynamics. Lastly, for linear systems a separation principle holds: if a state observer and a stabilizing controller are designed separately, it can be shown that the closed-loop system combining both is asymptotically stable. This separation principle does in general not hold for nonlinear systems.

State observer designs for nonlinear systems are problem specific and general design techniques do not exist. However, for some specific structures of smooth nonlinear systems, often referred to as normal forms, extensions of the Luenberger and Kalman observers exist. For example, systems with an additive triangular nonlinearity [49], state affine systems [14, 15] or a combination of both [14] all allow for such extensions. When designing state observers for nonlinear systems, it is therefore natural to seek state transformations that lead to one of the known normal forms in the transformed state. As a starting point, various conditions under which such a transformation exists are given in literature. Other techniques include linearization-based observer designs, where observers for linear systems are applied to the linearization of a nonlinear system. For example, the extended Kalman filter[1] [95] or the extended Luenberger observer [20] are often applied, but there is no guarantee for the asymptotic stability of the corresponding error dynamics. Next, for some classes of nonlinearities, such as Lipschitz continuous nonlinearities, it is possible to dominate the nonlinearities through correction terms with sufficiently large gains (in so-called high-gain observers [61]).

[1]The well-known Kalman filter was developed for linear systems with stochastic disturbances. However, the technique can be applied to deterministic systems as well, in which case it is referred to as a Kalman observer.

For mechanical systems with unilateral constraints and Coulomb friction, which are described by measure differential inclusions of the form (5.2), maximal monotone nonlinearities form an important class. First, Coulomb-type friction laws, typically formulated as normal cone inclusions, are maximal monotone for given normal contact forces. As it will be shown in the subsequent sections, the maximal monotonicity property, in combination with a suitable Lyapunov function, is favorable in the stability analysis of the estimation error dynamics. For the state observer design for impulsive systems with state jumps, an important aspect is whether or not the impact time instants (at which the state jumps occur) are known. Clearly, if impacts in the observer and the observed system do not occur at the same time instants, the estimation error dynamics is not asymptotically stable in the classical sense. For mechanical systems with unilateral constraints, the combined contact-impact law is typically formulated on position-switched velocity level. It turns out that for given positions, this law is also maximal monotone in many cases, which is very useful in the state observer design.

The goal of this chapter is to review and unify some state observer designs for mechanical systems with unilateral constraints, for known impact time instants, and to extend some designs to account for Coulomb-friction. It also serves as preparation for Chapter 6, which is concerned with the state observer design for such systems but with unknown impact time instants.

5.1 Switched unilateral constraints (Baumann & Leine)

One example of a state observer that only requires knowledge of the impact time instants in form of a boolean function was presented by Baumann and Leine in [11]. Therein, the concept of switched unilateral constraints is introduced, which is recalled in the following and extended to account for non-opening contacts with Coulomb friction. The starting point is a non-smooth mechanical system consisting of a linear, time-invariant structure subjected to geometric unilateral constraints (leading to impacts) and non-opening frictional contacts with a constant normal contact force. The system under consideration is typically a vibro-impact system with frictional linear guides. Hence, the dynamics is described by

$$\begin{aligned} \mathrm{d}\mathbf{q} &= \mathbf{u}\,\mathrm{d}t, \\ \mathbf{M}\,\mathrm{d}\mathbf{u} + (\mathbf{C}\mathbf{u} + \mathbf{K}\mathbf{q} - \mathbf{f}(t))\,\mathrm{d}t &= \mathbf{W}_N\,\mathrm{d}\mathbf{P}_N + \mathbf{W}_T\,\mathrm{d}\mathbf{P}_T \end{aligned} \tag{5.3}$$

where the system matrices $\mathbf{K} = \mathbf{K}^\mathsf{T} \succ 0$, $\mathbf{M} = \mathbf{M}^\mathsf{T} \succ 0$ and $\mathbf{C} \succ 0$ are assumed to be constant and positive definite. The matrix $\mathbf{W}_N$ is composed of the generalized force directions of each unilateral constraint and the matrix $\mathbf{W}_T$ contains the generalized force directions of the non-opening frictional contacts. Both $\mathbf{W}_N$ and $\mathbf{W}_T$ are assumed constant and all generalized force directions to be linearly independent. The system is excited by an external forcing $\mathbf{f}(t)$. The percussion measures $\mathrm{d}\mathbf{P}_N$ and $\mathrm{d}\mathbf{P}_T$ are composed of the non-impulsive constraint forces $\boldsymbol{\lambda}_N, \boldsymbol{\lambda}_T$ and the impulsive force $\boldsymbol{\Lambda}_N$ according to

$$\begin{aligned} \mathrm{d}\mathbf{P}_N &= \boldsymbol{\lambda}_N \,\mathrm{d}t + \boldsymbol{\Lambda}_N \,\mathrm{d}\eta\,, \\ \mathrm{d}\mathbf{P}_T &= \boldsymbol{\lambda}_T \,\mathrm{d}t. \end{aligned} \tag{5.4}$$

Note that all friction forces are non-impulsive, as only non-opening frictional contacts with a *constant* normal contact force are considered here. Furthermore, the Coulomb-type friction law (3.47) is assumed, which takes the form

$$\boldsymbol{\gamma}_T^i \in \mathcal{N}_{\mathcal{C}_T(\lambda_N^i)}(-\boldsymbol{\lambda}_T^i) \tag{5.5}$$

for the i-th friction contact, with a convex force reservoir $\mathcal{C}_T(\lambda_N^i)$. Assuming the generalized Newtonian impact law, recall that the force laws in normal direction are formulated component-wise on position-switched velocity level as

$$-\lambda_N^i \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(\gamma_N^i) & \text{if } g_N^i = 0 \\ 0 & \text{if } g_N^i > 0\,, \end{cases} \qquad -\Lambda_N^i \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(\xi^i) & \text{if } g_N^i = 0 \\ 0 & \text{if } g_N^i > 0\,, \end{cases} \tag{5.6}$$

with the kinematic variable $\xi^i = \gamma_N^{i+} + \varepsilon_i \gamma_N^{i-}$. However, if the time instants for which the contacts are closed are known in form of boolean functions $\chi_i(t)$, such that $\chi_i = 1$ for $g_N^i = 0$ and $\chi_i = 0$ for $g_N^i > 0$, the force laws can equivalently be written as

$$-\lambda_N^i \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(\gamma_N^i) & \text{if } \chi_i(t) = 1 \\ 0 & \text{if } \chi_i(t) = 0\,, \end{cases} \qquad -\Lambda_N^i \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(\xi^i) & \text{if } \chi_i(t) = 1 \\ 0 & \text{if } \chi_i(t) = 0\,. \end{cases} \tag{5.7}$$

The force laws (5.7) have been introduced as **switched unilateral constraint** in [11], since χ_i acts as a switching function which could be an external input.

A state observer for the system (5.3), (5.4), (5.5), (5.7) is obtained by a copy of the system with the only difference, that in the observer, the switching function is generated by the contact distances g_N^i of the observed system. Denoting all observer related variables with a circumflex symbol ($\hat{\ }$), this leads to

$$\begin{aligned} \mathrm{d}\hat{\mathbf{q}} &= \hat{\mathbf{u}}\,\mathrm{d}t, \\ \mathbf{M}\,\mathrm{d}\hat{\mathbf{u}} + (\mathbf{C}\hat{\mathbf{u}} + \mathbf{K}\hat{\mathbf{q}} - \mathbf{f}(t))\,\mathrm{d}t &= \mathbf{W}_N\,\mathrm{d}\hat{\mathbf{P}}_N + \mathbf{W}_T\,\mathrm{d}\hat{\mathbf{P}}_T \end{aligned} \tag{5.8}$$

with $\mathrm{d}\hat{\mathbf{P}}_N = \hat{\boldsymbol{\lambda}}_N \mathrm{d}t + \hat{\boldsymbol{\Lambda}}_N \mathrm{d}\eta$ and $\mathrm{d}\hat{\mathbf{P}}_T = \hat{\boldsymbol{\lambda}}_T \mathrm{d}t$. The idea is now to use the same switching functions $\chi_i(t)$ as in (5.7) when setting up the force laws for $\hat{\boldsymbol{\lambda}}_N$ and $\hat{\boldsymbol{\Lambda}}_N$.

In other words, the switching functions of the state observer are generated by the gap function of the observed system through measurements, written as

$$-\hat{\lambda}_N^i \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(\hat{\gamma}_N^i) & \text{if } \chi_i(t) = 1 \\ 0 & \text{if } \chi_i(t) = 0\,, \end{cases} \qquad -\hat{\Lambda}_N^i \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(\hat{\xi}^i) & \text{if } \chi_i(t) = 1 \\ 0 & \text{if } \chi_i(t) = 0\,. \end{cases} \tag{5.9}$$

It is readily verified that the force law for the normal contact force is (maximal) monotone, i.e. it holds that

$$(\lambda_N^i - \hat{\lambda}_N^i)(\gamma_N^i - \hat{\gamma}_N^i) \leq 0 \tag{5.10}$$

for any two λ_N^i and $\hat{\lambda}_N^i$ with corresponding γ_N^i and $\hat{\gamma}_N^i$. In analogy, for the impulsive normal contact forces, we have

$$(\Lambda_N^i - \hat{\Lambda}_N^i)(\xi^i - \hat{\xi}^i) \leq 0, \tag{5.11}$$

from which it directly follows for elastic impacts with $\varepsilon_i = 1\ \forall\, i$ that

$$(\Lambda_N^i - \hat{\Lambda}_N^i)^\mathsf{T}((\gamma_N^{i+} + \gamma_N^{i-}) - (\hat{\gamma}_N^{i+} + \hat{\gamma}_N^{i-})) \leq 0. \tag{5.12}$$

However, (5.12) holds for more general cases, such as for arbitrary but identical coefficients of restitution $\varepsilon_i = \varepsilon\ \forall\, i$, as shown in detail in [8].

For the tangential forces of the observer, the same force law is used as for the observed system, i.e.

$$\hat{\boldsymbol{\gamma}}_T^i \in \mathcal{N}_{\mathcal{C}_T(\lambda_N^i)}(-\hat{\boldsymbol{\lambda}}_T^i) \tag{5.13}$$

for the i-th friction contact. Therein, the normal forces λ_N^i in the frictional contacts are assumed to be known and are therefore chosen identical to (5.5). As a consequence, the corresponding normal cone inclusion for every friction contact is maximal monotone, such that it holds that

$$(\boldsymbol{\gamma}_T - \hat{\boldsymbol{\gamma}}_T)^\mathsf{T}(\boldsymbol{\lambda}_T - \hat{\boldsymbol{\lambda}}_T) \leq 0, \tag{5.14}$$

for any two $\boldsymbol{\lambda}_T$ and $\hat{\boldsymbol{\lambda}}_T$ obeying (5.5) and (5.13) with corresponding $\boldsymbol{\gamma}_T$ and $\hat{\boldsymbol{\gamma}}_T$.

To see how crucial the monotonicity properties (5.10), (5.12) and (5.14) are for showing attractive stability of the error dynamics by Lyapunov's direct method, introduce the estimation errors $\tilde{\mathbf{q}} = \mathbf{q} - \hat{\mathbf{q}}$ and $\tilde{\mathbf{u}} = \mathbf{u} - \hat{\mathbf{u}}$. The error dynamics then reads (in part)

$$\begin{aligned} \mathrm{d}\tilde{\mathbf{q}} &= \tilde{\mathbf{u}}\,\mathrm{d}t, \\ \mathbf{M}\,\mathrm{d}\tilde{\mathbf{u}} + (\mathbf{C}\tilde{\mathbf{u}} + \mathbf{K}\tilde{\mathbf{q}})\,\mathrm{d}t &= \mathbf{W}_N(\mathrm{d}\mathbf{P}_N - \mathrm{d}\hat{\mathbf{P}}_N) + \mathbf{W}_T(\mathrm{d}\mathbf{P}_T - \mathrm{d}\hat{\mathbf{P}}_T). \end{aligned} \tag{5.15}$$

Considering the quadratic energy-like Lyapunov function $V = \frac{1}{2}(\tilde{\mathbf{q}}^\mathsf{T}\mathbf{K}\tilde{\mathbf{q}} + \tilde{\mathbf{u}}^\mathsf{T}\mathbf{M}\tilde{\mathbf{u}})$, it follows for the non-impulsive motion that

$$\begin{aligned}\dot{V} &= \tilde{\mathbf{u}}^\mathsf{T}\mathbf{M}\dot{\tilde{\mathbf{u}}} + \tilde{\mathbf{q}}^\mathsf{T}\mathbf{K}\dot{\tilde{\mathbf{q}}} \\ &= \tilde{\mathbf{u}}^\mathsf{T}(-\mathbf{D}\tilde{\mathbf{u}} + \mathbf{W}_N(\boldsymbol{\lambda}_N - \hat{\boldsymbol{\lambda}}_N) + \mathbf{W}_T(\boldsymbol{\lambda}_T - \hat{\boldsymbol{\lambda}}_T)) \\ &= -\tilde{\mathbf{u}}^\mathsf{T}\mathbf{D}\tilde{\mathbf{u}} + (\boldsymbol{\gamma}_N - \hat{\boldsymbol{\gamma}}_N)^\mathsf{T}(\boldsymbol{\lambda}_N - \hat{\boldsymbol{\lambda}}_N) + (\boldsymbol{\gamma}_T - \hat{\boldsymbol{\gamma}}_T)^\mathsf{T}(\boldsymbol{\lambda}_T - \hat{\boldsymbol{\lambda}}_T),\end{aligned} \tag{5.16}$$

which is non-positive since $\mathbf{D}$ is positive definite and the force laws in normal and tangential direction are monotone. Furthermore, for the impulsive motion,

$$\begin{aligned}V^+ - V^- &= \frac{1}{2}(\tilde{\mathbf{u}}^+ + \tilde{\mathbf{u}}^-)^\mathsf{T}\mathbf{M}(\tilde{\mathbf{u}}^+ - \tilde{\mathbf{u}}^-) \\ &= (\tilde{\boldsymbol{\gamma}}_N^+ + \tilde{\boldsymbol{\gamma}}_N^-)^\mathsf{T}(\boldsymbol{\Lambda}_N - \hat{\boldsymbol{\Lambda}}_N)\end{aligned} \tag{5.17}$$

is also non-positive due to the monotonicity (5.12) of the impact law. Therefore, $\mathrm{d}V = \dot{V}\,\mathrm{d}t + (V^+ + V^-)\,\mathrm{d}\eta \leq 0$, which proves stability of the observer error dynamics. Moreover, under the assumption that the contact duration is always finite (i.e. the unilateral constraints open from time to time) and that the frictional contacts are never permanently in stick (i.e. they slide from time to time), the error dynamics is attractively stable, as shown in Appendix B in analogy to [8].

Remark 5.1. The utilization of the measured switching functions χ_i in the force laws (5.9) of the state observer can be seen as a Luenberger-type correction term inside the nonlinearity. In fact, $y_i = \chi_i$ are measured outputs. Now let the observer's switching functions $\hat{\chi}_i$ be such that $\hat{\chi}_i = 1$ for $g_N^i(\hat{\mathbf{q}}) = 0$ and $\hat{\chi}_i = 0$ for $g_N^i(\hat{\mathbf{q}}) > 0$. Then, with the predicted outputs $\hat{y}_i = \hat{\chi}_i$, it holds that

$$\chi_i = \hat{\chi}_i + (\chi_i - \hat{\chi}_i) = \hat{\chi}_i + L_i(y_i - \hat{y}_i), \tag{5.18}$$

which contains a correction term with a constant gain $L_i = 1$. However, unlike for typical correction terms, these gains L_i are fixed and cannot be used to tune the observer performance.

Remark 5.2. The measured switching functions contain more information than just the impact time instants. If a contact i is closed and $\chi_i(t) = 1$ is measured, then it is known that $g_N^i(\mathbf{q}) = 0$. Therefore, it was suggested in [8] to introduce (unphysical) position jumps in the observer, which includes augmenting (5.8) with an 'impact equation' of the form

$$\mathbf{K}(\hat{\mathbf{q}}^+ - \hat{\mathbf{q}}^-) = \mathbf{W}_N\boldsymbol{\Sigma} \tag{5.19}$$

for the positions. Therein, $\boldsymbol{\Sigma}$ are artificial impulsive forces that are chosen such that $g_N^i(\hat{\mathbf{q}}^+) = 0$ whenever $\chi_i(t) = 1$. For details, the reader is referred to [8, 10].

5.2 Full position measurement

The idea of using position measurements in the force laws of a state observer can also be applied to non-linear mechanical systems. Several such observers have been analyzed in literature for the case where all positions are measured.

Velocity observers (Tanwani et al. [98])

Observers that only estimate velocities have been investigated by Menini and Tornambè [75] in the hybrid systems framework and by Tanwani et al. [98] for MDIs of the form

$$\begin{aligned} \mathrm{d}\mathbf{q} &= \mathbf{u}\,\mathrm{d}t, \\ \mathbf{M}(\mathbf{q})\mathrm{d}\mathbf{u} - \mathbf{h}(t, \mathbf{q}, \mathbf{u})\,\mathrm{d}t &= \mathbf{W}_N\,\mathrm{d}\mathbf{P}_N \end{aligned} \tag{5.20}$$

together with $\mathrm{d}\mathbf{P}_N = \boldsymbol{\lambda}_N \mathrm{d}t + \boldsymbol{\Lambda}_N \mathrm{d}\eta$ and the contact laws

$$-\lambda_N^i \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(\gamma_N^i) & \text{if } g_N^i = 0 \\ 0 & \text{if } g_N^i > 0\,, \end{cases} \quad -\Lambda_N^i \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(\xi^i) & \text{if } g_N^i = 0 \\ 0 & \text{if } g_N^i > 0 \end{cases} \tag{5.21}$$

with the kinematic variable $\xi^i = \gamma_N^{i+} + \varepsilon_i \gamma_N^{i-}$. Since all positions $\mathbf{q}$ are continuously measured, this measurement can be used as an argument to all functions depending on $\mathbf{q}$ in the state observer.

Moreover, if the measurements are accurate enough to directly serve as a position estimate (for example, if the measurements are affected by only a low noise level) it is sufficient to only estimate the unknown velocities $\mathbf{u}$. This can be achieved by using a reduced-order velocity observer [2]of the form

$$\begin{aligned} \mathbf{M}(\mathbf{q})\,\mathrm{d}\mathbf{z} - \hat{\mathbf{h}}(t, \mathbf{q}, \hat{\mathbf{u}})\,\mathrm{d}t + \mathbf{M}(\mathbf{q})\mathbf{L}\hat{\mathbf{u}}\,\mathrm{d}t &= \mathbf{W}_N\,\mathrm{d}\hat{\mathbf{P}}_N, \\ \hat{\mathbf{u}} &= \mathbf{z} + \mathbf{L}\mathbf{q}\,, \end{aligned} \tag{5.22}$$

with the observer state $\mathbf{z}$, together with $\mathrm{d}\hat{\mathbf{P}}_N = \hat{\boldsymbol{\lambda}}_N \mathrm{d}t + \hat{\boldsymbol{\Lambda}}_N \mathrm{d}\eta$ and the contact laws

$$-\hat{\lambda}_N^i \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(\hat{\gamma}_N^i) & \text{if } g_N^i(\mathbf{q}) = 0 \\ 0 & \text{if } g_N^i(\mathbf{q}) > 0\,, \end{cases} \quad -\hat{\Lambda}_N^i \in \begin{cases} \mathcal{N}_{\mathbb{R}_0^+}(\hat{\xi}^i) & \text{if } g_N^i(\mathbf{q}) = 0 \\ 0 & \text{if } g_N^i(\mathbf{q}) > 0 \end{cases} \tag{5.23}$$

with the kinematic variable $\hat{\xi}^i = \hat{\gamma}_N^{i+} + \varepsilon_i \hat{\gamma}_N^{i-}$. Note that $\hat{\mathbf{h}}$ can differ from $\mathbf{h}$, giving more flexibility in the design process. However, if $\mathbf{h}(t, \mathbf{q}, \mathbf{u})$ is globally Lipschitz continuous in $\mathbf{u}$, it is sufficient to take $\hat{\mathbf{h}}(t, \mathbf{q}, \mathbf{u}) = \mathbf{h}(t, \mathbf{q}, \mathbf{u})$ [98].

[2]The order is reduced in the sense that the state of the observer is lower dimensional than the state of the observed system. The structure (5.22) can be seen as a generalization of reduced-order observers for linear system (which are treated for example in [83]).

In order to better understand the structure of (5.22), note that the second equation in (5.22) implies that $\mathrm{d}\hat{\mathbf{u}} = \mathrm{d}\mathbf{z} + \mathbf{L}\mathrm{d}\mathbf{q}$. Together with the first equation in (5.20) it follows $\mathrm{d}\hat{\mathbf{u}} = \mathrm{d}\mathbf{z} + \mathbf{L}\mathrm{d}\mathbf{q} = \mathrm{d}\mathbf{z} + \mathbf{L}\mathbf{u}\mathrm{d}t$, such that the first equation in (5.22) can be re-written as

$$\mathbf{M}(\mathbf{q})\mathrm{d}\hat{\mathbf{u}} - \hat{\mathbf{h}}(t, \mathbf{q}, \hat{\mathbf{u}})\mathrm{d}t - \mathbf{M}(\mathbf{q})\mathbf{L}(\mathbf{u} - \hat{\mathbf{u}})\mathrm{d}t = \mathbf{W}_N \mathrm{d}\hat{\mathbf{P}}_N \,. \tag{5.24}$$

The reformulation (5.24) reveals that the observer dynamics contains a Luenberger-type correction term $\mathbf{L}(\mathbf{u} - \hat{\mathbf{u}})$. However, since the velocities $\mathbf{u}$ are not directly measured, the observer structure (5.22) is more useful.

Attractive stability of the estimation error dynamics can be shown by using a quadratic Lyapunov function $V = \tilde{\mathbf{u}}^\mathsf{T}\mathbf{M}(\mathbf{q})\tilde{\mathbf{u}}$, under additional assumptions on the nonlinear terms $\mathbf{h}$, $\hat{\mathbf{h}}$ as well as $\mathbf{M}(\mathbf{q})$ (refer to [98] for a detailed listing of all assumptions). Similar to Section 5.1, the maximal monotonicity of the contact laws is crucial for attractive stability.

The main drawback of the observer (5.22) in practical implementations is that crucial information is lost by discretely sampling the measurement signal. If the positions are only known at discrete points in time, impacts are easily missed if an impact occurs between two sampling time instants. More comments on this problem will be made in Chapter 7.

Remark 5.3. Similar to Remark 5.1, using the measured positions $\mathbf{y} = \mathbf{q}$ in the force laws (5.23) of the state observer can be interpreted as a Luenberger-type correction term inside the nonlinearity. With the predicted output $\hat{\mathbf{y}} = \hat{\mathbf{q}}$, it holds that

$$g_N^i(\mathbf{q}) = g_N^i(\hat{\mathbf{q}} + (\mathbf{q} - \hat{\mathbf{q}})) = g_N^i(\hat{\mathbf{q}} + \mathbf{L}_i(\mathbf{y} - \hat{\mathbf{y}})), \tag{5.25}$$

which contains a correction term with a constant gain $\mathbf{L}_i = \mathbf{I}$, which cannot be tuned. However, in view of observers for unknown impact time instants, letting this gain be tunable is a possible staring point.

Remark 5.4. Similar to Section 5.1, the reduced order observer (5.22), (5.23) can be extended to include non-opening contacts with Coulomb friction and constant normal forces. This is also true for the full state observer discussed in the following and will be used in the experiments in Chapter 7.

Full state observers (Tanwani et al. [98])

Similar to the velocity observer above, a full state observer can be built for the system (5.20), (5.20), that makes use of a correction term based on the velocity estimation error, while only the positions are measured. Its working principle is similar to the velocity observer above, which is why it is only shortly stated here. The observer takes the form

$$\begin{aligned} \mathrm{d}\mathbf{z}_1 &= \mathbf{z}_2\,\mathrm{d}t + \mathbf{L}_a(\mathbf{q}-\mathbf{z}_1)\mathrm{d}t, \\ \mathbf{M}(\mathbf{q})\,\mathrm{d}\mathbf{z}_2 - \hat{\mathbf{h}}(t,\mathbf{q},\hat{\mathbf{u}})\,\mathrm{d}t - \mathbf{L}_b(\mathbf{q}-\mathbf{z}_1)\,\mathrm{d}t &= \mathbf{W}_N\,\mathrm{d}\hat{\mathbf{P}}_N, \end{aligned} \tag{5.26}$$

with the observer state $\mathbf{z} = (\mathbf{z}_1^\mathsf{T}\ \ \mathbf{z}_2^\mathsf{T})^\mathsf{T}$, together with the force laws (5.23) and the state estimates

$$\begin{aligned} \hat{\mathbf{q}} &= \mathbf{z}_1, \\ \hat{\mathbf{u}} &= \mathbf{z}_2 + \mathbf{L}(\mathbf{q}-\hat{\mathbf{q}}). \end{aligned} \tag{5.27}$$

Furthermore, the gains are defined through $\mathbf{L}_a = L\mathbf{I} + \mathbf{L}_1$ and $\mathbf{L}_b = \mathbf{L}_2 + \mathbf{M}L\mathbf{L}_1$, where $L > 0$, $\mathbf{L}_1 = \mathbf{L}_1^\mathsf{T} \succ 0$ and $\mathbf{L}_2 = \mathbf{L}_2^\mathsf{T} \succ 0$. The meaning of this structure becomes more clear after a reformulation, which is found as

$$\begin{aligned} \mathrm{d}\hat{\mathbf{q}} &= \hat{\mathbf{u}}\,\mathrm{d}t + \mathbf{L}_1(\mathbf{q}-\hat{\mathbf{q}})\,\mathrm{d}t, \\ \mathbf{M}(\mathbf{q})\mathrm{d}\hat{\mathbf{u}} - \hat{\mathbf{h}}(t,\mathbf{q},\hat{\mathbf{u}})\mathrm{d}t - \mathbf{L}_2(\mathbf{q}-\hat{\mathbf{q}})\mathrm{d}t - \mathbf{M}(\mathbf{q})L(\mathbf{u}-\hat{\mathbf{u}})\mathrm{d}t &= \mathbf{W}_N\mathrm{d}\hat{\mathbf{P}}_N. \end{aligned} \tag{5.28}$$

The reformulation (5.28) shows that the observer contains two Luenberger-type correction terms, depending on the position and velocity errors. As for the velocity observer, the formulation (5.26), (5.27) is more useful if only the positions are measured.

