Differential-Algebraic Equations Forum

DAE-F

Achim Ilchmann
Timo Reis *Editors*

Surveys in Differential-Algebraic Equations I

Springer

Differential-Algebraic Equations Forum

Differential-Algebraic Equations Forum

The series "Differential-Algebraic Equations Forum" is concerned with analytical, algebraic, control theoretic and numerical aspects of differential algebraic equations (DAEs) as well as their applications in science and engineering. It is aimed to contain survey and mathematically rigorous articles, research monographs and textbooks. Proposals are assigned to an Associate Editor, who recommends publication on the basis of a detailed and careful evaluation by at least two referees. The appraisals will be based on the substance and quality of the exposition.

For further volumes:
www.springer.com/series/11221

Achim Ilchmann • Timo Reis
Editors

Surveys in Differential-Algebraic Equations I

Springer

Editors
Achim Ilchmann
Institut für Mathematik
Technische Universität Ilmenau
Ilmenau, Germany

Timo Reis
Fachbereich Mathematik
Universität Hamburg
Hamburg, Germany

ISBN 978-3-642-34927-0 ISBN 978-3-642-34928-7 (eBook)
DOI 10.1007/978-3-642-34928-7
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013935149

Mathematics Subject Classification (2010): 34A08, 65L80, 93B05, 93D09

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

We are pleased to present the first of three volumes of survey articles in various fields of differential-algebraic equations (DAEs). In the last two decades, there has been a substantial research activity in the theory, applications, and computations of DAEs; our aim is to give an almost complete picture of these latest developments.

What are DAEs? They certainly belong to differential equations, but the terminology is not clear. In their most general form, DAEs are implicit differential equations. However, this is still too wide and in view of linearizations and the fact that most research is on linear DAEs, one uses the more narrow notion of differential-algebraic systems. This fact is reflected in the Mathematics Subject Classification (MSC 2010), which is a taxonomy on a first, second, and third level (2-, 3-, and 5-digit class, respectively). DAEs are mentioned twice on level three: *34 Ordinary differential equations, 34A General theory, 34A09 Implicit equations, differential-algebraic equations*, and *65 Numerical analysis, 65L Ordinary differential equations, 65L80 Methods for differential-algebraic equations*.

What is the history of DAEs? Although DAEs can be traced back earlier, it was not until the 1960s that mathematicians and engineers started to thoroughly study computational issues, mathematical theory, and applications of DAEs. There are many relationships with mathematical disciplines such as differential geometry, algebra, functional analysis, numerical analysis, stochastics, and control theory, to mention but a few; and there are extensive applications in electric circuit theory, chemical processes, constrained mechanics, as well as in economics. In addition to the intrinsic mathematical interest, there are two fundamental reasons for these advances: first, automatic modeling, which results in large dimensional DAEs, and second the advancement of computers and hence the feasibility of solving problems numerically. In quantitative terms, this development has lead to more than 1500 journal and conference papers on DAEs each year.

Is a level two rank, instead of the current level three, for DAEs appropriate? The MSC tries to rank the different levels hierarchically. However, terminological unities for different fields, so that they can be accurately separated from each other, do not necessarily exist. Moreover, fields and their importance vary in time: new fields arise, others become less important. One could imagine that DAEs are equally im-

portant as, for example, *34B Boundary value problems, 34G Differential equations in abstract spaces, 34K Functional-differential and differential-difference equations, 34L Ordinary differential operators*, to name but a few within the 34 ODEs class.

The immense number of papers on DAEs is certainly not a sufficient reason for any taxonomy, and the underlying methods in DAEs are very distinct: differential geometry, distributions, and linear algebra. But today's changing importance and relevance of DAEs have been shown by about ten research monographs in fields of DAEs in the last decade and, most importantly, recently the first textbooks on the mathematical theory of DAEs have been written. This may indicate a turning point: DAEs are becoming a field in their own right, beside other fields in ordinary differential equations.

The collection of survey articles in DAEs presented in the upcoming three volumes will include the topics

– Linear systems
– Nonlinear systems
– Solution theory
– Stability theory
– Control theory
– Model reduction
– Analytical methods
– Differential geometric methods
– Algebraic methods
– Numerical methods
– Coupled problems with partial differential equations
– Stochastic DAEs
– Chemical engineering
– Circuit modelling
– Mechanical engineering

This may show the depth and width of the recent progress in differential-algebraic equations and will possibly underpin the fact that differential-algebraic equations are in a state where they are no longer only a collection of results on the same topic, but a field within the class of ordinary differential equations.

Ilmenau, Germany Achim Ilchmann
Hamburg, Germany Timo Reis

Contents

Controllability of Linear Differential-Algebraic Systems—A Survey

Thomas Berger and Timo Reis

Abstract Different concepts related to controllability of differential-algebraic equations are described. The class of systems considered consists of linear differential-algebraic equations with constant coefficients. Regularity, which is, loosely speaking, a concept related to existence and uniqueness of solutions for any inhomogeneity, is not required in this article. The concepts of impulse controllability, controllability at infinity, behavioral controllability, and strong and complete controllability are described and defined in the time domain. Equivalent criteria that generalize the Hautus test are presented and proved.

Special emphasis is placed on normal forms under state space transformation and, further, under state space, input and feedback transformations. Special forms generalizing the Kalman decomposition and Brunovský form are presented. Consequences for state feedback design and geometric interpretation of the space of reachable states in terms of invariant subspaces are proved.

Keywords Differential-algebraic equations · Controllability · Stabilizability · Kalman decomposition · Canonical form · Feedback · Hautus criterion · Invariant subspaces

Mathematics Subject Classification (2010) 34A09 · 15A22 · 93B05 · 15A21 · 93B25 · 93B27 · 93B52

Thomas Berger was supported by DFG grant IL 25/9 and partially supported by the DAAD.

T. Berger (✉)
Institut für Mathematik, Technische Universität Ilmenau, Weimarer Straße 25, 98693 Ilmenau, Germany
e-mail: thomas.berger@tu-ilmenau.de

T. Reis
Fachbereich Mathematik, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany
e-mail: timo.reis@math.uni-hamburg.de

A. Ilchmann, T. Reis (eds.), *Surveys in Differential-Algebraic Equations I*,
Differential-Algebraic Equations Forum,
DOI 10.1007/978-3-642-34928-7_1,

1 Introduction

Controllability is, roughly speaking, the property of a system that any two trajectories can be concatenated by another admissible trajectory. The precise concept, however, depends on the specific framework, as quite a number of different concepts of controllability are present today.

Since the famous work by Kalman [81–83], who introduced the notion of controllability about 50 years ago, the field of mathematical control theory has been revived and rapidly growing ever since, emerging into an important area in applied mathematics, mainly due to its contributions to fields such as mechanical, electrical and chemical engineering (see e.g. [2, 47, 148]). For a good overview of standard mathematical control theory, i.e., involving ordinary differential equations (ODEs), and its history see e.g. [70, 76, 77, 80, 138, 142].

Just before mathematical control theory began to grow, Gantmacher published his famous book [60] and therewith laid the foundations for the rediscovery of differential-algebraic equations (DAEs), the first main theories of which have been developed by Weierstraß [158] and Kronecker [93] in terms of matrix pencils. DAEs have then been discovered to be appropriate for modeling a vast variety of problems in economics [111], demography [37], mechanical systems [7, 31, 59, 67, 127, 149], multibody dynamics [55, 67, 139, 141], electrical networks [7, 36, 54, 106, 117, 134, 135], fluid mechanics [7, 65, 106] and chemical engineering [48, 50–52, 126], which often cannot be modeled by standard ODE systems. Especially the tremendous effort in numerical analysis of DAEs [10, 96, 98] is responsible for DAEs being nowadays a powerful tool for modeling and simulation of the aforementioned dynamical processes.

In general, DAEs are implicit differential equations, and in the simplest case just a combination of differential equations along with algebraic constraints (from which the name DAE comes from). These algebraic constraints, however, may cause the solutions of initial value problems no longer to be unique, or solutions not to exist at all. Furthermore, when considering inhomogeneous problems, the inhomogeneity has to be "consistent" with the DAE in order for solutions to exist. Dealing with these problems a huge solution theory for DAEs has been developed, the most important contribution of which is the one by Wilkinson [159]. Nowadays, there are a lot of monographs [31, 37, 38, 49, 66, 98] and one textbook [96], where the whole theory can be looked up. A comprehensive representation of the solution theory of general linear time-invariant DAEs, along with possible distributional solutions based on the theory developed in [143, 144], is given in [22]. A good overview of DAE theory and a historical background can also be found in [99].

DAEs found its way into control theory ever since the famous book by Rosenbrock [136], in which he developed his ideas of the description of linear systems by polynomial system matrices. Then a rapid development followed with important contributions of Rosenbrock himself [137] and Luenberger [107–110], not to forget the work by Pugh et al. [131], Verghese et al. [151, 153–155], Pandolfi [124, 125], Cobb [42, 43, 45, 46], Yip et al. [169] and Bernard [27]. The most important of these contributions for the development of concepts of controllability are certainly [46, 155, 169]. Further developments were made by Lewis and

Özçaldiran [101, 102] and by Bender and Laub [19, 20]. The first monograph which summarizes the development of control theory for DAEs so far was the one by Dai [49]. All these contributions deal with regular systems, i.e., systems of the form

$$E\dot{x}(t) = Ax(t) + f(t), \qquad x(0) = x^0,$$

where for any inhomogeneity f there exist initial values x^0 for which the corresponding initial value problem has a solution and this solution is unique. This has been proved to be equivalent to the condition that E, A are square matrices and $\det(sE - A) \in \mathbb{R}[s] \setminus \{0\}$.

The aim of the present paper is to state the different concepts of controllability for differential-algebraic systems which are not necessarily regular, i.e., E and A may be non-square. Applications with the need for non-regular DAEs appear in the modeling of electrical circuits [54] for instance. Furthermore, a drawback in the consideration of regular systems arises when it comes to feedback: the class of regular DAE systems is not closed under the action of a feedback group [12]. This also rises the need for a complete and thorough investigation of non-regular DAE systems. We also like to stress that general, possibly non -regular, DAE systems are a subclass of the class of so-called differential behaviors, introduced by Polderman and Willems [128], see also [161]. In the present article we will pay a special attention to the behavioral setting, formulating most of the results and the concepts by using the underlying set of trajectories (behavior) of the system.

In this paper we do not treat controllability of time-varying DAEs, but refer to [40, 72–74, 156, 157]. We also do not treat controllability of discrete time DAEs, but refer to [13, 27, 99, 100, 168].

The paper is organized as follows.

2 Controllability Concepts, p. 5 The concepts of impulse controllability, controllability at infinity, R-controllability, controllability in the behavioral sense, strong and complete controllability, as well as strong and complete reachability and stabilizability in the behavioral sense, strong and complete stabilizability will be described and defined in the time domain in Sect. 2. In the more present DAE literature these notions are not consistently treated. We try to clarify this here. A comprehensive discussion of the introduced concepts as well as some first relations between them are also included in Sect. 2.

3 Solutions, Relations and Normal Forms, p. 15 In Sect. 3 we briefly revisit the solution theory of DAEs and then concentrate on normal forms under state space transformation and, further, under state space, input and feedback transformations. We introduce the concepts of system and feedback equivalence and state normal forms under these equivalences, which for instance generalize the Brunovský form. It is also discussed when these forms are canonical and what properties (regarding controllability and stabilizability) the appearing subsystems have.

4 Algebraic Criteria, p. 30 The generalized Brunovský form enables us to give short proofs of equivalent criteria, in particular generalizations of the Hautus test, for the controllability concepts in Sect. 4, the most of which are of course well-known—we discuss the relevant literature.

5 Feedback, Stability and Autonomous System p. 36 In Sect. 5 we revisit the concept of feedback for DAE systems and proof new results concerning the equivalence of stabilizability of DAE control systems and the existence of a feedback which stabilizes the closed-loop system.

6 Invariant Subspaces, p. 46 In Sect. 6 we give a brief summary of some selected results of the geometric theory using invariant subspaces which lead to a representation of the reachability space and criteria for controllability at infinity, impulse controllability, controllability in the behavioral sense, complete and strong controllability.

7 Kalman Decomposition, p. 50 Finally, in Sect. 7 the results regarding the Kalman decomposition for DAE systems are stated and it is shown how the controllability concepts can be related to certain properties of the Kalman decomposition.

We close the introduction with the nomenclature used in this paper:

$\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$	set of natural numbers, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, set of all integers, resp.
$\ell(\alpha)$, $\lvert\alpha\rvert$	length and absolute value of a multi-index $\alpha = (\alpha_1, \ldots, \alpha_l) \in \mathbb{N}^n$
$\mathbb{R}_{\geq 0}$ ($\mathbb{R}_{>0}$, $\mathbb{R}_{\leq 0}$, $\mathbb{R}_{<0}$)	$= [0, \infty)$ ($(0, \infty)$, $(-\infty, 0]$, $(-\infty, 0)$), resp.
$\mathbb{C}_+$, $\mathbb{C}_-$ ($\overline{\mathbb{C}}_+$, $\overline{\mathbb{C}}_-$)	the open (closed) set of complex numbers with positive, negative real part, resp.
$\mathbf{Gl}_n(\mathbb{R})$	the set of invertible real $n \times n$ matrices
$\mathbb{R}[s]$	the ring of polynomials with coefficients in $\mathbb{R}$
$\mathbb{R}(s)$	the quotient field of $\mathbb{R}[s]$
$R^{n,m}$	the set of $n \times m$ matrices with entries in a ring R
$\sigma(A)$	spectrum of the matrix $A \in \mathbb{R}^{n,n}$
$f\lvert_{\mathscr{I}}$	restriction of the function $f : \mathscr{T} \to \mathbb{R}^n$ to $\mathscr{I} \subseteq \mathscr{T}$,
$\mathscr{L}^1_{\mathrm{loc}}(\mathscr{T}; \mathbb{R}^n)$	locally Lebesgue integrable functions $f : \mathscr{T} \to \mathbb{R}^n$, see [1, Chap. 1]
$\dot{f}$ ($f^{(i)}$)	(ith) distributional derivative of $f \in \mathscr{L}^1_{\mathrm{loc}}(\mathscr{T}; \mathbb{R}^n)$, $i \in \mathbb{N}_0$
$\mathscr{W}^{k,1}_{\mathrm{loc}}(\mathscr{T}; \mathbb{R}^n)$	$:= \{ x \in \mathscr{L}^1_{\mathrm{loc}}(\mathscr{T}; \mathbb{R}^n) \lvert x^{(i)} \in \mathscr{L}^1_{\mathrm{loc}}(\mathscr{T}; \mathbb{R}^n)$ for $i = 0, \ldots, k\}$, $k \in \mathbb{N}_0$
σ_τ	the τ-shift operator, i.e., for $f : \mathscr{T} \to \mathbb{R}^n$, $\mathscr{T} \subseteq \mathbb{R}$, $\sigma_\tau f : \mathscr{T} - \tau \to \mathbb{R}^n$, $t \mapsto f(t + \tau)$
ρ	the reflection operator, i.e., for $f : \mathscr{T} \to \mathbb{R}^n$, $\mathscr{T} \subseteq \mathbb{R}$, $\rho f : -\mathscr{T} \to \mathbb{R}^n$, $t \mapsto f(-t)$

2 Controllability Concepts

We consider linear differential-algebraic control systems of the form

$$E\dot{x}(t) = Ax(t) + Bu(t), \tag{2.1}$$

with $E, A \in \mathbb{R}^{k,n}$, $B \in \mathbb{R}^{k,m}$; the set of these systems is denoted by $\Sigma_{k,n,m}$, and we write $[E, A, B] \in \Sigma_{k,n,m}$.

We do not assume that the pencil $sE - A \in \mathbb{R}[s]^{k,n}$ is regular, that is, $\operatorname{rk}_{\mathbb{R}(s)}(sE - A) = k = n$.

The function $u : \mathbb{R} \to \mathbb{R}^m$ is called *input*; $x : \mathbb{R} \to \mathbb{R}^n$ is called (*generalized*) *state*. Note that, strictly speaking, $x(t)$ is in general not a state in the sense that the free system (i.e., $u \equiv 0$) satisfies a semigroup property [89, Sect. 2.2]. We will, however, speak of the state $x(t)$ for sake of brevity, especially since $x(t)$ contains the full information about the system at time t. Furthermore, one might argue that (especially in the behavioral setting) it is not correct to call u "input", because due to the implicit nature of (2.1) it may be that actually some components of u are uniquely determined and some components of x are free, and only the free variables should be called inputs in the behavioral setting. However, the controllability concepts given in Definition 2.1 explicitly distinguish between x and u and not between free and determined variables. We feel that, in some cases, it might still be the choice of the designer to assign the input variables, that is, u, and if some of these are determined, then the input space has to be restricted in an appropriate way.

A trajectory $(x, u) : \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}^m$ is said to be a *solution* of (2.1) if, and only if, it belongs to the *behavior* of (2.1):

$$\mathfrak{B}_{[E,A,B]} := \left\{ (x,u) \in \mathscr{W}^{1,1}_{\mathrm{loc}}\big(\mathbb{R};\mathbb{R}^n\big) \times \mathscr{L}^1_{\mathrm{loc}}\big(\mathbb{R};\mathbb{R}^m\big) \;\middle|\; \begin{array}{l}(x,u) \text{ satisfies (2.1)}\\ \text{for almost all } t \in \mathbb{R}\end{array} \right\}. \tag{2.2}$$

Note that any function $x \in \mathscr{W}^{1,1}_{\mathrm{loc}}(\mathbb{R};\mathbb{R}^n)$ is continuous. Moreover, by linearity of (2.1), $\mathfrak{B}_{[E,A,B]}$ is a vector space. Further, since the matrices in (2.1) do not depend on t, the behavior is *shift-invariant*, that is, $(\sigma_\tau x, \sigma_\tau u) \in \mathfrak{B}_{[E,A,B]}$ for all $\tau \in \mathbb{R}$ and $(x, u) \in \mathfrak{B}_{[E,A,B]}$.

The following spaces play a fundamental role in this article:

(a) The *space of consistent initial states*

$$\mathscr{V}_{[E,A,B]} = \left\{ x^0 \in \mathbb{R}^n \;\middle|\; \exists (x,u) \in \mathfrak{B}_{[E,A,B]} : x(0) = x^0 \right\}.$$

(b) The *space of consistent initial differential variables*

$$\mathscr{V}^{\mathrm{diff}}_{[E,A,B]} = \left\{ x^0 \in \mathbb{R}^n \;\middle|\; \exists (x,u) \in \mathfrak{B}_{[E,A,B]} : Ex(0) = Ex^0 \right\}.$$

(c) The *reachability space at time* $t \geq 0$

$$\mathscr{R}^t_{[E,A,B]} = \left\{ x^0 \in \mathbb{R}^n \;\middle|\; \exists (x,u) \in \mathfrak{B}_{[E,A,B]} : x(0) = 0 \wedge x(t) = x^0 \right\}$$

and the *reachability space*

$$\mathscr{R}_{[E,A,B]} = \bigcup_{t\geq 0} \mathscr{R}^t_{[E,A,B]}.$$

(d) The *controllability space at time* $t \geq 0$

$$\mathscr{C}^t_{[E,A,B]} = \left\{x^0 \in \mathbb{R}^n \;\middle|\; \exists (x,u) \in \mathfrak{B}_{[E,A,B]} : x(0) = x^0 \wedge x(t) = 0\right\}$$

and the *controllability space*

$$\mathscr{C}_{[E,A,B]} = \bigcup_{t\geq 0} \mathscr{C}^t_{[E,A,B]}.$$

Note that, by linearity of the system, $\mathscr{V}_{[E,A,B]}$, $\mathscr{V}^{\mathrm{diff}}_{[E,A,B]}$, $\mathscr{R}^t_{[E,A,B]}$ and $\mathscr{C}^t_{[E,A,B]}$ are linear subspaces of $\mathbb{R}^n$. We will show that $\mathscr{R}^{t_1}_{[E,A,B]} = \mathscr{R}^{t_2}_{[E,A,B]} = \mathscr{C}^{t_1}_{[E,A,B]} = \mathscr{C}^{t_2}_{[E,A,B]}$ for all $t_1, t_2 \in \mathbb{R}_{>0}$, see Lemma 2.3. This implies $\mathscr{R}_{[E,A,B]} = \mathscr{R}^t_{[E,A,B]} = \mathscr{C}^t_{[E,A,B]} = \mathscr{C}_{[E,A,B]}$ for all $t \in \mathbb{R}_{>0}$. Note further that, by shift-invariance, we have for all $t \in \mathbb{R}$

$$\mathscr{V}_{[E,A,B]} = \left\{x^0 \in \mathbb{R}^n \;\middle|\; \exists (x,u) \in \mathfrak{B}_{[E,A,B]} : x(t) = x^0\right\}, \tag{2.3}$$

$$\mathscr{V}^{\mathrm{diff}}_{[E,A,B]} = \left\{x^0 \in \mathbb{R}^n \;\middle|\; \exists (x,u) \in \mathfrak{B}_{[E,A,B]} : Ex(t) = Ex^0\right\}. \tag{2.4}$$

In the following three lemmas we clarify some of the connections of the above defined spaces, before we state the controllability concepts.

Lemma 2.1 (Inclusions for reachability spaces) *For* $[E, A, B] \in \Sigma_{k,n,m}$ *and* $t_1, t_2 \in \mathbb{R}_{>0}$ *with* $t_1 < t_2$, *the following hold true*:

(a) $\mathscr{R}^{t_1}_{[E,A,B]} \subseteq \mathscr{R}^{t_2}_{[E,A,B]}$.
(b) *If* $\mathscr{R}^{t_1}_{[E,A,B]} = \mathscr{R}^{t_2}_{[E,A,B]}$, *then* $\mathscr{R}^{t_1}_{[E,A,B]} = \mathscr{R}^{t}_{[E,A,B]}$ *for all* $t \in \mathbb{R}$ *with* $t > t_1$.

Proof (a) Let $\bar{x} \in \mathscr{R}^{t_1}_{[E,A,B]}$. By definition, there exists some $(x,u) \in \mathfrak{B}_{[E,A,B]}$ with $x(0) = 0$ and $x(t_1) = \bar{x}$. Consider now $(x_1, u_1) : \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}^m$ with

$$\big(x_1(t), u_1(t)\big) = \begin{cases} (x(t - t_2 + t_1), u(t - t_2 + t_1)), & \text{if } t > t_2 - t_1, \\ (0,0), & \text{if } t \leq t_2 - t_1. \end{cases}$$

Then $x(0) = 0$ implies that x_1 is continuous at $t_2 - t_1$. Since, furthermore,

$$x_1|_{(-\infty, t_2-t_1]} \in \mathscr{W}^{1,1}_{\mathrm{loc}}\big((-\infty, t_2 - t_1]; \mathbb{R}^n\big) \quad \text{and}$$
$$x_1|_{[t_2-t_1, \infty)} \in \mathscr{W}^{1,1}_{\mathrm{loc}}\big([t_2 - t_1, \infty); \mathbb{R}^n\big),$$

we have $(x_1, u_1) \in \mathscr{W}^{1,1}_{\mathrm{loc}}(\mathbb{R}; \mathbb{R}^n) \times \mathscr{L}^1_{\mathrm{loc}}(\mathbb{R}; \mathbb{R}^m)$. By shift-invariance, $E\dot{x}_1(t) = Ax_1(t) + Bu_1(t)$ holds true for almost all $t \in \mathbb{R}$, i.e., $(x_1, u_1) \in \mathfrak{B}_{[E,A,B]}$. Then, due to $x_1(0) = 0$ and $\bar{x} = x(t_1) = x_1(t_2)$, we obtain $\bar{x} \in \mathscr{R}^{t_2}_{[E,A,B]}$.

(b) *Step 1:* We show that $\mathscr{R}^{t_1}_{[E,A,B]} = \mathscr{R}^{t_2}_{[E,A,B]}$ implies $\mathscr{R}^{t_1}_{[E,A,B]} = \mathscr{R}^{t_1+2(t_2-t_1)}_{[E,A,B]}$: By (a), it suffices to show the inclusion "$\supseteq$". Assume that $\bar{x} \in \mathscr{R}^{t_1+2(t_2-t_1)}_{[E,A,B]}$, i.e., there exists some $(x_1, u_1) \in \mathfrak{B}_{[E,A,B]}$ with $x_1(0) = 0$ and $x_1(t_1 + 2(t_2 - t_1)) = \bar{x}$. Since $x_1(t_2) \in \mathscr{R}^{t_2}_{[E,A,B]} = \mathscr{R}^{t_1}_{[E,A,B]}$, there exists some $(x_2, u_2) \in \mathfrak{B}_{[E,A,B]}$ with $x_2(0) = 0$ and $x_2(t_1) = x_1(t_2)$. Now consider the trajectory

$$\big(x(t), u(t)\big) = \begin{cases} (x_2(t), u_2(t)), & \text{if } t < t_1, \\ (x_1(t + (t_2 - t_1)), u_1(t + (t_2 - t_1))), & \text{if } t \geq t_1. \end{cases}$$

Since x is continuous at t_1, we can apply the same argumentation as in the proof of (a) to infer that $(x, u) \in \mathfrak{B}_{[E,A,B]}$. The result to be shown in this step is now a consequence of $x(0) = x_2(0) = 0$ and

$$\bar{x} = x_1\big(t_1 + 2(t_2 - t_1)\big) = x(t_2) \in \mathscr{R}^{t_2}_{[E,A,B]} = \mathscr{R}^{t_1}_{[E,A,B]}.$$

Step 2: We show (b): From the result shown in the first step, we may inductively conclude that $\mathscr{R}^{t_1}_{[E,A,B]} = \mathscr{R}^{t_2}_{[E,A,B]}$ implies $\mathscr{R}^{t_1}_{[E,A,B]} = \mathscr{R}^{t_1+l(t_2-t_1)}_{[E,A,B]}$ for all $l \in \mathbb{N}$. Let $t \in \mathbb{R}$ with $t > t_1$. Then there exists some $l \in \mathbb{N}$ with $t \leq t_1 + l(t_2 - t_1)$. Then statement (a) implies

$$\mathscr{R}^{t_1}_{[E,A,B]} \subseteq \mathscr{R}^{t}_{[E,A,B]} \subseteq \mathscr{R}^{t_1+l(t_2-t_1)}_{[E,A,B]},$$

and, by $\mathscr{R}^{t_1}_{[E,A,B]} = \mathscr{R}^{t_1+l(t_2-t_1)}_{[E,A,B]}$, we obtain the desired result. □

Now we present some relations between controllability and reachability spaces of $[E, A, B] \in \Sigma_{k,n,m}$ and its *backward system* $[-E, A, B] \in \Sigma_{k,n,m}$. It can be easily verified that

$$\mathfrak{B}_{[-E,A,B]} = \big\{(\rho x, \rho u) \;\big|\; (x, u) \in \mathfrak{B}_{[E,A,B]}\big\}. \tag{2.5}$$

Lemma 2.2 (Reachability and controllability spaces of the backward system) *For* $[E, A, B] \in \Sigma_{k,n,m}$ *and* $t \in \mathbb{R}_{>0}$, *we have*

$$\mathscr{R}^{t}_{[E,A,B]} = \mathscr{C}^{t}_{[-E,A,B]}, \quad \textit{and} \quad \mathscr{C}^{t}_{[E,A,B]} = \mathscr{R}^{t}_{[-E,A,B]}.$$

Proof Both assertions follow immediately from the fact that $(x, u) \in \mathfrak{B}_{[E,A,B]}$, if, and only if, $(\sigma_t(\rho x), \sigma_t(\rho u)) \in \mathfrak{B}_{[-E,A,B]}$. □

The previous lemma enables us to show that the controllability and reachability spaces of $[E, A, B] \in \Sigma_{k,n,m}$ are even equal. We further prove that both spaces do not depend on time $t \in \mathbb{R}_{>0}$.

Lemma 2.3 (Impulsive initial conditions and controllability spaces) *For* $[E, A, B] \in \Sigma_{k,n,m}$, *the following hold true*:

(a) $\mathscr{R}^{t_1}_{[E,A,B]} = \mathscr{R}^{t_2}_{[E,A,B]}$ *for all* $t_1, t_2 \in \mathbb{R}_{>0}$.

(b) $\mathscr{R}^t_{[E,A,B]} = \mathscr{C}^t_{[E,A,B]}$ *for all* $t \in \mathbb{R}_{>0}$.
(c) $\mathscr{V}^{\mathrm{diff}}_{[E,A,B]} = \mathscr{V}_{[E,A,B]} + \ker_{\mathbb{R}} E$.

Proof (a) By Lemma 2.1(a), we have

$$\mathscr{R}^{\frac{t_1}{n+1}}_{[E,A,B]} \subseteq \mathscr{R}^{\frac{2t_1}{n+1}}_{[E,A,B]} \subseteq \cdots \subseteq \mathscr{R}^{\frac{nt_1}{n+1}}_{[E,A,B]} \subseteq \mathscr{R}^{t_1}_{[E,A,B]} \subseteq \mathbb{R}^n,$$

and thus

$$0 \leq \dim \mathscr{R}^{\frac{t_1}{n+1}}_{[E,A,B]} \leq \dim \mathscr{R}^{\frac{2t_1}{n+1}}_{[E,A,B]} \leq \cdots \leq \dim \mathscr{R}^{\frac{nt_1}{n+1}}_{[E,A,B]} \leq \dim \mathscr{R}^{t_1}_{[E,A,B]} \leq n.$$

As a consequence, there has to exist some $j \in \{1, \ldots, n+1\}$ with

$$\dim \mathscr{R}^{\frac{jt_1}{n+1}}_{[E,A,B]} = \dim \mathscr{R}^{\frac{(j+1)t_1}{n+1}}_{[E,A,B]}.$$

Together with the subset inclusion, this yields

$$\mathscr{R}^{\frac{jt_1}{n+1}}_{[E,A,B]} = \mathscr{R}^{\frac{(j+1)t_1}{n+1}}_{[E,A,B]}.$$

Lemma 2.1(b) then implies the desired statement.

(b) Let $\bar{x} \in \mathscr{R}^t_{[E,A,B]}$. Then there exists some $(x_1, u_1) \in \mathfrak{B}_{[E,A,B]}$ with $x_1(0) = 0$ and $x_1(t) = \bar{x}$. Since, by (a), we have $x_1(2t) \in \mathscr{R}^t_{[E,A,B]}$, there also exists some $(x_2, u_2) \in \mathfrak{B}_{[E,A,B]}$ with $x_2(0) = 0$ and $x_2(t) = x_1(2t)$. By linearity and shift-invariance, we have

$$(x, u) := (\sigma_t x_1 - x_2, \sigma_t u_1 - u_2) \in \mathfrak{B}_{[E,A,B]}.$$

The inclusion $\mathscr{R}^t_{[E,A,B]} \subseteq \mathscr{C}^t_{[E,A,B]}$ then follows by

$$x(0) = x_1(t) - x_2(0) = \bar{x}, \qquad x(t) = x_1(2t) - x_2(t) = 0.$$

To prove the opposite inclusion, we make use of the previously shown subset relation and Lemma 2.2 to infer that

$$\mathscr{C}^t_{[E,A,B]} = \mathscr{R}^t_{[-E,A,B]} \subseteq \mathscr{C}^t_{[-E,A,B]} = \mathscr{R}^t_{[E,A,B]}.$$

(c) We first show that $\mathscr{V}^{\mathrm{diff}}_{[E,A,B]} \subseteq \mathscr{V}_{[E,A,B]} + \ker_{\mathbb{R}} E$: Assume that $x^0 \in \mathscr{V}^{\mathrm{diff}}_{[E,A,B]}$, i.e., $Ex^0 = Ex(0)$ for some $(x, u) \in \mathfrak{B}_{[E,A,B]}$. By $x(0) \in \mathscr{V}_{[E,A,B]}$, $x(0) - x^0 \in \ker_{\mathbb{R}} E$, we obtain

$$x^0 = x(0) + \big(x^0 - x(0)\big) \in \mathscr{V}_{[E,A,B]} + \ker_{\mathbb{R}} E.$$

To prove $\mathscr{V}_{[E,A,B]} + \ker_{\mathbb{R}} E \subseteq \mathscr{V}^{\mathrm{diff}}_{[E,A,B]}$, assume that $x^0 = x(0) + \bar{x}$ for some $(x, u) \in \mathfrak{B}_{[E,A,B]}$ and $\bar{x} \in \ker_{\mathbb{R}} E$. Then $x^0 \in \mathscr{V}^{\mathrm{diff}}_{[E,A,B]}$ is a consequence of $Ex^0 = E(x(0) + \bar{x}) = Ex(0)$. □

By Lemma 2.3 it is sufficient to only consider the spaces $\mathscr{V}_{[E,A,B]}$ and $\mathscr{R}_{[E,A,B]}$ in the following.

We are now in the position to define the central notions of controllability, reachability and stabilizability considered in this article.

Definition 2.1 The system $[E, A, B] \in \Sigma_{k,n,m}$ is called

(a) *controllable at infinity*

$$:\Leftrightarrow \forall x^0 \in \mathbb{R}^n \exists (x,u) \in \mathscr{B}_{[E,A,B]} : x(0) = x^0 \Leftrightarrow \mathscr{V}_{[E,A,B]} = \mathbb{R}^n.$$

(b) *impulse controllable*

$$:\Leftrightarrow \forall x^0 \in \mathbb{R}^n \exists (x,u) \in \mathscr{B}_{[E,A,B]} : Ex^0 = Ex(0) \Leftrightarrow \mathscr{V}^{\mathrm{diff}}_{[E,A,B]} = \mathbb{R}^n.$$

(c) *controllable within the set of reachable states (R-controllable)*

$$:\Leftrightarrow \forall x_0, x_f \in \mathscr{V}_{[E,A,B]} \exists t > 0 \exists (x,u) \in \mathscr{B}_{[E,A,B]} : x(0) = x_0 \wedge x(t) = x_f.$$

(d) *controllable in the behavioral sense*

$$:\Leftrightarrow \forall (x_1,u_1),(x_2,u_2) \in \mathscr{B}_{[E,A,B]}$$
$$\exists T > 0 \exists (x,u) \in \mathscr{B}_{[E,A,B]} : (x(t),u(t)) = \begin{cases} (x_1(t),u_1(t)), & \text{if } t < 0, \\ (x_2(t),u_2(t)), & \text{if } t > T. \end{cases}$$

(e) *stabilizable in the behavioral sense*

$$:\Leftrightarrow \forall (x,u) \in \mathscr{B}_{[E,A,B]} \exists (x_0,u_0) \in \mathscr{B}_{[E,A,B]} \cap \left(\mathscr{W}^{1,1}_{\mathrm{loc}}\left(\mathscr{T};\mathbb{R}^n\right) \times \mathscr{W}^{1,1}_{\mathrm{loc}}\left(\mathscr{T};\mathbb{R}^n\right)\right) :$$
$$\left(\forall t < 0 : \left(x(t),u(t)\right) = \left(x_0(t),u_0(t)\right)\right) \wedge \lim_{t\to\infty}\left(x_0(t),u_0(t)\right) = 0.$$

(f) *completely reachable*

$$:\Leftrightarrow \exists t \in \mathbb{R}_{>0} \forall x_f \in \mathbb{R}^n \exists (x,u) \in \mathscr{B}_{[E,A,B]} : x(0) = 0 \wedge x(t) = x_f$$
$$\Leftrightarrow \exists t \in \mathbb{R}_{>0} : \mathscr{R}^t_{[E,A,B]} = \mathbb{R}^n.$$

(g) *completely controllable*

$$:\Leftrightarrow \exists t \in \mathbb{R}_{>0} \forall x_0, x_f \in \mathbb{R}^n \exists (x,u) \in \mathscr{B}_{[E,A,B]} : x(0) = x_0 \wedge x(t) = x_f.$$

(h) *completely stabilizable*

$$:\Leftrightarrow \forall x_0 \in \mathbb{R}^n \exists (x,u) \in \mathscr{B}_{[E,A,B]} : x(0) = x_0 \wedge \lim_{t\to\infty} x(t) = 0.$$

(i) *strongly reachable*

$$:\Leftrightarrow \exists t \in \mathbb{R}_{>0} \forall x_f \in \mathbb{R}^n \exists (x,u) \in \mathscr{B}_{[E,A,B]} : Ex(0) = 0 \wedge Ex(t) = Ex_f.$$

(j) *strongly controllable*

$$:\Leftrightarrow \exists t \in \mathbb{R}_{>0} \forall x_0, x_f \in \mathbb{R}^n \exists (x,u) \in \mathscr{B}_{[E,A,B]} : Ex(0) = Ex_0 \wedge Ex(t) = Ex_f.$$

(k) *strongly stabilizable* (or merely *stabilizable*)

$$:\Leftrightarrow \forall x_0 \in \mathbb{R}^n \exists (x,u) \in \mathscr{B}_{[E,A,B]} : Ex(0) = Ex_0 \wedge \lim_{t\to\infty} Ex(t) = 0.$$

Some remarks on the definitions are warrant.

Remark 2.1

(i) The controllability concepts are not consistently treated in the literature. For instance, one has to pay attention if it is (tacitly) claimed that $[E, B] \in \mathbb{R}^{k,n+m}$ or $[E, A, B] \in \mathbb{R}^{k,2n+m}$ have full rank.
For regular systems we have the following:

concept	coincides with notion in	called [...] in
controllability at infinity	see item (2.1)	reachability at ∞ in [99]
impulse controllability	[46] and [73, Rem. 2]	controllability at ∞ in [99]; controllability at infinity in [5, 6, 155]
R-controllability	[41, 49, 169] and [73, Rem. 2]	–
complete controllability	[41, 49, 169]	controllability in [46]
strong controllability	[155] and [73, Rem. 2]	impulse controllability in [63]

Some of these aforementioned articles introduce the controllability by means of certain rank criteria for the matrix triple $[E, A, B]$. The connection of the concepts introduced in Definition 2.1 to linear algebraic properties of E, A and B will be highlighted in Sect. 4.

For general DAE systems we have

concept	coincides with notion in	called [...] in
controllability at infinity	–	–
impulse controllability	[61, 71, 75]	–
R-controllability	–	–
complete controllability	[120]	controllability in [58]
strong controllability	–	controllability in [120]

Our behavioral controllability coincides with the framework which is introduced in [128, Definition 5.2.2] for so-called *differential behaviors*, which

are general (possibly higher order) DAE systems with constant coefficients. Note that the concept of behavioral controllability does not require a distinction between input and state. The concepts of reachability and controllability in [11–14] coincide with our behavioral and complete controllability, resp. (see Sect. 4). Full controllability of [171] is our complete controllability together with the additional assumption that solutions have to be unique.

(ii) Stabilizability in the behavioral sense is introduced in [128, Definition 5.2.2]. For regular systems, stabilizability is usually defined either via linear algebraic properties of E, A and B, or by the existence of a stabilizing state feedback, see [33, 34, 57] and [49, Definition 3-1.2]. Our concepts of behavioral stabilizability and stabilizability coincide with the notions of internal stability and complete stabilizability, resp., defined in [114] for the system $\mathscr{E}\dot{z}(t) = \mathscr{A}z(t)$ with $\mathscr{E} = [E, 0]$, $\mathscr{A} = [A, B]$, $z(t) = [x^\top(t), u^\top(t)]^\top$.

(iii) Other concepts, not related to the ones considered in this article, are e.g. the instantaneous controllability (reachability) of order k in [120] or the impulsive mode controllability in [71]. Furthermore, the concept of strong controllability introduced in [147, Exercise 8.5] for ODE systems differs from the concepts considered in this article.

(iv) The notion of consistent initial conditions is the most important one for DAE systems and therefore the consideration of the space $\mathscr{V}_{[E,A,B]}$ (for $B = 0$ when no control systems were considered) is as old as the theory of DAEs itself, see e.g. [60]. $\mathscr{V}_{[E,A,B]}$ is sometimes called viability kernel [30], see also [8, 9]. The reachability and controllability space are some of the most important notions for (DAE) control systems and have been considered in [99] for regular systems. They are the fundamental subspaces considered in the geometric theory, see Sect. 6. Further usage of these concepts can be found in the following: in [122] generalized reachability and controllability subspaces of regular systems are considered; Eliopoulou and Karcanias [56] consider reachability and almost reachability subspaces of general DAE systems; Frankowska [58] considers the reachability subspace in terms of differential inclusions.

A nice formula for the reachability space of a regular system has been derived by Yip et al. [169] (and later been adopted by Cobb [46], however, called controllable subspace): Consider a regular system $[E, A, B] \in \Sigma_{n,n,m}$ in Weierstraß form [60], that is,

$$E = \begin{bmatrix} I_{n_1} & 0 \\ 0 & N \end{bmatrix}, \qquad A = \begin{bmatrix} J & 0 \\ 0 & I_{n_2} \end{bmatrix}, \qquad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix},$$

where N is nilpotent. Then [169, Thm. 2]

$$\mathscr{R}_{[E,A,B]} = \langle J | B_1 \rangle \times \langle N | B_2 \rangle,$$

where $\langle K | L \rangle := \operatorname{im}_{\mathbb{R}}[L, KL, \ldots, K^{n-1}L]$ for some matrices $K \in \mathbb{R}^{n\times n}$, $L \in \mathbb{R}^{n\times m}$. Furthermore, we have [169, Thm. 3]

$$\mathscr{V}_{[E,A,B]} = \mathbb{R}^{n_1} \times \langle N | B_2 \rangle.$$

This result has been improved later in [41] so that the Weierstraß form is no longer needed. Denoting by E^D the Drazin inverse of a given matrix $E \in \mathbb{R}^{n\times n}$ (see [39]), it is shown [41, Thm. 3.1] that, for $A = I$,

$$\mathscr{R}_{[E,A,B]} = E^D\langle E^D|B\rangle \oplus \left(I - EE^D\right)\langle E|B\rangle,$$

where the consideration of $A = I$ is justified by a certain (time-varying) transformation of the system [124]. We further have [41, Thm. 3.2]

$$\mathscr{V}_{[E,A,B]} = \operatorname{im}_{\mathbb{R}} E^D \oplus \left(I - EE^D\right)\langle E|B\rangle.$$

Yet another approach was followed by Cobb [42] who obtains that

$$\mathscr{R}_{[E,A,B]} = \left\langle (\alpha E - A)^{-1}E|(\alpha E - A)^{-1}B\right\rangle$$

for some $\alpha \in \mathbb{R}$ with $\det(\alpha E - A) \neq 0$. A simple proof of this result can also be found in [170].