5.3 Monotonicity, passivity and linear matrix inequalities

The state observer designs in Section 5.1 and Section 5.2 have been obtained by first selecting a specific dynamics for the observer and then showing attractive stability using Lyapunov-type stability theorems (as in Chapter 4). A system property that is closely related to Lyapunov stability and for linear systems allows to conclude stability based on linear matrix inequalities, is passivity. In the following, some basic relations are discussed for linear systems, which allow for a passivity interpretation of the previous observers and which will also be used in a discrete (and extended) form in Chapter 6.

Definition 5.5 ([25]). A linear time-invariant system of the form

$$\begin{aligned} \dot{\mathsf{x}} &= \mathsf{A}\mathsf{x} + \mathsf{B}\mathsf{w}, \\ \mathsf{z} &= \mathsf{C}\mathsf{x}, \end{aligned} \tag{5.29}$$

with the state $\mathsf{x} \in \mathbb{R}^n$, an input $\mathsf{w} \in \mathbb{R}^p$ and an output $\mathsf{z} \in \mathbb{R}^p$, is said to be **passive** if there exists a non-negative function $V : \mathbb{R}^n \to \mathbb{R}$ (called the storage function) with

$V(\mathbf{0}) = 0$ such that

$$V(\mathbf{x}^+(t)) - V(\mathbf{x}^-(t_0)) \leq \int_{t_0}^{t} \frac{1}{2}\mathbf{z}(\tau)^\mathsf{T}\mathbf{w}(\tau)\,\mathrm{d}\tau \tag{5.30}$$

$\forall\,\mathbf{w}$ and $\forall\,t \geq t_0$. It is moreover called **strictly passive** if

$$V(\mathbf{x}^+(t)) - V(\mathbf{x}^-(t_0)) \leq \int_{t_0}^{t} \frac{1}{2}\mathbf{z}(\tau)^\mathsf{T}\mathbf{w}(\tau)\,\mathrm{d}\tau - \int_{t_0}^{t} S(\mathbf{x}(\tau))\,\mathrm{d}\tau \tag{5.31}$$

$\forall\,\mathbf{w}$ and $\forall\,t \geq t_0$, with a positive definite function $S : \mathbb{R}^n \to \mathbb{R}$ (which ensures that the inequality (5.30) holds strictly, unless $\mathbf{x}(\tau) = 0\,\forall\,\tau \in [t_0, t]$).

The idea behind this definition is that if V is the internal energy, a passive system does not produce energy during its motion. The change in the internal energy is always lower or equal to the supplied energy, here given by the right hand side of (5.30).

Clearly, if the relation between the input and the output is such that $\mathbf{z}^\mathsf{T}\mathbf{w} \leq 0\,\forall\,t$, then the storage function V serves as a Lyapunov function for a stability proof. This is especially useful when analyzing an error dynamics such as

$$\begin{aligned}\dot{\tilde{\mathbf{x}}} &= \mathbf{A}\tilde{\mathbf{x}} + \mathbf{B}\tilde{\mathbf{w}},\\ \tilde{\mathbf{z}} &= \mathbf{C}\tilde{\mathbf{x}},\end{aligned} \tag{5.32}$$

with an error $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$, input difference $\tilde{\mathbf{w}} = \mathbf{w} - \hat{\mathbf{w}}$ and output difference $\tilde{\mathbf{z}} = \mathbf{z} - \hat{\mathbf{z}}$. Then, if the inputs are defined by a given monotone mapping $\mathcal{H} : \mathbb{R}^p \to \mathbb{R}^p$ through $-\mathbf{w} \in \mathcal{H}(\mathbf{z})$ and $-\hat{\mathbf{w}} \in \mathcal{H}(\hat{\mathbf{z}})$, the inequality $\tilde{\mathbf{z}}^\mathsf{T}\tilde{\mathbf{w}} = (\mathbf{z} - \hat{\mathbf{z}})^\mathsf{T}(\mathbf{w} - \hat{\mathbf{w}}) \leq 0$ directly follows from the monotonicity property of $\mathcal{H}$.

Passivity, being a system property, can conveniently be checked using the following theorem, in which the notation $\mathbf{M} \succeq 0$ and $\mathbf{M} \succ 0$ is used to express that a matrix $\mathbf{M}$ is positive semidefinite and positive definite, respectively. Likewise, $\mathbf{M} \preceq 0$ and $\mathbf{M} \prec 0$ express negative semidefiniteness and negative definiteness of $\mathbf{M}$.

Theorem 5.6 ([24]). System (5.32) is passive if and only if there exists a matrix $\mathbf{P} = \mathbf{P}^\mathsf{T} \succeq 0$ such that the following matrix conditions hold

$$\mathbf{A}^\mathsf{T}\mathbf{P} + \mathbf{P}\mathbf{A} \preceq 0, \tag{5.33a}$$

$$\mathbf{P}\mathbf{B} - \mathbf{C}^\mathsf{T} = \mathbf{0}. \tag{5.33b}$$

In that case $V = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{P}\mathbf{x}$ is a storage function. Moreover, system (5.32) is *strictly* passive if and only if the inequality (5.33a) holds strictly, i.e. there exists a matrix $\mathbf{Q} \succ 0$ such that

$$\mathbf{A}^\mathsf{T}\mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{Q} \tag{5.34}$$

Proof. Here, only the sufficiency part is shown, which is easily checked by calculation of $\dot{V}$. The time derivative of $V = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{P}\mathbf{x}$ is

$$\dot{V} = \mathbf{x}^\mathsf{T}\mathbf{P}\dot{\mathbf{x}} = \frac{1}{2}\mathbf{x}^\mathsf{T}(\mathbf{A}^\mathsf{T}\mathbf{P} + \mathbf{P}\mathbf{A})\mathbf{x} + \mathbf{x}^\mathsf{T}\mathbf{P}\mathbf{B}\mathbf{w}. \tag{5.35}$$

Therefore, with the conditions (5.33) it follows

$$\dot{V} \leq \mathbf{x}^\mathsf{T}\mathbf{P}\mathbf{B}\mathbf{w} = \mathbf{x}^\mathsf{T}\mathbf{C}^\mathsf{T}\mathbf{w} = \mathbf{z}^\mathsf{T}\mathbf{w}, \tag{5.36}$$

which is equivalent to (5.30). Similarly, if (5.33a) and (5.34) hold, we have

$$\dot{V} = -\mathbf{x}^\mathsf{T}\mathbf{Q}\mathbf{x} + \mathbf{x}^\mathsf{T}\mathbf{P}\mathbf{B}\mathbf{w} = \mathbf{x}^\mathsf{T}\mathbf{C}^\mathsf{T}\mathbf{w} - \mathbf{x}^\mathsf{T}\mathbf{Q}\mathbf{x} = \mathbf{z}^\mathsf{T}\mathbf{w} - \mathbf{x}^\mathsf{T}\mathbf{Q}\mathbf{x}, \tag{5.37}$$

from which it follows that (5.31) holds. The necessity part can be found in [24] (see p. 85). □

Theorem 5.6 is a useful tool in many design processes that involve stability. It therefore leads to the question, if similar results hold for non-smooth systems. However, it is not straightforward how to define passivity for such systems in the first place. The following definition, being proposed here for systems that are linear during non-impulsive motion, is motivated by energy considerations.

Definition 5.7. A time-invariant system of the form

$$\begin{aligned} \mathrm{d}\mathbf{x} &= (\mathbf{A}\mathbf{x})\,\mathrm{d}t + \mathbf{B}\,\mathrm{d}\boldsymbol{\omega}, \\ \mathbf{z} &= \mathbf{C}\mathbf{x}, \end{aligned} \tag{5.38}$$

with the state $\mathbf{x} \in \mathbb{R}^n$, an input $\mathrm{d}\boldsymbol{\omega} = \mathbf{w}\mathrm{d}t + \mathbf{W}\mathrm{d}\eta \in \mathbb{R}^p$ and an output $\mathbf{z} \in \mathbb{R}^p$, is said to be **passive** if there exists a non-negative function $V : \mathbb{R}^n \to \mathbb{R}$ (called the storage function) with $V(\mathbf{0}) = 0$ such that

$$V(\mathbf{x}^+(t)) - V(\mathbf{x}^-(t_0)) \leq \int_{[t_0,t]} \frac{1}{2}(\mathbf{z}^+ + \mathbf{z}^-)^\mathsf{T}\mathrm{d}\boldsymbol{\omega} \tag{5.39}$$

for all $\mathrm{d}\boldsymbol{\omega}$ and for all $t \geq t_0$. It is moreover called **strictly passive** if

$$V(\mathbf{x}^+(t)) - V(\mathbf{x}^-(t_0)) \leq \int_{[t_0,t]} \frac{1}{2}(\mathbf{z}^+ + \mathbf{z}^-)^\mathsf{T}\mathrm{d}\boldsymbol{\omega} - \int_{[t_0,t]} S(\mathbf{x}(\tau))\,\mathrm{d}\tau \tag{5.40}$$

for all $\mathrm{d}\boldsymbol{\omega}$ and for all $t \geq t_0$, with a positive definite function $S : \mathbb{R}^n \to \mathbb{R}$ the inequality (5.39) holds strictly, unless $\mathbf{x}(\tau) = 0\,\forall\,\tau \in [t_0, t]$.

Note that during non-impulsive motion, $\mathbf{z}^+ = \mathbf{z}^- = \mathbf{z}$, such that the inequality (5.39) is identical to (5.30). With Definition 5.7, the following can be stated.

Proposition 5.8. System (5.38) is passive if and only if there exists a matrix $\mathbf{P} = \mathbf{P}^\mathsf{T} \succeq 0$ such that the conditions (5.33) from Theorem 5.6 hold. In that case $V = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{P}\mathbf{x}$ is a storage function. Moreover, system (5.38) is strictly passive if and only if in addition (5.34) holds.

Proof. First, consider a time interval with non-impulsive motion. During that time interval the dynamics (5.38) is equivalent to the linear system (5.29) and the inequalities (5.39), (5.40) are equivalent to (5.30) and (5.31). Therefore, it follows from Theorem 5.6 that the conditions (5.33) are necessary for passivity and, in addition, (5.34) is necessary for strict passivity. Moreover (5.33) is sufficient for passivity during non-impulsive motion, and $V = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{P}\mathbf{x}$ is a storage function. Furthermore, (5.34) is sufficient for strict passivity. What is left to show, is that (5.39) holds during impulsive motion, which follows from calculating

$$\begin{aligned} V(\mathbf{x}^+) - V(\mathbf{x}^-) &= \frac{1}{2}(\mathbf{x}^+ + \mathbf{x}^-)^\mathsf{T}\mathbf{P}(\mathbf{x}^+ - \mathbf{x}^-) \\ &= \frac{1}{2}(\mathbf{x}^+ + \mathbf{x}^-)^\mathsf{T}\mathbf{P}\mathbf{B}\mathbf{W} \\ &= \frac{1}{2}(\mathbf{z}^+ + \mathbf{z}^-)^\mathsf{T}\mathbf{W}, \end{aligned} \tag{5.41}$$

where the impulsive dynamics $\mathbf{x}^- - \mathbf{x}^- = \mathbf{B}\mathbf{W}$ and conditions (5.33) have been used. $\square$

6

Discrete state observer with unknown impact time instants

In this chapter, the state observer problem is investigated for mechanical systems with impulsive motion, i.e. systems with state jumps caused by unilateral constraints, without explicitly measuring the impact time instants. It is an important aspect for the state observer design for systems with state jumps whether or not the time instants at which the state jumps occur are known. As discussed in Chapter 5, most proposed observers assume that these impact time instants can directly be extracted from measurements, for example by measuring all relevant positions in a system [74, 75, 98] or by directly measuring the presence of contact through contact or tactile sensors [11]. This allows for the design of a state observer that exhibits state jumps at the exact same time instants as the observed system. By exploiting the assumption of maximal monotonicity of the impact law, it is then possible to construct a Lyapunov function for the estimation error dynamics (i.e. the time evolution of the difference between the estimated state and the actual state) which does not increase over impacts. Simply put, the observer problem then reduces to stabilizing the error dynamics for the non-impulsive motion (restricted to the constructed Lyapunov function).

Only few attempts have been made to design state observers in the case of unknown impact time instants. The main difficulty for observer design with unknown impact time instants is that the state jumps of the observed system and the state observer do not coincide [62, 76]. This results in the peaking phenomenon: even if the observer state nearly matches the real state, a slight mismatch in the impact time instants leads

to a temporarily large Euclidean velocity error caused by velocity jumps [17,19,68,94]. The peaking phenomenon makes it difficult to show the asymptotic stability of the error dynamics based on Lyapunov's direct method. One approach for such systems is to find a state transformation that transforms the original system into a new system without state jumps (as discussed in Chapter 3), for which conventional state observer techniques can be applied [62,76]. However, such a transformation does not always exist and is in general difficult to find. Another approach is to introduce a distance metric that gives a distance between two states (such as from the observed system and a state observer), but does not change its value over state jumps [17]. Such a distance metric that is 'blind' to state jumps can be used together with suitable stability notions (see e.g. [18] for definitions of incremental stability for hybrid systems) to find a corresponding Lyapunov function. These approaches have been shown to be useful to solve the tracking problem [19] or for controlled synchronization [9], where all states are known. However, they do not imply a suitable state observer design and calculations can be cumbersome.
The starting point for a new approach is to recognize two main hurdles which are inherent to the problem of observer design for unilaterally constrained mechanical systems. The first problem is the simple fact that jumps in the state, occurring in a continuous-time system, generally form a hurdle for Lyapunov-type analysis and thereby for observer design. The second problem is related to the impact law describing the velocity jump in unilaterally constrained systems. As discussed in Chapter 3, instantaneous impact laws such as Newton's or Poisson's impact law [43,44] are formulated on velocity level, i.e. they directly relate post-impact relative velocities to pre-impact relative velocities. The generalized versions of these impact laws distinguish between superfluous unilateral constraints which, although closed, do not participate in the impact process, and actively participating unilateral constraints. The combined active-inactive behavior of generalized impact laws on velocity level is conveniently expressed through set-valued functions, e.g. normal cone inclusions, which enjoy the favorable property of maximal monotonicity (being related to, but in some sense more strong than, dissipativity of the impact law [102]). Maximal monotonicity of force laws or impact laws leads to contraction properties, which in essence are favorable for tracking or observer design. However, and her lies the problem, the generalized impact laws are only to be applied to *closed* unilateral constraints (i.e. when contact is present). Instantaneous impact laws for multibody systems are therefore formulated on position-switched velocity level. The switching on position level (from closed to open and vice versa) destroys the favorable properties of these impact laws, making the observer design of unilaterally constrained mechanical systems an incredibly difficult task. These two key problems, state jumps and loss

of maximal monotonicity of the impact law, explain why the observer design of this class of systems stands as a major problem.
In the following, a new approach is investigated, with the goal of circumventing the main difficulties. First, instead of analyzing the continuous-time problem with state jumps, a numerical scheme is used to transform the system to a discrete-time system, approximating the former depending on the chosen time step. The time discretization sidesteps the problem of state jumps as a discrete system only describes state updates over time steps. Moreover, a practical implementation of an observer on digital hardware requires the transformation to discrete-time. As discussed in Chapter 3, the time-stepping scheme of Moreau [79] (see also [1, 68]) is perhaps the most celebrated (velocity-impulse-based) scheme within the Nonsmooth Dynamics community as it can be applied to the simulation of systems with multiple unilateral constraints with Coulomb friction. Recall that the Moreau scheme directly discretizes the equality of measures, which describes the system dynamics, and a combined contact-impact law on position-switched velocity level. Regarding the second key problem, it is favorable to make use of the Paoli-Schatzman scheme, which involves an impact law directly formulated on position level, instead of on position-switched velocity level. Thereby the problem of switching of the impact law on position level is circumvented, giving access to the maximal monotonicity property and its related contraction property. The aim of this chapter is to investigate if this approach is useful to solve the observer design problem of unilaterally constrained mechanical systems without measuring impact time instants. It is based on the results in [89, 90] of the author.
A related area of research is the design of observers within the switched systems framework, i.e. for systems whose dynamics are described by a set of subsystems and a switching law describing how to switch between them. The switching can depend on the value of an external switching signal, in which case a distinction is made between known and unknown switching signals (with unknown switching signals making the design of observers more difficult). To name just a few examples, in [2, 103] systems with an external switching system are investigated. In other cases the switching is state dependent [57, 99]. Even though these publications are not concerned with impulsive motion (i.e. with state jumps), they are related to the approach taken here. More precisely, a discrete-time system will be used, which can be recast as a piecewise affine system where the switching is state and input dependent.

The outline of this chapter is as follows. In Section 6.1 the continuous-time observer problem is formulated for a specific system class. Based on the Paoli-Schatzman scheme, a suitable time discretization is then derived in Section 6.2. Subsequently, a deadbeat observer for the discrete observer problem is presented in Section 6.3. Furthermore, a passivity-based observer design for linear complementarity systems is transported to the discrete-time setting in Section 6.4. Finally, numerical results for an impact oscillator system are given in Section 6.5. The usefulness of the presented approach is discussed in Section 6.6.

6.1 Continuous-time state observer problem

Here, linear mechanical systems with unilateral constraints of the form

$$\begin{aligned} \mathrm{d}\mathbf{q} &= \mathbf{u}\mathrm{d}t, \\ \mathbf{M}\mathrm{d}\mathbf{u} + (\mathbf{K}\mathbf{q} + \mathbf{D}\mathbf{u} - \mathbf{f}(t))\mathrm{d}t &= \mathbf{W}\mathrm{d}\mathbf{P} \end{aligned} \tag{6.1}$$

are considered (similar to Section 5.1, but without considering friction). For the sake of simplicity, it is assumed that the mass matrix $\mathbf{M}$, the stiffness matrix $\mathbf{K}$ and the damping matrix $\mathbf{D}$ are constant. Furthermore, the unilateral constraints are described by linear inequality conditions $\boldsymbol{g}(\mathbf{q}) = \mathbf{W}^\mathsf{T}\mathbf{q} \geq 0$ and the generalized force directions, given by the columns of $\mathbf{W} = (\partial\boldsymbol{g}/\partial\mathbf{q})^\mathsf{T}$, are assumed to be constant and linearly independent ($\mathbf{W}$ has full rank). Note that since all contacts are assumed to be frictionless, the indices referring to the normal direction are omitted for easier readability, e.g. it is written $\mathrm{d}\mathbf{P}$ instead of $\mathrm{d}\mathbf{P}_N$. The system is subjected to a bounded, time-dependent external forcing $\mathbf{f}(t)$. As before, the differential contact effort measure is $\mathrm{d}\mathbf{P} = \boldsymbol{\lambda}\mathrm{d}t + \boldsymbol{\Lambda}\mathrm{d}\eta$ and for the components of the constraint forces $\boldsymbol{\lambda}$ the formulation (3.39) on position-switched velocity level is used. With $\boldsymbol{\gamma} = \mathbf{W}^\mathsf{T}\mathbf{u}$ it can be expressed component-wise as

$$g_i(\mathbf{q}) = 0 : 0 \leq \gamma_i \perp \lambda_i \geq 0\ , \quad g_i(\mathbf{q}) > 0 : \lambda_i = 0. \tag{6.2}$$

In addition, the generalized Newtonian impact law (3.50) is used, which is written component-wise as

$$g_i(\mathbf{q}) = 0 :\ 0 \leq \xi_i \perp \Lambda_i \geq 0\ , \quad g_i(\mathbf{q}) > 0 :\ \Lambda_i = 0, \tag{6.3}$$

with the kinematic variable $\xi_i := \gamma_i^+ + \varepsilon\gamma_i^-$. For the sake of simplification, a global coefficient of restitution $\varepsilon \in [0, 1]$ is assumed, although this is not essential.

6.2 Discrete-time state observer problem

In the following an approach is pursued, where first the dynamics is discretized and then a state observer for the discrete (and therefore approximate) system is designed. The main goal of this approach is to alleviate the problem of state jumps in the

observer design.
For system (6.1) the Paoli-Schatzman discretization scheme can be written as

$$\begin{aligned} \mathbf{q}_{k+1} &= \mathbf{q}_k + \Delta t \mathbf{u}_{k+1}, \\ \mathbf{M}(\mathbf{u}_{k+1} - \mathbf{u}_k) &+ (\mathbf{K}\mathbf{q}_k + \mathbf{D}\mathbf{u}_k - \mathbf{f}_k)\,\Delta t = \mathbf{W}\mathbf{P}_k, \end{aligned} \tag{6.4}$$

together with the discrete law

$$\begin{aligned} \boldsymbol{\zeta}_k &:= \boldsymbol{g}_{k+1} + \varepsilon \boldsymbol{g}_{k-1}, \\ 0 \le \boldsymbol{\zeta}_k &\perp \mathbf{P}_k \ge 0, \end{aligned} \tag{6.5}$$

where Δt is the (constant) time step size and an index k refers to the corresponding variable being evaluated (or approximated) at $t = t_k := k\Delta t$, e.g. $\mathbf{q}_k := \mathbf{q}(t_k)$. The above discretization has the form of a semi-implicit Euler scheme, directly applied to the MDI (6.1). Likewise the discrete contact distance is $\boldsymbol{g}_k = \mathbf{W}^\top \mathbf{q}_k$ and the corresponding discrete contact velocity is $\boldsymbol{\gamma}_k = \mathbf{W}^\top \mathbf{u}_k$.

Recall that what makes the Paoli-Schatzman scheme special is the fact that the discrete impact law (6.5), i.e. $0 \le \boldsymbol{g}_{k+1} + \varepsilon \boldsymbol{g}_{k-1} \perp \mathbf{P}_k \ge 0$, is formulated on position level, thereby circumventing the calculation of an index set. As discussed in Section 3.4, the somewhat heuristic impact law contains Newton's impact law in a discretized sense. Velocity jumps that occur instantaneously in continuous time take place over an interval of two time steps in the discretization, which can be seen as a regularization.

In the following, the state space representation of the discretized system (6.4), (6.5) is derived, with the discrete state $\mathsf{x}_k := \begin{pmatrix} \mathbf{q}_k^\top & \mathbf{u}_k^\top \end{pmatrix}^\top$. Recall that variables and matrices related to the state space description are written without serifs, whereas in the original description of the mechanical system serifs were used (therefore, variables denoted by the same letter, are assigned a different meaning depending on whether they are written with or without serifs). Rewriting the stepping equations (6.4) in matrix form yields

$$\begin{pmatrix} \mathbf{I} & -\Delta t \mathbf{I} \\ \mathbf{0} & \mathbf{M} \end{pmatrix} \mathsf{x}_{k+1} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\Delta t \mathbf{K} & \mathbf{M} - \Delta t \mathbf{D} \end{pmatrix} \mathsf{x}_k + \begin{pmatrix} \mathbf{0} \\ \mathbf{W} \end{pmatrix} \mathbf{P}_k + \begin{pmatrix} \mathbf{0} \\ \Delta t \mathbf{I} \end{pmatrix} \mathbf{f}_k. \tag{6.6}$$

Therein, $\mathbf{I}$ and $\mathbf{0}$ denote an identity matrix and a zero matrix of appropriate dimensions. The matrix on the left hand side can be inverted, which leads to an update rule for the state

$$\mathsf{x}_{k+1} = \begin{pmatrix} \mathbf{I} & \Delta t \mathbf{M}^{-1} \\ \mathbf{0} & \mathbf{M}^{-1} \end{pmatrix} \left[\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\Delta t \mathbf{K} & \mathbf{M} - \Delta t \mathbf{D} \end{pmatrix} \mathsf{x}_k + \begin{pmatrix} \mathbf{0} \\ \mathbf{W} \end{pmatrix} \mathbf{P}_k + \begin{pmatrix} \mathbf{0} \\ \Delta t \mathbf{I} \end{pmatrix} \mathbf{f}_k \right]. \tag{6.7}$$

Finally, after simple matrix multiplications, the discrete dynamics can be written as

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{P}_k + \mathbf{E}\mathbf{f}_k, \tag{6.8}$$

with the corresponding system matrix $\mathbf{A}$, given by

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} - \Delta t^2 \mathbf{M}^{-1}\mathbf{K} & \Delta t(\mathbf{I} - \Delta t \mathbf{M}^{-1}\mathbf{D}) \\ -\Delta t \mathbf{M}^{-1}\mathbf{K} & \mathbf{I} - \Delta t \mathbf{M}^{-1}\mathbf{D} \end{pmatrix}, \tag{6.9}$$

and the matrices $\mathbf{B}$ and $\mathbf{E}$

$$\mathbf{B} = \begin{pmatrix} \Delta t \mathbf{M}^{-1}\mathbf{W} \\ \mathbf{M}^{-1}\mathbf{W} \end{pmatrix}, \qquad \mathbf{E} = \begin{pmatrix} \Delta t^2 \mathbf{M}^{-1} \\ \Delta t \mathbf{M}^{-1} \end{pmatrix}. \tag{6.10}$$

To complete the state space description, the discrete impact law (6.5) has to be expressed in the state variables. This is achieved by using the contact distance $\boldsymbol{g}_k = \mathbf{W}^\mathsf{T}\mathbf{q}_k$ and substitution of the first equation of (6.4) in (6.5):

$$\begin{aligned} \boldsymbol{\zeta}_k &= \mathbf{W}^\mathsf{T}(\mathbf{q}_{k+1} + \varepsilon \mathbf{q}_{k-1}) = \mathbf{W}^\mathsf{T}(\mathbf{q}_{k+1} + \varepsilon(\mathbf{q}_k - \Delta t \mathbf{u}_k)) \\ &= \begin{pmatrix} \mathbf{W}^\mathsf{T} & \mathbf{0} \end{pmatrix} \mathbf{x}_{k+1} + \varepsilon \begin{pmatrix} \mathbf{W}^\mathsf{T} & -\Delta t \mathbf{W}^\mathsf{T} \end{pmatrix} \mathbf{x}_k \\ &= \begin{pmatrix} \mathbf{W}^\mathsf{T} & \mathbf{0} \end{pmatrix} [\mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{P}_k + \mathbf{E}\mathbf{f}_k] + \varepsilon \begin{pmatrix} \mathbf{W}^\mathsf{T} & -\Delta t \mathbf{W}^\mathsf{T} \end{pmatrix} \mathbf{x}_k \\ &= \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{P}_k + \mathbf{F}\mathbf{f}_k, \end{aligned} \tag{6.11}$$

with the matrix $\mathbf{C}$ defined by

$$\mathbf{C} = \begin{pmatrix} [(1+\varepsilon)\mathbf{I} - \Delta t^2 \mathbf{M}^{-1}\mathbf{K}]^\mathsf{T}\mathbf{W} \\ \Delta t[(1-\varepsilon)\mathbf{I} - \Delta t \mathbf{M}^{-1}\mathbf{D}]^\mathsf{T}\mathbf{W} \end{pmatrix}^\mathsf{T}, \tag{6.12}$$

and the matrices $\mathbf{D}$ and $\mathbf{F}$ given by

$$\mathbf{D} = \Delta t \mathbf{W}^\mathsf{T}\mathbf{M}^{-1}\mathbf{W}\,, \qquad \mathbf{F} = \Delta t^2 \mathbf{W}^\mathsf{T}\mathbf{M}^{-1}. \tag{6.13}$$

The matrix $\mathbf{D}$ above is recognized to be a scaled version of the so-called Delassus matrix $\mathbf{W}^\mathsf{T}\mathbf{M}^{-1}\mathbf{W}$ [22], which is symmetric and positive definite as $\mathbf{W}$ is assumed to have full column rank. Summarizing the discrete system dynamics (6.8), (6.11) and (6.5) and introducing an output equation $\mathbf{y}_k = \mathbf{G}\mathbf{x}_k$ (i.e. the available measurements) results in

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{P}_k + \mathbf{E}\mathbf{f}_k, \tag{6.14a}$$

$$\boldsymbol{\zeta}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{P}_k + \mathbf{F}\mathbf{f}_k, \tag{6.14b}$$

$$0 \leq \boldsymbol{\zeta}_k \perp \mathbf{P}_k \geq 0, \tag{6.14c}$$

$$\mathbf{y}_k = \mathbf{G}\mathbf{x}_k, \tag{6.14d}$$

which is a discrete **linear complementarity system** (LCS) [50, 53]. For a given state $\mathbf{x}_k$ and excitation $\mathbf{f}_k$, the equations (6.14b) and (6.14c) form together a **linear complementarity problem** (LCP) [33, 80], which has to be solved for $\boldsymbol{\zeta}_k$ and $\mathbf{P}_k$ in each time step.