(v) The notion $\mathscr{V}^{\mathrm{diff}}_{[E,A,B]}$ comes from the possible impulsive behavior of solutions of (2.1), i.e., x may have jumps, when distributional solutions are permitted, see e.g. [46] as a very early contribution in this regard. Since these jumps have no effect on the solutions if they occur at the initial time and within the kernel of E this leads to the definition of $\mathscr{V}^{\mathrm{diff}}_{[E,A,B]}$. See also the definition of impulse controllability.

(vi) Impulse controllability and controllability at infinity are usually defined by considering distributional solutions of (2.1), see e.g. [46, 61, 75], sometimes called impulsive modes, see e.g. [21, 71, 155]. For regular systems, impulse controllability has been introduced by Verghese et al. [155] (called controllability at infinity in this work) as controllability of the impulsive modes of the system, and later made more precise by Cobb [46], see also Armentano [5, 6] (who also calls it controllability at infinity) for a more geometric point of view. In [155] the authors do also develop the notion of strong controllability as impulse controllability with, additionally, controllability in the regular sense. Cobb [43] showed that under the condition of impulse controllability, the infinite eigenvalues of regular $sE - A$ can be assigned via a state feedback $u = Fx$ to arbitrary finite positions. Armentano [5] later showed how to calculate F. This topic has been further pursued in [94] in the form of invariant polynomial assignment.

The name "controllability at infinity" comes from the claim that the system has no infinite uncontrollable modes: Speaking in terms of rank criteria (see also Sect. 4) the system $[E, A, B] \in \Sigma_{k,n,m}$ is said to have an uncontrollable mode at $\frac{\alpha}{\beta}$ if, and only if, $\operatorname{rk}[\alpha E + \beta A, B] < \operatorname{rk}[E, A, B]$ for some $\alpha, \beta \in \mathbb{C}$. If $\beta = 0$, then the uncontrollable mode is infinite. Controllability at infinity has been introduced by Rosenbrock [137]—although he does not use this phrase—as controllability of the infinite frequency zeros. Later Cobb [46] compared the concepts of impulse controllability and controllability at infinity, see [46, Thm. 5]; the notions we use in the present article go back to the distinction in this work.

The concepts have later been generalized by Geerts [61] (see [61, Thm. 4.5 & Rem. 4.9], however, he does not use the name "controllability at infinity"). Controllability at infinity of (2.1) is equivalent to the strictness of the corresponding differential inclusion [58, Prop. 2.6]. The concept of impulsive mode controllability in [71] is even weaker than impulse controllability.

(vii) Controllability concepts with a distributional solution setup have been considered in [61, 120, 130] for instance, see also [46]. A typical argumentation in these works is that inconsistent initial values cause distributional solutions in a way that the state trajectory is composed of a continuous function and a linear combination of Dirac's delta impulse and some of its derivatives. However, some frequency domain considerations in [116] refute this approach (see [145] for an overview on inconsistent initialization). This justifies that we do only consider weakly differentiable solutions as defined in the behavior $\mathscr{B}_{[E,A,B]}$.
Distributional solutions for time-invariant DAEs have already been considered by Cobb [44] and Geerts [61, 62] and for time-varying DAEs by Rabier and Rheinboldt [132]. For a mathematically rigorous approach to distributional solution theory of linear DAEs we refer to [143, 144] by Trenn. The latter works introduce the notions of impulse controllability and jump controllability which coincide with our impulse controllability and behavioral controllability, resp.

(vii) R-controllability has been first defined in [169] for regular DAEs. Roughly speaking, R-controllability is the property that any consistent initial state x_0 can be steered to any reachable state x_f, where here x_f is reachable if, and only if, there exist $t > 0$ and $(x, u) \in \mathscr{B}_{[E,A,B]}$ such that $x(t) = x_f$; by (2.3) the latter is equivalent to $x_f \in \mathscr{V}_{[E,A,B]}$, as stated in Definition 2.1.

(viii) The concept of behavioral controllability has been introduced by Willems [160], see also [128]. This concept is very suitable for generalizations in various directions, see e.g. [35, 40, 72, 97, 133, 163, 167]. Having found the behavior of the considered control system one can take over the definition of behavioral controllability without the need for any further changes. From this point of view this appears to be the most natural of the controllability concepts. However, this concept also seems to be the least regarded in the DAE literature.

(ix) The controllability theory of DAE systems can also be treated with the theory of differential inclusions [8, 9] as showed by Frankowska [58].

(x) Karcanias and Hayton [85] pursued a special ansatz to simplify the system (2.1): provided that B has full column rank, we take a left annihilator N and a pseudoinverse $B^\dagger$ of B (i.e., $NB = 0$ and $B^\dagger B = I$) such that $W = \left[\begin{smallmatrix} N \\ B^\dagger \end{smallmatrix}\right]$ is invertible and then pre-multiply (2.1) by W, thus obtaining the equivalent system

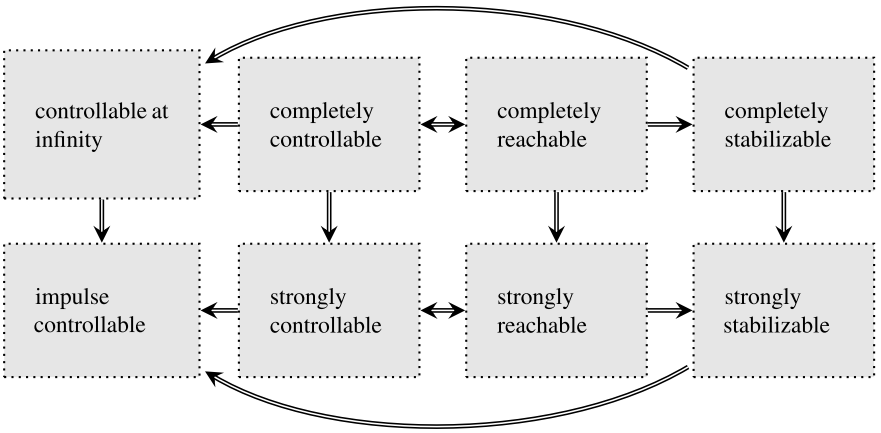
$$\begin{aligned} NE\dot{x} &= NAx, \\ u &= B^\dagger(E\dot{x} - Ax). \end{aligned}$$

The reachability (controllability) properties of (2.1) may now be studied in terms of the pencil $sNE - NA$, which is called the restriction pencil [78], first introduced as zero pencil for the investigation of system zeros of ODEs in [91, 92], see also [88]. For a comprehensive study of the properties of the pencil $sNE - NA$ see e.g. [84–87].

(xi) Banaszuk and Przyłuski [11] have considered perturbations of DAE control systems and obtained conditions under which the sets of all completely controllable systems (systems controllable in the behavioral sense) within the set of all systems $\Sigma_{k,n,m}$ contain an open and dense subset, or its complement contains an open and dense subset.

The following dependencies hold true between the concepts from Definition 2.1. Some further relations will be derived in Sect. 4.

Proposition 2.4 *For any* $[E, A, B] \in \Sigma_{k,n,m}$ *the following implications hold true*: *If "*$\Rightarrow$*" holds, then "*$\Leftarrow$*" does, in general, not hold.*



Proof Since it is easy to construct counterexamples for any direction where in the diagram only "$\Rightarrow$" holds, we skip their presentation. The following implications are immediate consequences of Definition 2.1:

completely controllable $\Rightarrow$ controllable at infinity $\Rightarrow$ impulse controllable,
completely controllable $\Rightarrow$ strongly controllable $\Rightarrow$ impulse controllable,
completely controllable $\Rightarrow$ completely reachable $\Rightarrow$ strongly reachable,
strongly controllable $\Rightarrow$ strongly reachable,
completely stabilizable $\Rightarrow$ controllable at infinity,
strongly stabilizable $\Rightarrow$ impulse controllable,
completely stabilizable $\Rightarrow$ strongly stabilizable.

It remains to prove the following assertions:

(a) completely reachable $\Rightarrow$ completely controllable,
(b) strongly reachable $\Rightarrow$ strongly controllable,
(c) completely reachable $\Rightarrow$ completely stabilizable,
(d) strongly reachable $\Rightarrow$ strongly stabilizable.

(a) Let $x_0, x_f \in \mathbb{R}^n$. Then, by complete reachability of $[E, A, B]$, there exist $t > 0$ and some $(x_1, u_1) \in \mathscr{B}_{[E,A,B]}$ with $x_1(0) = 0$ and $x_1(t) = x_0$. Further, there exists $(x_2, u_2) \in \mathscr{B}_{[E,A,B]}$ with $x_2(0) = 0$ and $x_2(t) = x_f - x_1(2t)$. By linearity and shift-invariance, we have

$$(x, u) := (\sigma_t x_1 + x_2, \sigma_t u_1 + u_2) \in \mathscr{B}_{[E,A,B]}.$$

On the other hand, this trajectory fulfills $x(0) = x_1(t) + x_2(0) = x_0$ and $x(t) = x_1(2t) + x_2(t) = x_f$.

(b) The proof of this statement is analogous to (a).

(c) By (a) it follows that the system is completely controllable. Complete controllability implies that there exists some $t > 0$, such that for all $x_0 \in \mathbb{R}^n$ there exists $(x_1, u_1) \in \mathscr{B}_{[E,A,B]}$ with $x_1(0) = x_0$ and $x_1(t) = 0$. Then, since (x, u) with

$$\big(x(\tau), u(\tau)\big) = \begin{cases} (x_1(\tau), u_1(\tau)), & \text{if } \tau \le t, \\ (0, 0), & \text{if } \tau \ge t \end{cases}$$

satisfies $(x, u) \in \mathscr{B}_{[E,A,B]}$ (cf. the proof of Lemma 2.1(a)), the system $[E, A, B]$ is completely stabilizable.

(d) The proof of this statement is analogous to (c). □

3 Solutions, Relations and Normal Forms

In this section we give the definitions for system and feedback equivalence of DAE control systems (see [63, 137, 155]), revisit the solution theory of DAEs (see [96, 159] and also [22]), and state a normal form under system and feedback equivalence (see [105]). For the definition of a canonical and a normal form see Remark 3.2.

3.1 System and Feedback Equivalence

We define the essential concepts of system and feedback equivalence. System equivalence was first studied by Rosenbrock [137] (called restricted system equivalence in his work, see also [155]) and later became a crucial concept in the control theory of DAEs [24, 25, 63, 64, 69]. Feedback equivalence for DAEs seems to have been first considered in [63] to derive a feedback canonical form for regular systems, little later also in [105] (for general DAEs) where additionally also derivative feedback was investigated and respective canonical forms derived, see also Sect. 3.3.

Definition 3.1 (System and feedback equivalence) Two systems $[E_i, A_i, B_i] \in \Sigma_{k,n,m}$, $i = 1, 2$, are called

- *system equivalent* if, and only if,

$$\exists W \in \mathbf{Gl}_k(\mathbb{R}), T \in \mathbf{Gl}_n(\mathbb{R}) : \begin{bmatrix} sE_1 - A_1 & B_1 \end{bmatrix} = W \begin{bmatrix} sE_2 - A_2 & B_2 \end{bmatrix} \begin{bmatrix} T & 0 \\ 0 & I_m \end{bmatrix};$$

  we write

$$[E_1, A_1, B_1] \overset{W,T}{\sim}_{se} [E_2, A_2, B_2];$$

- *feedback equivalent* if, and only if,

$$\begin{aligned} &\exists W \in \mathbf{Gl}_k(\mathbb{R}), T \in \mathbf{Gl}_n(\mathbb{R}), V \in \mathbf{Gl}_m(\mathbb{R}), F \in \mathbb{R}^{m,n} : \\ &\begin{bmatrix} sE_1 - A_1 & B_1 \end{bmatrix} = W \begin{bmatrix} sE_2 - A_2 & B_2 \end{bmatrix} \begin{bmatrix} T & 0 \\ -F & V \end{bmatrix}; \end{aligned} \tag{3.1}$$

  we write

$$[E_1, A_1, B_1] \overset{W,T,V,F}{\sim}_{fe} [E_2, A_2, B_2].$$

It is easy to observe that both system and feedback equivalence are equivalence relations on $\Sigma_{k,n,m}$. To see the latter, note that if $[E_1, A_1, B_1] \overset{W,T,V,F}{\sim}_{fe} [E_2, A_2, B_2]$, then

$$[E_2, A_2, B_2] \overset{W^{-1},T^{-1},V^{-1},-V^{-1}FT^{-1}}{\sim}_{fe} [E_1, A_1, B_1].$$

The behaviors of system and feedback equivalent systems are connected via

$$\begin{aligned} &\text{If} \quad [E_1, A_1, B_1] \overset{W,T}{\sim}_{se} [E_2, A_2, B_2], \quad \text{then} \\ &(x, u) \in \mathfrak{B}_{[E_1,A_1,B_1]} \Leftrightarrow (Tx, u) \in \mathfrak{B}_{[E_2,A_2,B_2]} \\ &\text{If} \quad [E_1, A_1, B_1] \overset{W,T,V,F}{\sim}_{fe} [E_2, A_2, B_2], \quad \text{then} \\ &(x, u) \in \mathfrak{B}_{[E_1,A_1,B_1]} \Leftrightarrow (Tx, Fx + Vu) \in \mathfrak{B}_{[E_2,A_2,B_2]}. \end{aligned} \tag{3.2}$$

In particular, if $[E_1, A_1, B_1] \overset{W,T}{\sim}_{se} [E_2, A_2, B_2]$, then

$$\mathscr{V}_{[E_1,A_1,B_1]} = T^{-1} \cdot \mathscr{V}_{[E_2,A_2,B_2]}, \qquad \mathscr{R}^t_{[E_1,A_1,B_1]} = T^{-1} \cdot \mathscr{R}^t_{[E_2,A_2,B_2]}.$$

Further, if $[E_1, A_1, B_1] \overset{W,T,V,F}{\sim}_{fe} [E_2, A_2, B_2]$, then

$$\mathscr{V}_{[E_1,A_1,B_1]} = T^{-1} \cdot \mathscr{V}_{[E_2,A_2,B_2]}, \qquad \mathscr{R}^t_{[E_1,A_1,B_1]} = T^{-1} \cdot \mathscr{R}^t_{[E_2,A_2,B_2]},$$

and properties of controllability at infinity, impulse controllability, R-controllability, behavioral controllability, behavioral stabilizability, complete controllability, complete stabilizability, strong controllability and strong stabilizability are invariant under system and feedback equivalence.

Remark 3.1 (Equivalence and minimality in the behavioral sense)

(i) Another equivalence concept has been introduced by Willems in [161] (see also [128, Def. 2.5.2]): Two systems $[E_i, A_i, B_i] \in \Sigma_{k_i,n,m}$, $i = 1, 2$, are called *equivalent in the behavioral sense*, if their behaviors coincide, i.e.,

$$\mathfrak{B}_{[E_1,A_1,B_1]} = \mathfrak{B}_{[E_2,A_2,B_2]}.$$

Note that, in particular, two systems being equivalent in the behavioral sense do not necessarily have the same number of equations. For instance, the following two systems are equivalent in the behavioral sense:

$$\big[[0], [1], [0]\big], \qquad \left[\begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\0\end{bmatrix}\right].$$

(ii) It is shown in [128, Thm. 2.5.4] that for a unimodular matrix $U(s) \in \mathbb{R}[s]^{k,k}$ (that is, $U(s)$ has a polynomial inverse), and $[E, A, B] \in \Sigma_{k,n,m}$, it holds $(x, u) \in \mathfrak{B}_{[E,A,B]}$ if, and only if,

$$U\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)E\dot{x}(t) = U\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)Ax(t) + U\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)Bu(t),$$

where the differential operator $U(\frac{\mathrm{d}}{\mathrm{d}t})$ has to be understood in the distributional sense. The unimodular matrix $U(s)$ can particularly been chosen in a way that

$$U(s) \cdot \big[\, sE - A, \quad -B \,\big] = \begin{bmatrix} R_x(s) & R_u(s) \\ 0 & 0 \end{bmatrix},$$

where $[R_x(s)\ \ R_u(s)] \in \mathbb{R}[s]^{l,n+m}$ has full row rank as a matrix in the field $\mathbb{R}(s)$ [128, Thm. 3.6.2]. It is shown that $R_x(\frac{\mathrm{d}}{\mathrm{d}t})x + R_u(\frac{\mathrm{d}}{\mathrm{d}t})u = 0$ is *minimal in the behavioral sense*, i.e., it describes the behavior by a minimal number of l differential equations among all behavioral descriptions of $\mathfrak{B}_{[E,A,B]}$. By using a special normal form, we will later remark that for any $[E, A, B] \in \Sigma_{k,n,m}$, there exists a unimodular transformation from the left such that the resulting differential-algebraic system is minimal in the behavioral sense.

(iii) Conversely, if two systems $[E_i, A_i, B_i] \in \Sigma_{k_i,n,m}$, $i = 1, 2$ are equivalent in the behavioral sense, and, moreover, $k_1 = k_2$, then there exists some unimodular $U(s) \in \mathbb{R}[s]^{k_1,k_1}$, such that

$$U(s) \cdot \big[\, sE_1 - A_1, \quad -B_1 \,\big] = \big[\, sE_2 - A_2, \quad -B_2 \,\big].$$

If $[E_i, A_i, B_i]$ $i = 1, 2$, contain different numbers of equations (such as, e.g., $k_1 > k_2$), then one can first add $k_1 - k_2$ equations of type "$0 = 0$" to the second system and, thereafter, perform a unimodular transformation leading from one system to another.

(iv) Provided that a unimodular transformation of $E\dot{x}(t) = Ax(t) + Bu(t)$ again leads to a differential-algebraic system (that is, neither a derivative of the input nor a higher derivative of the state occurs), the properties of controllability at infinity, R-controllability, behavioral controllability, behavioral stabilizability, complete controllability, complete stabilizability are invariant under this transformation. However, since the differential variables may be changed under a transformation of this kind, the properties of impulse controllability, strong controllability and strong stabilizability are not invariant. We will see in Remark 3.11 that any $[E, A, B] \in \Sigma_{k,n,m}$ is, in the behavioral sense, equivalent to a system that is controllable at infinity.

In order to study normal forms under system and feedback equivalence we introduce the following notation: For $k \in \mathbb{N}$ we introduce the matrices $N_k \in \mathbb{R}^{k,k}$, $K_k, L_k \in \mathbb{R}^{k-1,k}$ with

$$N_k = \begin{bmatrix} 0 & & \\ 1 & \ddots & \\ & \ddots & \ddots \\ & & 1 \quad 0 \end{bmatrix}, \qquad K_k = \begin{bmatrix} 1 & 0 & \\ & \ddots & \ddots \\ & & 1 \quad 0 \end{bmatrix}, \qquad L_k = \begin{bmatrix} 0 & 1 & \\ & \ddots & \ddots \\ & & 0 \quad 1 \end{bmatrix}.$$

Further, let $e_i^{[k]} \in \mathbb{R}^k$ be the ith canonical unit vector, and, for some multi-index $\alpha = (\alpha_1, \ldots, \alpha_l) \in \mathbb{N}^l$, we define

$$\begin{aligned} N_\alpha &= \operatorname{diag}(N_{\alpha_1}, \ldots, N_{\alpha_l}) \in \mathbb{R}^{|\alpha|,|\alpha|}, \\ K_\alpha &= \operatorname{diag}(K_{\alpha_1}, \ldots, K_{\alpha_l}) \in \mathbb{R}^{|\alpha|-l,|\alpha|}, \\ L_\alpha &= \operatorname{diag}(L_{\alpha_1}, \ldots, L_{\alpha_l}) \in \mathbb{R}^{|\alpha|-l,|\alpha|}, \\ E_\alpha &= \operatorname{diag}\big(e_{\alpha_1}^{[\alpha_1]}, \ldots, e_{\alpha_l}^{[\alpha_l]}\big) \in \mathbb{R}^{|\alpha|,l}. \end{aligned}$$

Kronecker proved [93] that any matrix pencil $sE - A \in \mathbb{R}[s]^{k,n}$ can be put into a certain canonical form, called Kronecker canonical form nowadays, of which a more comprehensive proof has been provided by Gantmacher [60]. In the following we may use the quasi-Kronecker form derived in [22, 23], since in general the Kronecker canonical form is complex-valued even though the given pencil $sE - A$ is real-valued, what we need to avoid. The obtained form then is not canonical anymore, but it is a normal form (see Remark 3.2).

Proposition 3.1 (Quasi-Kronecker form [22, 23, 60]) *For any matrix pencil $sE - A \in \mathbb{R}[s]^{k,n}$, there exist $W \in \mathbf{Gl}_k(\mathbb{R})$, $T \in \mathbf{Gl}_n(\mathbb{R})$ such that*

$$W(sE - A)T = \begin{bmatrix} sI_{n_s} - A_s & 0 & 0 & 0 \\ 0 & sN_\alpha - I_{|\alpha|} & 0 & 0 \\ 0 & 0 & sK_\beta - L_\beta & 0 \\ 0 & 0 & 0 & sK_\gamma^\top - L_\gamma^\top \end{bmatrix} \tag{3.3}$$

for some $A_s \in \mathbb{R}^{n_s,n_s}$ *and multi-indices* $\alpha \in \mathbb{N}^{n_\alpha}$, $\beta \in \mathbb{N}^{n_\gamma}$, $\gamma \in \mathbb{N}^{n_\gamma}$. *The multi-indices* α, β, γ *are uniquely determined by* $sE - A$. *Further, the matrix* A_s *is unique up to similarity.*

The (components of the) multi-indices α, β, γ are often called minimal indices and elementary divisors and play an important role in the analysis of matrix pencils, see e.g. [60, 104, 105, 113], where the components of α are the orders of the infinite elementary divisors, the components of β are the column minimal indices and the components of γ are the row minimal indices. In fact, the number of column (row) minimal indices equal to one corresponds to the dimension of $\ker_{\mathbb{R}} E \cap \ker_{\mathbb{R}} A$ ($\ker_{\mathbb{R}} E^\top \cap \ker_{\mathbb{R}} A^\top$), or, equivalently, the number of zero columns (rows) in a quasi-Kronecker form of $sE - A$. Further, note that $sI_{n_s} - A_s$ may be further transformed into Jordan canonical form to obtain the finite elementary divisors.

Since the multi-indices $\alpha \in \mathbb{N}^{n_\alpha}$, $\beta \in \mathbb{N}^{n_\gamma}$, $\gamma \in \mathbb{N}^{n_\gamma}$ are well-defined by means of the pencil $sE - A$ and, furthermore, the matrix A_s is unique up to similarity, this justifies the introduction of the following quantities.

Definition 3.2 (Index of $sE - A$) Let the matrix pencil $sE - A \in \mathbb{R}[s]^{k,n}$ be given with quasi-Kronecker form (3.3). Then the *index* $\nu \in \mathbb{N}_0$ of $sE - A$ is defined as

$$\nu = \max\{\alpha_1, \ldots, \alpha_{\ell(\alpha)}, \gamma_1, \ldots, \gamma_{\ell(\gamma)}\}.$$

The index is larger or equal to the index of nilpotency ζ of N_α, i.e., $\zeta \leq \nu$, $N_\alpha^\zeta = 0$ and $N_\alpha^{\zeta-1} \neq 0$. By means of the quasi-Kronecker form (3.3) it can be seen that the index of $sE - A$ does not exceed one if, and only if,

$$\operatorname{im}_{\mathbb{R}} A \subseteq \operatorname{im}_{\mathbb{R}} E + A \cdot \ker_{\mathbb{R}} E. \tag{3.4}$$

This is moreover equivalent to the fact that for some (and hence any) real matrix Z with $\operatorname{im}_{\mathbb{R}} Z = \ker_{\mathbb{R}} E$, we have

$$\operatorname{im}_{\mathbb{R}}[E, AZ] = \operatorname{im}_{\mathbb{R}}[E, A]. \tag{3.5}$$

Since each block in $sK_\beta - L_\beta$ ($sK_\gamma^\top - L_\gamma^\top$) causes a single drop of the column (row) rank of $sE - A$, we have

$$\ell(\beta) = n - \operatorname{rk}_{\mathbb{R}(s)}(sE - A), \qquad \ell(\gamma) = k - \operatorname{rk}_{\mathbb{R}(s)}(sE - A). \tag{3.6}$$

Further, $\lambda \in \mathbb{C}$ is a generalized eigenvalue of $sE - A$ if, and only if,

$$\operatorname{rk}_{\mathbb{C}}(\lambda E - A) < \operatorname{rk}_{\mathbb{R}(s)}(sE - A).$$

3.2 A Normal Form Under System Equivalence

Using Proposition 3.1 it is easy to determine a normal form under system equivalence. For regular systems this normal form was first discovered by Rosenbrock [137].

Corollary 3.2 (Decoupled DAE) *Let* $[E, A, B] \in \Sigma_{k,n,m}$. *Then there exist* $W \in \mathbf{Gl}_k(\mathbb{R})$, $T \in \mathbf{Gl}_n(\mathbb{R})$ *such that*

$$[E, A, B] \overset{W,T}{\sim}_{se} \left[\begin{bmatrix} I_{n_s} & 0 & 0 & 0 \\ 0 & N_\alpha & 0 & 0 \\ 0 & 0 & K_\beta & 0 \\ 0 & 0 & 0 & K_\gamma^\top \end{bmatrix}, \begin{bmatrix} A_s & 0 & 0 & 0 \\ 0 & I_{|\alpha|} & 0 & 0 \\ 0 & 0 & L_\beta & 0 \\ 0 & 0 & 0 & L_\gamma^\top \end{bmatrix}, \begin{bmatrix} B_s \\ B_f \\ B_u \\ B_o \end{bmatrix} \right], \tag{3.7}$$

for some $B_s \in \mathbb{R}^{n_s,m}$, $B_f \in \mathbb{R}^{|\alpha|,m}$, $B_o \in \mathbb{R}^{|\beta|-\ell(\beta),m}$, $B_u \in \mathbb{R}^{|\gamma|,m}$, $A_s \in \mathbb{R}^{n_s,n_s}$ *and multi-indices* $\alpha \in \mathbb{N}^{n_\alpha}$, $\beta \in \mathbb{N}^{n_\beta}$, $\gamma \in \mathbb{N}^{n_\gamma}$. *This is interpreted, in terms of the DAE* (2.1), *as follows*: $(x, u) \in \mathfrak{B}_{[E,A,B]}$ *if, and only if,*

$$\left(x_s(\cdot)^\top, x_f(\cdot)^\top, x_u(\cdot)^\top, x_o(\cdot)^\top\right)^\top := Tx(\cdot)$$

with

$$x_f(\cdot) = \begin{pmatrix} x_{f[1]}(\cdot) \\ \vdots \\ x_{f[\ell(\alpha)]}(\cdot) \end{pmatrix}, \qquad x_u(\cdot) = \begin{pmatrix} x_{u[1]}(\cdot) \\ \vdots \\ x_{u[\ell(\beta)]}(\cdot) \end{pmatrix}, \qquad x_o(\cdot) = \begin{pmatrix} x_{o[1]}(\cdot) \\ \vdots \\ x_{o[\ell(\gamma)]}(\cdot) \end{pmatrix}$$

solves the decoupled DAEs

$$\dot{x}_s(t) = A_s x_s(t) + B_s u(t), \tag{3.8a}$$

$$N_{\alpha_i} \dot{x}_{f[i]}(t) = x_{f[i]}(t) + B_{f[i]} u(t) \quad \text{for } i = 1, \ldots, \ell(\alpha), \tag{3.8b}$$

$$K_{\beta_i} \dot{x}_{u[i]}(t) = L_{\beta_i} x_{u[i]}(t) + B_{u[i]} u(t) \quad \text{for } i = 1, \ldots, \ell(\beta), \tag{3.8c}$$

$$K_{\gamma_i}^\top \dot{x}_{o[i]}(t) = L_{\gamma_i}^\top x_{o[i]}(t) + B_{o[i]} u(t) \quad \text{for } i = 1, \ldots, \ell(\gamma) \tag{3.8d}$$

with suitably labeled partitions of B_f, B_u *and* B_o.

Remark 3.2 (Canonical and normal form) Recall the definition of a canonical form: given a group G, a set $\mathscr{S}$, and a group action $\alpha : G \times \mathscr{S} \to \mathscr{S}$ which defines an equivalence relation $s \overset{\alpha}{\sim} s'$ if, and only if, $\exists U \in G : \alpha(U, s) = s'$. Then a map $\gamma : \mathscr{S} \to \mathscr{S}$ is called a *canonical form for* α [28] if, and only if,

$$\forall s, s' \in \mathscr{S} : \gamma(s) \overset{\alpha}{\sim} s \wedge \left[s \overset{\alpha}{\sim} s' \Leftrightarrow \gamma(s) = \gamma\left(s'\right) \right].$$

Therefore, the set $\mathscr{S}$ is divided into disjoint orbits (i.e., equivalence classes) and the mapping γ picks a unique representative in each equivalence class. In the setup of system equivalence, the group is $G = \mathbf{Gl}_n(\mathbb{R}) \times \mathbf{Gl}_n(\mathbb{R})$, the considered set is $\mathscr{S} = \Sigma_{k,n,m}$ and the group action $\alpha((W, T), [E, A, B]) = [WET, WAT, WB]$ corresponds to $\overset{W^{-1},T^{-1}}{\sim}$. However, Corollary 3.2 does not provide a mapping γ. That means that the form (3.7) is not a unique representative within the equivalence class and hence it is not a canonical form. Nevertheless, we may call it a *normal form*, since every entry is (at least) unique up to similarity.

Remark 3.3 (Canonical forms for regular systems) For regular systems which are completely controllable two actual canonical forms of $[E, A, B] \in \Sigma_{n,n,m}$ under system equivalence have been obtained: the Jordan control canonical form in [64] and, later, the more simple canonical form in [69] based on the Hermite canonical form for controllable ODEs $[I, A, B]$.

Remark 3.4 (DAEs corresponding to the blocks in the quasi-Kronecker form) Corollary 3.2 leads to the separate consideration of the differential-algebraic equations (3.8a)–(3.8c):

(i) (3.8a) is an ordinary differential equation whose solution satisfies

$$x_s(t) = e^{A_s t} x_s(0) + \int_0^t e^{A_s(t-\tau)} B_s u(\tau)\, \mathrm{d}\tau, \quad t \in \mathbb{R}.$$

In particular, solvability is guaranteed by $u \in \mathscr{L}^1_{\mathrm{loc}}(\mathbb{R}; \mathbb{R}^m)$. The initial value $x_s(0) \in \mathbb{R}^n$ can be chosen arbitrarily; the prescription of $u \in \mathscr{L}^1_{\mathrm{loc}}(\mathbb{R}; \mathbb{R}^m)$ and $x_s(0) \in \mathbb{R}^n$ guarantees uniqueness of the solution.

(ii) The solutions of (3.8b) can be calculated by successive differentiation and premultiplication with N_{α_i}, hence we have

$$\begin{aligned} 0 = N_{\alpha_i}^{\alpha_i} x_{f[i]}^{(\alpha_i)}(t) &\overset{(3.8b)}{=} N_{\alpha_i}^{\alpha_i - 1} x_{f[i]}(t)^{(\alpha_i - 1)} + N_{\alpha_i}^{\alpha_i - 1} B_{f[i]} u^{(\alpha_i - 1)}(t) \\ &= \cdots = x_{f[i]}(t) + \sum_{j=0}^{\alpha_i - 1} N_{\alpha_i}^{j} B_{f[i]} u^{(j)}(t), \end{aligned}$$

where $u^{(j)}$ denotes the jth distributional derivative of u. As a consequence, the solution requires a certain smoothness of the input, expressed by

$$\sum_{j=0}^{\alpha_i - 1} N_{\alpha_i}^{j} B_{f[i]} u^{(j)} \in \mathscr{W}^{1,1}_{\mathrm{loc}}\left(\mathbb{R}; \mathbb{R}^{\alpha_i}\right).$$

In particular, condition $u \in \mathscr{W}^{\alpha_i, 1}_{\mathrm{loc}}(\mathbb{R}; \mathbb{R}^{\alpha_i})$ guarantees solvability of the DAE (3.8b). Note that the initial value $x_{f[i]}(0)$ cannot be chosen at all: It is fixed by u via the relation

$$x_{f[i]}(0) = -\left(\sum_{j=0}^{\alpha_i - 1} N_{\alpha_i}^{j} B_{f[i]} u^{(j)} \right)(0).$$

On the other hand, for any (sufficiently smooth) input there exists a unique solution for appropriately chosen initial value.

(iii) Writing

$$x_{u[i]-} = \begin{bmatrix} x_{u[i],1} \\ \vdots \\ x_{u[i],\beta_i-1} \end{bmatrix},$$

(3.8c) is equivalent to

$$\dot{x}_{u[i]-} = N_{\beta_i-1}^{\top} x_{u[i]-} + e_{\beta_i-1}^{[\beta_i-1]} x_{u[i],\beta_i} + B_{u[i]} u(t).$$

Hence, a solution exists for all inputs $u \in \mathscr{L}^1_{\mathrm{loc}}(\mathbb{R};\mathbb{R}^m)$ and all $x_{u[i],\beta_i} \in \mathscr{W}^{1,1}_{\mathrm{loc}}(\mathbb{R};\mathbb{R})$ as well as $x_{u[i],1}(0), \ldots, x_{u[i],\beta_i-1}(0)$. This system is therefore underdetermined in the sense that one component as well as all initial values can be freely chosen. Hence any existing solution for fixed input u and fixed initial value $x_{u[i]}(0)$ is far from being unique.

(iv) Denoting

$$x_{o[i]+} = \begin{bmatrix} 0_{1,1} \\ x_{o[i]} \end{bmatrix},$$

(3.8d) can be rewritten as

$$N_{\gamma_i}^{\top} \dot{x}_{o[i]+} = x_{o[i]+} + B_{o[i]} u(t).$$

Hence we obtain $x_{o[i]+}(t) = -\sum_{j=0}^{\gamma_i-1} (N_{\gamma_i}^{\top})^j B_{o[i]} u^{(j)}(t)$, which gives

$$x_{o[i]}(t) = -[0_{(\gamma_i-1),1}, I_{\gamma_i-1}] \sum_{j=0}^{\gamma_i-1} \left(N_{\gamma_i}^{\top}\right)^j B_{o[i]} u^{(j)}(t)$$

together with the consistency condition on the input:

$$\left(e_1^{[\gamma_i]}\right)^{\top} \sum_{j=0}^{\gamma_i-1} \left(N_{\gamma_i}^{\top}\right)^j B_{o[i]} u^{(j)}(t) = 0. \tag{3.9}$$

The smoothness condition

$$\sum_{j=0}^{\gamma_i-1} \left(N_{\gamma_i}^{\top}\right)^j B_{o[i]} u^{(j)} \in \mathscr{W}^{1,1}_{\mathrm{loc}}\left(\mathbb{R};\mathbb{R}^{\gamma_i}\right)$$

is therefore not enough to guarantee existence of a solution; the additional constraint formed by (3.9) has to be satisfied, too. Furthermore, as in (ii), the initial value $x_{o[i]}(0)$ is fixed by the input u. Hence, a solution does only exist if the consistency conditions on the input and initial value are satisfied, but then the solution is unique.

Remark 3.5 (Solutions on (finite) time intervals) The solution of a DAE $[E, A, B] \in \Sigma_{k,n,m}$ on some time interval $I \subsetneq \mathbb{R}$ can be defined in a straightforward manner (compare (2.2)). By the considerations in Remark 3.4, we can infer that any solution (x, u) on some finite time interval $I \subsetneq \mathbb{R}$ can be extended to a solution on the whole real axis. Consequently, all concepts which have been defined in Sect. 2 could be also made based on solutions on intervals I including zero.

3.3 A Normal Form under Feedback Equivalence

A normal form under feedback transformation (3.1) was first studied for systems governed by ordinary differential equations by Brunovský [32]. In this section we present a generalization of the Brunovský form for general DAE systems $[E, A, B] \in \Sigma_{k,n,m}$ from [105]. For more details of the feedback form and a more geometric point of view on feedback invariants and feedback canonical forms see [87, 105].

Remark 3.6 (Feedback for regular systems) It is known [12, 63] that the class of regular DAE systems is not closed under the action of state feedback. Therefore, in [140] the class of regular systems is divided into the families

$$\Sigma_\theta := \left\{(E, A, B) \in \Sigma_{n,n,m} \mid \det(\cos\theta E - \sin\theta A) \neq 0\right\}, \quad \theta \in [0, \pi),$$

and it is shown that any of these families is dense in the set of regular systems and the union of these families is exactly the set of regular systems. The authors of [140] then introduce the "constant-ratio proportional and derivative" feedback on Σ_θ, i.e.

$$u = F(\cos\theta x - \sin\theta \dot{x}) + v.$$

This feedback leads to a group action and enables them to obtain a generalization of Brunovský's theorem [32] on each of the subsets of completely controllable systems in Σ_θ, see [140, Thm. 6].

Glüsing-Lüerßen [63] derived a canonical form under the unchanged feedback equivalence (3.1) on the set of strongly controllable (called impulse controllability in [63]) regular systems, see [63, Thm. 4.7]. In particular it was shown that this set is closed under the action of a feedback group.

Theorem 3.3 (Normal form under feedback equivalence [105]) *Let* $[E, A, B] \in \Sigma_{k,n,m}$. *Then there exist* $W \in \mathbf{Gl}_k(\mathbb{R})$, $T \in \mathbf{Gl}_n(\mathbb{R})$, $V \in \mathbf{Gl}_m(\mathbb{R})$, $F \in \mathbb{R}^{m,n}$ *such that*

$$[E, A, B] \overset{W,T,V,F}{\underset{fe}{\sim}} \left[\begin{bmatrix} I_{|\alpha|} & 0 & 0 & 0 & 0 & 0 \\ 0 & K_\beta & 0 & 0 & 0 & 0 \\ 0 & 0 & L_\gamma^\top & 0 & 0 & 0 \\ 0 & 0 & 0 & K_\delta^\top & 0 & 0 \\ 0 & 0 & 0 & 0 & N_\kappa & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{n_{\overline{c}}} \end{bmatrix}, \right.$$
$$\left. \begin{bmatrix} N_\alpha^\top & 0 & 0 & 0 & 0 & 0 \\ 0 & L_\beta & 0 & 0 & 0 & 0 \\ 0 & 0 & K_\gamma^\top & 0 & 0 & 0 \\ 0 & 0 & 0 & L_\delta^\top & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{|\kappa|} & 0 \\ 0 & 0 & 0 & 0 & 0 & A_{\overline{c}} \end{bmatrix}, \begin{bmatrix} E_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & E_\gamma & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right], \tag{3.10}$$

for some multi-indices $\alpha, \beta, \gamma, \delta, \kappa$ and a matrix $A_{\overline{c}} \in \mathbb{R}^{n_{\overline{c}}, n_{\overline{c}}}$. This is interpreted, in terms of the DAE (2.1), as follows: $(x, u) \in \mathfrak{B}_{[E,A,B]}$ *if, and only if,*

$$\begin{aligned} \big(x_c(\cdot)^\top, x_u(\cdot)^\top, x_{ob}(\cdot)^\top, x_o(\cdot)^\top, x_f(\cdot)^\top, x_{\overline{c}}(\cdot)^\top\big)^\top &:= Tx(\cdot), \\ \big(u_c(\cdot)^\top, u_{ob}(\cdot)^\top, u_s(\cdot)^\top\big)^\top &:= V\big(u(\cdot) - Fx(\cdot)\big), \end{aligned}$$

with

$$x_c(\cdot) = \begin{pmatrix} x_{c[1]}(\cdot) \\ \vdots \\ x_{c[\ell(\alpha)]}(\cdot) \end{pmatrix}, \quad u_c(\cdot) = \begin{pmatrix} u_{c[1]}(\cdot) \\ \vdots \\ x_{c[\ell(\alpha)]}(\cdot) \end{pmatrix}, \quad x_u(\cdot) = \begin{pmatrix} x_{u[1]}(\cdot) \\ \vdots \\ x_{u[\ell(\beta)]}(\cdot) \end{pmatrix},$$
$$x_{ob}(\cdot) = \begin{pmatrix} x_{ob[1]}(\cdot) \\ \vdots \\ x_{ob[\ell(\gamma)]}(\cdot) \end{pmatrix}, \quad u_{ob}(\cdot) = \begin{pmatrix} u_{ob[1]}(\cdot) \\ \vdots \\ u_{ob[\ell(\gamma)]}(\cdot) \end{pmatrix}, \quad x_o(\cdot) = \begin{pmatrix} x_{o[1]}(\cdot) \\ \vdots \\ x_{o[\ell(\delta)]}(\cdot) \end{pmatrix},$$
$$x_f(\cdot) = \begin{pmatrix} x_{f[1]}(\cdot) \\ \vdots \\ x_{f[\ell(\kappa)]}(\cdot) \end{pmatrix}$$

solves the decoupled DAEs

$$\dot{x}_{c[i]}(t) = N_{\alpha_i}^\top x_c(t) + e_{\alpha_i}^{[\alpha_i]} u_{c[i]}(t) \quad \text{for } i = 1, \ldots, \ell(\alpha), \tag{3.11a}$$
$$K_{\beta_i} \dot{x}_{u[i]}(t) = L_{\beta_i} x_{u[i]}(t) \quad \text{for } i = 1, \ldots, \ell(\beta), \tag{3.11b}$$
$$L_{\gamma_i}^\top \dot{x}_{ob[i]}(t) = K_{\gamma_i}^\top x_{ob[i]}(t) + e_{\gamma_i}^{[\gamma_i]} u_{ob[i]} \quad \text{for } i = 1, \ldots, \ell(\gamma), \tag{3.11c}$$
$$K_{\delta_i}^\top \dot{x}_{o[i]}(t) = L_{\delta_i}^\top x_{o[i]}(t) \quad \text{for } i = 1, \ldots, \ell(\delta), \tag{3.11d}$$

$$N_{\kappa_i}\dot{x}_{f[i]}(t) = x_c(t) \quad \text{for } i = 1, \ldots, \ell(\kappa), \tag{3.11e}$$

$$\dot{x}_{\bar{c}}(t) = A_{\bar{c}}x_{\bar{c}}(t). \tag{3.11f}$$

Note that by Remark 3.2 the form (3.10) is a normal form. However, if we apply an additional state space transformation to the block $[I_{n_{\bar{c}}}, A_{\bar{c}}, 0]$ which puts $A_{\bar{c}}$ into Jordan canonical form, and then prescribe the order of the blocks of each type, e.g. from largest dimension to lowest (what would mean $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_{\ell(\alpha)}$ for α for instance), then (3.10) becomes a canonical form.