Remark 6.1. As noted in [85], the time-stepping scheme of Paoli-Schatzman admits a unique solution if the set $\mathcal{A} := \{\mathbf{q} \in \mathbb{R}^f \mid \boldsymbol{g}(\mathbf{q}) \geq 0\}$ of admissible positions is convex. Here, the inequality constraints are restricted to linear constraints of the form $\boldsymbol{g}(\mathbf{q}) = \mathbf{W}^\mathsf{T}\mathbf{q}$. It is therefore straightforward to verify that $\mathcal{A}$ is always convex in this setting. Also, the LCP (6.14b), (6.14c) has a unique solution if all principal minors of the matrix $\mathbf{D}$ are strictly positive (i.e. it is a so-called $\mathcal{P}$-matrix, see for example [33]), which is fulfilled since $\mathbf{D}$ is symmetric and positive definite.

6.3 A discrete-time deadbeat observer

A state observer that is able to reconstruct the exact state in finite time is commonly called a deadbeat observer. One way to obtain such a deadbeat observer is to simply calculate the state from a collection of known system outputs. Clearly, this method requires exact output measurements and a perfectly accurate model. For discrete linear time invariant systems, one way to calculate the initial state is to propagate the discrete dynamics over $n-1$ steps and to relate it to the measured output in each step. This results in a system of linear equations that can be solved for the initial condition. In the following, it will be shown that for discrete LCS, it is possible to reconstruct the current state from a number of outputs, by solving a linear complementarity problem. Sufficient conditions guaranteeing the existence of a unique solution to this LCP then serve as an observability condition. Unsurprisingly, one of these conditions is that the unconstrained motion (which is described by a linear system) is observable. In order to use a more standard notation, consider discrete LCS of the form

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{w}_k + \mathbf{E}\mathbf{v}_k, \\ \mathbf{z}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{w}_k + \mathbf{F}\mathbf{v}_k, \\ 0 \leq \mathbf{z}_k &\perp \mathbf{w}_k \geq 0, \\ \mathbf{y}_k &= \mathbf{G}\mathbf{x}_k, \end{aligned} \tag{6.15}$$

with the state $\mathbf{x}_k$, some external input $\mathbf{v}_k$, the output $\mathbf{y}_k$ and the complementary variables $\mathbf{z}_k$ and $\mathbf{w}_k$, which play the role of the kinematic variable $\boldsymbol{\zeta}_k$ and the discrete percussion $\mathbf{P}_k$ in (6.14). For the sake of simplicity, the treatment is restricted for the moment to the case without external inputs, i.e. $\mathbf{v}_k = \mathbf{0}\,\forall\, k$. However, all subsequent steps can straightforwardly be extended to include inputs, as will be discussed in

Remark 6.3 below. The first step is to connect the outputs to the initial condition by successively calculating the outputs using (6.15), which results in

$$\begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{pmatrix} = \begin{bmatrix} \mathbf{G} \\ \mathbf{GA} \\ \vdots \\ \mathbf{GA}^k \end{bmatrix} \mathbf{x}_0 + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{GB} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{GAB} & \mathbf{GB} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & & & & \vdots \\ \mathbf{GA}^{k-1}\mathbf{B} & \mathbf{GA}^{k-2}\mathbf{B} & \dots & \mathbf{GB} & \mathbf{0} \end{bmatrix} \begin{pmatrix} \mathbf{w}_0 \\ \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_k \end{pmatrix}. \tag{6.16}$$

For a more compact notation, it is convenient to summarize (6.16) with $\mathbf{Y}_k := \left(\mathbf{y}_0^\mathsf{T} \dots \mathbf{y}_k^\mathsf{T}\right)^\mathsf{T}$ and $\mathbf{W}_k := \left(\mathbf{w}_0^\mathsf{T} \dots \mathbf{w}_k^\mathsf{T}\right)^\mathsf{T}$ in

$$\mathbf{Y}_k = \mathcal{O}_k \mathbf{x}_0 + \mathbf{M}_k \mathbf{W}_k, \tag{6.17}$$

where $\mathcal{O}_k$ and $\mathbf{M}_k$ represent the matrices in square brackets in (6.16). In a similar fashion, the corresponding sequence

$$\begin{pmatrix} \mathbf{z}_0 \\ \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_k \end{pmatrix} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^k \end{bmatrix} \mathbf{x}_0 + \begin{bmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{CB} & \mathbf{D} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{CAB} & \mathbf{CB} & \mathbf{D} & \dots & \mathbf{0} \\ \vdots & & & & \\ \mathbf{CA}^{k-1}\mathbf{B} & \mathbf{CA}^{k-2}\mathbf{B} & \dots & \mathbf{CB} & \mathbf{D} \end{bmatrix} \begin{pmatrix} \mathbf{w}_0 \\ \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_k \end{pmatrix}, \tag{6.18}$$

is summarized with $\mathbf{Z}_k = \left(\mathbf{z}_0^\mathsf{T} \dots \mathbf{z}_k^\mathsf{T}\right)^\mathsf{T}$ in the compact form

$$\mathbf{Z}_k = \overline{\mathcal{O}}_k \mathbf{x}_0 + \overline{\mathbf{M}}_k \mathbf{W}_k. \tag{6.19}$$

Note that matrices with a similar structure are denoted by the same letter, but distinguished by overlines. Since $\mathbf{z}_i$ and $\mathbf{w}_i$ satisfy the inequality complementarity $0 \leq \mathbf{z}_i \perp \mathbf{w}_i \geq 0$ for all i, it directly follows that

$$0 \leq \mathbf{Z}_k \perp \mathbf{W}_k \geq 0. \tag{6.20}$$

Finally, after propagating the first equation in (6.15) over k time steps with $\mathbf{v}_k = \mathbf{0}\,\forall\, k$, $\mathbf{x}_k$ can be expressed as

$$\mathbf{x}_k = \mathbf{A}^k \mathbf{x}_0 + \begin{bmatrix} \mathbf{A}^{k-1}\mathbf{B} & \mathbf{A}^{k-2}\mathbf{B} & \dots & \mathbf{B} & \mathbf{0} \end{bmatrix} \mathbf{W}_k =: \mathbf{A}^k \mathbf{x}_0 + \mathbf{Q}_k \mathbf{W}_k. \tag{6.21}$$

Now let $k = n - 1$ with the number of states n. For a better readability, all indices are omitted if they equal $n - 1$. Then, the equations (6.17), (6.19) and (6.20) with known outputs $\mathbf{Y} := \mathbf{Y}_{n-1}$ and the unknown initial state $\mathbf{x}_0$ and contact efforts $\mathbf{W} := \mathbf{W}_{n-1}$ are

$$\begin{aligned} \mathbf{Y} &= \mathcal{O}\mathbf{x}_0 + \mathbf{M}\mathbf{W}, \\ \mathbf{Z} &= \overline{\mathcal{O}}\mathbf{x}_0 + \overline{\mathbf{M}}\mathbf{W}, \\ 0 \leq \mathbf{Z} &\perp \mathbf{W} \geq 0, \end{aligned} \tag{6.22}$$

and form a **mixed linear complementarity problem** (MLCP) [46]. Therein, the matrices $\mathcal{O}, \mathbf{M}, \overline{\mathcal{O}}$ and $\overline{\mathbf{M}}$ are given by the system properties. The goal is to calculate the initial state $\mathbf{x}_0$ and complementarity variables $\mathbf{W}$ for given outputs $\mathbf{Y}$. It is worth mentioning that the MLCP (6.22) does have a solution, because it is generated by (6.14), which does admit a solution. The matrix $\mathcal{O}$ is the well known observability matrix for the non-impulsive motion. Therefore, if the system is observable in the absence of impacts, $\mathcal{O}$ has full column rank and the first equation of (6.22) can uniquely be solved for

$$\mathbf{x}_0 = \mathcal{O}^\dagger \left[\mathbf{Y} - \mathbf{M}\mathbf{W}\right], \tag{6.23}$$

with the left inverse $\mathcal{O}^\dagger$ (which is equal to the inverse $\mathcal{O}^{-1}$ if $\mathcal{O}$ is square). By inserting (6.23) in the remaining equations of the MLCP one arrives at

$$\begin{aligned} \mathbf{Z} &= \left[\overline{\mathbf{M}} - \overline{\mathcal{O}}\mathcal{O}^\dagger\mathbf{M}\right]\mathbf{W} + \overline{\mathcal{O}}\mathcal{O}^\dagger\mathbf{Y}, \\ 0 \leq \mathbf{Z} &\perp \mathbf{W} \geq 0, \end{aligned} \tag{6.24}$$

being a **linear complementarity problem** (LCP) [80]. This LCP (and with it the MLCP) is guaranteed to have a unique solution if all principal minors of the matrix $\left[\overline{\mathbf{M}} - \overline{\mathcal{O}}\mathcal{O}^\dagger\mathbf{M}\right]$ are strictly positive [80], which can not easily be checked in this general form. However, it has to be checked for a specific system at hand and serves, together with the rank condition for $\mathcal{O}$, as a sufficient observability condition. Once $\mathbf{W}$ is known from the LCP solution, the current state can be calculated with (6.23) and (6.21), i.e.

$$\mathbf{x}_{n-1} = \mathbf{A}^{n-1}\mathcal{O}^\dagger\mathbf{Y} + [\mathbf{Q} - \mathbf{A}^{n-1}\mathcal{O}^\dagger\mathbf{M}]\mathbf{W}, \tag{6.25}$$

where $\mathbf{Q} := \mathbf{Q}_{n-1}$.

Remark 6.2. In order to obtain a state estimate at every time step k one applies (6.24),(6.25) to a moving time window with a length of n time steps, i.e.

$$\mathbf{x}_k = \mathbf{A}^{n-1}\mathcal{O}^\dagger\mathbf{Y}_{k,n-1} + [\mathbf{Q} - \mathbf{A}^{n-1}\mathcal{O}^\dagger\mathbf{M}]\mathbf{W}_{k,n-1}. \tag{6.26}$$

where $\mathbf{Y}_{k,n-1} := \left(\mathbf{y}_{k-n+1}^\mathsf{T} \dots \mathbf{y}_k^\mathsf{T}\right)^\mathsf{T}$ and $\mathbf{W}_{k,n-1} := \left(\mathbf{w}_{k-n+1}^\mathsf{T} \dots \mathbf{w}_k^\mathsf{T}\right)^\mathsf{T}$.

Remark 6.3. It is easy to verify that if inputs are taken into consideration, the moving window deadbeat observer reads

$$\begin{aligned}
\mathbf{x}_k &= \mathbf{A}^{n-1}\mathcal{O}^{\dagger}\mathbf{Y}_{k,n-1} + [\mathbf{Q} - \mathbf{A}^k\mathcal{O}^{\dagger}\mathbf{M}]\mathbf{W}_{k,n-1} + [\mathbf{R} - \mathbf{A}^k\mathcal{O}^{\dagger}\mathbf{N}]\mathbf{V}_{k,n-1},\\
\mathbf{Z}_{k,n-1} &= \left[\overline{\mathbf{M}} - \overline{\mathcal{O}}\mathcal{O}^{\dagger}\mathbf{M}\right]\mathbf{W}_{k,n-1} + \overline{\mathcal{O}}\mathcal{O}^{\dagger}\mathbf{Y}_{k,n-1} + \left[\overline{\mathbf{N}} - \overline{\mathcal{O}}\mathcal{O}^{\dagger}\mathbf{N}\right]\mathbf{V}_{k,n-1},\\
0 &\leq \mathbf{Z}_{k,n-1} \perp \mathbf{W}_{k,n-1} \geq 0,
\end{aligned} \tag{6.27}$$

where $\mathbf{V}_{k,n-1} := \left(\mathbf{v}_{k-n+1}^{\mathsf{T}} \dots \mathbf{v}_k^{\mathsf{T}}\right)^{\mathsf{T}}$ are the collected inputs. Furthermore, the matrix $\mathbf{R} := \begin{bmatrix}\mathbf{A}^{k-1}\mathbf{E} & \mathbf{A}^{k-2}\mathbf{E} & \dots & \mathbf{E} & \mathbf{0}\end{bmatrix}$ has been introduced, and the two remaining matrices are

$$\begin{aligned}
\mathbf{N} &:= \begin{bmatrix}
\mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0}\\
\mathbf{GE} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0}\\
\mathbf{GAE} & \mathbf{GE} & \mathbf{0} & \dots & \mathbf{0}\\
\vdots & & & & \vdots\\
\mathbf{GA}^{k-1}\mathbf{E} & \mathbf{GA}^{k-2}\mathbf{E} & \dots & \mathbf{GE} & \mathbf{0}
\end{bmatrix},\\
\overline{\mathbf{N}} &:= \begin{bmatrix}
\mathbf{F} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0}\\
\mathbf{CE} & \mathbf{F} & \mathbf{0} & \dots & \mathbf{0}\\
\mathbf{CAE} & \mathbf{CE} & \mathbf{F} & \dots & \mathbf{0}\\
\vdots & & & & \vdots\\
\mathbf{CA}^{k-1}\mathbf{E} & \mathbf{CA}^{k-2}\mathbf{E} & \dots & \mathbf{CE} & \mathbf{F}
\end{bmatrix}.
\end{aligned} \tag{6.28}$$

6.4 Passivity-based observers for discrete LCS

The deadbeat approach presented in the last section suffers from one main drawback: the state estimate is highly sensitive on measurement and model errors. One reason is that the observability matrix $\mathcal{O}$ is often ill-conditioned and, therefore, taking the (left) inverse, strongly amplifies measurement noise. Therefore an asymptotic state observer is much more desirable in practical applications. For continuous-time linear complementarity systems, Heemels et al. [51] suggest a Luenberger-type state observer, where the observer gains are determined based on a linear matrix inequality related to system passivity. The equivalent procedure is presented here for discrete linear complementarity systems. In analogy to Section 5.3, passivity for discrete-time systems is defined as follows (see also [26]).

Definition 6.4. A linear time-invariant discrete-time system

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{w}_k, \\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{w}_k, \end{aligned} \tag{6.29}$$

written in short as system $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$, is said to be **passive** if there exists a non-negative function $V : \mathbb{R}^n \to \mathbb{R}_0^+$ (called the storage function) with $V(\mathbf{0}) = 0$ such that

$$V(\mathbf{x}_{k+1}) - V(\mathbf{x}_k) \leq \mathbf{y}_k^\mathsf{T}\mathbf{w}_k \tag{6.30}$$

for all $\mathbf{w}_k$ and $\forall\, k$. Furthermore, it is called **strictly passive** if there exists a positive definite function $S : \mathbb{R}^n \to \mathbb{R}$ such that

$$V(\mathbf{x}_{k+1}) - V(\mathbf{x}_k) \leq \mathbf{y}_k^\mathsf{T}\mathbf{w}_k - S(\mathbf{x}_k) \tag{6.31}$$

for all $\mathbf{w}_k$ and for all k.

As in the continuous-time case, linear matrix inequalities serve as necessary and sufficient conditions for passivity that can be checked numerically. A well-known result, adapted from [24] to discrete-time systems, is the following.

Theorem 6.5. System (6.29) is passive if and only if there exists a matrix $\mathbf{P} = \mathbf{P}^\mathsf{T} \succeq 0$ such that the following matrix inequality holds

$$\begin{pmatrix} \mathbf{A}^\mathsf{T}\mathbf{P}\mathbf{A} - \mathbf{P} & \mathbf{A}^\mathsf{T}\mathbf{P}\mathbf{B} - \mathbf{C}^\mathsf{T} \\ \mathbf{B}^\mathsf{T}\mathbf{P}\mathbf{A} - \mathbf{C} & \mathbf{B}^\mathsf{T}\mathbf{P}\mathbf{B} - (\mathbf{D} + \mathbf{D}^\mathsf{T}) \end{pmatrix} \preceq 0. \tag{6.32}$$

Proof. The proof is a direct adaptation of the continuous-time counterpart found in [24]. □

For the slightly stronger property of strict passivity the following holds.

Proposition 6.6. System (6.29) is **strictly passive** if there exists a matrix $\mathbf{P} = \mathbf{P}^\mathsf{T} > 0$ and a constant $\mu > 0$ such that the matrix inequality

$$\begin{pmatrix} \mathbf{A}^\mathsf{T}\mathbf{P}\mathbf{A} - \mathbf{P} + \mu\mathbf{P} & \mathbf{A}^\mathsf{T}\mathbf{P}\mathbf{B} - \mathbf{C}^\mathsf{T} \\ \mathbf{B}^\mathsf{T}\mathbf{P}\mathbf{A} - \mathbf{C} & \mathbf{B}^\mathsf{T}\mathbf{P}\mathbf{B} - (\mathbf{D} + \mathbf{D}^\mathsf{T}) \end{pmatrix} \preceq 0 \tag{6.33}$$

holds.

Proof. The proof is a slight modification of the sufficiency part in 6.5, and a direct adaptation of the continuous-time counterpart found in [24]. □

Note that proof of the converse statement in Proposition 6.6 has been given in [72] for continuous-time systems.

Now consider a general discrete LCS (not necessarily the discretization of the dynamics of a mechanical system) of the form (6.15). The proposed Luenberger-type state observer for the discrete LCS (6.15) is in analogy to [51]

$$\begin{aligned}\hat{\mathbf{x}}_{k+1} &= \mathbf{A}\hat{\mathbf{x}}_k + \mathbf{B}\hat{\mathbf{w}}_k + \mathbf{E}\mathbf{v}_k + \mathbf{L}_1(\mathbf{y}_k - \hat{\mathbf{y}}_k),\\ \hat{\mathbf{z}}_k &= \mathbf{C}\hat{\mathbf{x}}_k + \mathbf{D}\hat{\mathbf{w}}_k + \mathbf{F}\mathbf{v}_k + \mathbf{L}_2(\mathbf{y}_k - \hat{\mathbf{y}}_k),\\ 0 \leq \hat{\mathbf{z}}_k &\perp \hat{\mathbf{w}}_k \geq 0,\\ \hat{\mathbf{y}}_k &= \mathbf{G}\hat{\mathbf{x}}_k,\end{aligned} \tag{6.34}$$

where all observer related quantities are written with a circumflex ($\hat{}$). The observer consists of a copy of the original system, augmented by two correction terms, both linear in the output difference. Defining the observation errors as $\tilde{\mathbf{x}}_k := \mathbf{x}_k - \hat{\mathbf{x}}_k$, $\tilde{\mathbf{z}}_k := \mathbf{z}_k - \hat{\mathbf{z}}_k$ and $\tilde{\mathbf{w}}_k := \mathbf{w}_k - \hat{\mathbf{w}}_k$, it follows that

$$\begin{aligned}\tilde{\mathbf{x}}_{k+1} &= (\mathbf{A} - \mathbf{L}_1\mathbf{G})\tilde{\mathbf{x}}_k + \mathbf{B}\tilde{\mathbf{w}}_k,\\ \tilde{\mathbf{z}}_k &= (\mathbf{C} - \mathbf{L}_2\mathbf{G})\tilde{\mathbf{x}}_k + \mathbf{D}\tilde{\mathbf{w}}_k,\\ \tilde{\mathbf{z}}_k^\mathsf{T}\tilde{\mathbf{w}}_k &\leq 0.\end{aligned} \tag{6.35}$$

The last relation in (6.35) is easily checked by expanding

$$\tilde{\mathbf{z}}_k^\mathsf{T}\tilde{\mathbf{w}}_k = (\mathbf{z}_k - \hat{\mathbf{z}}_k)^\mathsf{T}(\mathbf{w}_k - \hat{\mathbf{w}}_k) = \mathbf{z}_k^\mathsf{T}\mathbf{w}_k - \mathbf{z}_k^\mathsf{T}\hat{\mathbf{w}}_k - \hat{\mathbf{z}}_k^\mathsf{T}\mathbf{w}_k + \hat{\mathbf{z}}_k^\mathsf{T}\hat{\mathbf{w}}_k. \tag{6.36}$$

Therein, the first and the last term vanish and the two other terms are non-positive due to the inequality complementarities in (6.15) and (6.34). Note that the inequality $\tilde{\mathbf{z}}_k^\mathsf{T}\tilde{\mathbf{w}}_k \leq 0$ represents the monotonicity property of the discrete impact law for mechanical systems. It is, however, not an inequality complementarity. The equations (6.35) do therefore not form a full description of the error dynamics, because $\tilde{\mathbf{w}}_k$ cannot be expressed as a function of the estimation error $\tilde{\mathbf{x}}_k$. Instead, the last three lines of (6.15) and (6.34) have to be used. As a consequence, $\tilde{\mathbf{w}}_k$ depends on $\mathbf{x}_k, \hat{\mathbf{x}}_k$ and $\mathbf{v}_k$, where $\hat{\mathbf{x}}_k$ can be replaced by $\mathbf{x}_k - \tilde{\mathbf{x}}_k$ (or the other way around). As pointed out in [51] for the continuous-time case, the error dynamics is therefore non-autonomous and has two states, $\tilde{\mathbf{x}}_k$ and $\mathbf{x}_k$ (or alternatively $\tilde{\mathbf{x}}_k$ and $\hat{\mathbf{x}}_k$). However, only the estimation error $\tilde{\mathbf{x}}_k$ has to tend to zero as k increases. From the inequality in (6.35), it follows that if system $(\mathbf{A} - \mathbf{L}_1\mathbf{G}, \mathbf{B}, \mathbf{C} - \mathbf{L}_2\mathbf{G}, \mathbf{D})$ is strictly passive, the corresponding storage function serves as a Lyapunov function to show asymptotic stability of the estimation error. Indeed, by selecting a quadratic Lyapunov function candidate $V(\tilde{\mathbf{x}}_k) = \tilde{\mathbf{x}}_k^\mathsf{T}\mathbf{P}\tilde{\mathbf{x}}_k$

with $\mathbf{P} = \mathbf{P}^\top > 0$, it follows

$$\begin{aligned} V(\tilde{\mathbf{x}}_{k+1}) - V(\tilde{\mathbf{x}}_k) &= \tilde{\mathbf{x}}_{k+1}^\top \mathbf{P} \tilde{\mathbf{x}}_{k+1} - \tilde{\mathbf{x}}_k^\top \mathbf{P} \tilde{\mathbf{x}}_k \\ &= (\tilde{\mathbf{x}}_{k+1} + \tilde{\mathbf{x}}_k)^\top \mathbf{P} (\tilde{\mathbf{x}}_{k+1} - \tilde{\mathbf{x}}_k) \\ &= ((\mathbf{A} - \mathbf{L}_1 \mathbf{G})\tilde{\mathbf{x}}_k + \mathbf{B}\tilde{\mathbf{w}}_k + \tilde{\mathbf{x}}_k)^\top \mathbf{P}((\mathbf{A} - \mathbf{L}_1 \mathbf{G})\tilde{\mathbf{x}}_k + \mathbf{B}\tilde{\mathbf{w}}_k - \tilde{\mathbf{x}}_k), \end{aligned} \tag{6.37}$$

which, after first subtracting and then again adding the term $2\tilde{\mathbf{z}}_k^\top \tilde{\mathbf{w}}_k$, can be written as

$$V(\tilde{\mathbf{x}}_{k+1}) - V(\tilde{\mathbf{x}}_k) = \begin{pmatrix} \tilde{\mathbf{x}}_k \\ \tilde{\mathbf{w}}_k \end{pmatrix}^\top \mathbf{H} \begin{pmatrix} \tilde{\mathbf{x}}_k \\ \tilde{\mathbf{w}}_k \end{pmatrix} + 2\tilde{\mathbf{z}}_k^\top \tilde{\mathbf{w}}_k, \tag{6.38}$$

where the matrix $\mathbf{H}$ is given by

$$\mathbf{H} = \begin{pmatrix} (\mathbf{A} - \mathbf{L}_1\mathbf{G})^\top \mathbf{P}(\mathbf{A} - \mathbf{L}_1\mathbf{G}) - \mathbf{P} & (\mathbf{A} - \mathbf{L}_1\mathbf{G})^\top \mathbf{P}\mathbf{B} - (\mathbf{C} - \mathbf{L}_2\mathbf{G})^\top \\ \mathbf{B}^\top \mathbf{P}(\mathbf{A} - \mathbf{L}_1\mathbf{G}) - (\mathbf{C} - \mathbf{L}_2\mathbf{G}) & \mathbf{B}^\top \mathbf{P}\mathbf{B} - (\mathbf{D} + \mathbf{D}^\top) \end{pmatrix}. \tag{6.39}$$

Because $\tilde{\mathbf{z}}_k^\top \tilde{\mathbf{w}}_k \leq 0$, it follows that if the system $(\mathbf{A} - \mathbf{L}_1\mathbf{G}, \mathbf{B}, \mathbf{C} - \mathbf{L}_2\mathbf{G}, \mathbf{D})$ is passive, it holds that $\mathbf{H} \preceq 0$ and therefore $V(\tilde{\mathbf{x}}_{k+1}) - V(\tilde{\mathbf{x}}_k) \leq 0$. More strongly, it holds that $V(\tilde{\mathbf{x}}_{k+1}) - V(\tilde{\mathbf{x}}_k) \leq -\mu V(\tilde{\mathbf{x}}_k)$ if the system $(\mathbf{A} - \mathbf{L}_1\mathbf{G}, \mathbf{B}, \mathbf{C} - \mathbf{L}_2\mathbf{G}, \mathbf{D})$ is strictly passive. In that case the non-autonomous estimation error dynamics is exponentially stable. The matrix inequality

$$\begin{pmatrix} (\mathbf{A} - \mathbf{L}_1\mathbf{G})^\top \mathbf{P}(\mathbf{A} - \mathbf{L}_1\mathbf{G}) - \mathbf{P} + \mu\mathbf{P} & (\mathbf{A} - \mathbf{L}_1\mathbf{G})^\top \mathbf{P}\mathbf{B} - (\mathbf{C} - \mathbf{L}_2\mathbf{G})^\top \\ \mathbf{B}^\top \mathbf{P}(\mathbf{A} - \mathbf{L}_1\mathbf{G}) - (\mathbf{C} - \mathbf{L}_2\mathbf{G}) & \mathbf{B}^\top \mathbf{P}\mathbf{B} - (\mathbf{D} + \mathbf{D}^\top) \end{pmatrix} \preceq 0, \tag{6.40}$$

that that needs to be fulfilled in order to ensure strict passivity, is nonlinear in the unknowns $\mathbf{L}_1, \mathbf{L}_2$ and $\mathbf{P}$. However, by introducing $\mathbf{S} := \mathbf{P}\mathbf{L}_1$ and applying the Schur complement lemma (see Proposition B.2), it can be checked that (6.40) is equivalent to the linear matrix inequality (LMI)

$$\begin{pmatrix} -\mathbf{P} + \mu\mathbf{P} & -(\mathbf{C} - \mathbf{L}_2\mathbf{G})^\top & \mathbf{A}^\top\mathbf{P} - \mathbf{G}^\top\mathbf{S}^\top \\ -(\mathbf{C} - \mathbf{L}_2\mathbf{G}) & -(\mathbf{D} + \mathbf{D}^\top) & \mathbf{B}^\top\mathbf{P} \\ \mathbf{P}\mathbf{A} - \mathbf{S}\mathbf{G} & \mathbf{P}\mathbf{B} & -\mathbf{P} \end{pmatrix} \preceq 0. \tag{6.41}$$

Since $\mathbf{P}$ is invertible, $\mathbf{L}_1$ can be recovered in a second step as $\mathbf{L}_1 = \mathbf{P}^{-1}\mathbf{S}$. The transformation from the nonlinear matrix inequality (6.40) to the LMI (6.41) is very useful, since efficient numerical LMI solvers are available. Note, however, that imposing additional constraints on the solutions of (6.41) could lead again to a nonlinear matrix inequality.

An extended version for position measurements

For a mechanical system, $\mathbf{z}_k$ represents the kinematic variable $\boldsymbol{\zeta}_k = \boldsymbol{g}_{k+1} + \varepsilon \boldsymbol{g}_{k-1}$ in (6.5). Therefore $\mathbf{z}_k$, and with it $\tilde{\mathbf{z}}_k$, depend on one past value $\boldsymbol{g}_{k-1}$ of the contact distance and one future value $\boldsymbol{g}_{k+1}$. It therefore makes sense to extended the observer (6.34) to include past and future measurements at t_{k-1} and t_{k+1}, respectively. This can be achieved by adding more correction terms, summarized in

$$\begin{aligned} \hat{\mathbf{x}}_{k+1} &= \mathbf{A}\hat{\mathbf{x}}_k + \mathbf{B}\hat{\mathbf{w}}_k + \mathbf{E}\mathbf{v}_k + \bar{\mathbf{L}}_1(\bar{\mathbf{y}}_k - \hat{\bar{\mathbf{y}}}_k^{(-)}), \\ \hat{\mathbf{z}}_k &= \mathbf{C}\hat{\mathbf{x}}_k + \mathbf{D}\hat{\mathbf{w}}_k + \mathbf{F}\mathbf{v}_k + \bar{\mathbf{L}}_2(\bar{\mathbf{y}}_k - \hat{\bar{\mathbf{y}}}_k^{(-)}), \\ \hat{\bar{\mathbf{y}}}_k &= \bar{\mathbf{G}}\hat{\mathbf{x}}_k, \end{aligned} \tag{6.42}$$

where $\bar{\mathbf{y}}_k$ and $\hat{\bar{\mathbf{y}}}_k$ are extended outputs and $\bar{\mathbf{G}}$ is the corresponding extended output matrix according to

$$\bar{\mathbf{y}}_k = \begin{pmatrix} \mathbf{y}_k \\ \mathbf{y}_{k+1} \\ \mathbf{y}_{k-1} \end{pmatrix}, \quad \hat{\bar{\mathbf{y}}}_k^{(-)} = \begin{pmatrix} \hat{\mathbf{y}}_k \\ \hat{\mathbf{y}}_{k+1}^{(-)} \\ \hat{\mathbf{y}}_{k-1}^{(-)} \end{pmatrix}, \quad \bar{\mathbf{G}} = \begin{pmatrix} \mathbf{G} \\ \mathbf{G}\mathbf{A} \\ \mathbf{G}\tilde{\mathbf{A}} \end{pmatrix}. \tag{6.43}$$

Therein, $\hat{\mathbf{y}}_{k+1}^{(-)} := \mathbf{G}(\mathbf{A}\hat{\mathbf{x}}_k + \mathbf{B}\hat{\mathbf{w}}_k + \mathbf{E}\mathbf{v}_k)$ is the predicted future output without considering the correction terms. Similarly, $\mathbf{y}_{k-1}^{(-)}$ refers to the past output, obtained by back-propagation without considering correction terms. As it was done in (6.11), positions can be back-propagated with the kinematic equation $\mathbf{q}_{k-1} = \mathbf{q}_k - \Delta t \mathbf{u}_k$. Here, we restrict ourselves to the case where the output $\mathbf{y}_k$ depends only on positions $\mathbf{q}_k$. The reason for this is that obtaining past velocities $\mathbf{u}_{k-1}$ by back-propagation would involve past values $\mathbf{w}_{k-1}$ and $\mathbf{v}_{k-1}$, such that the estimation error dynamics would become structurally different from (6.35). Furthermore, in cases where $\mathbf{A}$ is not invertible, the back-propagation might not have a (unique) solution. Past outputs are written as $\mathbf{y}_{k-1} = \mathbf{G}\mathbf{x}_{k-1} = \mathbf{G}\tilde{\mathbf{A}}\mathbf{x}_k$ with the back-propagation matrix $\tilde{\mathbf{A}} := [(\mathbf{I} \quad -\Delta t\mathbf{I})^\mathsf{T} \quad (\mathbf{0} \quad \mathbf{0})^\mathsf{T}]^\mathsf{T}$, where $\mathbf{I}$ is a unit matrix of appropriate dimensions. With $\tilde{\mathbf{x}}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$ the error dynamics can be written as

$$\begin{aligned} \tilde{\mathbf{x}}_{k+1} &= (\mathbf{A} - \bar{\mathbf{L}}_1\bar{\mathbf{G}})\tilde{\mathbf{x}}_k + \bar{\mathbf{B}}\tilde{\mathbf{w}}_k, \\ \tilde{\mathbf{z}}_k &= (\mathbf{C} - \bar{\mathbf{L}}_2\bar{\mathbf{G}})\tilde{\mathbf{x}}_k + \bar{\mathbf{D}}\tilde{\mathbf{w}}_k, \\ \tilde{\mathbf{z}}_k^\mathsf{T}\tilde{\mathbf{w}}_k &\le 0. \end{aligned} \tag{6.44}$$

Therein, with $\bar{\mathbf{L}}_1 = (\mathbf{L}_1 \ \mathbf{L}_3 \ \mathbf{L}_5)$ and $\bar{\mathbf{L}}_2 = (\mathbf{L}_2 \ \mathbf{L}_4 \ \mathbf{L}_6)$, the matrices are defined as $\bar{\mathbf{B}} = \mathbf{B} - \mathbf{L}_3\mathbf{G}\mathbf{B}$ and $\bar{\mathbf{D}} = \mathbf{D} - \mathbf{L}_4\mathbf{G}\mathbf{B}$. Clearly (6.44) is of the same form as (6.35) but with more design variables. Therefore, the same steps as in (6.37) to (6.41) lead to an LMI.

Remark 6.7. The observer dynamics (6.42) can alternatively be written as

$$\begin{aligned}\hat{\mathbf{x}}_{k+1} &= (\mathbf{A} - \mathbf{L}_1\mathbf{G} - \mathbf{L}_3\mathbf{G}\mathbf{A} - \mathbf{L}_5\mathbf{G}\tilde{\mathbf{A}})\hat{\mathbf{x}}_k + (\mathbf{B} - \mathbf{L}_3\mathbf{G}\mathbf{B})\hat{\mathbf{w}}_k + (\mathbf{E} - \mathbf{L}_3\mathbf{G}\mathbf{E})\mathbf{v}_k \\ &\quad + \mathbf{L}_1\mathbf{y}_k + \mathbf{L}_3\mathbf{y}_{k+1} + \mathbf{L}_5\mathbf{y}_{k-1}, \\ \hat{\mathbf{z}}_k &= (\mathbf{C} - \mathbf{L}_2\mathbf{G} - \mathbf{L}_4\mathbf{G}\mathbf{A} - \mathbf{L}_6\mathbf{G}\tilde{\mathbf{A}})\hat{\mathbf{x}}_k + (\mathbf{D} - \mathbf{L}_4\mathbf{G}\mathbf{B})\hat{\mathbf{w}}_k + (\mathbf{F} - \mathbf{L}_4\mathbf{G}\mathbf{E})\mathbf{v}_k \\ &\quad + \mathbf{L}_2\mathbf{y}_k + \mathbf{L}_4\mathbf{y}_{k+1} + \mathbf{L}_6\mathbf{y}_{k-1}, \\ 0 &\le \hat{\mathbf{z}}_k \perp \hat{\mathbf{w}}_k \ge 0.\end{aligned} \tag{6.45}$$

To ensure the existence of a unique solution, the observer gain $\mathbf{L}_4$ has therefore to be designed such that $\mathbf{D} - \mathbf{L}_4\mathbf{G}\mathbf{B}$ is a $\mathcal{P}$-matrix.

Remark 6.8. If $\mathbf{L}_3$ can be designed such that $\mathbf{B} - \mathbf{L}_3\mathbf{G}\mathbf{B} = \mathbf{0}$, the state error in (6.44) becomes independent of $\tilde{\mathbf{w}}_k$. We are then left with designing $\mathbf{L}_1$ and $\mathbf{L}_5$ such that the error dynamics is asymptotically stable. This corresponds to an *unknown input observer* as it is known for linear systems. However, as in the numerical example below, such a gain $\mathbf{L}_3$ often does not exist.

6.5 Numerical example

As an example system, consider the two-mass oscillator depicted in Figure 6.1. The oscillator with masses $m_1 = m_2 = m$, spring constants k and damping ratios d is excited by an external force $F(t)$ applied to the first mass. The movement of the second mass is restricted by a motion limiting stop. The positions of the two masses, relative to the equilibrium positions for $F = 0$, are described by the coordinates q_1 and q_2. The system dynamics is described by (6.1) with the constant system matrices

$$\mathbf{M} = \begin{pmatrix} m & 0 \\ 0 & m \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} 2d & -d \\ -d & d \end{pmatrix}, \quad \mathbf{K} = \begin{pmatrix} 2k & -k \\ -k & k \end{pmatrix}, \tag{6.46}$$

as well as

$$\mathbf{W} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} F(t). \tag{6.47}$$

The parameters are given by $m = 1$ kg, $k = 1500$ N/m, $d = 0.5$ Ns/m and the coefficient of restitution is $\varepsilon = 0.8$. Furthermore, a periodic excitation $F(t) = a\sin(\omega t)$ is used, with an amplitude of $a = 10$ N and a frequency $\omega = 5.25 \cdot 2\pi$ rad/s.

To reduce large differences in the order of magnitude of numerical values during the solution, $\mathbf{u}_k$ and $\mathbf{w}_k$ are scaled by the time step length Δt prior to solving the LMI. More precisely, in (6.15), the velocity $\mathbf{u}_k$ is replaced by $\Delta\mathbf{q}_k := \mathbf{q}_k - \mathbf{q}_{k-1}$ and

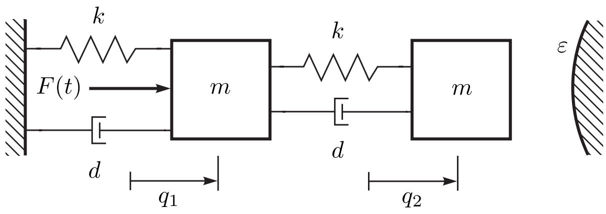

Figure 6.1: An example system with one unilateral constraint.

$\mathbf{w}_k$ by $\Delta t\overline{\mathbf{w}}_k$. As a result, scaled system matrix entries are obtained (which are not given here). The LMI (6.41) and its extended counterpart are then solved with the scaled system matrices using the Matlab integrated LMI solver *feasp*, which is based on a projective method [81].

In the following, three cases of different difficulty will be inspected:

- *Case I* with output $\mathbf{y}_k = \mathbf{q}_k$, i.e. all positions are measured with $\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$
- *Case II* with output $\mathbf{y}_k = q_{1k}$, i.e. only the position of the non-colliding mass is measured with $\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}$
- *Case III* with $\mathbf{y}_k = q_{2k} - q_{1k}$, i.e. the difference between the two mass positions are measured with $\mathbf{G} = \begin{pmatrix} -1 & 1 & 0 & 0 \end{pmatrix}$

Even though in all three cases positions are measured, these cases are different in their difficulty because they do not fulfill the same observability conditions. *Case I* is the easiest, since we directly know from measurements whether or not the contact is open or closed. In that case the observer designs discussed in Chapter 5 are applicable. The observability matrix $\mathcal{O}$ has full column rank in all three cases, indicating that the non-impulsive dynamics is observable in all cases. For the (discretized) impulsive dynamics, we introduced a sufficient observability condition in Section 6.3: if the matrix $\left[\overline{\mathbf{M}} - \overline{\mathcal{O}}\mathcal{O}^{\dagger}\mathbf{M}\right]$ is a $\mathcal{P}$-matrix, then the LCP (6.24) has a unique solution. This observability condition is fulfilled in *case I* and *case II*, but not in *case III*. Therefore, for the cases *I* and *II*, observability is confirmed and the deadbeat observer (6.24), (6.25) is applicable, while for *case III* observability remains undetermined and the deadbeat observer is not applicable. Regarding the passivity-based observers from Section 6.4, for *case I* both the LMI (6.41) and its extended counterpart admit a solution. For *case II*, only the LMI of the extended observer version admits a solution, which shows the necessity to include past and future measurements in the observer in the presence of unilateral constraints. Finally, for *case III*, both LMIs do not admit a solution. Therefore, the passivity-based observer is not applicable in *case III* which we will exclude from further discussion. In the following the discussion will concentrate on *case II*, since it was observed in the numerical analysis that *case I* shows a similar behavior, but *case II* is more difficult and therefore more interesting. The first thing observed from the numerical LMI solution of *case II* is that the entries of the observer gains $\mathbf{L}_i$ $(i = 1, \ldots, 6)$ increase as the step size Δt decreases. Figure 6.3 shows the maximum absolute value of all observer gain entries as a function of Δt^2. The observer gains are roughly inversely proportional to the squared time step size. As a result of the high observer gains for small time steps, the initial observer error is often strongly amplified at the beginning. In Figure 6.4 the trajectories of the impacting mass are plotted for both, the discretization of the observed system and the corresponding extended passivity-based observer, with $\Delta t = 10^{-4}$ s and selected initial conditions. The velocity estimation strongly deviates from the true trajectory for a short period of time. This phenomenon is also known as 'peaking' in the literature on high-gain observers [61], but is not to be confused with the peaking phenomenon of impulsive systems, which refers to the fact that the Euclidean error between trajectories can jump to high values, even if the trajectories are arbitrarily close. The Lyapunov function (and with it the state estimation error), however, decreases quickly, as shown in Figure 6.2. A real-time implementation of the state observer is challenging for very small time steps. In the numerical example, the average computation time per time step on a standard desktop PC is $2.5 \cdot 10^{-5}$ s. Here, $\Delta t = 10^{-4}$ s was chosen for presentation. For other time steps Δt, the qualitative behavior of the extended passivity-based observer is similar, with higher deflections in the transient phase for lower Δt.

Large entries in the observer gains $\mathbf{L}_i$ $(i = 1, \ldots, 6)$ have a negative influence on the observer's robustness against measurement noise, since in the observer dynamics (6.45), they are multiplied by the measurements. Because the observer gain entries are roughly inversely proportional to the squared step size Δt^2, one might therefore be tempted to select a large step size. However, a large time step would cause a pronounced deviation between the continuous-time system and the discrete-time model. Furthermore, the ability of the time discretization to describe collisions which are separated by a short period of time also depends on the chosen time step. Compared to the observer design for linear systems, where the designer is usually facing a performance-robustness tradeoff when selecting the observer gains, the situation is more complicated here. For each system at hand, the observer's

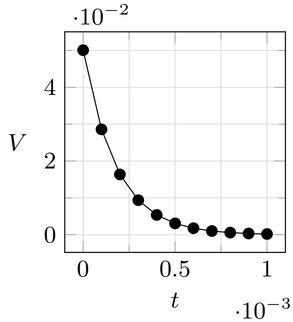

Figure 6.2: Lyapunov function $V(\tilde{\mathbf{x}}_k) = \tilde{\mathbf{x}}_k^\top \mathbf{P}\tilde{\mathbf{x}}_k$ over time t (in s) for the example in Figure 6.4.

sensitivity to variations in the step size has to be analyzed and it has to be decided, if a certain step size leads to admissible noise levels.
Interestingly, in our example, measurement noise has a marginal influence on the observer's impact law, compared to the strong influence on the non-impulsive motion due to the high observer gains. In Figure 6.5, the first two plots show the true and estimated velocities u_2 and $\hat{u}_2$ for the extended observer with additive, normally distributed noise on the measurements. More precisely, we use $y_k = \mathbf{G}\mathbf{x}_k + d_k$ with $d_k \sim \mathcal{N}(0, \sigma^2)$, i.e. d_k is a discrete random variable drawn from a normal distribution with zero mean and standard deviation σ. It is observed, that even small standard deviations lead to strong noise levels on the velocity estimation $\hat{u}_2$. The step hight during the impulsive motion however is only marginally altered by the measurement noise, as is shown in the zoom plots on the right hand side of Figure 6.5. Conversely, a time step Δt which is long compared to the impact duration, mainly affects the step height of the observer during the impulsive motion, whereas the non-impulsive motion is only marginally altered. This is shown in the last plot of Figure 6.5. Therein, the true trajectory is generated by simulating an LCS with a time step Δt_{sim} which is an order of magnitude smaller than the time step Δt used for the observer design. From the zoom on the right hand side, it can be observed that due to the difference in the time steps, the observer's step height during the impulsive motion is smaller than the actual step height.

It is worth mentioning that the chosen time step does not have an immediate effect on the impact law of the discrete-time model. Furthermore, without measurement noise and modeling errors, the impact law of the passivity-based state observer is

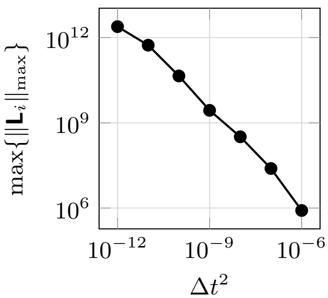

Figure 6.3: *Case II*: Logarithmic plot of the maximum observer gain entry $\max\{\|\mathbf{L}_i\|_{\max}\}$ $(i = 1, \ldots, 6)$ as a function of the squared step size Δt^2 (in s^2). Here $\|\mathbf{L}_i\|_{\max} := \max_{i,j} |\ell_{ij}|$ if ℓ_{ij} are the scalar entries of $\mathbf{L}_i$.

identical to the impact law of the discretized observed system after the estimation error converged to zero. However, before the estimation error converged to zero (i.e. if the output of the observer is not identical to the measurements), the observer's impact law is not physical and the correction terms in the observer dynamics have an influence on how the contact velocity changes during phases of contact. In other words, the chosen time step does have an effect on the observer's impact law, but since this impact law is not physical, this does not impose any constraints on the time step.

6.6 Conclusion

This chapter is an attempt to give the research on developing an observer design for unilaterally constrained mechanical systems *without* using contact measurement a new impulse. As the results show again, this remains a difficult task. In the following the merits are summarized and identified.

The first important step has been to consider a discrete approximation of the original continuous-time system. It has been shown that the deliberate choice of the Paoli-Schatzman scheme leads to a discrete linear complementarity system (whereas other schemes do not). This can be seen as a result, which may also be useful outside the scope of observer design.

The formulation as discrete LCS opens the way for the derivation of a deadbeat observer for this type of systems. It has been shown that a deadbeat observer for a discrete LCS leads to a mixed LCP. In addition, a sufficient observability condition

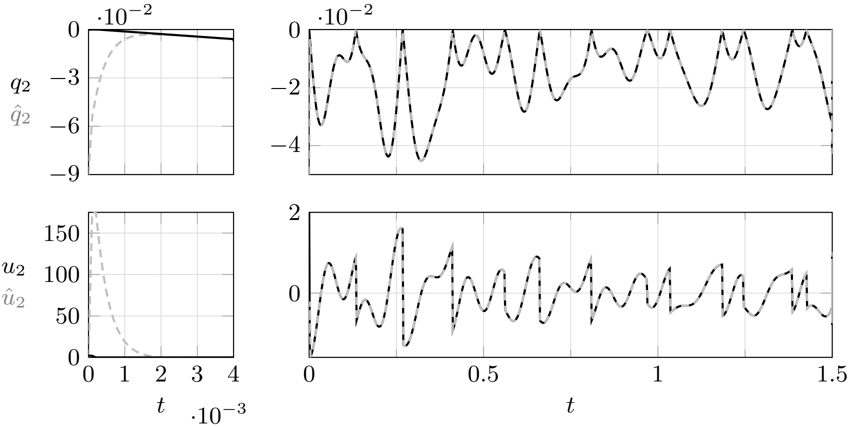

Figure 6.4: Example *case II* with $\Delta t = 10^{-4}$ s and the initial conditions $\mathbf{x}_0 = (0, -10^{-4}, 1, 2)$ and $\hat{\mathbf{x}}_0 = (0, -10^{-1}, 10^{-4}, 10^{-4})$ (in m and m/s resp.). Trajectories of the impacting mass (solid black: observed system (discretized), dashed gray: extended observer) over 10 periods of excitation (t in s, q_2 in m, u_2 in m/s). On the left, the transient phase is shown with a different zoom: the extended observer strongly deviates from the true trajectory in the beginning, due to the large observer gains.

has been obtained, by requiring that the mixed LCP has a unique solution. The deadbeat observer, however, is not robust with respect to measurement noise, but may serve as starting point for future research.

Furthermore, the formulation as discrete LCS allows to use existing observer design techniques for LCS as developed by [51]. Hereto, the existing results for continuous-time LCS have been transported to discrete LCS. The observer consists of a copy of the observed system, augmented by Luenberger-type correction terms. Here, we consider the observed system and the observer both to be discrete-time systems with matching time steps. It turns out, that the observer gains which follow from a LMI are inversely proportional to the squared time step Δt^2.

Because a small Δt leads to high observer gains, lowering the time step increases the sensitivity with respect to measurement noise. Conversely, as we are using a discrete version of the continuous time problem, increasing the time step yields a larger modeling error, since the discretization is an approximation. Compared to the performance-robustness tradeoff encountered in many other observer designs,

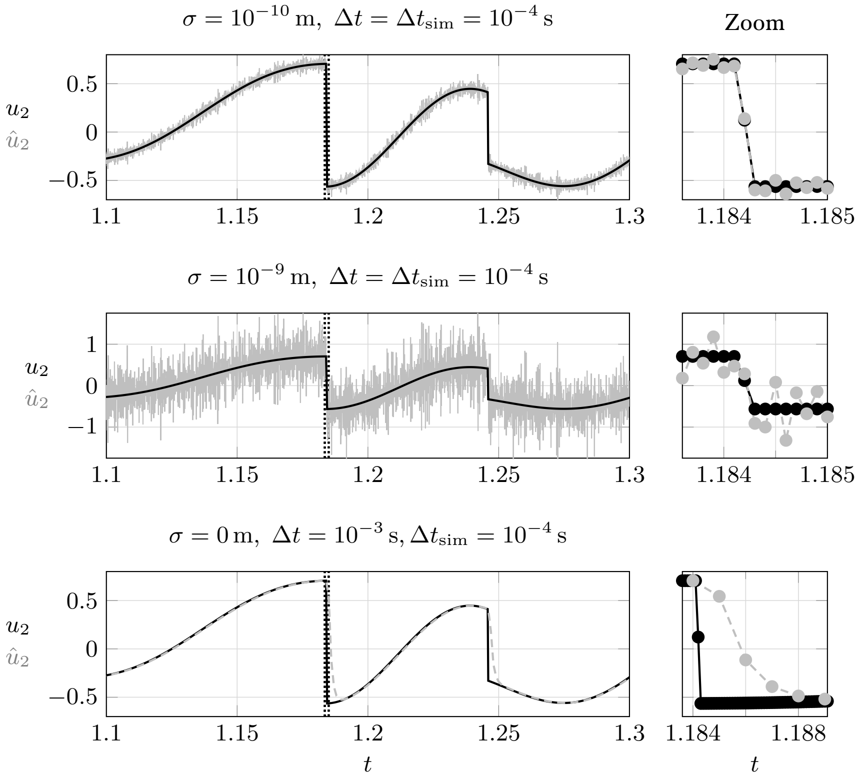

Figure 6.5: Influence of measurement noise and a time step error on the velocity estimation $\hat{u}_2$ of the impacting mass for *case II* (initial conditions identical to Figure 6.4). Solid black: observed system (discretized), gray: extended observer (u_2 in m/s, t in s). The first plot shows the influence of a normally distributed measurement noise with zero mean and a standard deviation $\sigma = 10^{-10}$ m. In the second plot, the standard deviation is 10^{-9} m. In the third plot, the true trajectory is generated by simulating the LCS with Δt_{sim} which is smaller than Δt of the observer.

the selection of the time step here is more complicated. For every system at hand it has to be checked if an admissible choice of the step size can lead to an acceptable performance.