Remark 3.7 (DAEs corresponding to the blocks in the feedback form) The form in Theorem 3.3 again leads to the separate consideration of the differential-algebraic equations (3.11a)–(3.11f):

(i) (3.11a) is given by $[I_{\alpha_i}, N_{\alpha_i}^\top, e_{\alpha_i}^{[\alpha_i]}]$, and is completely controllable by the classical results for ODE systems (see e.g. [147, Sect. 3.2]). This system has furthermore the properties of being R-controllable, and both controllable and stabilizable in the behavioral sense.

(ii) (3.11b) corresponds to an underdetermined system with zero dimensional input space. Since $x_{u[i]}$ satisfies (3.11b) if, and only if, there exists some $v_i \in \mathscr{L}^1_{\mathrm{loc}}(\mathbb{R};\mathbb{R})$ with

$$\dot{x}_{u[i]}(t) = N_{\beta_i}^\top x_{u[i]}(t) + e_{\beta_i}^{[\beta_i]} v_i(t),$$

this system has the same properties as (3.11a).

(iii) Denoting

$$z_{ob[i]} = \begin{bmatrix} x_{ob[i]} \\ u_{ob[i]} \end{bmatrix},$$

then (3.11c) can be rewritten as

$$N_{\gamma_i}\dot{z}_{ob[i]}(t) = z_{ob[i]}(t),$$

which has, by (ii) in Remark 3.4, the unique solution $z_{ob[i]} = 0$. Hence,

$$\mathfrak{B}_{[L_{\gamma_i}^\top, K_{\gamma_i}^\top, e_{\gamma_i}^{[\gamma_i]}]} = \{0\}.$$

The system $[L_{\gamma_i}^\top, K_{\gamma_i}^\top, e_{\gamma_i}^{[\gamma_i]}]$ is therefore completely controllable if, and only if, $\gamma_i = 1$. In the case where $\gamma_i > 1$, this system is not even impulse controllable. However, independent of γ_i, $[L_{\gamma_i}^\top, K_{\gamma_i}^\top, e_{\gamma_i}^{[\gamma_i]}]$ is R-controllable, and both controllable and stabilizable in the behavioral sense.

(iv) Again, we have

$$\mathfrak{B}_{[K_{\delta_i}^\top, L_{\delta_i}^\top, 0_{\delta_i,0}]} = \{0\},$$

whence, in dependence on δ_i, we can infer the same properties as in (iii).

(v) Due to

$$\mathfrak{B}_{[N_{\kappa_i}, I_{\kappa_i}, 0_{\kappa_i,0}]} = \{0\},$$

the system $[N_{\kappa_i}, I_{\kappa_i}, 0_{\kappa_i,0}]$ is never controllable at infinity, but always R-controllable and both controllable and stabilizable in the behavioral sense. $[N_{\kappa_i}, I_{\kappa_i}, 0_{\kappa_i,0}]$ is strongly controllable if, and only if, $\kappa_i = 1$.

(vi) The system $[I_{n_{\overline{c}}}, A_{\overline{c}}, 0_{\overline{c},0}]$ satisfies

$$\mathfrak{B}_{[I_{n_{\overline{c}}}, A_{\overline{c}}, 0_{n_{\overline{c}},0}]} = \left\{ e^{A_{\overline{c}} \cdot} x^0 \;\middle|\; x^0 \in \mathbb{R}^{n_{\overline{c}}} \right\},$$

whence it is controllable at infinity, but neither strongly controllable nor controllable in the behavioral sense nor R-controllable. The properties of being complete and strong stabilizability and stabilizability in the behavioral sense are attained if, and only if, $\sigma(A_{\overline{c}}) \subseteq \mathbb{C}_-$.

By using the implications shown in Proposition 2.4, we can deduce the following for the systems arising in the feedback form:

	$[I_{\alpha_i}, N_{\alpha_i}^\top, e_{\alpha_i}^{[\alpha_i]}]$	$[K_{\beta_i}, L_{\beta_i}, 0_{\beta_i-1,0}]$	$[L_{\gamma_i}^\top, K_{\gamma_i}^\top, e_{\gamma_i}^{[\gamma_i]}]$	$[K_{\delta_i}^\top, L_{\delta_i}^\top, 0_{\delta_i,0}]$	$[N_{\kappa_i}, I_{\kappa_i}, 0_{\kappa_i,0}]$	$[I_{n_{\overline{c}}}, A_{\overline{c}}, 0_{\overline{c},0}]$
controllable at infinity	✓	✓	$\Leftrightarrow \gamma_i = 1$	$\Leftrightarrow \delta_i = 1$	×	✓
impulse controllable	✓	✓	$\Leftrightarrow \gamma_i = 1$	$\Leftrightarrow \delta_i = 1$	$\Leftrightarrow \kappa_i = 1$	✓
completely controllable	✓	✓	$\Leftrightarrow \gamma_i = 1$	$\Leftrightarrow \delta_i = 1$	×	×
completely reachable	✓	✓	$\Leftrightarrow \gamma_i = 1$	$\Leftrightarrow \delta_i = 1$	×	×
strongly controllable	✓	✓	$\Leftrightarrow \gamma_i = 1$	$\Leftrightarrow \delta_i = 1$	$\Leftrightarrow \kappa_i = 1$	×
strongly reachable	✓	✓	$\Leftrightarrow \gamma_i = 1$	$\Leftrightarrow \delta_i = 1$	$\Leftrightarrow \kappa_i = 1$	×
completely stabilizable	✓	✓	$\Leftrightarrow \gamma_i = 1$	$\Leftrightarrow \delta_i = 1$	×	$\Leftrightarrow \sigma(A_{\overline{c}}) \subseteq \mathbb{C}_-$
strongly stabilizable	✓	✓	$\Leftrightarrow \gamma_i = 1$	$\Leftrightarrow \delta_i = 1$	$\Leftrightarrow \kappa_i = 1$	$\Leftrightarrow \sigma(A_{\overline{c}}) \subseteq \mathbb{C}_-$
R-controllable	✓	✓	✓	✓	✓	×
controllable in the behavioral sense	✓	✓	✓	✓	✓	×
stabilizable in the behavioral sense	✓	✓	✓	✓	✓	$\Leftrightarrow \sigma(A_{\overline{c}}) \subseteq \mathbb{C}_-$

Corollary 3.4 *A system* $[E, A, B] \in \Sigma_{k,n,m}$ *with feedback form* (3.10) *is*

(a) *controllable at infinity if, and only if,* $\gamma = (1, \ldots, 1)$, $\delta = (1, \ldots, 1)$ *and* $\ell(\kappa) = 0$;

(b) *impulse controllable if, and only if,* $\gamma = (1, \ldots, 1)$, $\delta = (1, \ldots, 1)$ *and* $\kappa = (1, \ldots, 1)$;

(c) *strongly controllable* (*and thus also strongly reachable*) *if, and only if,* $\gamma = (1, \ldots, 1)$, $\delta = (1, \ldots, 1)$, $\kappa = (1, \ldots, 1)$ *and* $n_{\overline{c}} = 0$;

(d) *completely controllable* (*and thus also completely reachable*) *if, and only if,* $\gamma = (1, \dots, 1)$, $\delta = (1, \dots, 1)$ *and* $\ell(\kappa) = n_{\overline{c}} = 0$;
(e) *R-controllable if, and only if,* $n_{\overline{c}} = 0$;
(f) *controllable in the behavioral sense if, and only if,* $n_{\overline{c}} = 0$;
(g) *strongly stabilizable if, and only if,* $\gamma = (1, \dots, 1)$, $\delta = (1, \dots, 1)$, $\ell(\kappa) = 0$, *and* $\sigma(A_{\overline{c}}) \subseteq \mathbb{C}_-$;
(h) *completely stabilizable if and only if,* $\gamma = (1, \dots, 1)$, $\delta = (1, \dots, 1)$, $\kappa = (1, \dots, 1)$, *and* $\sigma(A_{\overline{c}}) \subseteq \mathbb{C}_-$;
(i) *stabilizable in the behavioral sense if, and only if,* $\sigma(A_{\overline{c}}) \subseteq \mathbb{C}_-$.

Remark 3.8 (Parametrization of the behavior of systems in feedback form) With the findings in Remark 3.7, we may explicitly characterize the behavior of systems in feedback form. Define

$$V_k(s) = \left[1, s, \dots, s^k\right]^\top \in \mathbb{R}[s]^{k,1}$$

and, for some multi-index $\mu = (\mu_1, \dots, \mu_l) \in \mathbb{N}^l$,

$$V_\mu(s) = \operatorname{diag}\left(V_{\mu_1}(s), \dots, V_{\mu_l}(s)\right) \in \mathbb{R}[s]^{|\mu|, \ell(\mu)},$$
$$W_\mu(s) = \operatorname{diag}\left(s^{\mu_1}, \dots, s^{\mu_l}\right) \in \mathbb{R}[s]^{\ell(\mu), \ell(\mu)}.$$

Further let $\mu + k := (\mu_1 + k, \dots, \mu_l + k)$ for $k \in \mathbb{Z}$, and

$$\mathscr{W}_{\mathrm{loc}}^{\mu,1}(\mathbb{R}; \mathbb{R}) := \mathscr{W}_{\mathrm{loc}}^{\mu_1,1}(\mathbb{R}; \mathbb{R}) \times \cdots \times \mathscr{W}_{\mathrm{loc}}^{\mu_{\ell(\mu)},1}(\mathbb{R}; \mathbb{R}).$$

Then the behavior of a system in feedback form can, formally, be written as

$$\mathfrak{B}_{[E,A,B]} = \left[\begin{array}{cccc} V_{\alpha-1}(\frac{\mathrm{d}}{\mathrm{d}t}) & 0 & 0 & 0 \\ 0 & V_{\beta-1}(\frac{\mathrm{d}}{\mathrm{d}t}) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & e^{A_{\overline{c}}\cdot} & 0 \\ \hline W_\alpha(\frac{\mathrm{d}}{\mathrm{d}t}) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I \end{array}\right] \cdot \left[\begin{array}{c} \mathscr{W}_{\mathrm{loc}}^{\alpha,1}(\mathbb{R}; \mathbb{R}) \\ \mathscr{W}_{\mathrm{loc}}^{\beta,1}(\mathbb{R}; \mathbb{R}) \\ \mathbb{R}^{n_{\overline{c}}} \\ \mathscr{L}_{\mathrm{loc}}^1(\mathbb{R}; \mathbb{R}^{m-\ell(\alpha)-\ell(\gamma)}) \end{array}\right],$$

where the sizes of the blocks are according to the block structure in the feedback form (3.10) and the horizontal line is the dividing line between x- and u-variables. If the system $[E, A, B] \in \Sigma_{k,n,m}$ is not in feedback form, then a parametrization of the behavior can be found by using the above representation and relation (3.2) expressing the connection between behaviors of feedback equivalent systems.

For general differential behaviors, a parametrization of the above kind is called *image representation* [128, Sect. 6.6].

Remark 3.9 (Derivative feedback) A canonical form under proportional and derivative feedback (PD feedback) was derived in [105] as well (note that PD feedback defines an equivalence relation on $\Sigma_{k,n,m}$). The main tool for doing this is the restriction pencil (see Remark 2.1(xi)): Clearly, the system

$$\begin{aligned} NE\dot{x} &= NAx, \\ u &= B^{\dagger}(E\dot{x} - Ax) \end{aligned}$$

is equivalent, via PD feedback, to the system

$$\begin{aligned} NE\dot{x} &= NAx, \\ u &= 0. \end{aligned}$$

Then putting $sNE - NA$ into Kronecker canonical form yields a PD canonical form for the DAE system with a 5×4-block structure.

We may, however, directly derive this PD canonical form from the normal form (3.10). To this end we may observe that the system $[I_{\alpha_i}, N_{\alpha_i}^{\top}, e_{\alpha_i}^{[\alpha_i]}]$ can be written as

$$K_{\alpha_i}\dot{x}_{c[i]}(t) = L_{\alpha_i}x_{c[i]}(t), \qquad \dot{x}_{c[i],\alpha_i}(t) = u_{c[i]}(t),$$

and hence is, via PD feedback, equivalent to the system

$$\left[\begin{bmatrix} K_{\alpha_i} \\ 0 \end{bmatrix}, \begin{bmatrix} L_{\alpha_i} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right].$$

On the other hand, the system $[L_{\gamma_i}^{\top}, K_{\gamma_i}^{\top}, e_{\gamma_i}^{[\gamma_i]}]$ can be written as

$$N_{\gamma_i-1}\dot{x}_{ob[i]}(t) = x_{ob[i]}(t), \qquad \dot{x}_{ob[i],\gamma_i-1}(t) = u_{ob[i]}(t),$$

and hence is, via PD feedback, equivalent to the system

$$\left[\begin{bmatrix} N_{\gamma_i-1} \\ 0 \end{bmatrix}, \begin{bmatrix} I_{\gamma_i-1} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right].$$

A canonical form for $[E, A, B] \in \Sigma_{k,n,m}$ under PD feedback is therefore given by

$$[E, A, B] \sim_{PD} \left[\begin{bmatrix} K_\beta & 0 & 0 & 0 \\ 0 & K_\delta^{\top} & 0 & 0 \\ 0 & 0 & N_\kappa & 0 \\ 0 & 0 & 0 & I_{n_{\bar{c}}} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} L_\beta & 0 & 0 & 0 \\ 0 & L_\delta^{\top} & 0 & 0 \\ 0 & 0 & I_{|\kappa|} & 0 \\ 0 & 0 & 0 & A_{\bar{c}} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ I_\zeta & 0 \end{bmatrix}\right],$$

where $A_{\bar{c}}$ is in Jordan canonical form, and the blocks of each type are ordered from largest dimension to lowest.

Note that the properties of complete controllability, controllability at infinity and controllability in the behavioral sense are invariant under PD feedback. However,

since derivative feedback changes the set of differential variables, the properties of strong controllability as well as impulse controllability may be lost/gained after PD feedback.

Remark 3.10 (Connection to Kronecker form) We may observe from (3.1) that feedback transformation may be alternatively considered as a transformation of the extended pencil

$$s\mathscr{E} - \mathscr{A} = \left[sE - A, \quad -B \right], \tag{3.12}$$

that is based on a multiplication from the left by $\mathscr{W} = W \in \mathbf{Gl}_k(\mathbb{R})$, and from the right by

$$\mathscr{T} = \begin{bmatrix} T & 0 \\ F & V \end{bmatrix} \in \mathbf{Gl}_{n+m}(\mathbb{R}).$$

This equivalence is therefore a subclass of the class which is induced by the pre- and post-multiplication of $s\mathscr{E} - \mathscr{A}$ by arbitrary invertible matrices. Loosely speaking, one can hence expect a normal form under feedback equivalence which specializes the quasi-Kronecker form of $s\mathscr{E} - \mathscr{A}$. Indeed, the latter form may be obtained from the feedback form of $[E, A, B]$ by several simple row transformations $s\mathscr{E} - \mathscr{A}$ which are not interpretable as feedback group actions anymore. More precisely, simple permutations of columns lead to the separate consideration of the extended pencils corresponding to the systems (3.11a)–(3.11f): The extended pencils corresponding to $[I_{\alpha_i}, N_{\alpha_i}^\top, e_{\alpha_i}^{[\alpha_i]}]$ and $[K_{\beta_i}, L_{\beta_i}, 0_{\alpha_i,0}]$ are $sK_{\alpha_i} - L_{\alpha_i}$ and $sK_{\beta_i} - L_{\beta_i}$, resp. The extended matrix pencil corresponding to the system $[L_{\gamma_i}^\top, K_{\gamma_i}^\top, e_{\gamma_i}^{[\gamma_i]}]$ is given by $sN_{\gamma_i} - I_{\gamma_i}$. The extended matrix pencils corresponding to the systems $[K_{\delta_i}^\top, L_{\delta_i}^\top, 0_{\delta_i,0}]$, $[N_{\kappa_i}, I_{\kappa_i}, 0_{\kappa_i,0}]$ and $[I_{n_{\overline{c}}}, A_{\overline{c}}, 0_{\overline{c},0}]$ are obviously given by $sK_{\delta_i}^\top - L_{\delta_i}^\top$, $sN_{\kappa_i} - I_{\kappa_i}$ and $sI_{n_{\overline{c}}} - A_{\overline{c}}$, respectively. In particular, $\lambda \in \mathbb{C}$ is a generalized eigenvalue of $s\mathscr{E} - \mathscr{A}$, if, and only if, $\lambda \in \sigma(A_{\overline{c}})$.

Remark 3.11 (Minimality in the behavioral sense)

(i) According to Remark 3.1, a differential-algebraic system $[E, A, B] \in \Sigma_{k,n,m}$ is minimal in the behavioral sense, if, and only if, the extended pencil $s\mathscr{E} - \mathscr{A}$ as in (3.12) has full row rank as a matrix with entries in the field $\mathbb{R}(s)$. On the other hand, a system $[E, A, B] \in \Sigma_{k,n,m}$ with feedback form (3.10) satisfies

$$\mathrm{rk}_{\mathbb{R}(s)}(s\mathscr{E} - \mathscr{A}) = k - \ell(\delta).$$

Using that $\mathrm{rk}_{\mathbb{R}(s)}(s\mathscr{E} - \mathscr{A})$ is invariant under feedback transformation (3.1), we can conclude that minimality of $[E, A, B] \in \Sigma_{k,n,m}$ in the behavioral sense corresponds to the absence of blocks of type (3.11d) in its feedback form.

(ii) The findings in Remark 3.4 imply that a system in feedback form is, in the behavioral sense, equivalent to

$$\left[\begin{bmatrix} I_{|\alpha|} & 0 & 0\,0\,0 & 0 \\ 0 & K_\beta & 0\,0\,0 & 0 \\ 0 & 0 & 0\,0\,0 & 0 \\ 0 & 0 & 0\,0\,0 & 0 \\ 0 & 0 & 0\,0\,0 & 0 \\ 0 & 0 & 0\,0\,0 & I_{n_{\overline{c}}} \end{bmatrix}, \begin{bmatrix} N_\alpha^\top & 0 & 0 & 0 & 0 & 0 \\ 0 & L_\beta & 0 & 0 & 0 & 0 \\ 0 & 0 & K_\gamma^\top & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{|\delta|-\ell(\delta)} & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{|\kappa|} & 0 \\ 0 & 0 & 0 & 0 & 0 & A_{\overline{c}} \end{bmatrix}, \begin{bmatrix} E_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & E_\gamma & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\right].$$

This system can alternatively be achieved by multiplying the extended pencil (3.12) in feedback form (3.10) from the left with the polynomial matrix

$$Z(s) = \operatorname{diag}\left(I_{|\alpha|}, I_{|\beta|-\ell(\beta)}, -\sum_{k=0}^{\nu_\gamma-1} s^k N_\gamma^k, P_\delta(s), -\sum_{k=0}^{\nu_\kappa-1} s^k N_\kappa^k, I_{n_{\overline{c}}}\right),$$

where $\nu_\gamma = \max\{\gamma_1, \ldots, \gamma_{\ell(\gamma)}\}$, $\nu_\kappa = \max\{\kappa_1, \ldots, \kappa_{\ell(\kappa)}\}$, and

$$P_\delta(s) = \operatorname{diag}\left(\left[\begin{array}{cc} 0_{\delta_i-1,1}, & -\sum_{k=0}^{\delta_i-2} s^k (N_{\delta_i-1}^\top)^k \end{array}\right]\right)_{j=1,\ldots,\ell(\delta)}.$$

(iii) Let a differential-algebraic system $[E, A, B] \in \Sigma_{k,n,m}$ be given. Using the notation from (3.10) and the previous item, a behaviorally equivalent and minimal system $[E_M, A_M, B_M] \in \Sigma_{k-\ell(\delta),n,m}$ can be constructed by

$$\begin{bmatrix} sE_M - A_M, & -B_M \end{bmatrix} = Z(s) W \begin{bmatrix} sE - A, & -B \end{bmatrix}.$$

It can be seen that this representation is furthermore controllable at infinity. As well, it minimizes, among all differential-algebraic equations representing the same behavior, the index and the rank of the matrix in front of the state derivative (that is, loosely speaking, the number of differential variables). This procedure is very closely related to *index reduction* [96, Sect. 6.1].

4 Criteria of Hautus Type

In this section we derive equivalent criteria on the matrices $E, A \in \mathbb{R}^{k,n}$, $B \in \mathbb{R}^{k,m}$ for the controllability and stabilizability concepts of Definition 2.1. The criteria are generalizations of the Hautus test (also called Popov–Belevitch–Hautus test, since independently developed by Popov [129], Belevitch [17] and Hautus [68]) in terms of rank criteria on the involved matrices. Note that these conditions are not new—we refer to the relevant literature. However, we provide new proofs using only the feedback normal form (3.10).

First we show that certain rank criteria on the matrices involved in control systems are invariant under feedback equivalence. After that, we relate these rank criteria to the feedback form (3.10).