Although the presented observer design is not inherently robust, we come to a fundamental insight: The discretization using the Paoli-Schatzman scheme can be viewed as a regularization which reveals that the peaking phenomenon of impulsive mechanical systems is in fact a singularity with respect to the time step Δt.

Lastly, one can conclude that the body of methods that has been presented here links, somewhat unexpectedly, different research topics: measure differential inclusions, mixed LCPs and linear complementarity systems. The Paoli-Schatzman scheme plays a crucial role in establishing these links.

7

Experimental observer performance analysis

In this chapter, several state observers are implemented for an experimental setup. The setup consists of a two degrees of freedom impact oscillator with a single unilateral constraint. After discussing the setup in detail, its mechanical model is presented and the model parameters are identified in from experimental data. Based on this model, observer gains are designed for several state observers that have been discussed in Chapter 5 and the estimation performances are analyzed in experiments. Finally, the strengths and weaknesses are discussed and the performances are compared (as far as possible). Earlier experimental results of the author can also be found in [88].

7.1 Experimental setup

The experimental setup, for which the observers will be designed, consists of an impact oscillator, which is a mass-spring-damper system with a single unilateral constraint. More precisely, two steel blocks are mounted on a carriage of a linear roller guide (Smalltec SMLFS-20), as shown in Figure 7.1. Each steel block is additionally attached to four linear coil tension springs that connect the blocks to both ends of the linear guide as well as to each other through eyelets. All springs are under pretension to prevent buckling during operation. A unilateral constraint is implemented as a limiting stop in form of a massive aluminium block that can be freely positioned

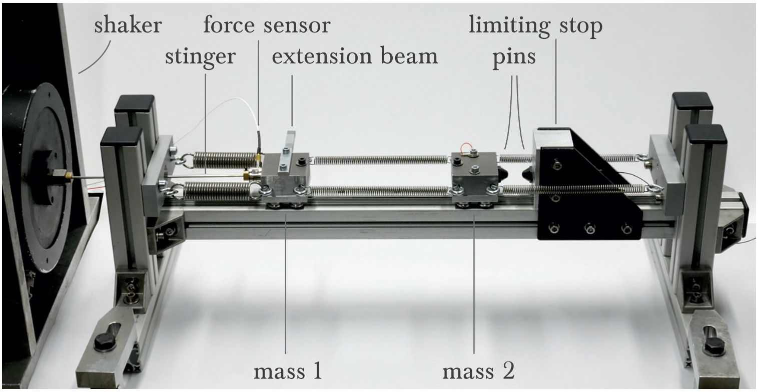

Figure 7.1: The experimental setup. Two masses on a linear roller guide are connected to pre-stressed coil springs. The motion of mass 2 (on the right) is restricted by a limiting stop. Mass 1 (on the left) is excited by an electrodynamic shaker.

at any location on the linear guide. The linear guide and all surrounding parts are mounted on an adjustable frame consisting of aluminium profiles (40×40 mm).

In order to reduce plastic deformation due to the impulsive impact forces, two tempered steel support pins are installed in the contact points of both mass 2 and the aluminium block. As shown in Figure 7.2, one of the support pins is electrically isolated from the rest of the setup. This is achieved by a thin isolating plastic foil between the support pin and the aluminium block as well as a synthetic washer under the screw head of the connected screw. Both the electrically isolated pin and the colliding block are connected to a generic adjustable laboratory DC voltage source, generating a voltage difference of $V_c = 5$V between the colliding parts. The contacting pins are therefore part of an open electric circuit and are acting as an electric switch, such that the circuit closes when contact occurs. In that case, a voltage drop between the two pins can be measured. This voltage measurement signal is then used to generate a switching function $\chi(t)$ indicating whether the contact is open or closed. For the performance assessment of the state observers, the velocities and positions of both masses are measured using two laser Doppler vibrometers (the specific sensing and actuation instrumentation is listed in Table 7.5). While one laser is pointing directly to mass 2 through bores in the surrounding parts, an extension beam is used for mass 1 because of spacial restrictions. Finally, an electrodynamic

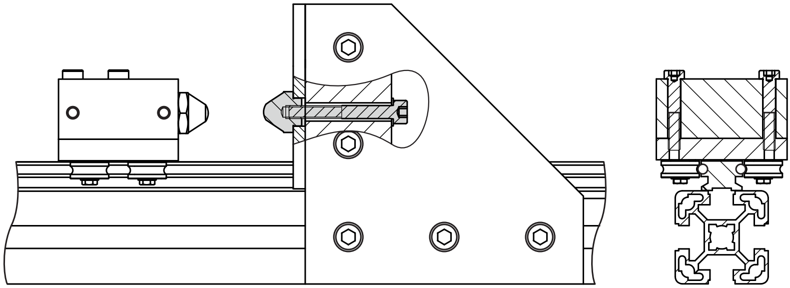

Figure 7.2: *Left*: The rigid motion stop with changeable position on the linear guide. The support pin and the connecting screw (highlighted in gray) are electrically isolated from all other parts through an isolating foil and an isolating washer. *Right*: Sectional drawing of a carriage on the linear roller guide, with mounted steel block.

shaker (permanent magnetic vibration exciter) induces an excitation force on mass 1 (on the left). The shaker itself is excited by a time dependent voltage input $V(t)$, which is generated by a function generator and amplified by a power amplifier. The excitation force, which is also a required input for the state observers, is measured with a force sensor mounted on mass 1, being connected to the shaker over a stinger. Finally, all measurement signals are recorded through USB connected data acquisition hardware and a matching data recording software, as depicted in Figure 7.3.

Shaker	Brüel & Kjaer, type 4808
Function generator	Agilent 33210A
Power amplifier	Brüel & Kjaer, type 2712
Laser Doppler vibrometer	Polytec OFV-505 and OFV-353 (sensor heads)
	OFV-5000 and OFV-3001 (controllers)
Force sensor	Endevco, model 2311
Data acquisition	Dewetron DEWE-50-USB2-8
	with DAQP-LV and DAQP-ACC-A modules

Table 7.1: Instrumentation for actuation and sensing

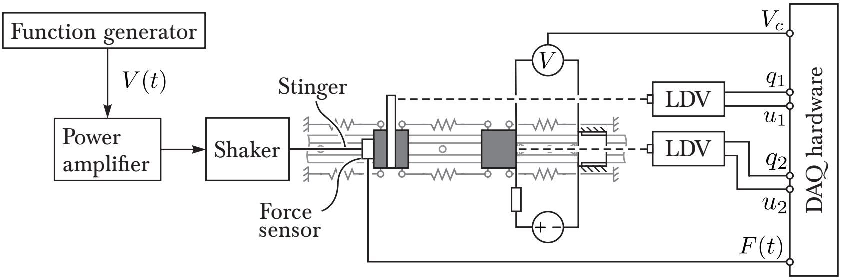

Figure 7.3: Schematic measurement setup. Two laser Doppler vibrometers (LDV) measure the positions and velocities of the two masses. Contact is detected by measuring the voltage between mass 2 and the stop. Furthermore, a piezoelectric force sensor measures the excitation force which is applied by an electrodynamic shaker.

7.2 Mechanical model and parameters

The experimental setup is modeled as a two degrees of freedom oscillator with linear springs and damping elements, as depicted in Figure 7.4. The model dynamics is of the form (5.3) with the constant system matrices

$$\mathbf{M} = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}, \mathbf{D} = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}, \mathbf{K} = \begin{pmatrix} k_1 + k_3 & -k_3 \\ -k_3 & k_2 + k_3 \end{pmatrix}, \tag{7.1}$$

and the generalized force directions and the external forcing according to

$$\mathbf{W}_N = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \mathbf{W}_T = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \mathbf{f}(t) = \begin{pmatrix} F(t) \\ 0 \end{pmatrix}. \tag{7.2}$$

The generalized Newtonian impact law is assumed, with a constant coefficient of restitution (COR) $\varepsilon < 1$, which yields a maximal monotone impact law. Note that no damper between the two masses is included in this model, as a preliminary parameter identification suggested a vanishing damping coefficient.

The model parameters for the experimental setup have been identified in two stages. In a first step, dry friction was neglected and the the spring stiffnesses and damping constants were estimated in the sense of a linear experimental modal analysis. More precisely, the electrodynamic shaker was excited by a harmonic voltage input with a constant amplitude and a frequency sweep between 0.5 Hz and 150 Hz. The

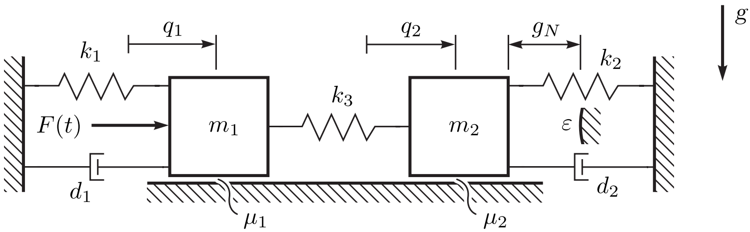

Figure 7.4: Mechanical model of the experimental setup.

parameters were then obtained by a least squares fit of the position and velocity trajectories in the time domain. The resulting parameters are $m_1 = 0.760\,\text{kg}$, $m_2 = 0.696\,\text{kg}$, $k_1 = 5195\,\text{N/m}$, $k_2 = 1004\,\text{N/m}$, $k_3 = 953\,\text{N/m}$, $d_1 = 1.00\,\text{Ns/m}$ and $d_2 = 1.15\,\text{Ns/m}$. In a second step, the friction coefficients were tuned in order to better match experiments with an excitation of constant frequency but for various amplitudes, resulting in $\mu_1 = 0.04$ and $\mu_2 = 0.035$. The COR has been identified in an experiment containing 316 impacts, with pre-impact relative velocities in the range $\gamma_N^- \in [-0.5, 0]$ m/s. Since it is known at all times from measurements whether the contact is open or closed, a COR can be calculated for each impact, by dividing the post-impact velocity by the corresponding negative pre-impact velocity. The resulting CORs are shown in Figure 7.5, where each cross represents one impact. It is observed, that ε is not constant, but decreases as the absolute value $|\gamma_N^-|$ of the pre-impact velocity increases. Above 0.2 m/s the relation is fairly linear with a minor slope, but ε strongly increases for low pre-impact relative velocities. As discussed in Chapter 5, it is essential for the observer design, that the impact law is maximal monotone, which requires the COR to be constant. In the experiments that will be used to evaluate the observers, the pre-impact velocities will predominantly be located in the range $\gamma_{N,i}^- \in [-0.5, -0.2]$ m/s. Therefore, the COR is approximated as a constant $\varepsilon = 0.056$ (which corresponds to the mean value identified for $\gamma_{N,i}^- \in [-0.5, -0.2]$ m/s).

7.3 State observer design

In the following, the state observers presented in Chapter 5 are designed and implemented for the experimental setup. The state observer gains need to fulfill all assumptions that were made in the derivation of the respective observer. Furthermore,

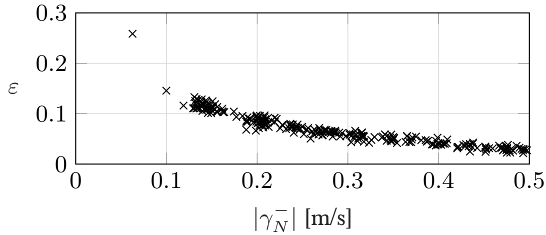

Figure 7.5: Measured COR ε as a function of the pre-impact relative velocity γ_N^-.

if a range of observer gains is admissible, the gains have to be selected such that the sensitivity with respect to measurement noise and modeling errors is low enough for the practical application.

Observer with switched unilateral constraints (Baumann & Leine)

The state observer (5.8), (5.9) will be used including the position jumps (5.19) in Remark 5.2, and will henceforth be referred to as 'Baumann & Leine' observer. The design does not include any gains that have to be tuned. The state estimation only relies on an accurate model (including the location of the unilateral constraint) and a precise measurement of the impact time instants.

Velocity observer (Tanwani et al.)

The observer (5.22), (5.23) will subsequently be called 'Tanwani (reduced)'. Here, a scalar observer gain will be used, i.e. $\mathbf{L} = L\mathbf{I}$ with $L \in \mathbb{R}$ and $\mathbf{I}$ an identity matrix of appropriate dimensions. Then, using the formulation (5.24) and the fact that the model of the experimental system is linear, the estimation error dynamics can be written as

$$\mathbf{M}\mathrm{d}\tilde{\mathbf{u}} + (\mathbf{D} + \mathbf{M}L)\mathrm{d}t = \mathbf{W}(\mathrm{d}\mathbf{P} - \mathrm{d}\hat{\mathbf{P}})\,, \tag{7.3}$$

with $\mathbf{W} = (\mathbf{W}_N \quad \mathbf{W}_T)$, $\mathrm{d}\mathbf{P} = (\mathrm{d}\mathbf{P}_N^\mathsf{T} \quad \mathrm{d}\mathbf{P}_T^\mathsf{T})^\mathsf{T}$ and $\mathrm{d}\hat{\mathbf{P}} = (\mathrm{d}\hat{\mathbf{P}}_N^\mathsf{T} \quad \mathrm{d}\hat{\mathbf{P}}_T^\mathsf{T})^\mathsf{T}$. Finally, using $V = \tilde{\mathbf{u}}^\mathsf{T}\mathbf{M}\tilde{\mathbf{u}}$ as a Lyapunov function[1]reveals that the error dynamics is attractively stable if L satisfies the LMI

$$\mathbf{D} + \mathbf{M}L = \begin{pmatrix} d_1 + m_1 L & 0 \\ 0 & d_2 + m_2 L \end{pmatrix} \succ 0. \tag{7.4}$$

[1]Alternatively, applying Proposition 5.8 with $\mathbf{P} = \mathbf{M}$ and $\mathbf{C} = (\mathbf{0} \quad \mathbf{W}^\mathsf{T})$ leads to the same result.

Therefore the gain has to be chosen such that $L > \max\{-d_1/m_1,\ -d_2/m_2\}$. For the experiment, a value of $L = 150$ has been selected (after some manual tuning).

Full state observer (Tanwani et al.)

In analogy to the velocity observer, the full state observer (5.26), (5.27), (5.23) will be referred to as 'Tanwani (full)'. Again, its state estimate converges globally to the true state if the observer gains fulfill an LMI. Considering the quadratic Lyapunov function $V = \tilde{\mathbf{q}}^\mathsf{T}\mathbf{R}\tilde{\mathbf{q}} + \tilde{\mathbf{u}}^\mathsf{T}\mathbf{M}\tilde{\mathbf{u}}$ with $\mathbf{R} = \mathbf{L}_2 = \mathbf{L}_2^\mathsf{T} \succ 0$, it follows that the error dynamics is attractively stable if the LMI

$$\begin{pmatrix} \mathbf{L}_1\mathbf{L}_2 & \mathbf{0} \\ \mathbf{0} & (\mathbf{D} + \mathbf{M}L) \end{pmatrix} \succ 0. \tag{7.5}$$

is satisfied. Moreover, recall that $L > 0$ and $\mathbf{L}_1 = \mathbf{L}_1^\mathsf{T} \succ 0$ in the given observer structure. In the experiment, the gains $L = 150$ and $\mathbf{L}_1 = \mathbf{L}_2 = L\mathbf{I}$ will be used, where $\mathbf{I}$ is an identity matrix of appropriate dimensions.

7.4 Performance measures

An objective, quantitative comparison of different state observers is difficult for several reasons. First, if the observers are not using the same measurements as inputs, any direct comparison is unfair, since the number and types of measurements has a strong influence on the information content of the measurements. Second, most observer designs include tunable gains. The selection of these gains has a direct influence on the observer performance. Therefore, the performance of observers is only directly comparable if these gains are designed according to some common criterion or method. Nevertheless, standard error measures allow for a comparison of the abilities of the given designs.

In the following, the error measures for positions and velocities are kept separate, since some designs only estimate the velocities. In particular, the mean estimation error norms will be compared, which are defined for the position and velocity errors as

$$e_{\tilde{\mathbf{q}}} := \frac{1}{N}\sum_{k=0}^{N} \|\tilde{\mathbf{q}}_k\| \qquad \text{and} \qquad e_{\tilde{\mathbf{u}}} := \frac{1}{N}\sum_{k=0}^{N} \|\tilde{\mathbf{u}}_k\|, \tag{7.6}$$

where the Euclidean norm $\|\mathbf{x}\| := \sqrt{\mathbf{x}^\mathsf{T}\mathbf{x}}$ is used. Furthermore, $\tilde{\mathbf{q}}_k = \tilde{\mathbf{q}}(k\Delta t)$ and $\tilde{\mathbf{u}}_k = \tilde{\mathbf{u}}(k\Delta t)$ are the discrete measurements, sampled uniformly with a time step Δt, and the sums are taken over all measurement points $k \in \{0, \cdots, N\}$. As it is more intuitive, these measures will also be stated as relative values with respect to the maximum measured position and velocity norms, i.e.

$$e_{\tilde{\mathbf{q}}}^{\mathrm{rel}} := \frac{e_{\tilde{\mathbf{q}}}}{\max_k \|\mathbf{q}_k\|} \qquad \text{and} \qquad e_{\tilde{\mathbf{u}}}^{\mathrm{rel}} := \frac{e_{\tilde{\mathbf{u}}}}{\max_k \|\mathbf{u}_k\|}. \tag{7.7}$$

Moreover, for a more refined comparison it will be convenient to visualize the mean error norms with respect to a moving time window with a length of $n < N$ data points, referred to as

$$e_{\tilde{\mathbf{q}}}^{\mathrm{mov}} := \frac{1}{n}\sum_{k=0}^{n} \|\tilde{\mathbf{q}}_k\| \qquad \text{and} \qquad e_{\tilde{\mathbf{u}}}^{\mathrm{mov}} := \frac{1}{n}\sum_{k=0}^{n} \|\tilde{\mathbf{u}}_k\|. \tag{7.8}$$

7.5 Performance evaluation and comparison

For a comparison, all state observers are applied to the same experiments, in which non-periodic motions were generated by feeding a harmonic input voltage signal with a modulated frequency and a modulated amplitude to the excitation shaker. Note that even though this voltage input is harmonic, the resulting excitation force acting on the impact oscillator is generally non-harmonic, since it is affected by the response of the impact oscillator as well as various non-modeled influences such as the shaker internal dynamics.

In the experiment, an input of the form

$$V(t) = a(t)\sin(\omega(t)\, t) \tag{7.9}$$

is generated, with a modulated excitation frequency in the range $\omega(t) \in [15\pi, 19\pi]\,\frac{\mathrm{rad}}{\mathrm{s}}$. More precisely, $\omega(t) = 17\pi + \sin(0.2\pi t)\,\frac{\mathrm{rad}}{\mathrm{s}}$, resulting in an excitation frequency centered at $8.5\,\mathrm{Hz}$ (which is near the resonance frequency), modulated with a cycle duration of $10\,\mathrm{s}$. In addition, the amplitude was manually modulated in the range $a(t) \in [0.7, 1.3]\,\mathrm{V}$. The system's state, the excitation force and the voltage between the impacting parts have been measured over a time interval of 100 s and all measurement signals were sampled at a rate of 20 kHz.

Quantitative results

The initial conditions for all state observers (as well as for the simulation) are set to zero, $\hat{\mathbf{q}}(0) = \mathbf{0}$ and $\hat{\mathbf{u}}(0) = \mathbf{0}$. Furthermore, the measurement signals start at a selected time instant, which is such that for $t = 0$, the measured velocities are large and therefore it can be observed at the beginning of the trajectories how quickly the estimates are converging to the true state. The measured initial conditions are

$\mathbf{q}(0) = (1.88, -4.19)^\mathsf{T}$ [mm] and $\mathbf{u}(0) = (0.26, 0.87)^\mathsf{T}$ [m/s]. The discretization time step Δt for all observers has been set to $\Delta t = 10^{-3}$s, which is 20 times slower than the measurement sampling rate. This discretization time step ensures that velocity jumps due to impacts occur within one observer time step. In fact, it can be observed from the measurement signals, that the velocity changes at impacts take place over a time interval of about $30 \cdot 10^{-5}$s, which is 6 times longer than the measurement sampling time interval.

Figure 7.6 shows the measured and estimated trajectories for all observers during an initial transient phase (left) as well as during an interval with converged state estimations (right). The minimal order observer 'Tanwani (reduced)', only estimates velocities. Therefore, its positions agree with the measured positions at all times. For the selected observer gain, the estimated velocities converge fairly quickly to the true velocities and show a good match over the entire measurement interval.

The full state observer 'Tanwani (full)', estimates both positions and velocities. With the selected observer gains, the estimated positions and velocities converge to the true velocities quickly.

The observer based on switched unilateral constraints and position jumps ('Baumann & Leine') converges more slowly in comparison. The reason for this is that the convergence rate cannot be tuned as no adjustable correction terms are present in the observer. As shown in Figure 7.6 (on the left), the state estimation initially corresponds to a pure simulation, but strongly improves as soon as contact is detected at the first impact (especially the positions that undergo unphysical jumps). The overall estimation accuracy is lower in comparison to tunable observers that include correction terms, as might have been expected. Especially in longer time intervals without impacts, the state estimation strongly relies on the model accuracy. Figure 7.7 shows the end of such an impact free interval, with pronounced estimation errors that are reduced as soon as more contacts are detected. For comparison, a pure simulation, which does not include any correction based on measurements but makes use of the measured excitation force as an input is also shown in Figure 7.6 and 7.7.

Table 7.8 shows the mean Euclidean norm of the position and velocity estimation errors as performance measures for all state observers. While they are all within the same order of magnitude, the Baumann & Leine observer comes with the largest estimation errors. This is no surprise, as it only uses knowledge of the collision time instants, while the other observers contain a permanent correction based on position measurements. The mean Euclidean error norms of the Baumann & Leine observer, relative to the maximal measured position and velocity norms, are at 4.8 % for the position estimates and 4.2 % for the velocity estimates. In comparison, the full Tanwani observer achieves a mean relative velocity error norm of 2.0 %. The reduced Tanwani observer yields the most accurate velocity estimates with the chosen observer gains, with a mean relative error norm of 1.8 %, and its position is zero since the position measurements are directly used as estimates.

For a more refined comparison, Figure 7.8 shows the mean error norms with respect to a moving time window are shown for a representative time interval (with a moving window length of about 2 excitation cycles). It can be observed, that the estimation error of the Tanwani observers strongly increases in some time intervals, for example around $t = 8\,\mathrm{s}$. These are time intervals with a high density of impacts. The impacts are exciting unmodeled spring-internal vibrations, which lead to a superimposed velocity oscillation at a frequency of about 100 Hz. This can be observed in Figure 7.6 on the right, in the plot for u_2, looking at the gray measurement signal. Since these vibrations are not contained in the mechanical model, the estimation error strongly increases. In addition, since in the Tanwani observers the impact time instants are detected using a threshold, slight mismatches between the observer's and the true impact time instants occur, which also result in increased estimation errors, although only during very short time intervals.

		Baumann & Leine	Tanwani (reduced)	Tanwani (full)	Simulation
Inputs		$\mathbf{f}(t), \chi(t)$	$\mathbf{f}(t), \mathbf{q}(t)$	$\mathbf{f}(t), \mathbf{q}(t)$	$\mathbf{f}(t)$
Estimates		$\hat{\mathbf{q}}, \hat{\mathbf{u}}$	$\hat{\mathbf{u}}$	$\hat{\mathbf{q}}, \hat{\mathbf{u}}$	$\hat{\mathbf{q}}, \hat{\mathbf{u}}$
$e^{\mathrm{rel}}_{\tilde{\mathbf{q}}}$	[%]	4.8	0	0.5	26.4
$e^{\mathrm{rel}}_{\tilde{\mathbf{u}}}$	[%]	4.2	1.8	2.0	23.7
$e_{\tilde{\mathbf{q}}}$	[mm]	0.9005	0	0.0838	4.9639
$e_{\tilde{\mathbf{u}}}$	[m/s]	0.0489	0.0210	0.0228	0.2749

Table 7.2: Performance comparison

Remark 7.1. Here, only the performance of the state observers for known impact time instants have been shown. As discussed in Chapter 6, the discretization-based approach for the case of unknown impact time instants resulted in large observer gains (especially for small time steps, which are desired). In the practical application, the noise sensitivity turned out to be unacceptably large. Furthermore, an ad-hoc adaptation to achieve a practical observer is not trivial.

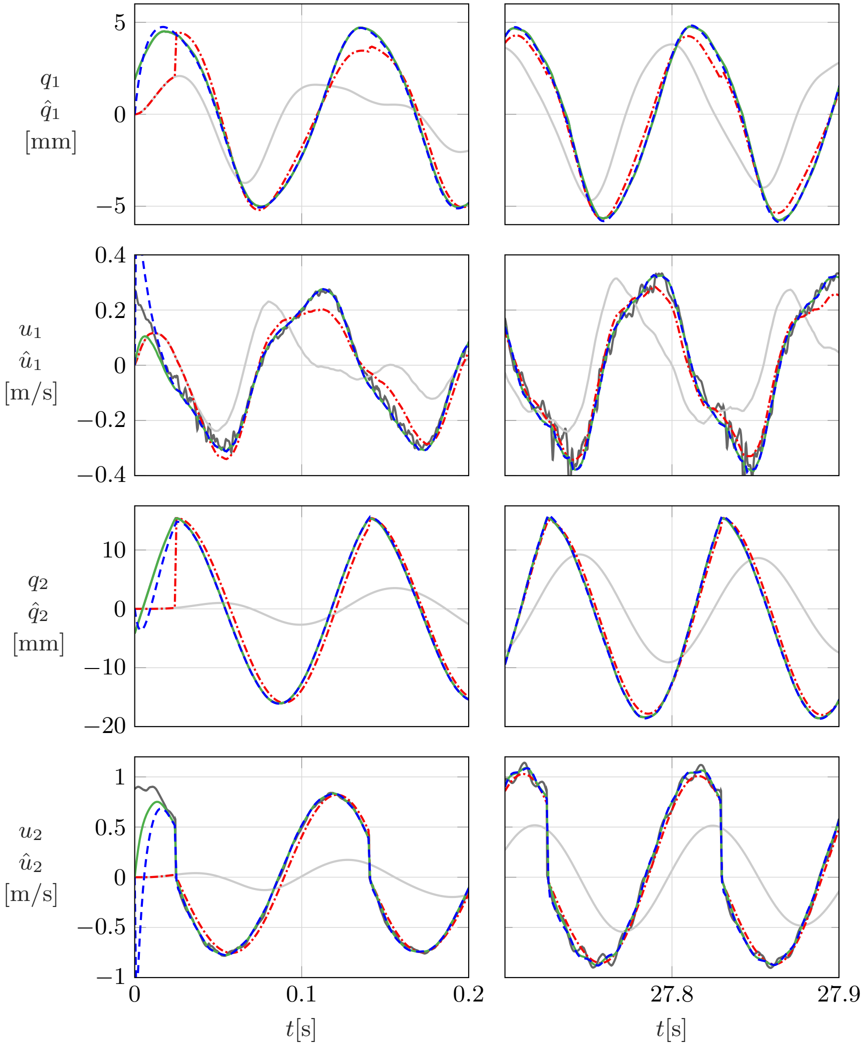

Figure 7.6: Extract from the time evolution of the positions and velocities of the measured signals (dark gray), the state estimations (Baumann & Leine: dashdotted red, Tanwani (reduced): solid green, Tanwani (full): dashed blue) and a pure simulation that only uses the measured excitation force as an input (solid light gray). *Left:* Transient phase, *right:* converged state estimates.

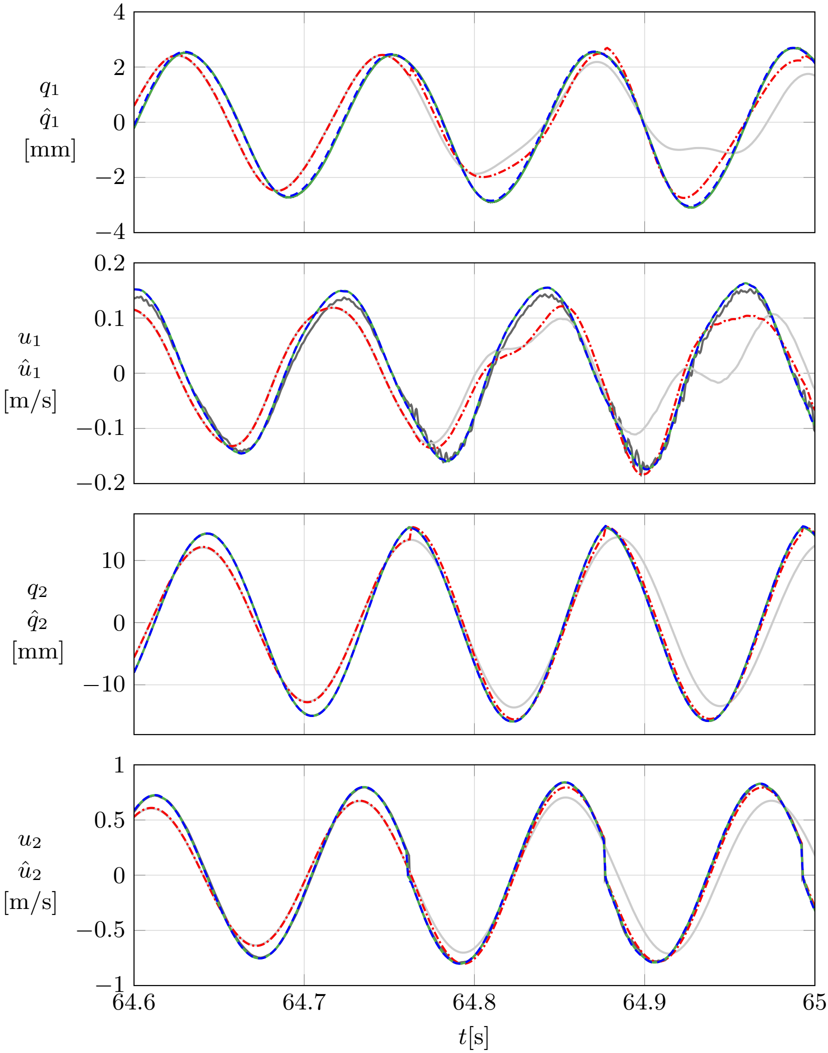

Figure 7.7: Extract from the time evolution of the positions and velocities of the measured signals (dark gray), the state estimations (Baumann & Leine: dashdotted red, Tanwani (reduced): solid green, Tanwani (full): dashed blue) and a pure simulation that only uses the measured excitation force as an input (solid light gray).

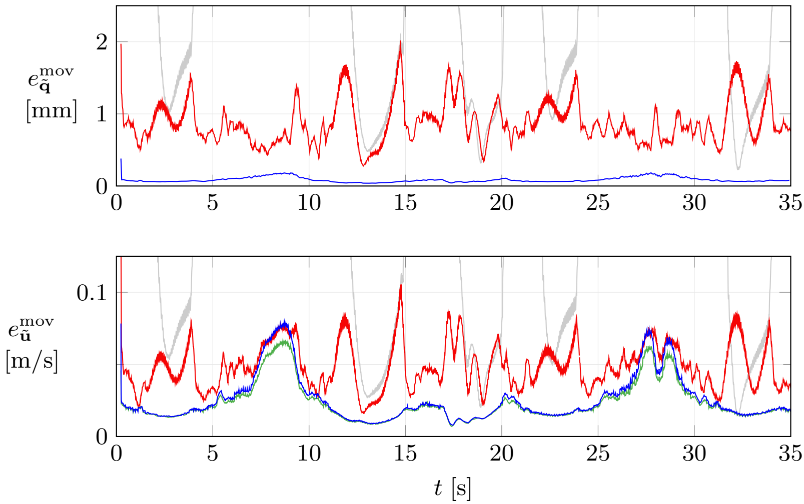

Figure 7.8: Position and velocity estimation error in a selected representative time interval (Baumann & Leine: red, Tanwani (reduced): green, Tanwani (full): blue). For comparison, the pure simulation is shown in light gray.

8

Conclusion

In this monograph, the state observer design problem was investigated for mechanical systems with unilateral constraints, with a special focus on the case where it cannot instantaneously be concluded from measurements whether contacts are open or closed. The main insights have been provided by the analysis of a discretization-based approach to the state observer problem and by the experimental performance analysis of selected state observers for impulsive systems.

8.1 Summary and contributions

In Chapter 2, a short introduction to the formalism of measure differential inclusions was given and its capabilities have been discussed. MDIs are especially well suited for modeling mechanical systems with unilateral constraints and an equality of measures that describes such systems has been derived in Chapter 3. The derivation follows a classical path, but since the equations of motion for smooth mechanical systems, as well as the axioms they are derived from, require the time evolution to be continuously differentiable, a generalization to differential measures is required to allow for a discontinuous or non-differentiable motion. This generalization can be made starting from the equations of motion, as it can be found in the works of Moreau [79], or by generalizing the underlying axioms. However, in Chapter 3, the principle of virtual action has been chosen as a starting point, as suggested by Capobianco [27].

Partial stability results for MDIs were presented in Chapter 4, which are a direct extension of their smooth counterpart. They provide a fundamental tool for problems that involve the stabilization of an error dynamics, such as state observer design, tracking control or synchronization problems. By discussing the known peaking phenomenon, it was highlighted that further generalizations or alternative approaches are necessary for analyzing an error dynamics.

Existing state observer designs were reviewed in a unified framework in Chapter 5. Furthermore, the extension to account for Coulomb friction was discussed for the special case of non-opening frictional contacts with known normal contact forces. Under these restrictions the friction force laws are maximal monotone, which fits into the stability proofs for the estimation errors, which rely on the maximal monotonicity of the force laws for the normal contact forces.

A new approach to the observer design problem for the case of *unknown* impact time instants was investigated in Chapter 6 for the class of impulsive mechanical systems which are linear in the absence of contacts. A key point was the use of the specific time discretization scheme of Paoli and Schatzman in order to side-step the main difficulties involved with the stability analysis of the estimation error dynamics: first, peaking of the estimation error due to velocity jumps does not occur in the time discretization, as the discrete system only describes updates over time steps. Second, normal contact force and impact laws, which are generally formulated on (position-switched) velocity level, loose their favorable property of maximal monotonicity (which holds for closed contacts) due to the switching on position level. The Paoli-Schatzman scheme however discretizes the normal contact force laws directly on position level. For the discretized system, sufficient observability conditions have been obtained from deriving a dead-beat observer by directly calculating initial conditions from a set of output measurements. Another useful result was, that for linear constraint equations, the chosen discretization leads to a discrete linear complementarity problem. Therefore, the discrete adaptation and extension of an existing passivity-based observer design could be applied. In a numerical case study, the usefulness of the taken approach was analyzed and limitations were highlighted. Specifically, it was shown that for a two degree-of-freedom impact oscillator, where only the position of the non-impacting mass is measured, a state observer can indeed be designed. It turned out however, that the contained observer gains have to be chosen inversely proportional to the chosen discretization time step, which results in a high noise sensitivity and is therefore problematic in real-world applications.

In Chapter 7, selected state observer designs (for known impact time instants) have been realized and tested on an experimental setup. The setup consists of a two degree-of-freedom impact oscillator with a single unilateral constraint. It was found that all the implemented observers can provide useful state estimations. However, tunable observers that use continuous measurements are, as one might expect, more accurate.

8.2 Recommendations

As this thesis shows, many problems related to state observation of non-smooth mechanical systems still remain unsolved. Several topics for future research that are of particular interest can be identified.

First, the state observers in Chapter 5, that make use of the explicit knowledge of contact time instants, have been shown to be applicable in experiments. However, only the accuracy of the state estimation has been analyzed. One of the main purposes of the observers is to provide a state estimate for state-feedback control. In general, when designing such a controller, the closed-loop dynamics including the observer has to be analyzed. In some cases a 'separation principle' holds, that allows to design the controller and the observer separately. It is an open question for what control applications such a separation principle could be provided with the observers in Chapter 5. Furthermore, the observers have not yet been tested experimentally in a control setup.

The approach in Chapter 6 relies on using the Paoli-Schatzman time discretization scheme for constructing an observer. With this specific scheme, the calculation of an index set (indicating whether a contact is open or closed at a given time instant) is avoided. This property might also be useful in optimization-based state estimation methods, where the initial state at the beginning of a moving time window is estimated by minimizing a certain functional of the estimation error (between a model-based predicted output and the measurements) over that window. Since in most cases an online state estimation is required, the computation time is a limiting factor in this approach.

For the discrete observer problem, sufficient observability conditions have been provided in Chapter 6. Roughly speaking, observability conditions guarantee that enough information is contained in the measurements to recover the initial state (and with it the current state, which can be obtained by propagating the system model from the recovered initial condition). Such observability conditions are an important aspect for the observer design. First, before trying to find a new observer structure, one wants to know if the observer problem is theoretically solvable. Second, with a given observer design at hand in practice, it is necessary to know what kinds of measurements are required and how many sensors have to be used. However, for non-smooth systems (and also for smooth non-linear systems), it is difficult to provide verifiable observability conditions. This problem has not explicitly been addressed in this thesis, but further research in this direction is essential for an observer theory for non-smooth systems. Furthermore, it might provide more insight of how to construct observers.

From the experimental analysis in Chapter 7 it became clear, that observers that only use contact measurements strongly rely on the model accuracy in time intervals without contact. Moreover, observers that use position measurements require a threshold for detecting contact, which causes slight mismatches between the observer's impact time instants and the true impact time instants. In practice, it can be difficult to set these thresholds properly. For those reasons, the possibility of a combination of contact measurements and continuous position measurements should be investigated to overcome these shortcomings.

A

Stieltjes integrals and measures in Non-smooth Dynamics

In Nonsmooth Dynamics, an equality of measures takes the place of the usual equations of motion to describe the dynamics of a mechanical system (see Chapter 3). Its derivation and analysis require basic knowledge on Riemann-Stieltjes integrals and their connections to measure theory. The interested reader who is unfamiliar with measure and integration theory may find it difficult to extract the essential points for its application in mechanics from literature. This is due to several reasons. Firstly, the terminology is often not completely identical among different texts. Definitions of the same objects may slightly vary and it is difficult to oversee the consequences of such differences. Furthermore, there exist two competing expositions of measure theory in literature. For example Moreau [78], whose works constitute a foundation for most of Nonsmooth Dynamics, understands measures as linear functionals on the vector space of continuous functions with compact support. Conversely, most textbooks on measure theory introduce measures as functions mapping subsets of a given set to the positive real numbers. Moreover, measure theory is a vast subject, such that it is easy to lose track of what is important for the application in mechanics.

In this appendix, a short overview over the key points is given while introducing a minimum number of concepts. All subjects are based on the lecture notes [47] and standard textbooks [35, 37, 65] as well as some papers [78, 79]. Similar treatments with more depth can be found in [41, 68]. Whenever possible, the set of real numbers $\mathbb{R}$ is used instead of the most general sets treated in measure theory. Theorems are stated without proof for a more compact exposition.

A.1 Function classes of special interest in measure theory

First, some classes of functions are introduced that possess favourable properties with respect to the specific measures and integrals that will be introduced subsequently. Let $\mathcal{I} \subset \mathbb{R}$ be a real interval and $\|\cdot\|$ be a norm on $\mathbb{R}^n$.

Definition A.1. A real-valued function $\mathbf{f} : \mathcal{I} \to \mathbb{R}^n$ is called **absolutely continuous** on $\mathcal{I}$ if for every $\varepsilon > 0$ there exists a $\delta(\varepsilon) > 0$ such that

$$\sum_{i=1}^{N}(b_i - a_i) < \delta \quad \Rightarrow \quad \sum_{i=1}^{N} \|\mathbf{f}(b_i) - \mathbf{f}(a_i)\| < \varepsilon$$

for all finite sequences of pairwise disjunct subintervals (a_i, b_i) of $\mathcal{I}$ with $a_i < b_i$.

Definition A.2. Let $\mathbf{f} : \mathcal{I} \to \mathbb{R}^n$ and let $\mathcal{J} \subseteq \mathcal{I}$ be a subinterval. Then the **variation** of $\mathbf{f}$ on $\mathcal{J}$ is defined as the nonnegative extended real number

$$\mathrm{var}(\mathbf{f}, \mathcal{J}) := \sup \left\{ \sum_{i=1}^{N} \|\mathbf{f}(x_i) - \mathbf{f}(x_{i-1})\| \right\},$$

where the suppremum is taken over all strictly increasing finite sequences $x_0 < x_2 < \cdots < x_N$ of points on $\mathcal{J}$.

The function $\mathbf{f}$ is said to be of **bounded variation** on $\mathcal{I}$ if and only if $\mathrm{var}(\mathbf{f}, \mathcal{I}) < \infty$, which is expressed as $\mathbf{f} \in \mathrm{bv}(\mathcal{I}, \mathbb{R}^n)$. As a weaker condition, $\mathbf{f}$ is of **locally bounded variation** on $\mathcal{I}$ if $\mathrm{var}(\mathbf{f}, [a, b]) < \infty$ for every compact subinterval $[a, b]$ of $\mathcal{I}$, which is expressed as $\mathbf{f} \in \mathrm{lbv}(\mathcal{I}, \mathbb{R}^n)$.

For functions of bounded variation, two useful decompositions exist:

i. *Jordan decomposition*: If $\mathbf{f} \in \mathrm{bv}(\mathcal{I}, \mathbb{R}^n)$, then there exists a pair of functions $\mathbf{f}_\mathrm{p}$ and $\mathbf{f}_\mathrm{n}$, both with non-decreasing entries, such that

$$\mathbf{f} = \mathbf{f}_\mathrm{p} - \mathbf{f}_\mathrm{n},$$

and the pair is unique up to addition of a constant.

ii. *Lebesgue decomposition*: Every function $\mathbf{f} \in \mathrm{bv}(\mathcal{I}, \mathbb{R}^n)$ can be decomposed into a sum of three functions as

$$\mathbf{f} = \mathbf{f}_\mathrm{ac} + \mathbf{f}_\mathrm{step} + \mathbf{f}_\mathrm{sing}$$

such that $\mathbf{f}_{\mathrm{ac}}$ is an absolutely continuous function, $\mathbf{f}_{\mathrm{step}}$ is a step function and $\mathbf{f}_{\mathrm{sing}}$ is a so-called singular function, which can be characterized as a function of bounded variation whose classical derivative vanishes almost everywhere. Moreover, these functions are unique up to a constant.

Functions $\mathbf{f} \in \mathrm{lbv}(\mathcal{I}, \mathbb{R}^n)$, for which the singular part $\mathbf{f}_{\mathrm{sing}}$ vanishes in the Lebesgue decomposition, are called **special functions of locally bounded variation**[1], written as $\mathbf{f} \in \mathrm{slbv}(\mathcal{I}, \mathbb{R}^n)$.

A.2 The Riemann-Stieltjes integral

In order to capture discontinuity events in the description of a mechanical system, the equations of motion have to be understood in an integral sense. To this end, it is instrumental to introduce in the following a notion of an integral, which is a generalization of the classical Riemann integral and will be called the Riemann-Stieltjes integral. Its definition is taken from Moreau [78].
Let $\mathcal{I}$ be a real interval of any form and let $\varphi : \mathcal{I} \to \mathbb{R}$ be a continuous real-valued function on the interval $\mathcal{I}$ with compact support in $\mathcal{I}$ (i.e. its support is contained in $\mathcal{I}$), written as $\varphi \in \mathrm{C}_0(\mathcal{I}, \mathbb{R})$. For any partition

$$S_m : a = \tau_0 < \tau_1 < \ldots < \tau_m = b$$

of an interval $[a, b] \in \mathcal{I}$, we introduce its size

$$|S_m| := \max_{1 \le i \le m} \{\tau_i - \tau_{i-1}\}.$$

Let $\mathbf{f} : \mathcal{I} \to \mathbb{R}^n$ be a given function of bounded variation. Then, for selected values $\xi_i^{S_m} \in [\tau_{i-1}, \tau_i]$ in every subinterval of a given partition S_m, a Riemann-Stieltjes sum can then be built as

$$\mathbf{H}(S_m, \xi^{S_m}, \varphi, \mathbf{f}) := \sum_{i=1}^{m} \varphi(\xi_i^{S_m})[\mathbf{f}(\tau_i) - \mathbf{f}(\tau_{i-1})].$$

It can be shown[2]that for any sequence of partitions with $|S_m| \to 0$ as $m \to \infty$, the Riemann-Stieltjes sum converges to a limit, independent of the selected values ξ^{S_m}. This limit is denoted as

$$\int_{[a,b]} \varphi \, \mathrm{d}\mathbf{f} := \lim_{|S_m| \to 0} \mathbf{H}(S_m, \xi^{S_m}, \varphi, \mathbf{f}) \tag{A.1}$$

[1] In accordance with [3].
[2] [78], Proposition 6.1.

and is called the **Riemann-Stieltjes integral** of φ with respect to $\mathbf{f}$. If $\mathbf{f}$ is the identity function, the classical Riemann integral is obtained.

Notation A.3. The short writing $\mathrm{d}\mathbf{f}[\varphi]$ is used here to mean the the bounded functional $\varphi \mapsto \int_{[a,b]} \varphi \,\mathrm{d}\mathbf{f}$, which is referred to as the **differential measure**[3]**of** $\mathbf{f}$.

Remark A.4. The Riemann-Stieltjes integral is also defined for more general bounded functions φ and $\mathbf{f}$. For example, if the integrand φ is discontinuous, the Riemann-Stieltjes integral is still defined as in (A.1), but *only if* the limit exists. In that case it is said that φ is Riemann-Stieltjes integrable with respect to $\mathbf{f}$. Functions $\varphi \in \mathrm{C}_0(\mathcal{I}, \mathbb{R})$ are always integrable with respect to $\mathbf{f} \in \mathrm{lbv}(\mathcal{I}, \mathbb{R}^n)$.

Remark A.5. For vector-valued functions $\boldsymbol{\varphi} : \mathcal{I} \to \mathbb{R}^n$ the Riemann-Stieltjes integral is defined as follows. For scalar functions $f : \mathcal{I} \to \mathbb{R}$ it is defined as

$$\int_{[a,b]} \boldsymbol{\varphi} \,\mathrm{d}f := \left(\int_{[a,b]} \varphi_1 \,\mathrm{d}f \,, \cdots , \int_{[a,b]} \varphi_n \,\mathrm{d}f \right)^{\mathsf{T}} , \tag{A.2}$$

and for vector-valued functions $\mathbf{f} : \mathcal{I} \to \mathbb{R}^n$ it is

$$\int_{[a,b]} \boldsymbol{\varphi}^{\mathsf{T}} \mathrm{d}\mathbf{f} := \int_{[a,b]} \sum_{i=1}^{n} \varphi_i \,\mathrm{d}f_i . \tag{A.3}$$

The differential measures for these cases are written as $\mathrm{d}f[\boldsymbol{\varphi}]$ and $\mathrm{d}\mathbf{f}[\boldsymbol{\varphi}^{\mathsf{T}}]$ respectively.

Linearity

An immediate consequence of the definition of the Riemann-Stieltjes integral is its linearity with respect to $\mathbf{f}$ and φ, i.e. for any $c \in \mathbb{R}$ the following properties hold:

$$\begin{aligned}
\int_{[a,b]} \varphi \,\mathrm{d}(\mathbf{f}_1 + \mathbf{f}_2) &= \int_{[a,b]} \varphi \,\mathrm{d}\mathbf{f}_1 + \int_{[a,b]} \varphi \,\mathrm{d}\mathbf{f}_2, & \int_{[a,b]} \varphi \,\mathrm{d}(c\mathbf{f}) &= c \int_{[a,b]} \varphi \,\mathrm{d}\mathbf{f}, \\
\int_{[a,b]} (\varphi_1 + \varphi_2) \,\mathrm{d}\mathbf{f} &= \int_{[a,b]} \varphi_1 \,\mathrm{d}\mathbf{f} + \int_{[a,b]} \varphi_2 \,\mathrm{d}\mathbf{f}, & \int_{[a,b]} (c\varphi) \,\mathrm{d}\mathbf{f} &= c \int_{[a,b]} \varphi \,\mathrm{d}\mathbf{f}.
\end{aligned}$$

[3]In literature on Nonsmooth Dynamics it is often referred to as *differential measure*, owing to the fact that one branch of measure theory defines measures as continuous linear functionals on the vector space of continuous functions with compact support.

Subintervals

It is common that the notation $\int_{\mathcal{I}} \mathrm{d}\mathbf{f}$ stands for $\int \chi_{\mathcal{I}}\, \mathrm{d}\mathbf{f}$ with the characteristic function $\chi_{\mathcal{I}} \in \mathrm{C}_0(\mathcal{I}, \mathbb{R})$ of the interval $\mathcal{I}$, which is defined as

$$\chi_{\mathcal{I}}(t) := \begin{cases} 1 & \text{if } t \in \mathcal{I} \\ 0 & \text{if } t \notin \mathcal{I}. \end{cases} \tag{A.4}$$

For any interval $\mathcal{I}$, the mapping $\boldsymbol{\nu}_{\mathbf{f}} : \mathcal{I} \mapsto \int_{\mathcal{I}} \mathrm{d}\mathbf{f}$ defines the Lebesgue-Stieltjes vector measure for $\mathbf{f}$, which is introduced in Section A.3. Indeed, as shown in [78], we have the following properties:

$$\begin{aligned} \int_{[a,b]} \mathrm{d}\mathbf{f} &= \mathbf{f}^{+}(b) - \mathbf{f}^{-}(a), & \int_{[a,b)} \mathrm{d}\mathbf{f} &= \mathbf{f}^{-}(b) - \mathbf{f}^{-}(a), \\ \int_{(a,b)} \mathrm{d}\mathbf{f} &= \mathbf{f}^{-}(b) - \mathbf{f}^{+}(a), & \int_{(a,b]} \mathrm{d}\mathbf{f} &= \mathbf{f}^{+}(b) - \mathbf{f}^{+}(a), \end{aligned} \tag{A.5}$$

which correspond to (the vector form of) the properties (A.10).
As a direct consequence, it follows that for an interval that consists of a singleton $\{a\}$ it holds that

$$\int_{\{a\}} \mathrm{d}\mathbf{f} = \mathbf{f}^{+}(a) - \mathbf{f}^{-}(a)\,. \tag{A.6}$$

Densities

Let $\mathcal{I}$ be a real interval, $h : \mathcal{I} \to \mathbb{R}$ be a continuous function and $\mathbf{f} \in \mathrm{slbv}(\mathcal{I}, \mathbb{R}^n)$. Then for every $\varphi \in \mathrm{C}_0(\mathcal{I}, \mathbb{R})$, i.e. for every continuous function with compact support on $\mathcal{I}$, it holds that $\varphi h \in \mathrm{C}_0(\mathcal{I}, \mathbb{R})$. The **product of the differential measure** $\mathrm{d}\mathbf{f}[\varphi] : \varphi \mapsto \int_{\mathcal{I}} \varphi\, \mathrm{d}\mathbf{f}$ **with** h is declared to be the functional

$$h\, \mathrm{d}\mathbf{f}[\varphi] : \varphi \mapsto \int_{\mathcal{I}} \varphi h\, \mathrm{d}\mathbf{f}. \tag{A.7}$$

If the differential measure $\mathrm{d}\boldsymbol{g}[\varphi]$ of a function $\boldsymbol{g} \in \mathrm{slbv}(\mathcal{I}, \mathbb{R})$ can be written as $\mathrm{d}\boldsymbol{g}[\varphi] = h\, \mathrm{d}\mathbf{f}[\varphi]$, i.e. it is such that

$$\mathrm{d}\boldsymbol{g}[\varphi] : \varphi \mapsto \int_{\mathcal{I}} \varphi h\, \mathrm{d}\mathbf{f} \tag{A.8}$$

holds, then it is said that h is the **density** of $\mathrm{d}\boldsymbol{g}$ with respect to $\mathrm{d}\mathbf{f}$, which is written in short notation as $\mathrm{d}\boldsymbol{g} = h\, \mathrm{d}\mathbf{f}$.