Lemma 4.1 *Let* $[E_1, A_1, B_1], [E_2, A_2, B_2] \in \Sigma_{k,n,m}$ *be given such that for* $W \in \mathbf{Gl}_k(\mathbb{R})$, $T \in \mathbf{Gl}_n(\mathbb{R})$, $V \in \mathbf{Gl}_m(\mathbb{R})$ *and* $F \in \mathbb{R}^{m,n}$, *we have*

$$[E_1, A_1, B_1] \overset{W,T,V,F}{\underset{fe}{\sim}} [E_2, A_2, B_2].$$

Then

$$\begin{aligned}
\operatorname{im}_{\mathbb{R}} E_1 + \operatorname{im}_{\mathbb{R}} A_1 + \operatorname{im}_{\mathbb{R}} B_1 &= W \cdot (\operatorname{im}_{\mathbb{R}} E_2 + \operatorname{im}_{\mathbb{R}} A_2 + \operatorname{im}_{\mathbb{R}} B_2),\\
\operatorname{im}_{\mathbb{R}} E_1 + A_1 \cdot \ker_{\mathbb{R}} E_1 + \operatorname{im}_{\mathbb{R}} B_1 &= W \cdot (\operatorname{im}_{\mathbb{R}} E_2 + A_2 \cdot \ker_{\mathbb{R}} E_2 + \operatorname{im}_{\mathbb{R}} B_2),\\
\operatorname{im}_{\mathbb{R}} E_1 + \operatorname{im}_{\mathbb{R}} B_1 &= W \cdot (\operatorname{im}_{\mathbb{R}} E_2 + \operatorname{im}_{\mathbb{R}} B_2),\\
\operatorname{im}_{\mathbb{C}}(\lambda E_1 - A_1) + \operatorname{im}_{\mathbb{C}} B_1 &= W \cdot \big(\operatorname{im}_{\mathbb{C}}(\lambda E_2 - A_2) + \operatorname{im}_{\mathbb{C}} B_2\big) \quad \textit{for all } \lambda \in \mathbb{C},\\
\operatorname{im}_{\mathbb{R}(s)}(sE_1 - A_1) + \operatorname{im}_{\mathbb{R}(s)} B_1 &= W \cdot \big(\operatorname{im}_{\mathbb{R}(s)}(sE_2 - A_2) + \operatorname{im}_{\mathbb{R}(s)} B_2\big).
\end{aligned}$$

Proof Immediate from (3.1). □

Lemma 4.2 (Algebraic criteria via feedback form) *For a system* $[E, A, B] \in \Sigma_{k,n,m}$ *with feedback form* (3.10) *the following statements hold true*:

(a)

$$\begin{aligned}
&\operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} A + \operatorname{im}_{\mathbb{R}} B = \operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} B\\
\iff &\gamma = (1, \ldots, 1), \delta = (1, \ldots, 1), \ell(\kappa) = 0.
\end{aligned}$$

(b)

$$\begin{aligned}
&\operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} A + \operatorname{im}_{\mathbb{R}} B = \operatorname{im}_{\mathbb{R}} E + A \cdot \ker_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} B\\
\iff &\gamma = (1, \ldots, 1), \delta = (1, \ldots, 1), \kappa = (1, \ldots, 1).
\end{aligned}$$

(c)

$$\begin{aligned}
&\operatorname{im}_{\mathbb{C}} E + \operatorname{im}_{\mathbb{C}} A + \operatorname{im}_{\mathbb{R}} B = \operatorname{im}_{\mathbb{C}}(\lambda E - A) + \operatorname{im}_{\mathbb{C}} B\\
\iff &\delta = (1, \ldots, 1), \lambda \notin \sigma(A_{\overline{c}}).
\end{aligned}$$

(d) *For* $\lambda \in \mathbb{C}$ *we have*

$$\begin{aligned}
&\dim\big(\operatorname{im}_{\mathbb{R}(s)}(sE - A) + \operatorname{im}_{\mathbb{R}(s)} B\big) = \dim\big(\operatorname{im}_{\mathbb{C}}(\lambda E - A) + \operatorname{im}_{\mathbb{C}} B\big)\\
\iff &\lambda \notin \sigma(A_{\overline{c}}).
\end{aligned}$$

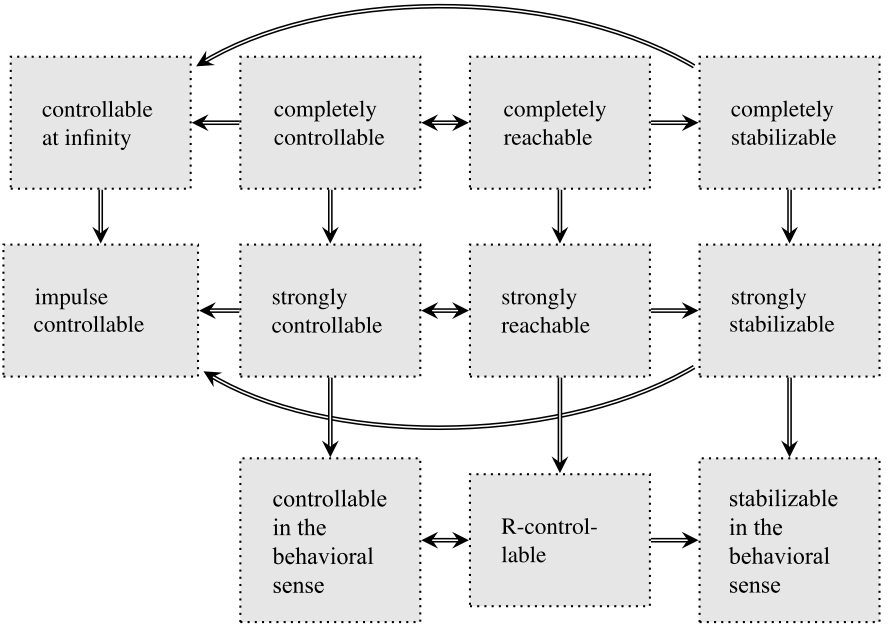
Proof It is, by Lemma 4.1, no loss of generality to assume that $[E, A, B]$ is already in feedback normal form. The results then follow by a simple verification of the above statements by means of the feedback form. □

Combining Lemmas 4.1 and 4.2 with Corollary 3.4, we may deduce the following criteria for the controllability and stabilizability concepts introduced in Definition 2.1.

Corollary 4.3 (Algebraic criteria for controllability/stabilizability) *Let a system* $[E, A, B] \in \Sigma_{k,n,m}$ *be given. Then the following holds*:

$[E, A, B]$ *is*	*if, and only if,*
controllable at infinity	$\operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} A + \operatorname{im}_{\mathbb{R}} B = \operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} B.$
impulse controllable	$\operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} A + \operatorname{im}_{\mathbb{R}} B = \operatorname{im}_{\mathbb{R}} E + A \cdot \ker_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} B.$
completely controllable	$\operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} A + \operatorname{im}_{\mathbb{R}} B = \operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} B$ $\wedge \operatorname{im}_{\mathbb{C}} E + \operatorname{im}_{\mathbb{C}} A + \operatorname{im}_{\mathbb{C}} B = \operatorname{im}_{\mathbb{C}}(\lambda E - A) + \operatorname{im}_{\mathbb{C}} B \ \forall \lambda \in \mathbb{C}.$
strongly controllable	$\operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} A + \operatorname{im}_{\mathbb{R}} B = A \cdot \ker_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} B$ $\wedge \operatorname{im}_{\mathbb{C}} E + \operatorname{im}_{\mathbb{C}} A + \operatorname{im}_{\mathbb{C}} B = \operatorname{im}_{\mathbb{C}}(\lambda E - A) + \operatorname{im}_{\mathbb{C}} B \ \forall \lambda \in \mathbb{C}.$
completely stabilizable	$\operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} A + \operatorname{im}_{\mathbb{R}} B = \operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} B$ $\wedge \operatorname{im}_{\mathbb{C}} E + \operatorname{im}_{\mathbb{C}} A + \operatorname{im}_{\mathbb{C}} B = \operatorname{im}_{\mathbb{C}}(\lambda E - A) + \operatorname{im}_{\mathbb{C}} B \ \forall \lambda \in \overline{\mathbb{C}}_+.$
strongly stabilizable	$\operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} A + \operatorname{im}_{\mathbb{R}} B = \operatorname{im}_{\mathbb{R}} E + A \cdot \ker_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} B$ $\wedge \operatorname{im}_{\mathbb{C}} E + \operatorname{im}_{\mathbb{C}} A + \operatorname{im}_{\mathbb{C}} B = \operatorname{im}_{\mathbb{C}}(\lambda E - A) + \operatorname{im}_{\mathbb{C}} B \ \forall \lambda \in \overline{\mathbb{C}}_+.$
controllable in the behavioral sense	$\operatorname{rk}_{\mathbb{R}(s)}[sE - A, B] = \operatorname{rk}_{\mathbb{C}}[\lambda E - A, B] \ \forall \lambda \in \mathbb{C}.$
stabilizable in the behavioral sense	$\operatorname{rk}_{\mathbb{R}(s)}[sE - A, B] = \operatorname{rk}_{\mathbb{C}}[\lambda E - A, B] \ \forall \lambda \in \overline{\mathbb{C}}_+.$

The above result leads to the following extension of the diagram in Proposition 2.4. Note that the equivalence of R-controllability and controllability in the behavioral sense was already shown in Corollary 3.4.



In the following we will consider further criteria for the concepts introduced in Definition 2.1.

Remark 4.1 (Controllability at infinity) Corollary 4.3 immediately implies that controllability at infinity is equivalent to

$$\operatorname{im}_{\mathbb{R}} A \subseteq \operatorname{im}_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} B.$$

In terms of a rank criterion, this is the same as

$$\operatorname{rk}_{\mathbb{R}}[E, A, B] = \operatorname{rk}_{\mathbb{R}}[E, B]. \tag{4.1}$$

Criterion (4.1) has first been derived by Geerts [61, Thm. 4.5] for the case $\operatorname{rk}[E, A, B] = k$, although he does not use the name "controllability at infinity".

In the case of regular $sE - A \in \mathbb{R}[s]^{n,n}$, condition (4.1) reduces to

$$\operatorname{rk}_{\mathbb{R}}[E, B] = n.$$

Remark 4.2 (Impulse controllability) By Corollary 4.3, impulse controllability of $[E, A, B] \in \Sigma_{k,n,m}$ is equivalent to

$$\operatorname{im}_{\mathbb{R}} A \subseteq \operatorname{im}_{\mathbb{R}} E + A \cdot \ker_{\mathbb{R}} E + \operatorname{im}_{\mathbb{R}} B.$$

Another equivalent characterization is that, for one (and hence any) matrix Z with $\operatorname{im}_{\mathbb{R}}(Z) = \ker_{\mathbb{R}}(E)$, we have

$$\operatorname{rk}_{\mathbb{R}}[E, A, B] = \operatorname{rk}_{\mathbb{R}}[E, AZ, B]. \tag{4.2}$$

This has first been derived by Geerts [61, Rem. 4.9], again for the case $\mathrm{rk}[E, A, B] = k$. In [75, Thm. 3] and [71] the result has been obtained that impulse controllability is equivalent to

$$\mathrm{rk}_{\mathbb{R}} \begin{bmatrix} E & 0 & 0 \\ A & E & B \end{bmatrix} = \mathrm{rk}_{\mathbb{R}}[E, A, B] + \mathrm{rk}_{\mathbb{R}} E,$$

which is in fact equivalent to (4.2). It has also been shown in [75, p. 1] that impulse controllability is equivalent to

$$\mathrm{rk}_{\mathbb{R}(s)}(s\mathscr{E} - \mathscr{A}) = \mathrm{rk}_{\mathbb{R}}[E, A, B].$$

This criterion can be alternatively shown by using the feedback form (3.10). Using condition (3.5) we may also infer that this is equivalent to the index of the extended pencil $s\mathscr{E} - \mathscr{A} \in \mathbb{R}[s]^{k,n+m}$ being at most one.

If the pencil $sE - A$ is regular, then condition (4.2) reduces to

$$\mathrm{rk}_{\mathbb{R}}[E, AZ, B] = n.$$

This condition can be also inferred from [49, Th. 2-2.3].

Remark 4.3 (Controllability in the behavioral sense and R-controllability) The concepts of controllability in the behavioral sense and R-controllability are equivalent by Corollary 3.4. The algebraic criterion for behavioral controllability in Corollary 4.3 is equivalent to the extended matrix pencil $s\mathscr{E} - \mathscr{A} \in \mathbb{R}[s]^{k,n+m}$ having no generalized eigenvalues, or, equivalently, in the feedback form (3.10) it holds $n_{\overline{c}} = 0$.

The criterion for controllability in the behavioral sense is shown in [128, Thm. 5.2.10] for the larger class of linear differential behaviors. R-controllability for systems with regular $sE - A$ was considered in [49, Thm. 2-2.2], where the condition

$$\mathrm{rk}_{\mathbb{C}}[\lambda E - A, B] = n \quad \forall \lambda \in \mathbb{C}$$

was derived. This is, for regular $sE - A$, in fact equivalent to the criterion for behavioral stabilizability in Corollary 4.3.

Remark 4.4 (Complete controllability and strong controllability) By Corollary 4.3, complete controllability of $[E, A, B] \in \Sigma_{k,n,m}$ is equivalent to $[E, A, B]$ being R-controllable and controllable at infinity, whereas strong controllability of $[E, A, B] \in \Sigma_{k,n,m}$ is equivalent to $[E, A, B]$ being R-controllable and impulse controllable.

Banaszuk et al. [12] already obtained the condition in Corollary 4.3 for complete controllability considering discrete systems. Complete controllability is called $\mathscr{H}$-controllability in [12]. Recently, Zubova [171] considered full controllability, which is just complete controllability with the additional assumption that solutions have to be unique, and obtained three equivalent criteria [171, Sect. 7], where the first one

characterizes the uniqueness and the other two are equivalent to the condition for complete controllability in Corollary 4.3.

For regular systems, the conditions in Corollary 4.3 for complete and strong controllability are also derived in [49, Thm. 2-2.1 & Thm. 2-2.3].

Remark 4.5 (Stabilizability) By Corollary 4.3, complete stabilizability of $[E, A, B] \in \Sigma_{k,n,m}$ is equivalent to $[E, A, B]$ being stabilizable in the behavioral sense and controllable at infinity, whereas strong stabilizability of $[E, A, B] \in \Sigma_{k,n,m}$ is equivalent to $[E, A, B]$ being stabilizable in the behavioral sense and impulse controllable.

The criterion for stabilizability in the behavioral sense is shown in [128, Thm. 5.2.30] for the class of linear differential behaviors.

Remark 4.6 (Kalman criterion for regular systems) For regular systems $[E, A, B] \in \Sigma_{n,n,m}$ with $\det(sE - A) \in \mathbb{R}[s] \setminus \{0\}$ the usual Hautus and Kalman criteria can be found in a summarized form e.g. in [49]. Other approaches to derive controllability criteria rely on the expansion of $(sE - A)^{-1}$ as a power series in s, which is only feasible in the regular case. For instance, in [115] the numerator matrices of this expansion, i.e., the coefficients of the polynomial $\operatorname{adj}(sE - A)$, are used to derive a rank criterion for complete controllability. Then again, in [90] Kalman rank criteria for complete controllability, R-controllability and controllability at infinity are derived in terms of the coefficients of the power series expansion of $(sE - A)^{-1}$. The advantage of these criteria, especially the last one, is that no transformation of the system needs to be performed as it is usually necessary in order to derive Kalman rank criteria for DAEs, see e.g. [49].

However, simple criteria can be obtained using only a left transformation of little impact: if $\alpha \in \mathbb{R}$ is chosen such that $\det(\alpha E - A) \neq 0$ then the system is complete controllable if, and only if, [170, Cor. 1]

$$\operatorname{rk}_{\mathbb{R}}\big[(\alpha E - A)^{-1}B, \big((\alpha E - A)^{-1}E\big)(\alpha E - A)^{-1}B, \ldots \\ \ldots, \big((\alpha E - A)^{-1}E\big)^{n-1}(\alpha E - A)^{-1}B\big] = n,$$

and it is impulse controllable if, and only if, [170, Thm. 2]

$$\operatorname{im}_{\mathbb{R}}(\alpha E - A)^{-1}E + \ker(\alpha E - A)^{-1}E + \operatorname{im}_{\mathbb{R}}(\alpha E - A)^{-1}B = \mathbb{R}^n.$$

The result concerning complete controllability has also been obtained in [41, Thm. 4.1] for the case $A = I$ and $\alpha = 0$.

Yet another approach was followed by Kučera and Zagalak [94] who introduced controllability indices and characterized strong controllability in terms of an equation for these indices.

5 Feedback, Stability and Autonomous Systems

State feedback is, roughly speaking, the special choice of the input being a function of the state. Due to the mutual dependence of state and input in a feedback system, this is often referred to as *closed-loop control*. In the linear case, feedback is the imposition of the additional relation $u(t) = Fx(t)$ for some $F \in \mathbb{R}^{m,n}$. This results in the system

$$E\dot{x}(t) = (A + BF)x(t).$$

Feedback for linear ODE systems was studied by Wonham [165], where it is shown that controllability of $[I, A, B] \in \Sigma_{n,n,m}$ is equivalent to any set $\Lambda \subseteq \mathbb{C}$ which has at most n elements and is symmetric with respect to the imaginary axis (that is, $\lambda \in \Lambda \Leftrightarrow \overline{\lambda} \in \Lambda$) being achievable by a suitable feedback, i.e., there exists some $F \in \mathbb{R}^{m,n}$ with the property that $\sigma(A + BF) = \Gamma$. In particular, the input may be chosen in a way that the closed-loop system is stable, i.e., any state trajectory tends to zero. Using the *Kalman decomposition* [82] (see also Sect. 7), it can be shown for ODE systems that stabilizability is equivalent to the existence of a feedback such that the resulting system is stable.

These results have been generalized to regular DAE systems by Cobb [43], see also [49, 57, 102, 103, 121, 123]. Note that, for DAE systems, not only the problem of assignment of eigenvalues occurs, but also the index may be changed by imposing feedback.

The crucial ingredient for the treatment of DAE systems with non-regular pencil $sE - A$ will be the feedback form by Loiseau et al. [105] (see Thm. 3.3).

5.1 Stabilizability, Autonomy and Stability

The feedback law $u(t) = Fx(t)$ applied to (2.1) results in a DAE in which the input is completely eliminated. We now focus on DAEs without input, and we introduce several properties and concepts. For matrices $E, A \in \mathbb{R}^{k,n}$, consider a DAE

$$E\dot{x}(t) = Ax(t). \tag{5.1}$$

Its *behavior* is given by

$$\mathfrak{B}_{[E,A]} := \left\{ x \in \mathscr{W}^{1,1}_{\mathrm{loc}}\left(\mathbb{R}; \mathbb{R}^n\right) \;\middle|\; x \text{ satisfies (5.1) for almost all } t \in \mathbb{R} \right\}.$$

Definition 5.1 (Stability/Stabilizability concepts for DAEs, autonomous DAEs) A linear time-invariant DAE $[E, A] \in \Sigma_{k,n}$ is called

(a) *completely stabilizable*

$$:\Leftrightarrow \forall x^0 \in \mathbb{R}^n \exists x \in \mathfrak{B}_{[E,A]} : x(0) = x^0 \wedge \lim_{t\to\infty} x(t) = 0;$$

(b) *strongly stabilizable*

$$:\Leftrightarrow \forall x^0 \in \mathbb{R}^n \exists x \in \mathfrak{B}_{[E,A]} : Ex(0) = Ex^0 \wedge \lim_{t\to\infty} x(t) = 0;$$

(c) *stabilizable in the behavioral sense*

$$:\Leftrightarrow \forall x \in \mathscr{B}_{[E,A]} \exists x_0 \in \mathscr{B}_{[E,A]} : \big(\forall t < 0 : x(t) = x_0(t)\big) \wedge \lim_{t\to\infty} x_0(t) = 0;$$

(d) *autonomous*

$$:\Leftrightarrow \forall x_1, x_2 \in \mathfrak{B}_{[E,A]} : \big(\forall t < 0 : x_1(t) = x_2(t)\big) \Rightarrow \big(\forall t \in \mathbb{R} : x_1(t) = x_2(t)\big);$$

(e) *completely stable*

$$:\Leftrightarrow \big\{x(0) \;\big|\; x \in \mathfrak{B}_{[E,A]}\big\} = \mathbb{R}^n \wedge \forall x \in \mathfrak{B}_{[E,A]} : \lim_{t\to\infty} x(t) = 0;$$

(d) *strongly stable*

$$:\Leftrightarrow \big\{Ex(0) \;\big|\; x \in \mathfrak{B}_{[E,A]}\big\} = \operatorname{im}_{\mathbb{R}} E \wedge \forall x \in \mathfrak{B}_{[E,A]} : \lim_{t\to\infty} x(t) = 0;$$

(g) *stable in the behavioral sense*

$$:\Leftrightarrow \forall x \in \mathfrak{B}_{[E,A]} : \lim_{t\to\infty} x(t) = 0.$$

Remark 5.1 (Stabilizable and autonomous DAEs are stable) The notion of autonomy is introduced by Polderman and Willems in [128, Sect. 3.2] for general behaviors. For DAE systems $E\dot{x}(t) = Ax(t)$ we can further conclude that autonomy is equivalent to any $x \in \mathfrak{B}_{[E,A]}$ being uniquely determined by $x(0)$. This gives also rise to the fact that autonomy is equivalent to $\dim_{\mathbb{R}} \mathfrak{B}_{[E,A]} \leq n$ which is, on the other hand, equivalent to $\dim_{\mathbb{R}} \mathfrak{B}_{[E,A]} < \infty$. Autonomy indeed means that the DAE is not underdetermined.

Moreover, due to possible underdetermined blocks of type $[K_\beta, L_\beta, 0_{|\beta|-\ell(\beta),0}]$, in general there are solutions $x \in \mathfrak{B}_{[E,A]}$ which grow unboundedly. As a consequence, for a quasi-Kronecker form of any completely stable, strongly stable or behavioral stable DAE, $\ell(\beta) = 0$ holds. Hence, systems of this type are autonomous. In fact, complete, strong and behavioral stability are equivalent to the respective stabilizability notion together with autonomy, cf. also Corollary 5.1.

In regard of Remark 3.4 we can infer the following:

Corollary 5.1 (Stability/Stabilizability criteria and quasi-Kronecker form) *Let* $[E, A] \in \Sigma_{k,n}$ *and assume that the quasi-Kronecker form of* $sE - A$ *is given by* (3.3). *Then the following holds true*:

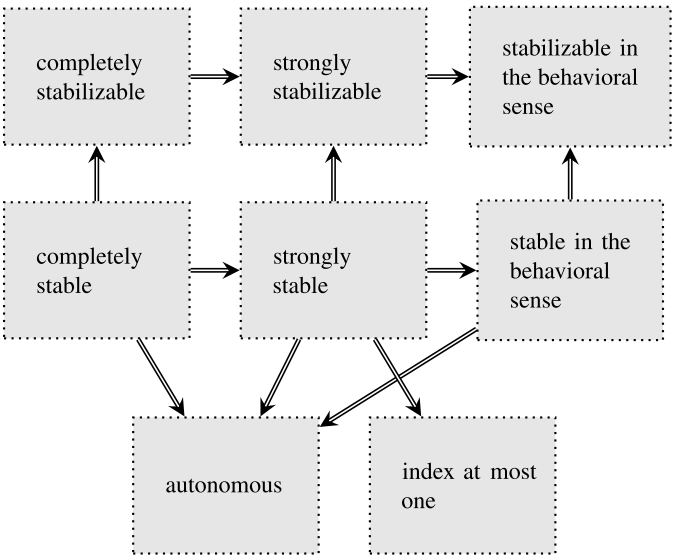
$[E, A]$ *is*	*if, and only if,*
completely stabilizable	$\ell(\alpha) = 0$, $\gamma = (1, \ldots, 1)$ *and* $\sigma(A_s) \subseteq \mathbb{C}_-$.
strongly stabilizable	$\alpha = (1, \ldots, 1)$, $\gamma = (1, \ldots, 1)$ *and* $\sigma(A_s) \subseteq \mathbb{C}_-$.
stabilizable in the behavioral sense	$\sigma(A_s) \subseteq \mathbb{C}_-$.
autonomous	$\ell(\beta) = 0$.
completely stable	$\ell(\alpha) = 0$, $\ell(\beta) = 0$, $\gamma = (1, \ldots, 1)$ *and* $\sigma(A_s) \subseteq \mathbb{C}_-$.
strongly stable	$\alpha = (1, \ldots, 1)$, $\ell(\beta) = 0$, $\gamma = (1, \ldots, 1)$ *and* $\sigma(A_s) \subseteq \mathbb{C}_-$.
stable in the behavioral sense	$\ell(\beta) = 0$, $\sigma(A_s) \subseteq \mathbb{C}_-$.

The subsequent algebraic criteria for the previously defined notions of stabilizability and autonomy can be inferred from Corollary 5.1 by using further arguments similar to the ones of Sect. 4.

Corollary 5.2 (Algebraic criteria for stabilizability) *Let* $[E, A] \in \Sigma_{k,n}$. *Then the following holds true*:

$[E, A]$ *is*	*if, and only if,*
completely stabilizable	$\operatorname{im}_{\mathbb{R}} A \subseteq \operatorname{im}_{\mathbb{R}} E$ *and* $\operatorname{rk}_{\mathbb{R}(s)}(sE - A) = \operatorname{rk}_{\mathbb{C}}(\lambda E - A)$ *for all* $\lambda \in \overline{\mathbb{C}}_+$.
strongly stabilizable	$\operatorname{im}_{\mathbb{R}} A \subseteq \operatorname{im}_{\mathbb{R}} E + A \cdot \ker_{\mathbb{R}} E$ *and* $\operatorname{rk}_{\mathbb{R}(s)}(sE - A) = \operatorname{rk}_{\mathbb{C}}(\lambda E - A)$ *for all* $\lambda \in \overline{\mathbb{C}}_+$.
stabilizable in the behavioral sense	$\operatorname{rk}_{\mathbb{R}(s)}(sE - A) = \operatorname{rk}_{\mathbb{C}}(\lambda E - A)$ *for all* $\lambda \in \overline{\mathbb{C}}_+$.
autonomous	$\ker_{\mathbb{R}(s)}(sE - A) = \{0\}$.

Corollary 5.2 leads to the following implications:



Remark 5.2

(i) Strong stabilizability implies that the index of $sE - A$ is at most one. In the case where the matrix $[E, A] \in \mathbb{R}^{k,2n}$ has full row rank, complete stabilizability is sufficient for the index of $sE - A$ being zero.
On the other hand, behavioral stabilizability of $[E, A]$ together with the index of $sE - A$ being not greater than one implies strong stabilizability of $[E, A]$. Furthermore, for systems $[E, A] \in \Sigma_{k,n}$ with $\mathrm{rk}_{\mathbb{R}}[E, A] = k$, complete stabilizability is equivalent to behavioral stabilizability together with the property that the index of $sE - A$ is zero.
For ODEs the notions of complete stabilizability, strong stabilizability, stabilizability in the behavioral sense, complete stability, strong stability and stability in the behavioral sense are equivalent.

(ii) The behavior of an autonomous system $[E, A]$ satisfies $\dim_{\mathbb{R}} \mathfrak{B}_{[E,A]} = n_s$, where n_s denotes the number of rows of the matrix A_s in the quasi-Kronecker form (3.3) of $sE - A$. Note that regularity of $sE - A$ is sufficient for autonomy of $[E, A]$.

(iii) Autonomy has been algebraically characterized for linear differential behaviors in [128, Sect. 3.2]. The characterization of autonomy in Corollary 5.2 can indeed be generalized to a larger class of linear differential equations.

5.2 Stabilization by Feedback

A system $[E, A, B] \in \Sigma_{k,n,m}$ can, via state feedback with some $F \in \mathbb{R}^{m,n}$, be turned into a DAE $[E, A + BF] \in \Sigma_{k,n}$. We now present some properties of

$[E, A + BF] \in \Sigma_{k,n}$ that can be achieved by a suitable feedback matrix $F \in \mathbb{R}^{m,n}$. Recall that the stabilizability concepts for a system $[E, A, B] \in \Sigma_{k,n,m}$ have been defined in Definition 2.1.

Theorem 5.3 (Stabilizing feedback) *For a system $[E, A, B] \in \Sigma_{k,n,m}$ the following holds true*:

(a) $[E, A, B]$ *is impulse controllable if, and only if, there exists $F \in \mathbb{R}^{m,n}$ such that the index of $sE - (A + BF)$ is at most one.*
(b) $[E, A, B]$ *is completely stabilizable if, and only if, there exists $F \in \mathbb{R}^{m,n}$ such that $[E, A + BF]$ is completely stabilizable.*
(c) $[E, A, B]$ *is strongly stabilizable if, and only if, there exists $F \in \mathbb{R}^{m,n}$ such that $[E, A + BF]$ is strongly stabilizable.*

Proof (a) Let $[E, A, B]$ be impulse controllable. Then $[E, A, B]$ can be put into feedback form (3.10), i.e., there exist $W \in \mathbf{Gl}_k(\mathbb{R})$, $T \in \mathbf{Gl}_n(\mathbb{R})$ and $\tilde{F} \in \mathbb{R}^{m,n}$ such that

$$W\big(sE - \big(A + B\tilde{F}T^{-1}\big)\big)T = \begin{bmatrix} sI_{|\alpha|} - N_\alpha^\top & 0 & 0 & 0 & 0 & 0 \\ 0 & sK_\beta - L_\beta & 0 & 0 & 0 & 0 \\ 0 & 0 & sL_\gamma^\top - K_\gamma^\top & 0 & 0 & 0 \\ 0 & 0 & 0 & sK_\delta^\top - L_\delta^\top & 0 & 0 \\ 0 & 0 & 0 & 0 & sN_\kappa - I_{|\kappa|} & 0 \\ 0 & 0 & 0 & 0 & 0 & sI_{n_{\bar{c}}} - A_{\bar{c}} \end{bmatrix}. \tag{5.2}$$

By Corollary 3.4(b) the impulse controllability of $[E, A, B]$ implies that $\gamma = (1, \ldots, 1)$, $\delta = (1, \ldots, 1)$ and $\kappa = (1, \ldots, 1)$. Therefore, we see that, with $F = \tilde{F}T^{-1}$, the pencil $sE - (A + BF)$ has index at most one as the index is preserved under system equivalence.

Conversely, assume that $[E, A, B]$ is not impulse controllable. We show that for all $F \in \mathbb{R}^{m,n}$ the index of $sE - (A + BF)$ is greater than one. To this end, let $F \in \mathbb{R}^{m,n}$ and choose $W \in \mathbf{Gl}_k(\mathbb{R})$, $T \in \mathbf{Gl}_n(\mathbb{R})$ and $\tilde{F} \in \mathbb{R}^{m,n}$ such that (3.10) holds. Then, partitioning $V^{-1}FT = [F_{ij}]_{i=1,\ldots,3,\, j=1,\ldots,6}$ accordingly, we obtain

$$\begin{aligned} s\tilde{E} - \tilde{A} &:= W\big(sE - \big(A + BF + B\tilde{F}T^{-1}\big)\big)T \\ &= W\big(sE - \big(A + B\tilde{F}T^{-1}\big)\big)T - WBVV^{-1}FT \\ &= \begin{bmatrix} sI_{|\alpha|} - (N_\alpha^\top + E_\alpha F_{11}) & -E_\alpha F_{12} & -E_\alpha F_{13} & -E_\alpha F_{14} & -E_\alpha F_{15} & -E_\alpha F_{16} \\ 0 & sK_\beta - L_\beta & 0 & 0 & 0 & 0 \\ -E_\gamma F_{21} & -E_\gamma F_{22} & sL_\gamma^\top - (K_\gamma^\top + E_\gamma F_{23}) & -E_\gamma F_{24} & -E_\gamma F_{25} & -E_\gamma F_{26} \\ 0 & 0 & 0 & sK_\delta^\top - L_\delta^\top & 0 & 0 \\ 0 & 0 & 0 & 0 & sN_\kappa - I_{|\kappa|} & 0 \\ 0 & 0 & 0 & 0 & 0 & sI_{n_{\bar{c}}} - A_{\bar{c}} \end{bmatrix}. \end{aligned} \tag{5.3}$$

Now the assumption that $[E, A, B]$ is not impulse controllable leads to $\gamma \neq (1, \ldots, 1)$, $\delta \neq (1, \ldots, 1)$ or $\kappa \neq (1, \ldots, 1)$. We will now show that the index of

$sE - (A + BF + B\tilde{F}T^{-1})$ is greater than one by showing this for the equivalent pencil in (5.3) via applying the condition in (3.5): Let Z be a real matrix with $\operatorname{im}_{\mathbb{R}} Z = \ker_{\mathbb{R}} \tilde{E}$. Then

$$Z = \begin{bmatrix} 0 & Z_1^\top & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & Z_2^\top & 0 \end{bmatrix}^\top,$$

where $\operatorname{im} Z_1 = \ker K_\beta = \operatorname{im} E_\beta$ and $\operatorname{im} Z_2 = \ker N_\kappa = \operatorname{im} E_\kappa$. Taking into account that $\operatorname{im}_{\mathbb{R}} E_\gamma \subseteq \operatorname{im}_{\mathbb{R}} L_\gamma^\top$, we obtain

$$\begin{aligned} &\operatorname{im}_{\mathbb{R}} \begin{bmatrix} 0_{|\alpha|-\ell(\alpha)+|\beta|-\ell(\beta),k} & I_{|\gamma|+|\delta|+|\kappa|} & 0_{k,n_{\bar{c}}} \end{bmatrix} \begin{bmatrix} \tilde{E} & \tilde{A}Z \end{bmatrix} \\ &\quad = \operatorname{im}_{\mathbb{R}} \begin{bmatrix} L_\gamma^\top & 0 & 0 & E_\gamma F_{25} Z_2 \\ 0 & K_\delta^\top & 0 & 0 \\ 0 & 0 & N_\kappa & Z_2 \end{bmatrix}. \end{aligned}$$

On the other hand, we have

$$\begin{aligned} &\operatorname{im}_{\mathbb{R}} \begin{bmatrix} 0_{|\alpha|-\ell(\alpha)+|\beta|-\ell(\beta),k} & I_{|\gamma|+|\delta|+|\kappa|} & 0_{k,n_{\bar{c}}} \end{bmatrix} \begin{bmatrix} \tilde{E} & \tilde{A} \end{bmatrix} \\ &\quad = \operatorname{im}_{\mathbb{R}} \begin{bmatrix} L_\gamma^\top & 0 & 0 & K_\gamma^\top + E_\gamma F_{23} & E_\gamma F_{24} & E_\gamma F_{25} \\ 0 & K_\delta^\top & 0 & 0 & L_\delta^\top & 0 \\ 0 & 0 & N_\kappa & 0 & 0 & I_{|\kappa|} \end{bmatrix}. \end{aligned}$$

Since the assumption that at least one of the multi-indices satisfies $\gamma \neq (1, \ldots, 1)$, $\delta \neq (1, \ldots, 1)$, or $\kappa \neq (1, \ldots, 1)$ and the fact that $\operatorname{im} Z_2 = \operatorname{im} E_\kappa$ lead to

$$\begin{aligned} &\operatorname{im}_{\mathbb{R}} \begin{bmatrix} L_\gamma^\top & 0 & 0 & E_\gamma F_{25} Z_2 \\ 0 & K_\delta^\top & 0 & 0 \\ 0 & 0 & N_\kappa & Z_2 \end{bmatrix} \\ &\quad \subsetneq \operatorname{im}_{\mathbb{R}} \begin{bmatrix} L_\gamma^\top & 0 & 0 & K_\gamma^\top + E_\gamma F_{23} & E_\gamma F_{24} & E_\gamma F_{25} \\ 0 & K_\delta^\top & 0 & 0 & L_\delta^\top & 0 \\ 0 & 0 & N_\kappa & 0 & 0 & I_{|\kappa|} \end{bmatrix}, \end{aligned}$$

and thus

$$\operatorname{im}_{\mathbb{R}} \begin{bmatrix} \tilde{E} & \tilde{A}Z \end{bmatrix} \subsetneq \operatorname{im}_{\mathbb{R}} \begin{bmatrix} \tilde{E} & \tilde{A} \end{bmatrix},$$

we find that, by condition (3.5), the index of $sE - (A + BF + B\tilde{F}T^{-1})$ has to be greater than one. Since F was chosen arbitrarily we may conclude that $sE - (A + BF)$ has index greater than one for all $F \in \mathbb{R}^{m,n}$, which completes the proof of (a).

(b) If $[E, A, B]$ is completely stabilizable, then we may transform the system into feedback form (5.2). Corollary 3.4(h) implies $\gamma = (1, \ldots, 1)$, $\delta = (1, \ldots, 1)$, $\ell(\kappa) = 0$, and $\sigma(A_{\bar{c}}) \subseteq \mathbb{C}_-$. Further, by [147, Thm. 4.20], there exists some $F_{11} \in \mathbb{R}^{|\alpha|,\ell(\alpha)}$ such that $\sigma(N_\alpha + E_\alpha F_{11}) \subseteq \mathbb{C}_-$. Setting $\hat{F} := [F_{ij}]_{i=1,\ldots,3,\, j=1,\ldots,6}$ with $F_{ij} = 0$ for $i \neq 1$ or $j \neq 1$, we find that with $F = \tilde{F}T^{-1} + V\hat{F}T^{-1}$ the system

$[E, A + BF]$ is completely stabilizable by Corollary 5.1 as complete stabilizability is preserved under system equivalence.

On the other hand, assume that $[E, A, B]$ is not completely stabilizable. We show that for all $F \in \mathbb{R}^{m,n}$ the system $[E, A + BF]$ is not completely stabilizable. To this end, let $F \in \mathbb{R}^{m,n}$ and observe that we may do a transformation as in (5.3). Then the assumption that $[E, A, B]$ is not completely stabilizable yields $\gamma \neq (1, \ldots, 1)$, $\delta \neq (1, \ldots, 1)$, $\ell(\kappa) > 0$, or $\sigma(A_{\overline{c}}) \not\subseteq \mathbb{C}_-$. If $\gamma \neq (1, \ldots, 1)$, $\delta \neq (1, \ldots, 1)$ or $\ell(\kappa) > 0$, then $\operatorname{im}_{\mathbb{R}} \tilde{A} \not\subseteq \operatorname{im}_{\mathbb{R}} \tilde{E}$, and by Corollary 5.2 the system $[\tilde{E}, \tilde{A}]$ is not completely stabilizable. On the other hand, if $\gamma = (1, \ldots, 1)$, $\delta = (1, \ldots, 1)$, $\ell(\kappa) = 0$, and $\lambda \in \sigma(A_{\overline{c}}) \cap \overline{\mathbb{C}}_+$, we find $\operatorname{im}_{\mathbb{C}}(\lambda\tilde{E} - \tilde{A}) \subsetneq \operatorname{im}_{\mathbb{C}} \tilde{E}$, which implies

$$\operatorname{rk}_{\mathbb{C}}(\lambda\tilde{E} - \tilde{A}) < \operatorname{rk}_{\mathbb{C}} \tilde{E} = n - \ell(\beta) - \ell(\kappa) = n - \ell(\beta) \overset{(3.6)}{=} \operatorname{rk}_{\mathbb{R}(s)}(s\tilde{E} - \tilde{A}).$$

Hence, applying Corollary 5.2 again, the system $[\tilde{E}, \tilde{A}]$ is not completely stabilizable. As complete stabilizability is invariant under system equivalence it follows that $[E, A + BF + B\tilde{F}T^{-1}]$ is not completely stabilizable. Since F was chosen arbitrarily we may conclude that $[E, A + BF]$ is not completely stabilizable for all $F \in \mathbb{R}^{m,n}$, which completes the proof of (b).

(c) The proof is analogous to (b). □

Remark 5.3 (State feedback)

(i) If the pencil $sE - A$ is regular and $[E, A, B]$ is impulse controllable, then a feedback $F \in \mathbb{R}^{m,n}$ can be constructed such that the pencil $sE - (A + BF)$ is regular and its index does not exceed one: First we choose W, T, $\tilde{F}$ such that we can put the system into the form (5.2). Now, impulse controllability implies that $\gamma = (1, \ldots, 1)$, $\delta = (1, \ldots, 1)$ and $\kappa = (1, \ldots, 1)$. Assuming $\ell(\delta) > 0$ implies that any quasi-Kronecker form of the pencil $sE - (A + B\tilde{F}T^{-1} + B\hat{F})$ fulfills $\ell(\gamma) > 0$ (in the form (3.3)), a feedback $\hat{F} \in \mathbb{R}^{m,n}$ as the feedback cannot act on this block, which contradicts regularity of $sE - A$. Hence it holds $\ell(\delta) = 0$ and from $k = n$ we further obtain $\ell(\gamma) = \ell(\beta)$. Now applying another feedback as in (5.3), where we choose $F_{22} = E_\beta^\top \in \mathbb{R}^{\ell(\beta),|\beta|}$ and $F_{ij} = 0$ otherwise, we obtain, taking into account that $E_\gamma = I_{ell(\gamma)}$ and that the pencil $\left[\begin{smallmatrix} sK_\beta - L_\beta \\ -E_\beta^\top \end{smallmatrix}\right]$ is regular, the result that $sE - (A + BF)$ is indeed regular with index at most one.

(ii) The matrix F_{11} in the proof of Theorem 5.3(b) can be constructed as follows: For $j = 1, \ldots, \ell(\alpha)$, consider vectors

$$a_j = -[a_{j\alpha_j - 1}, \ldots, a_{j0}] \in \mathbb{R}^{1,\alpha_j}.$$

Then, for

$$F_{11} = \operatorname{diag}(a_1, \ldots, a_{\ell(\alpha)}) \in \mathbb{R}^{\ell(\alpha),|\alpha|}$$

the matrix $N_\alpha + E_\alpha F_{11}$ is diagonally composed of companion matrices, whence, for

$$p_j(s) = s^{\alpha_j} + a_{j\alpha_j - 1}s^{\alpha_j - 1} + \cdots + a_{j0} \in \mathbb{R}[s]$$

the characteristic polynomial of $N_\alpha + E_\alpha F_{11}$ is given by

$$\det\big(sI_{|\alpha|} - (N_\alpha + E_\alpha F_{11})\big) = \prod_{j=1}^{\ell(\alpha)} p_j(s).$$

Hence, choosing the coefficients a_{ji}, $j = 1, \ldots, \ell(\alpha)$, $i = 0, \ldots, \alpha_j$ such that the polynomials $p_1(s), \ldots, p_{\ell(\alpha)}(s) \in \mathbb{R}[s]$ are all Hurwitz, i.e., all roots of $p_1(s), \ldots, p_{\ell(\alpha)}(s)$ are in $\mathbb{C}_-$, we obtain stability.