In view of the application in mechanics, the densities of the differential measures of two main function classes are crucial.

1. If $\mathcal{I}$ is a real interval and $\mathbf{f}_{\mathrm{ac}} : \mathcal{I} \to \mathbb{R}^n$ is absolutely continuous, then $\mathbf{f}_{\mathrm{ac}}$ is differentiable almost everywhere (i.e. everywhere, except for a set of Lebesgue measure[4] zero) on $\mathcal{I}$ and it holds that

$$\mathbf{f}_{\mathrm{ac}}(b) = \mathbf{f}_{\mathrm{ac}}(a) + \int_{(a,b]} \dot{\mathbf{f}}_{\mathrm{ac}} \, \mathrm{d}t \quad \forall \, a, b \in \mathcal{I},$$

with the derivative $\dot{\mathbf{f}}_{\mathrm{ac}} = \mathrm{d}\mathbf{f}_{\mathrm{ac}}/\mathrm{d}t$. As a consequence, for the Stieltjes functional of $\mathbf{f}_{\mathrm{ac}}$ it holds that $\mathrm{d}\mathbf{f}_{\mathrm{ac}}[\varphi] = \dot{\mathbf{f}}_{\mathrm{ac}} \, \mathrm{d}t[\varphi]$, i.e. it has a density $\dot{\mathbf{f}}_{\mathrm{ac}}$ w.r.t. $\mathrm{d}t$. In short, $\mathrm{d}\mathbf{f}_{\mathrm{ac}} = \dot{\mathbf{f}}_{\mathrm{ac}} \, \mathrm{d}t$.

2. A second important class are step functions, which are piecewise constant functions with a countable number of discontinuities. Taking a unit step function

$$h_{t_k}(t) = \begin{cases} 1 & \text{if } \ t \geq t_k \\ 0 & \text{if } \ t < t_k, \end{cases}$$

with a discontinuity at a given $t = t_k$, it follows with the definition (A.1) and a given time interval $\mathcal{I}$ that

$$\int_{\mathcal{I}} \varphi \, \mathrm{d}h_{t_k} = \begin{cases} \varphi(t_k) & \text{if } \ t_k \in \mathcal{I} \\ 0 & \text{if } \ t_k \notin \mathcal{I}. \end{cases}$$

The differential measure $\mathrm{d}h_{t_k}[\varphi]$ is commonly written as $\mathrm{d}\delta_{t_k}[\varphi]$ and referred to as **Dirac (differential) measure**. Now, take a step function $\mathbf{f}_{\mathrm{s}} : \mathcal{I} \to \mathbb{R}^n$ for which $\mathbf{f}_{\mathrm{s}}^+(t_i) \neq \mathbf{f}_s^-(t_i)$ at a countable number of discontinuity points $t_i, i = 1, 2, \cdots$ and $\mathbf{f}_{\mathrm{s}}^+(t) = \mathbf{f}_{\mathrm{s}}^-(t) = \mathbf{f}_{\mathrm{s}}(t) \; \forall \; t \neq t_i$. It can then be verified that

$$\mathbf{f}_{\mathrm{s}}^+(b) = \mathbf{f}_{\mathrm{s}}^-(a) + \int_{[a,b]} \left(\mathbf{f}_{\mathrm{s}}^+ - \mathbf{f}_{\mathrm{s}}^-\right) \sum_i \mathrm{d}\delta_{t_i},$$

showing that step functions $\mathbf{f}_{\mathrm{s}}$ have a density $(\mathbf{f}_{\mathrm{s}}^+ - \mathbf{f}_{\mathrm{s}}^-)$ w.r.t. a sum of Dirac measures. For brevity, it is typically written $\mathrm{d}\eta := \sum_i \mathrm{d}\delta_{t_i}$ such that $\mathrm{d}\mathbf{f}_{\mathrm{s}}[\varphi] = (\mathbf{f}_{\mathrm{s}}^+ - \mathbf{f}_{\mathrm{s}}^-)\mathrm{d}\eta[\varphi]$ (or $\mathrm{d}\mathbf{f}_{\mathrm{s}} = (\mathbf{f}_{\mathrm{s}}^+ - \mathbf{f}_{\mathrm{s}}^-)\mathrm{d}\eta$ in short).

Partial integration

For the Riemann-Stieltjes integral, a very useful integration by parts formula exists. Namely, if $\mathcal{I}$ is a real interval and $\mathbf{u} : \mathcal{I} \mapsto \mathbb{R}^n$ and $\mathbf{v} : \mathcal{I} \mapsto \mathbb{R}^n$ are two functions of

[4] Measures, and particularly the Lebesgue measure, will be defined in Section A.3

bounded variation, then it holds that

$$\begin{aligned} \int_{\mathcal{I}} \mathrm{d}(\mathbf{u} \cdot \mathbf{v}) &= \int_{\mathcal{I}} \mathbf{u}^{+} \mathrm{d}\mathbf{v} + \int_{\mathcal{I}} \mathbf{v}^{-} \mathrm{d}\mathbf{u} \\ &= \int_{\mathcal{I}} \mathbf{u}^{-} \mathrm{d}\mathbf{v} + \int_{\mathcal{I}} \mathbf{v}^{+} \mathrm{d}\mathbf{u}, \end{aligned} \tag{A.9}$$

which is written in short as $\mathrm{d}(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u}^{+}\mathrm{d}\mathbf{v} + \mathbf{v}^{-}\mathrm{d}\mathbf{u} = \mathbf{u}^{-}\mathrm{d}\mathbf{v} + \mathbf{v}^{+}\mathrm{d}\mathbf{u}$ [78].

A.3 Measures on $\mathbb{R}$

In the following, some basic measure theoretic definitions are introduced. In view of the application in mechanics, the treatment is restricted to the set $\mathbb{R}$ of real numbers (or $\mathbb{R}^n$ if necessary) wherever possible.

The use of measures aims at generalizing concepts such as length, area or volume, which assign a non-negative number to all subsets of a given base set.

Definition A.6. A **σ-Algebra** on $\mathbb{R}$[5] is a system $\mathfrak{A}$ of subsets of $\mathbb{R}$ with the following properties:

i. $\mathbb{R} \in \mathfrak{A}$

ii. $A \in \mathfrak{A} \Rightarrow \mathbb{R} \backslash A \in \mathfrak{A}$

iii. $A_1, A_2, \ldots \in \mathfrak{A} \Rightarrow \bigcup\limits_{i=1}^{\infty} A_i \in \mathfrak{A}$

Of special interest is the **Borel σ-Algebra** $\mathfrak{B}(\mathbb{R})$ on $\mathbb{R}$, which is implicitly defined as the smallest σ-Algebra containing all open intervals with rational endpoints, i.e. all elements of the set $\mathcal{E} = \{(a, b) \subset \mathbb{R} \mid a < b \text{ and } a, b \in \mathbb{Q}\}$. Due to this implicit definition, it is said that $\mathfrak{B}(\mathbb{R})$ is generated by the generator $\mathcal{E}$. Note that generators are not unique, i.e. various generators can generate the identical σ-Algebra. Elements of $\mathfrak{B}(\mathbb{R})$ are called Borel sets and particularly include all real intervals of the form $[a, b], [a, b), (a, b], (a, b)$ with $a \leq b$ and $a, b \in \mathbb{R}$.

Definition A.7. A **measure**[6] on $\mathbb{R}$ is a map $\mu : \mathfrak{B}(\mathbb{R}) \to [0, \infty]$ which fulfills

i. $\mu(\emptyset) = 0$

ii. $\mu\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i} \mu(A_i)$ for any sequence of disjoint sets $A_i \in \mathfrak{B}(\mathbb{R})$

[5] More generally, in measure theory, an arbitrary non-empty set Ω is used in the definition of a σ-Algebra.

Property ii. is referred to as σ-additivity. Furthermore, if $\mu(A) < \infty$ for all $A \in \mathfrak{B}(\mathbb{R})$, then μ is called a finite measure.

As a slight generalization, so-called signed measures are introduced to allow for a notion of measures which can take negative values.

Definition A.8. A **signed measure** on $\mathbb{R}$ is a map $\nu : \mathfrak{B}(\mathbb{R}) \to (-\infty, \infty]$ for which $\mu(\emptyset) = 0$ and for which σ-additivity holds.

For signed measures there exists a *Jordan decomposition:* If ν is a signed measure on $\mathbb{R}$, then there exist two unique (non-negative) measures μ_+ and μ_- on $\mathbb{R}$, such that $\nu = \mu_+ - \mu_-$.

In the sequel, tuples $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)$ of measures $\mu_1, \cdots, \mu_n$ on $\mathbb{R}$ are referred to as **vector measures**[7]on $\mathbb{R}$. Similarly, tuples $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_n)$ of signed measures $\nu_1, \cdots, \nu_n$ on $\mathbb{R}$ are called **signed vector measures** on $\mathbb{R}$.

Lebesgue-Stieltjes measure

In the following, specific measures are introduced, which are based on functions of bounded variation.

Theorem A.9. If $f : \mathbb{R} \to \mathbb{R}$ is a monotonically increasing, right continuous function, then there exists a unique measure μ_f on $\mathbb{R}$, called the **Lebesgue-Stieltjes measure** of f, for which it holds that $\mu_f((a, b]) = f(b) - f(a)$ for all $a, b \in \mathbb{R}$.

Note that the set $\mathcal{J} := \{(a, b] | a, b \in \mathbb{R})\}$ is not a σ-algebra. Therefore the map $\mu : \mathcal{J} \to [0, \infty]$ defined by $\mu((a, b]) = f(b) - f(a) \,\forall\, a, b \in \mathbb{R}$ is not a measure. However, $\mathcal{J}$ is a subset of the Borel σ-algebra $\mathfrak{B}(\mathbb{R})$ and a well known extension theorem of Carathéodory guarantees the existence of a measure μ_f on $\mathbb{R}$ such that $\mu_f(A) = \mu(A) \,\forall\, A \in \mathcal{J}$.

Several cases of the Lebesgue-Stieltjes measure for specific functions f are of special interest.

i. For the identity function, i.e. $f(t) = t$, we have $\mu_f((a, b]) = b - a$, which defines the **Lebesgue measure** and is equal to the length of the measured interval. It is customary that for this case, μ_f is denoted as $\mathrm{d}t$.

[6]The domain of a measure in literature is typically required to be any σ-Algebra on a given non-empty set Ω. However, especially in older texts, some authors use other spaces for the domain, such as so-called σ-rings. Measures with $\mathfrak{B}(\mathbb{R})$ as the domain are sometimes called *Borel measures* on $\mathbb{R}$.

[7]In literature, vector measures are typically introduced as measures that map to a Banach space. However, both the domain and the codomain of a vector measure are not standardized in the literature on vector measures. A detailed treatment can be found in the very readable book [35].

ii. The unit step function $f(t) = \begin{cases} 1 & \text{if } t \geq t_k \\ 0 & \text{if } t < t_k \end{cases}$ leads to the **Dirac measure**
$\mu_f((a,b]) = \begin{cases} 1 & \text{if } t_k \in (a,b] \\ 0 & \text{if } t_k \notin (a,b] \end{cases} =: \delta_{t_k}((a,b]),$
which indicates whether a given point t_k lies within the measured interval.

iii. If f is a monotonically increasing, continuously differentiable function, it holds that $f^+(t) = f^-(t) = f(t)$ and $\mu_f((a,b]) = f(b) - f(a) = \int_a^b \dot{f}(t)\mathrm{d}t$.

The Lebesgue-Stieltjes measure can be generalized to a signed measure. Using the Jordan decomposition of a function $f \in \mathrm{bv}(\mathcal{I}, \mathbb{R})$, i.e.

$$f = f_\mathrm{p} - f_\mathrm{n}$$

with non-decreasing functions f_p and f_n, the **signed Lebesgue-Stieltjes measure** of f can be defined as

$$\nu_f(A) = \mu_{f_\mathrm{p}}(A) - \mu_{f_\mathrm{n}}(A) \;\; \forall A \in \mathfrak{B}(\mathbb{R}).$$

As a consequence, it holds that

$$\begin{aligned} \nu_f([a,b]) &= f^+(b) - f^-(a), \\ \nu_f([a,b)) &= f^-(b) - f^-(a), \\ \nu_f((a,b]) &= f^+(b) - f^+(a), \\ \nu_f((a,b)) &= f^-(b) - f^-(a). \end{aligned} \tag{A.10}$$

Notation A.10. In view of the Lebesgue measure, which is commonly denoted as $\mathrm{d}t$, the notation $\mathrm{d}f$ is henceforth used to mean the signed Lebesgue-Stieltjes measure of f. Accordingly, it is written $\mathrm{d}f(A)$ to mean $\nu_f(A)$ and for signed Lebesgue-Stieltjes vector measures of $\mathbb{R}^n$-valued functions $\mathbf{f}$ we write $\mathrm{d}\mathbf{f}(A)$ to mean $\boldsymbol{\nu}_\mathbf{f}(A)$.

A.4 The Lebesgue integral

In the following, it will be defined what it means to integrate a function with respect to a given measure. The introduced integral allows for a broader class of functions to be integrated, compared to the well-known Riemann-Integral. The construction of this so-called Lebesgue integral will follow several steps. First the integral is defined for non-negative simple functions. Second, the integral is defined for non-negative functions. Third, the integral of signed functions is introduced. Lastly, the notion of the integral is extended to integrals with respect to signed measures.

First, the definition of the characteristic function is re-stated as

$$\chi_A(t) := \begin{cases} 1 & \text{if } t \in A \\ 0 & \text{if } t \notin A, \end{cases} \tag{A.11}$$

where A is any subset of the real numbers, $A \subseteq \mathbb{R}$. A **simple function** $s : \mathcal{I} \to \mathbb{R}$ takes finitely many values $\alpha_1, \cdots, \alpha_N$ on the interval $\mathcal{I} \subset \mathbb{R}$ and can be written as

$$s = \sum_{i=1}^{N} \alpha_i \chi_{A_i},$$

where $A_i = s^{-1}(\alpha_i)$ are pairwise disjunct.
The **Lebesgue Integral** of a non-negative simple function ($\alpha_i > 0$) with respect to a measure μ on $\mathbb{R}$ is defined as

$$\int_{\mathcal{I}} s \, \mathrm{d}\mu := \sum_{i=1}^{N} \alpha_i \mu(\mathcal{I}_i).$$

More generally, the Lebesgue integral of a non-negative (measurable) function f is defined as

$$\int_{\mathcal{I}} f \, \mathrm{d}\mu := \sup_{s} \left\{ \int_{\mathcal{I}} s \, \mathrm{d}\mu \mid s \leq f \, \forall \, t \right\},$$

where the supremum is taken over all simple functions and $s \leq f$ refers to a pointwise inequality, $s(t) \leq f(t) \, \forall \, t$. A signed function f can be decomposed as $f = f_+ - f_-$, where f_+ and f_- are non-negative functions, and its integral is defined as

$$\int_{\mathcal{I}} f \, \mathrm{d}\mu := \int_{\mathcal{I}} f_+ \, \mathrm{d}\mu - \int_{\mathcal{I}} f_- \, \mathrm{d}\mu.$$

Finally, the integral of a (measurable) function f with respect to a signed measure ν on $\mathbb{R}$ is defined using the Jordan decomposition $\nu = \mu_+ - \mu_-$, where μ_+ and μ_- are (non-negative) measures on $\mathbb{R}$, as

$$\int_{\mathcal{I}} f \, \mathrm{d}\nu := \int_{\mathcal{I}} f \, \mathrm{d}\mu_+ - \int_{\mathcal{I}} f \, \mathrm{d}\mu_-.$$

Densities

It is common that the notation $\int_A f \, \mathrm{d}\mu$ stands for the integral $\int f \chi_A \, \mathrm{d}\mu$ with the characteristic function χ_A of the set A. Clearly, the characteristic function is a simple function if A consists of a finite number of intervals. In that case it therefore holds that

$$\int_A \mathrm{d}\mu := \int \chi_A \,\mathrm{d}\mu = \mu(A).$$

The same is true for signed measures ν, i.e. $\int_A \mathrm{d}\nu = \nu(A)$. Moreover, if f is a μ-integrable function, then

$$\nu(A) := \int_A f \,\mathrm{d}\mu \quad , \quad A \in \mathfrak{B}(\mathbb{R})$$

is a signed measure on $\mathbb{R}$. In that case, it is said that ν has **density**[8] f with respect to μ and we use the short notation $\mathrm{d}\nu = f\,\mathrm{d}\mu$ (also $\mathrm{d}\nu(A) = f\,\mathrm{d}\mu(A)$).

Theorem A.11. A function g is integrable w.r.t. such a signed measure ν if gf is integrable w.r.t. μ and in that case it holds that

$$\int_A g\,\mathrm{d}\nu = \int_A gf\,\mathrm{d}\mu,$$

or, written in short, $g\,\mathrm{d}\nu = gf\,\mathrm{d}\mu$ (also $g\,\mathrm{d}\nu(A) = gf\,\mathrm{d}\mu(A)$).

With regard to the application in mechanics, the densities of the Lebesgue-Stieltjes measure of two main function classes are crucial.

1. If $\mathcal{I}$ is a real interval and $f_{\mathrm{ac}} : \mathcal{I} \to \mathbb{R}$ is absolutely continuous, then it is differentiable almost everywhere on $\mathcal{I}$ and it holds that

   $$f_{\mathrm{ac}}(b) = f_{\mathrm{ac}}(a) + \int_{(a,b]} \dot{f}_{\mathrm{ac}} \mathrm{d}t \quad \forall\, a, b \in \mathcal{I},$$

   with the derivative $\dot{f}_{\mathrm{ac}} = \mathrm{d}f_{\mathrm{ac}}/\mathrm{d}t$. The signed Lebesgue-Stieltjes measure of f_{ac}, for which it holds that $\nu_{f_{\mathrm{ac}}}((a,b]) = f_{\mathrm{ac}}(b) - f_{\mathrm{ac}}(a) = \int_a^b \dot{f}_{\mathrm{ac}}\,\mathrm{d}t$, therefore has a density $\dot{f}_{\mathrm{ac}}$ w.r.t. the Lebesgue measure $\mathrm{d}t$. In short, $\mathrm{d}f_{\mathrm{ac}} = \dot{f}_{\mathrm{ac}}\,\mathrm{d}t$ (meaning $\mathrm{d}\nu_{f_{\mathrm{ac}}} = \dot{f}_{\mathrm{ac}}\,\mathrm{d}t$).

2. We have already seen, that the Lebesgue-Stieltjes measure of a unit step function is equal to the Dirac measure δ_{t_k} with respect to its discontinuity point t_k. Now, take a step function f_{s} for which $f_{\mathrm{s}}^+(t_i) \neq f_{\mathrm{s}}^-(t_i)$ at a countable number of discontinuity points $t_i, i = 1, 2, \cdots$ and $f_{\mathrm{s}}^+(t) = f_{\mathrm{s}}^-(t) = f_{\mathrm{s}}(t)\ \forall\, t \neq t_i$. It can then be verified that

   $$\mathrm{d}f_{\mathrm{s}}((a,b]) := \nu_{f_{\mathrm{s}}}((a,b]) = \int_{(a,b]} \left(f_{\mathrm{s}}^+ - f_{\mathrm{s}}^-\right) \sum_i \mathrm{d}\delta_{t_i},$$

[8] The density is sometimes called Radon-Nikodym derivative of ν w.r.t. μ and is then denoted as $\mathrm{d}\nu/\mathrm{d}\mu$.

which requires the fact that the (Lebesgue) integral of a given function $g(t)$ with respect to δ_{t_k} over an interval $(a,b]$ is

$$\int_{(a,b]} g \mathrm{d}\delta_{t_k} = \begin{cases} g(t_k) & \text{if } t_k \in (a,b] \\ 0 & \text{if } t_k \notin (a,b], \end{cases} = g(t_k)\delta_{t_k}((a,b]).$$

Eventually, one can conclude that step functions f_s have a density $f_\mathrm{s}^+ - f_\mathrm{s}^-$ with respect to the sum of Dirac measures. For brevity, this sum will from now on be denoted as $\mathrm{d}\eta := \sum_i \mathrm{d}\delta_{t_i}$ and we have $\mathrm{d}f_\mathrm{s} = (f_\mathrm{s}^+ - f_\mathrm{s}^-)\,\mathrm{d}\eta$.

A.5 Overview

The preceding sections mainly described the relations between three objects, namely functions of bounded variation, the Riemann-Stieltjes integral and the signed Lebesgue-Stieltjes measure. For a quick overview the following diagram shows the most important relations.

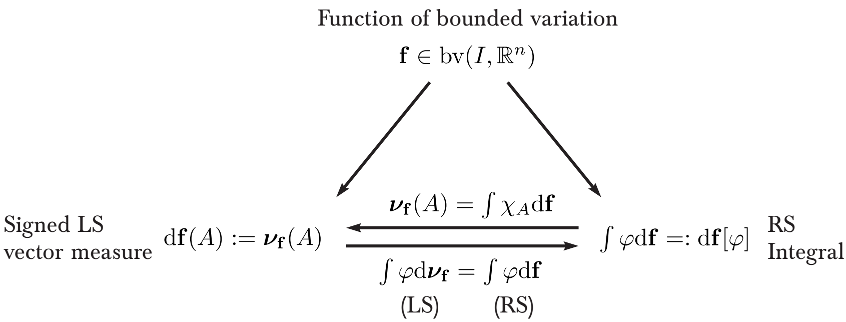

Figure A.1: An overview over the most important relations. LS stands for Lebesgue-Stieltjes and RS stands for Riemann-Stieltjes.

B

Some tools and proofs

B.1 Schur complement and passivity LMIs

The negative (semi-)definiteness of a matrix is preserved under congruence transformations. To see this, let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a square, positive definite matrix, $\mathbf{T} \in \mathbb{R}^{n \times n}$ be an invertible matrix and $\mathbf{B} = \mathbf{T}^\top \mathbf{A} \mathbf{T}$. Then it follows from $\mathbf{A} \prec 0$ that $\mathbf{x}^\top (\mathbf{A} + \mathbf{A}^\top)\mathbf{x} < 0$ for all $\mathbf{x} \in \mathbb{R}^n$, which implies that

$$\begin{aligned} \mathbf{x}^\top (\mathbf{B} + \mathbf{B}^\top)\mathbf{x} &= \mathbf{x}^\top (\mathbf{T}^\top \mathbf{A} \mathbf{T} + \mathbf{T}^\top \mathbf{A}^\top \mathbf{T})\mathbf{x} \\ &= \mathbf{x}^\top \mathbf{T}^\top (\mathbf{A} + \mathbf{A}^\top) \mathbf{T} \mathbf{x} \\ &< 0 \ \forall \ \mathbf{x} \in \mathbb{R}^n. \end{aligned} \tag{B.1}$$

Therefore, $\mathbf{A} \prec 0$ implies $\mathbf{B} \prec 0$. Furthermore, since $\mathbf{T}$ is invertible, for every $\mathbf{y} \in \mathbb{R}^n$ there exists an $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{y} = \mathbf{T}\mathbf{x}$. Therefore it follows from (B.1) that $\mathbf{B} \prec 0$ implies $\mathbf{A} \prec 0$. These facts lead to the following.

Lemma B.1. (Schur complement) Let $\mathbf{A}$ be a symmetric partitioned matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^\top & \mathbf{A}_{22} \end{pmatrix},$$

with $\mathbf{A}_{22}$ square and invertible. Then the following conditions hold:

1. $\mathbf{A} \prec 0$ if and only if $\mathbf{A}_{22} \prec 0$ and $\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^\top \prec 0$
2. $\mathbf{A} \preceq 0$ if and only if $\mathbf{A}_{22} \prec 0$ and $\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^\top \preceq 0$.

The expression $\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^{\top}$ is referred to as **Schur complement** of the block $\mathbf{A}_{22}$.

Proof. The conditions 1. and 2. directly follow from (B.1) with

$$\mathbf{T} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^{\top} & \mathbf{I} \end{pmatrix},$$

where $\mathbf{I}$ and $\mathbf{0}$ are an identity matrix and a zero matrix of appropriate dimensions. □

Lemma (B.1) leads to the following consequence, which can be applied to passivity based LMIs.

Corollary B.2. For given matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ and $\alpha \in \mathbb{R}$, the following matrix inequalities are equivalent, if all blocks in the partitioned matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ are symmetric.

1. $\mathbf{H}_1 := \begin{pmatrix} \mathbf{A}^{\top}\mathbf{P}\mathbf{A} - \alpha\mathbf{P} & \mathbf{A}^{\top}\mathbf{P}\mathbf{B} - \mathbf{C}^{\top} \\ \mathbf{B}^{\top}\mathbf{P}\mathbf{A} - \mathbf{C} & \mathbf{B}^{\top}\mathbf{P}\mathbf{B} - (\mathbf{D} + \mathbf{D}^{\top}) \end{pmatrix} \preceq 0 \quad \text{and} \quad \mathbf{P} \succ 0$

2. $\mathbf{H}_2 := \begin{pmatrix} -\alpha\mathbf{P} & -\mathbf{C}^{\top} & \mathbf{A}^{\top}\mathbf{P} \\ -\mathbf{C} & -(\mathbf{D} + \mathbf{D}^{\top}) & \mathbf{B}^{\top}\mathbf{P} \\ \mathbf{P}\mathbf{A} & \mathbf{P}\mathbf{B} & -\mathbf{P} \end{pmatrix} \preceq 0$

Proof. The proposition directly follows from applying condition 2. of Lemma B.1 to $\mathbf{H}_2$. Indeed, the Schur complement of the lower right block $-\mathbf{P}$ in $\mathbf{H}_2$ equates to

$$\begin{aligned} &\begin{pmatrix} -\alpha\mathbf{P} & -\mathbf{C}^{\top} \\ -\mathbf{C} & -(\mathbf{D} + \mathbf{D}^{\top}) \end{pmatrix} - \begin{pmatrix} \mathbf{A}^{\top}\mathbf{P} \\ \mathbf{A}^{\top}\mathbf{B} \end{pmatrix} (-\mathbf{P})^{-1} \begin{pmatrix} \mathbf{P}\mathbf{A} & \mathbf{P}\mathbf{B} \end{pmatrix} \\ &= \begin{pmatrix} -\alpha\mathbf{P} & -\mathbf{C}^{\top} \\ -\mathbf{C} & -(\mathbf{D} + \mathbf{D}^{\top}) \end{pmatrix} + \begin{pmatrix} \mathbf{A}^{\top}\mathbf{P}\mathbf{A} & \mathbf{A}^{\top}\mathbf{P}\mathbf{B} \\ \mathbf{B}^{\top}\mathbf{P}\mathbf{A} & \mathbf{B}^{\top}\mathbf{P}\mathbf{B} \end{pmatrix} \\ &= \mathbf{H}_1. \end{aligned}$$

Therefore, the equivalence of 1. and 2. in Corollary B.2 follows. □

B.2 Non-opening contacts with Coulomb friction and a known normal contact force

In Section 5.1, it was shown that the error dynamics of a state observer with switched unilateral constraints is stable. This was achieved by using the quadratic Lyapunov function $V = \frac{1}{2}(\tilde{\mathbf{q}}^\mathsf{T}\mathbf{K}\tilde{\mathbf{q}} + \tilde{\mathbf{u}}^\mathsf{T}\mathbf{M}\tilde{\mathbf{u}})$ and verifying that $\mathrm{d}V \leq 0$. What is left to show, is that the error dynamics is attractively stable, which requires the following steps that are adapted from Baumann [8, p. 62] to include Coulomb friction forces.

1. The first step is to show that $\lim_{t\to\infty} \tilde{\mathbf{q}}$ is constant and $\lim_{t\to\infty} \tilde{\mathbf{u}} = \mathbf{0}$, which is unaffected by Coulomb friction and therefore is identical to the proof in [8]. In summary: Since V is bounded from below and non-increasing, it converges to a limit $V_\infty := \lim_{t\to\infty} V(\tilde{\mathbf{q}}(t), \tilde{\mathbf{u}}(t)) \leq V(\tilde{\mathbf{q}}(t_0), \tilde{\mathbf{u}}(t_0))$. Next, since $\dot{V} \leq -\tilde{\mathbf{u}}^\mathsf{T}\mathbf{D}\tilde{\mathbf{u}}$ and $V^+ - V^- \leq 0$, it follows that $V_\infty - V(\tilde{\mathbf{q}}(t_0), \tilde{\mathbf{u}}(t_0)) \leq \lim_{t\to\infty} -\int_{t_0}^t \tilde{\mathbf{u}}(\tau)^\mathsf{T}\mathbf{D}\tilde{\mathbf{u}}(\tau)\,\mathrm{d}\tau$. The left side of this inequality is finite, which implies that $\lim_{t\to\infty} \int_{t_0}^t \tilde{\mathbf{u}}(\tau)^\mathsf{T}\mathbf{D}\tilde{\mathbf{u}}(\tau)\,\mathrm{d}\tau$ is finite as well, and $\lim_{t\to\infty} \int_{t_0}^t \tilde{\mathbf{u}}(\tau)^\mathsf{T}\mathbf{M}\tilde{\mathbf{u}}(\tau)\mathrm{d}\tau < \infty$ due to the equivalence of norms and $\mathbf{D} \succ 0$, $\mathbf{M} \succ 0$. Furthermore, the term $\tilde{\mathbf{u}}^\mathsf{T}\mathbf{M}\tilde{\mathbf{u}}\,\mathrm{d}\tau$ tends to an absolutely continuous function (because V_∞ is constant and $\tilde{\mathbf{e}}$ is absolutely continuous). Therefore, by an extension of the Lemma of Barbalat (see [8, Proposition A.3]), $\lim_{t\to\infty} \tilde{\mathbf{u}}(t)^\mathsf{T}\mathbf{M}\tilde{\mathbf{u}}(t) = 0$, which implies

$$\lim_{t\to\infty} \tilde{\mathbf{u}}(t) = \mathbf{0} \quad \text{and} \quad \lim_{t\to\infty} \tilde{\mathbf{q}}(t) = \mathbf{c} \tag{B.2}$$

with a constant $\mathbf{c}$.

2. Since $\lim_{t\to\infty} \tilde{\mathbf{u}} = \mathbf{0}$, it follows that the contact velocity error in the frictional contacts tends to zero as well, i.e. $\lim_{t\to\infty} \tilde{\boldsymbol{\gamma}}_T^i = \lim_{t\to\infty} \mathbf{W}_T^i\tilde{\mathbf{u}} = \mathbf{0}$ for all i. Therefore, the corresponding friction force error $\boldsymbol{\lambda}_T^i - \hat{\boldsymbol{\lambda}}_T^i = \mathbf{0}$ vanishes whenever $\boldsymbol{\gamma}_T^i \neq \mathbf{0}$ (the frictional contact is in slip), because the normal forces of the frictional contacts in the observer and the observed system are assumed to be identical).

3. Integrating the equality of measures of the error dynamics

$$\mathbf{M}\,\mathrm{d}\tilde{\mathbf{u}} + (\mathbf{C}\tilde{\mathbf{u}} + \mathbf{K}\tilde{\mathbf{q}})\,\mathrm{d}t = \mathbf{W}_N(\mathrm{d}\mathbf{P}_N - \mathrm{d}\hat{\mathbf{P}}_N) + \mathbf{W}_T(\mathrm{d}\mathbf{P}_T - \mathrm{d}\hat{\mathbf{P}}_T) \tag{B.3}$$

over a time interval $[t, t+\Delta t]$, taking the limit $t \to \infty$, and inserting (B.2) results in

$$\begin{aligned}\Delta t\mathbf{K}\mathbf{c} = \lim_{t\to\infty} \mathbf{W}_N \int_{[t,t+\Delta t]} (\mathrm{d}\mathbf{P}_N - \mathrm{d}\hat{\mathbf{P}}_N) \\ + \lim_{t\to\infty} \mathbf{W}_T \int_{[t,t+\Delta t]} (\mathrm{d}\mathbf{P}_T - \mathrm{d}\hat{\mathbf{P}}_T).\end{aligned} \tag{B.4}$$

Next, since all generalized force directions, i.e. all columns of $(\mathbf{W}_N \;\; \mathbf{W}_T)$ are assumed to be linearly independent, one can conclude that the limits on the right side of (B.4) exist. Since the unilateral constraints are assumed to never be permanently closed, there exists for every t and every constraint j an interval $[t_j, t_j + \Delta t]$ with $t < t_j$, such that $\lambda_N^j = \hat{\lambda}_N^j = 0$ and $\Lambda_N^j = \hat{\Lambda}_N^j = 0$ in that interval (if Δt is chosen small enough). Similarly, since the frictional contacts are never permanently in stick, there exists for every t and every friction contact i an interval $[t_i, t_i + \Delta t]$ with $t < t_i$, in which $\boldsymbol{\lambda}_T^i = \hat{\boldsymbol{\lambda}}_T^i = \mathbf{0}$ (if Δt is chosen small enough). Therefore, one can conclude that the limits on the right side of (B.4) vanish, such that $\Delta t \mathbf{K} \mathbf{c} = \mathbf{0}$. Finally, since $\Delta t \mathbf{K}$ is positive definite, $\tilde{\mathbf{q}} = \mathbf{c} = \mathbf{0}$. $\square$

Bibliography

[1] V. Acary and B. Brogliato. *Numerical Methods for Nonsmooth Dynamical Systems; Applications in Mechanics and Electronics*, volume 35 of *Lecture Notes in Applied and Computational Mechanics.* Springer-Verlag, Berlin, 2008.

[2] A. Alessandri and P. Coletta. Design of Luenberger observers for a class of hybrid linear systems. In M. D. Di Benedetto and A. Sangiovanni-Vincentelli, editors, *Hybrid Systems: Computation and Control*, pages 7–18, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

[3] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems.* Oxford University Press, New York, 2000.

[4] M. Arcak and P. Kokotovic. Observer-based control of systems with slope-restricted nonlinearities. *IEEE Transactions on Automatic Control*, 46(7):1146–1150, 2001.

[5] J.-P. Aubin and A. Cellina. *Differential Inclusions.* Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg, 1984.

[6] J.-P. Aubin and I. Ekeland. *Applied Nonlinear Analysis.* John Wiley & Sons, New York, 1984.

[7] P. Ballard. The dynamics of discrete mechanical systems with perfect unilateral constraints. *Archive for Rational Mechanics and Analysis*, 154(3):199–274, 2000.

[8] M. Baumann. *Synchronization of Nonsmooth Mechanical Systems with Impulsive Motion.* PhD thesis, ETH Zurich, 2016.

[9] M. Baumann, J. J. B. Biemond, R. I. Leine, and N. van de Wouw. Synchronization of impacting mechanical systems with a single constraint. *Physica D: Nonlinear Phenomena*, 362:9 – 23, 2018.

[10] M. Baumann and R. I. Leine. Synchronization-based state observer for impacting multibody systems using switched geometric unilateral constraints. In *Proceedings*

of the ECCOMAS Thematic Conference on Multibody Dynamics (ECCOMAS 2015), Barcelona, Spain, 2015.

[11] M. Baumann and R. I. Leine. A synchronization-based state observer for impact oscillators using only collision time information. *International Journal of Robust and Nonlinear Control*, 26(12):2542–2563, 2016.

[12] A. Bemporad and M. Morari. Control of systems integrating logic, dynamics, and constraints. *Automatica*, 35(3):407–427, 1999.

[13] P. Bernard and R. G. Sanfelice. A local hybrid observer for a class of hybrid dynamical systems with linear maps and unknown jump times. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 5362–5367, 2021.

[14] G. Besançon. An overview on observer tools for nonlinear systems. In *Nonlinear Observers and Applications*, number 363 in Lecture Notes in Control and Information Sciences. Springer, Berlin, Heidelberg, 2007.

[15] G. Besançon, G. Bornard, and H. Hammouri. Observer synthesis for a class of nonlinear control systems. *European Journal of Control*, 2(3):176–192, 1996.

[16] J. J. B. Biemond. *Nonsmooth dynamical systems: on stability of hybrid trajectories and bifurcations of discontinuous systems.* PhD thesis, KU Leuven, Mar. 2013. Nijmeijer, H. (supervisor), van de Wouw, N. (cosupervisor).

[17] J. J. B. Biemond, W. P. M. H. Heemels, R. G. Sanfelice, and N. van de Wouw. Distance function design and Lyapunov techniques for the stability of hybrid trajectories. *Automatica*, 73:38–46, 2016.

[18] J. J. B. Biemond, R. Postoyan, W. P. M. H. Heemels, and N. van de Wouw. Incremental stability of hybrid dynamical systems. *IEEE Transactions on Automatic Control*, 63(12):4094–4109, 2018.

[19] J. J. B. Biemond, N. van de Wouw, W. P. M. H. Heemels, and H. Nijmeijer. Tracking control for hybrid systems with state-triggered jumps. *IEEE Transactions on Automatic Control*, 58(4):876–890, 2013.

[20] J. Birk and M. Zeitz. Extended Luenberger observer for non-linear multivariable systems. *International Journal of Control*, 47(6):1823–1836, 1988.

[21] H. Bremer. *Dynamik und Regelung mechanischer Systeme.* Teubner, Stuttgart, 1988.

[22] B. Brogliato. *Nonsmooth Mechanics. Models, Dynamics and Control.* Communications and Control Engineering. Springer-Verlag, International Publishing, 3rd edition, 2016.

[23] B. Brogliato and W. P. M. H. Heemels. Observer design for Lur'e systems with multivalued mappings: a passivity approach. *IEEE Transactions on Automatic Control*, 54(8):1996–2001, 2009.

[24] B. Brogliato, R. Lozano, B. Maschke, and O. Egeland. *Dissipative Systems Analysis and Control: Theory and Applications.* Springer-Verlag, Cham, 3rd edition, 2020.

[25] C. I. Byrnes, A. Isidori, and J. C. Willems. Passivity, feedback equivalence, and the global stabilization of minimum phase nonlinear systems. *IEEE Transactions on Automatic Control*, 36(11):1228–1240, 1991.

[26] C. I. Byrnes and Wei Lin. Losslessness, feedback equivalence, and the global stabilization of discrete-time nonlinear systems. *IEEE Transactions on Automatic Control*, 39(1):83–98, 1994.

[27] G. Capobianco. *Mechanical systems with frictional contact: Geometric theory and time discretization methods.* PhD thesis, University of Stuttgart, 2021.

[28] G. Capobianco and S. R. Eugster. Time finite element based Moreau-type integrators. *International Journal for Numerical Methods in Engineering*, 114(3):215–231, 2018.

[29] V. Chellaboina and W. Haddad. A unification between partial stability and stability theory for time-varying systems. *IEEE Control Systems Magazine*, 22(6):66–75, 2002.

[30] C. T. Chen. *Linear System Theory and Design.* The Oxford Series in Electrical and Computer Engineering. Oxford University Press, 3 edition, 2014.

[31] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and R. R. Wolenski. *Nonsmooth Analysis and Control Theory*, volume 178 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1998.

[32] J. Cortes. Discontinuous dynamical systems. *IEEE Control Systems Magazine*, 28(3):36–73, 2008.

[33] R. W. Cottle, J. S. Pang, and R. E. Stone. *The Linear Complementarity Problem.* Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), 1992.

[34] B. De Schutter and B. De Moor. The extended linear complementarity problem and the modeling and analysis of hybrid systems. In P. Antsaklis, M. Lemmon, W. Kohn, A. Nerode, and S. Sastry, editors, *Hybrid Systems V*, pages 70–85, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.

[35] N. Dinculeanu. *Vector Measures.* Pergamon, 1967.

[36] A. Doris, A. L. Juloski, N. Mihajlovic, W. P. M. H. Heemels, N. van de Wouw, and H. Nijmeijer. Observer designs for experimental non-smooth and discontinuous systems. *IEEE Transactions on Control Systems Technology*, 16(6):1323–1332, 2008.

[37] J. Elstrodt. *Maß- und Integrationstheorie.* Springer-Verlag, Berlin, Heidelberg, New York, 2 edition, 1999.

[38] A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides.* Kluwer Academic, Dordrecht, 1988.

[39] D. Fiore, M. Coraggio, and M. di Bernardo. Observer design for piecewise smooth and switched systems via contraction theory. *IFAC-PapersOnLine*, 50(1):2959–2964, 2017. 20th IFAC World Congress.

[40] Ch. Glocker. On frictionless impact models in rigid-body systems. *Philosophical Transactions of the Royal Society of London A*, 359:2385–2404, 2001.

[41] Ch. Glocker. *Set-valued Force Laws, Dynamics of Non-smooth Systems*, volume 1 of *Lecture Notes in Applied Mechanics.* Springer-Verlag, 2001.

[42] Ch. Glocker. An introduction to impacts. In *Nonsmooth mechanics of solids*, pages 45–101. Springer-Verlag, 2006.

[43] Ch. 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.

[44] Ch. Glocker. Energetic consistency conditions for standard impacts. Part II: Poisson-type inequality impact laws. *Multibody System Dynamics*, 32(4):1–65, 2013.

[45] R. Goebel, R. G. Sanfelice, and A. R. Teel. *Hybrid Dynamical Systems: Modeling, Stability, and Robustness.* Princeton University Press, Princeton, 2012.

[46] M. S. Gowda and J.-S. Pang. Stability analysis of variational inequalities and nonlinear complementarity problems, via the mixed linear complementarity problem and degree theory. *Mathematics of Operations Research*, 19(4):831–879, 1994.

[47] J. Gärtner. *Mass- und Integrationstheorie.* Vorlesungsskript, 2008.

[48] W. M. Haddad, V. Chellaboina, and S. G. Nersesov. *Impulsive and Hybrid Dynamical Systems: Stability, Dissipativity, and Control.* Princeton Series in Applied Mathematics. Princeton University Press, Princeton, 2006.

[49] H. Hammouri, B. Targui, and F. Armanet. High gain observer based on a triangular structure. *International Journal of Robust and Nonlinear Control*, 12(6):497–518, 2002.

[50] W. P. M. H. Heemels and B. Brogliato. The complementarity class of hybrid dynamical systems. *European Journal of Control*, 9(2-3):322–360, 2003.

[51] W. P. M. H. Heemels, M. K. Camlibel, J. M. Schumacher, and B. Brogliato. Observer-based control of linear complementarity systems. *International Journal of Robust and Nonlinear Control*, 21(10):1193–1218, 2011.

[52] W. P. M. H. Heemels, B. De Schutter, and A. Bemporad. Equivalence of hybrid dynamical models. *Automatica*, 37(7):1085–1091, 2001.

[53] W. P. M. H. Heemels, Heemels, J. Schumacher, and S. Weiland. Linear complementarity systems. *SIAM J. Appl. Math*, 60:2000, 1997.

[54] A. P. Ivanov. Analytical methods in the theory of vibro-impact systems. *Journal of Applied Mathematics and Mechanics*, 57(2):221–236, 1993.

[55] M. Jean. The non smooth contact dynamics method. *Computer Methods in Applied Mechanics and Engineering*, 177:235–257, 1999.

[56] M. Jean and J. J. Moreau. Unilaterality and dry friction in the dynamics of rigid body collections. In *Proceedings of the Contact Mechanics International Symposium*, pages 31–48, Lausanne, 1992. Presses Polytechniques et Universitaires Romandes.

[57] A. Juloski, W. P. M. H. Heemels, and S. Weiland. Observer design for a class of piecewise linear systems. *International Journal of Robust and Nonlinear Control*, 17(15):1387 – 1404, Oct. 2007.

[58] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, 82:35 – 45, 1960.

[59] R. E. Kalman and R. S. Bucy. New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering*, 83(1):95–108, 03 1961.

[60] H. K. Khalil. *Nonlinear Systems.* Prentice Hall, third edition, 2002.

[61] H. K. Khalil and L. Praly. High-gain observers in nonlinear feedback control. *International Journal of Robust and Nonlinear Control*, 24(6):993–1015, 2014.

[62] J. Kim, H. Cho, A. Shamsuarov, H. Shim, and J. H. Seo. State estimation strategy without jump detection for hybrid systems using gluing function. In *53rd IEEE Conference on Decision and Control*, pages 139–144, Dec 2014.

[63] J. Kim, H. Shim, and J. H. Seo. State estimation and tracking control for hybrid systems by gluing the domains. *IEEE Transactions on Automatic Control*, 64:3026–3033, 2019.

[64] A. J. Krener and A. Isidori. Linearization by output injection and nonlinear observers. *Systems & Control Letters*, 3(1):47–52, 1983.

[65] S. Lang. *Real and Functional Analysis.* Springer, 1993.

[66] R. I. Leine. The historical development of classical stability concepts: Lagrange, Poisson and Lyapunov stability. *Nonlinear Dynamics*, 59:173–182, 2010.

[67] 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 Verlag, Berlin, 2004.

[68] R. I. Leine and N. van de Wouw. *Stability and Convergence of Mechanical Systems with Unilateral Constraints.* Lecture Notes in Applied and Computational Mechanics. Springer, Germany, 2008.

[69] D. Liberzon. *Switching in Systems and Control.* Birkhäuser, Boston, 2003.

[70] D. Luenberger. An introduction to observers. *IEEE Transactions on Automatic Control*, 16(6):596–602, 1971.

[71] D. G. Luenberger. Observers for multivariable systems. *IEEE Transactions on Automatic Control*, 11(2):190 – 197, 1966.

[72] D. d. S. Madeira and J. Adamy. On the equivalence between strict positive realness and strict passivity of linear systems. *IEEE Transactions on Automatic Control*, 61(10):3091–3095, 2016.

[73] F. Martinelli, L. Menini, and A. Tornambè. Observability, reconstructibility and observer design for linear mechanical systems unobservable in absence of impacts. *Journal of Dynamic Systems, Measurement, and Control*, 125(4):549–562, 2004.

[74] L. Menini and A. Tornambè. Velocity observers for linear mechanical systems subject to single non-smooth impacts. *Systems & Control Letters*, 43(3):193–202, 2001.

[75] L. Menini and A. Tornambè. Velocity observers for non-linear mechanical systems subject to non-smooth impacts. *Automatica*, 38(12):2169–2175, 2002.

[76] L. Menini and A. Tornambè. State immersion observers for mechanical systems with impacts. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 7117–7122, Dec 2016.

[77] M. D. P. Monteiro Marques. *Differential Inclusions in Nonsmooth Mechanical Problems. Shocks and Dry Friction.*, volume 9. Birkhäuser, Basel, 1993.

[78] J. J. Moreau. Bounded variation in time. In J. J. Moreau, P. D. Panagiotopoulos, and G. Strang, editors, *Topics in Nonsmooth Mechanics*, pages 1–74. Birkhäuser, Basel, Boston, Berlin, 1988.

[79] J. J. Moreau. Unilateral contact and dry friction in finite freedom dynamics. In *Non-Smooth Mechanics and Applications*, pages 1–82. Springer-Verlag, 1988.

[80] K. Murty. *Linear Complementarity, Linear and Nonlinear Programming.* Sigma Series in Applied Mathematics. Heldermann, 1988.

[81] A. Nemirovskii and P. Gahinet. The projective method for solving linear matrix inequalities. In *Proceedings of 1994 American Control Conference - ACC '94*, volume 1, pages 840–844 vol.1, 1994.

[82] V. R. Nosov and V. D. Furasov. Stability of discrete processes with respect to specified variables, and the convergence of some optimization algorithms. *USSR Computational Mathematics and Mathematical Physics*, 19(2):44–58, 1979 (in Russian).

[83] J. O'Reilly. *Observers for Linear Systems.* Academic Press, London, 1983.

[84] H. B. Oza, Y. V. Orlov, and S. K. Spurgeon. Finite time stabilization of a perturbed double integrator with unilateral constraints. *Mathematics and Computers in Simulation*, 95:200–212, 2014. Discontinuous Differential Systems : Theory and Numerical Methods.

[85] L. Paoli. Mathematical aspects of vibro-impact problems. In R. I. Leine, V. Acary, and O. Brüls, editors, *Advanced Topics in Nonsmooth Dynamics: Transactions of the European Network for Nonsmooth Dynamics*, pages 135–189. Springer International Publishing, Cham, 2018.

[86] L. Paoli and M. Schatzman. A Numerical Scheme for Impact Problems I: The One-Dimensional Case. *SIAM Journal on Numerical Analysis*, 40(2):702 – 733, Jan. 2002.

[87] F. Pfeiffer and Ch. Glocker. *Multibody Dynamics with Unilateral Contacts.* Wiley, New York, 1996.

[88] P. V. Preiswerk and R. I. Leine. Experimental performance verification of a synchronization based state observer using only collision time information. In *Proceedings of the ASME 2018 International Design Engineering Technical Conferences (IDETC)*, 2018.

[89] P. V. Preiswerk and R. I. Leine. A nonsmooth dynamics perspective on state observer design for mechanical systems with unilateral constraints. In *Proceedings of the 10th ECCOMAS Thematic Conference on Multibody Dynamics*, pages 48–54, 2021.

[90] P. V. Preiswerk and R. I. Leine. State observers for the time discretization of a class of impulsive mechanical systems. *International Journal of Robust and Nonlinear Control*, 2022.

[91] M. Rijnen. *Enabling motions with impacts in robotic and mechatronic systems.* PhD thesis, Technische Universiteit Eindhoven, 2018.

[92] M. Rijnen, A. Saccon, and H. Nijmeijer. Motion signals with velocity jumps: Velocity estimation employing only quantized position data. *IEEE Robotics and Automation Letters*, 3(3):1498–1505, 2018.

[93] R. T. Rockafellar. *Convex Analysis.* Princeton Landmarks in Mathematics. Princeton University Press, Princeton, New Jersey, 1970.

[94] R. G. Sanfelice, J. J. B. Biemond, N. van de Wouw, and W. P. M. H. Heemels. An embedding approach for the design of state-feedback tracking controllers for references with jumps. *International Journal of Robust and Nonlinear Control*, 24(11):1585–1608, 2014.

[95] D. Simon. *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches.* John Wiley & Sons, Inc., 2006.

[96] J. S. Spraker. A comparison of the Carathéodory and Filippov solution sets. *Journal of Mathematical Analysis and Applications*, 198(2):571–580, 1996.

[97] A. Tanwani, B. Brogliato, and C. Prieur. Stability and observer design for Lur'e systems with multivalued, nonmonotone, time-varying nonlinearities and state jumps. *SIAM Journal on Control and Optimization*, 52(6):3639–3672, 2014.

[98] A. Tanwani, B. Brogliato, and C. Prieur. Observer design for unilaterally constrained Lagrangian systems: A passivity-based approach. *IEEE Transactions on Automatic Control*, 61(9):2386–2401, 2016.

[99] N. van de Wouw and A. V. Pavlov. Tracking and synchronisation for a class of PWA systems. *Automatica*, 44(11):2909–2915, 2008.

[100] V. I. Vorotnikov. *Partial Stability and Control.* Birkhäuser, 1998.

[101] V. I. Vorotnikov. Partial stability and control: The state-of-the-art and development prospects. *Automation Remote Control*, 66:511–561, 2005.

[102] T. Winandy, M. Baumann, and R. I. Leine. Variational analysis of inequality impact laws for perfect unilateral constraints. In R. I. Leine, V. Acary, and O. Brüls, editors, *Advanced Topics in Nonsmooth Dynamics: Transactions of the European Network for Nonsmooth Dynamics*, pages 47–92. Springer International Publishing, Cham, 2018.

[103] J. Zhang, X. Zhao, F. Zhu, and H. R. Karimi. Reduced-order observer design for switched descriptor systems with unknown inputs. *IEEE Transactions on Automatic Control*, 65(1):287–294, 2020.

[104] V. F. Zhuravlev. Equations of motion of mechanical systems with ideal onesided links. *Journal of Applied Mathematics and Mechanics*, 42(5):839–847, 1978.