5.3 Control in the Behavioral Sense

The hitherto presented feedback concept consists of the additional application of the relation $u(t) = Fx(t)$ to the system $E\dot{x}(t) = Ax(t) + Bu(t)$. Feedback can therefore be seen as an additional algebraic constraint that can be resolved for the input. Control in the behavioral sense, or, also called, *control via interconnection* [163] generalizes this approach by also allowing further algebraic relations in which the state not necessarily uniquely determines the input. That is, for given (or to be determined) $K = [K_x, K_u]$ with $K_x \in \mathbb{R}^{l,n}$, $K_u \in \mathbb{R}^{l,m}$, we consider

$$\begin{aligned}\mathfrak{B}^K_{[E,A,B]} :=& \big\{(x,u) \in \mathfrak{B}_{[E,A,B]} \;\big|\; \forall t \in \mathbb{R} : \big(x(t)^\top, u(t)^\top\big)^\top \in \ker_{\mathbb{R}}(K)\big\} \\ =& \mathfrak{B}_{[E,A,B]} \cap \mathfrak{B}_{[0_{l,n},K_x,K_u]}.\end{aligned}$$

We can alternatively write

$$\mathfrak{B}^K_{[E,A,B]} = \mathfrak{B}_{[E^K,A^K]},$$

where

$$\big[E^K, A^K\big] = \left[\begin{bmatrix} E & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} A & B \\ K_x & K_u \end{bmatrix}\right].$$

The concept of control in the behavioral sense has its origin in the works by Willems, Polderman and Trentelman [18, 128, 146, 162, 163], where differential behaviors and their stabilization via *control by interconnection* is considered. The latter means a systematic addition of some further (differential) equations in a way that a desired behavior is achieved. In contrast to these works we only add equations which are purely algebraic. This justifies to speak of *control by interconnection using static control laws*. We will give equivalent conditions for this type of generalized feedback stabilizing the system. Note that, in principle, one could make the extreme choice $K = I_{n+m}$ to end up with a behavior $\mathfrak{B}^K_{[E,A,B]} = \{0\}$ which is obviously autonomous and stable. This, however, is not suitable from a practical point of view, since in this interconnection, the space of consistent initial differential variables is a proper subset of the initial differential variables which are consistent with the original system $[E, A, B]$. Consequently, the interconnected system does not have the

causality property—that is, the implementation of the controller at a certain time $t \in \mathbb{R}$ is not possible, since this causes jumps in the differential variables. To avoid this, we introduce the concept of *compatibility*.

Definition 5.2 (Compatible and stabilizing control) The static control $K = [K_x, K_u]$, defined by $K_x \in \mathbb{R}^{l,n}$, $K_u \in \mathbb{R}^{l,m}$, is called

(a) *compatible*, if for any $x^0 \in \mathscr{V}^{\mathrm{diff}}_{[E,A,B]}$, there exists some $(x, u) \in \mathfrak{B}^K_{[E,A,B]}$ with $Ex(0) = Ex^0$.
(b) *stabilizing*, if $[E^K, A^K] \in \Sigma_{k+l,n}$ is stabilizable in the behavioral sense.

Remark 5.4 (Compatible control) Our definition of compatible control is a slight modification of the concept introduced by Julius and van der Schaft in [79] where an interconnection is called compatible, if any trajectory of the system without control law can be concatenated with a trajectory of the interconnected system. This certainly implies that the space of initial differential variables of the interconnected system cannot be smaller than the corresponding set for the nominal system.

Theorem 5.4 (Stabilizing control in the behavioral sense) *Let $[E, A, B] \in \Sigma_{k,n,m}$ be given. Then there exists a compatible and stabilizing control $K = [K_x, K_u]$ with $K_x \in \mathbb{R}^{l,n}$, $K_u \in \mathbb{R}^{l,m}$, if, and only if, $[E, A, B]$ is stabilizable in the behavioral sense. In case of $[E, A, B]$ being stabilizable in the behavioral sense, the compatible and stabilizing control K can moreover be chosen such that $[E^K, A^K]$ is autonomous, i.e., the interconnected system $[E^K, A^K]$ is stable in the behavioral sense.*

Proof Since, by definition, $[E, A, B] \in \Sigma_{k,n,m}$ is stabilizable in the behavioral sense if, and only if, for $s\mathscr{E} - \mathscr{A} = [sE - A, -B]$, the DAE $[\mathscr{E}, \mathscr{A}] \in \Sigma_{k,n+m}$ is stabilizable in the behavioral sense, necessity follows from setting $l = 0$.

In order to show sufficiency, let $K = [K_x, K_u]$ with $K_x \in \mathbb{R}^{l,n}$, $K_u \in \mathbb{R}^{l,m}$, be a compatible and stabilizing control for $[E, A, B]$. Now the system can be put into feedback form, i.e., there exist $W \in \mathbf{Gl}_k(\mathbb{R})$, $T \in \mathbf{Gl}_n(\mathbb{R})$, $V \in \mathbf{Gl}_m(\mathbb{R})$ and $F \in \mathbb{R}^{m,n}$ such that

$$\begin{bmatrix} s\tilde{E} - \tilde{A} & \tilde{B} \\ -\tilde{K}_x & \tilde{K}_u \end{bmatrix} = \begin{bmatrix} W & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} sE - A & B \\ -K_x & K_u \end{bmatrix} \begin{bmatrix} T & 0 \\ -F & V \end{bmatrix},$$

where $[\tilde{E}, \tilde{A}, \tilde{B}]$ is in the form (3.10). Now the behavioral stabilizability of $[E^K, A^K]$ implies that the system $[\tilde{E}^K, \tilde{A}^K] := \left[\left[\begin{smallmatrix} \tilde{E} & 0 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} \tilde{A} & \tilde{B} \\ \tilde{K}_x & \tilde{K}_u \end{smallmatrix}\right]\right]$ is stabilizable in the behavioral sense as well. Assume that $[E, A, B]$ is not stabilizable in the behavioral sense, that is, by Corollary 3.4(i), there exists $\lambda \in \sigma(A_{\overline{c}}) \cap \overline{\mathbb{C}}_+$. Hence we find $x^0_6 \in \mathbb{R}^{n_{\overline{c}}} \setminus \{0\}$ such that $A_{\overline{c}} x^0_6 = \lambda x^0_6$. Then, with $x(\cdot) := (0, \ldots, 0, (e^{\lambda \cdot} x^0_6)^\top)^\top$, we have $(x, 0) \in \mathscr{B}_{[\tilde{E},\tilde{A},\tilde{B}]}$. As $x(0) \in \mathscr{V}^{\mathrm{diff}}_{[\tilde{E},\tilde{A},\tilde{B}]} = T^{-1} \cdot \mathscr{V}^{\mathrm{diff}}_{[E,A,B]}$, the compatibility of the control K implies that there exists

$(\tilde{x}, \tilde{u}) \in \mathfrak{B}^K_{[E,A,B]}$ with $E\tilde{x}(0) = ETx(0)$. This gives $(WET)T^{-1}\tilde{x}(0) = WETx(0)$ and writing $T^{-1}\tilde{x}(t) = (\tilde{x}_1(t)^\top, \ldots, \tilde{x}_6(t)^\top)^\top$ with vectors of appropriate size, we obtain $\tilde{x}_6(0) = x_6^0$. Since the solution of the initial value problem $\dot{y} = A_{\bar{c}}y$, $y(0) = x_6^0$, is unique, we find $\tilde{x}_6(t) = e^{\lambda t}x_6^0$ for all $t \in \mathbb{R}$. Now $(T^{-1}\tilde{x}, -V^{-1}FT^{-1}\tilde{x} + V^{-1}\tilde{u}) \in \mathscr{B}_{[\tilde{E}^K, \tilde{A}^K]}$ and as for all $(\hat{x}, \hat{u}) \in \mathscr{B}_{[\tilde{E}^K, \tilde{A}^K]}$ with $(\hat{x}(t), \hat{u}(t)) = (T^{-1}\tilde{x}(t), -V^{-1}FT^{-1}\tilde{x} + V^{-1}\tilde{u}(t))$ for all $t < 0$ we have $\hat{x}_6(t) = \tilde{x}_6(t)$ for all $t \in \mathbb{R}$, and $\tilde{x}_6(t) \not\to_{t\to\infty} 0$ since $\lambda \in \overline{\mathbb{C}}_+$, this contradicts that $[\tilde{E}^K, \tilde{A}^K]$ is stabilizable in the behavioral sense.

It remains to show the second assertion, that is, for a system $[E, A, B] \in \Sigma_{k,n,m}$ that is stabilizable in the behavioral sense, there exists some compatible and stabilizing control K such that $[E^K, A^K]$ is autonomous: Since, for $[E_1, A_1, B_1], [E_2, A_2, B_2] \in \Sigma_{k,n,m}$ with

$$[E_1, A_1, B_1] \overset{W,T,V,F}{\sim_{fe}} [E_2, A_2, B_2], \quad K_2 \in \mathbb{R}^{l,n+m} \quad \text{and} \quad K_1 = K_2 \begin{bmatrix} T & 0 \\ F & V \end{bmatrix},$$

the behaviors of the interconnected systems are related by

$$\begin{bmatrix} T & 0 \\ F & V \end{bmatrix} \mathfrak{B}^{K_1}_{[E_1,A_1,B_1]} = \mathfrak{B}^{K_2}_{[E_2,A_2,B_2]},$$

it is no loss of generality to assume that $[E, A, B]$ is in feedback form (3.10), i.e.,

$$sE - A = \begin{bmatrix} sI_{|\alpha|} - N_\alpha & 0 & 0 & 0 & 0 & 0 \\ 0 & sK_\beta - L_\beta & 0 & 0 & 0 & 0 \\ 0 & 0 & sK_\gamma^\top - L_\gamma^\top & 0 & 0 & 0 \\ 0 & 0 & 0 & sK_\delta^\top - L_\delta^\top & 0 & 0 \\ 0 & 0 & 0 & 0 & sN_\kappa - I_{|\kappa|} & 0 \\ 0 & 0 & 0 & 0 & 0 & sI_{n_{\bar{c}}} - A_{\bar{c}} \end{bmatrix},$$

$$B = \begin{bmatrix} E_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & E_\gamma & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Let $F_{11} \in \mathbb{R}^{\ell(\alpha),|\alpha|}$ such that $\det(sI_{|\alpha|} - (N_\alpha + E_\alpha F_{11}))$ is Hurwitz. Then the DAE

$$\begin{bmatrix} I_{|\alpha|} & 0 \\ 0 & 0 \end{bmatrix} \dot{z}(t) = \begin{bmatrix} N_\alpha & E_\alpha \\ F_{11} & -I_{\ell(\alpha)} \end{bmatrix} z(t)$$

is stable in the behavioral sense. Furthermore, by reasoning as in Remark 5.3(ii), for

$$a_j = [a_{j\beta_j-2}, \ldots, a_{j0}, 1] \in \mathbb{R}^{1,\beta_j}$$

with the property that the polynomials

$$p_j(s) = s^{\beta_j} + a_{j\beta_j-1}s^{\beta_j-1} + \cdots + a_{j0} \in \mathbb{R}[s]$$

are Hurwitz for $j = 1, \ldots, \ell(\alpha)$, the choice

$$K_x = \operatorname{diag}(a_1, \ldots, a_{\ell(\beta)}) \in \mathbb{R}^{\ell(\beta),|\beta|}$$

leads to an autonomous system

$$\begin{bmatrix} K_\beta \\ 0 \end{bmatrix} \dot{z}(t) = \begin{bmatrix} L_\beta \\ K_x \end{bmatrix} z(t),$$

which is also stable in the behavioral sense. Since, moreover, by Corollary 3.4(i), we have $\sigma(A_{\overline{c}}) \subseteq \mathbb{C}_-$, the choice

$$K = \begin{bmatrix} F_{11} & 0 & 0 & 0 & 0 & 0 & -I_{\ell(\alpha)} & 0 & 0 \\ 0 & K_x & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

leads to a behavioral stable (in particular autonomous) system. Since the differential variables can be arbitrarily initialized in any of the previously discussed subsystems, the constructed control is also compatible. □

6 Invariant Subspaces

This section is dedicated to some selected results of the geometric theory of differential-algebraic control systems. Geometric theory plays a fundamental role in standard ODE system theory and has been introduced independently by Wonham and Morse and Basile and Marro, see the famous books [16, 166] and also [147], which are the three standard textbooks on geometric control theory. In [100] Lewis gave an up-to-date overview of the geometric theory of DAEs. As we will do here he put special emphasis on the two fundamental sequences of subspaces $\mathscr{V}_i$ and $\mathscr{W}_i$ defined as follows:

$$\mathscr{V}_0 := \mathbb{R}^n, \qquad \mathscr{V}_{i+1} := A^{-1}(E\mathscr{V}_i + \operatorname{im}_{\mathbb{R}} B) \subseteq \mathbb{R}^n, \qquad \mathscr{V}^* := \bigcap_{i\in\mathbb{N}_0} \mathscr{V}_i,$$

$$\mathscr{W}_0 := \{0\}, \qquad \mathscr{W}_{i+1} := E^{-1}(A\mathscr{W}_i + \operatorname{im}_{\mathbb{R}} B) \subseteq \mathbb{R}^n, \qquad \mathscr{W}^* := \bigcup_{i\in\mathbb{N}_0} \mathscr{W}_i.$$

The sequences $(\mathscr{V}_i)_{i\in\mathbb{N}}$ and $(\mathscr{W}_i)_{i\in\mathbb{N}}$ are called *augmented Wong sequences*. In [22, 23, 26] the Wong sequences for matrix pencils (i.e., $B = 0$) are investigated, the name chosen this way since Wong [164] was the first one who used both sequences for the analysis of matrix pencils. The sequences $(\mathscr{V}_i)_{i\in\mathbb{N}}$ and $(\mathscr{W}_i)_{i\in\mathbb{N}}$ are no Wong sequences corresponding to any matrix pencils, which is why we call them augmented Wong sequences with respect to control systems (2.1). In fact, the

Wong sequences (with $B = 0$) can be traced back to Dieudonné [53], who focused on the first of the two Wong sequences. Bernhard [27] and Armentano [6] used the Wong sequences to carry out a geometric analysis of matrix pencils. They appear also in [3, 4, 95, 150].

In control theory, that is, when $B \neq 0$, the augmented Wong sequences have been extensively studied by several authors, see e.g. [99, 112, 113, 118, 119, 121, 122, 152] for regular systems and [3, 13–15, 29, 30, 56, 100, 105, 120, 130] for general DAE systems. Frankowska [58] did a nice investigation of systems (2.1) in terms of differential inclusions [8, 9], however, requiring controllability at infinity (see [58, Prop. 2.6]). Nevertheless, she is the first to derive a formula for the reachability space [58, Thm. 3.1], which was later generalized by Przyłuski and Sosnowski [130, Sect. 4] (in fact, the same generalization has been announced in [105, p. 296], [100, Sect. 5] and [15, p. 1510], however, without proof); it also occurred in [56, Thm. 2.5].

Proposition 6.1 (Reachability space [130, Sect. 4]) *For* $[E, A, B] \in \Sigma_{k,n,m}$ *and limits* $\mathscr{V}^*$ *and* $\mathscr{W}^*$ *of the augmented Wong sequences we have*

$$\mathscr{R}_{[E,A,B]} = \mathscr{V}^* \cap \mathscr{W}^*.$$

It has been shown in [13] (for discrete systems), see also [14, 15, 30, 120], that the limit $\mathscr{V}^*$ of the first augmented Wong sequence is the space of consistent initial states. For regular systems this was proved in [99].

Proposition 6.2 (Consistent initial states [13]) *For* $[E, A, B] \in \Sigma_{k,n,m}$ *and limit* $\mathscr{V}^*$ *of the first augmented Wong sequence we have*

$$\mathscr{V}_{[E,A,B]} = \mathscr{V}^*.$$

Various other properties of $\mathscr{V}^*$ and $\mathscr{W}^*$ have been derived in [13] in the context of discrete systems.

A characterization of the spaces $\mathscr{V}^*$ and $\mathscr{W}^*$ in terms of distributions is also given in [130]: $\mathscr{V}^* + \ker_{\mathbb{R}} E$ is the set of all initial values such that the distributional initial value problem [130, (3)] has a smooth solution (x, u); $\mathscr{W}^*$ is the set of all initial values such that [130, (3)] has an impulsive solution (x, u); $\mathscr{V}^* + \mathscr{W}^*$ is the set of all initial values such that [130, (3)] has an impulsive-smooth solution (x, u).

For regular systems Özçaldiran [119] showed that $\mathscr{V}^*$ is the supremal $(A, E; \operatorname{im}_{\mathbb{R}} B)$-invariant subspace of $\mathbb{R}^n$ and $\mathscr{W}^*$ is the infimal restricted $(E, A; \operatorname{im}_{\mathbb{R}} B)$-invariant subspace of $\mathbb{R}^n$. These concepts, which have also been used in [3, 13, 99, 113] are defined as follows.

Definition 6.1 ($(A, E; \operatorname{im}_{\mathbb{R}} B)$- and $(E, A; \operatorname{im}_{\mathbb{R}} B)$-invariance [119]) Let $[E, A, B] \in \Sigma_{k,n,m}$. A subspace $\mathscr{V} \subseteq \mathbb{R}^n$ is called $(A, E; \operatorname{im}_{\mathbb{R}} B)$-*invariant* if, and only if,

$$A\mathscr{V} \subseteq E\mathscr{V} + \operatorname{im}_{\mathbb{R}} B.$$

A subspace $\mathscr{W} \subseteq \mathbb{R}^n$ is called *restricted* $(E, A; \operatorname{im}_{\mathbb{R}} B)$*-invariant* if, and only if,

$$\mathscr{W} = E^{-1}(A\mathscr{W} + \operatorname{im}_{\mathbb{R}} B).$$

It is easy to verify that the proofs given in [119, Lems. 2.1 & 2.2] remain the same for general $E, A \in \mathbb{R}^{k,n}$ and $B \in \mathbb{R}^{n,m}$—this was shown in [13] as well. For $\mathscr{V}^*$ this can be found in [3], see also [113]. So we have the following proposition.

Proposition 6.3 (Augmented Wong sequences as invariant subspaces) *Consider* $[E, A, B] \in \Sigma_{k,n,m}$ *and the limits* $\mathscr{V}^*$ *and* $\mathscr{W}^*$ *of the augmented Wong sequences. Then the following statements hold true.*

(a) $\mathscr{V}^*$ *is* $(A, E; \operatorname{im}_{\mathbb{R}} B)$*-invariant and for any* $\mathscr{V} \subseteq \mathbb{R}^n$ *which is* $(A, E; \operatorname{im}_{\mathbb{R}} B)$*-invariant it holds* $\mathscr{V} \subseteq \mathscr{V}^*$;
(b) $\mathscr{W}^*$ *is restricted* $(E, A; \operatorname{im}_{\mathbb{R}} B)$*-invariant and for any* $\mathscr{W} \subseteq \mathbb{R}^n$ *which is restricted* $(E, A; \operatorname{im}_{\mathbb{R}} B)$*-invariant it holds* $\mathscr{W}^* \subseteq \mathscr{W}$.

It is now clear how the controllability concepts can be characterized in terms of the invariant subspaces $\mathscr{V}^*$ and $\mathscr{W}^*$. However, the statement about R-controllability (behavioral controllability) seems to be new. The only other appearance of a subspace inclusion as a characterization of R-controllability that the authors are aware of occurs in [41] for regular systems: if $A = I$, then the system is R-controllable if, and only if, $\operatorname{im}_{\mathbb{R}} E^D \subseteq \langle E^D | B \rangle$, where E^D is the Drazin inverse of E, see Remark 2.1(iv).

Theorem 6.4 (Geometric criteria for controllability) *Consider* $[E, A, B] \in \Sigma_{k,n,m}$ *and the limits* $\mathscr{V}^*$ *and* $\mathscr{W}^*$ *of the augmented Wong sequences. Then* $[E, A, B]$ *is*

(a) *controllable at infinity if, and only if,* $\mathscr{V}^* = \mathbb{R}^n$;
(b) *impulse controllable if, and only if,* $\mathscr{V}^* + \ker_{\mathbb{R}} E = \mathbb{R}^n$ *or, equivalently,* $E\mathscr{V}^* = \operatorname{im}_{\mathbb{R}} E$;
(c) *controllable in the behavioral sense if, and only if,* $\mathscr{V}^* \subseteq \mathscr{W}^*$;
(d) *completely controllable if, and only if,* $\mathscr{V}^* \cap \mathscr{W}^* = \mathbb{R}^n$;
(e) *strongly controllable if, and only if,* $(\mathscr{V}^* \cap \mathscr{W}^*) + \ker_{\mathbb{R}} E = \mathbb{R}^n$ *or, equivalently,* $E(\mathscr{V}^* \cap \mathscr{W}^*) = \operatorname{im}_{\mathbb{R}} E$.

Proof By Propositions 6.1 and 6.2 it is clear that it only remains to prove (c). We proceed in several steps.

Step 1: Let $[E_1, A_1, B_1], [E_2, A_2, B_2] \in \Sigma_{k,n,m}$ such that for some $W \in \mathbf{Gl}_k(\mathbb{R})$, $T \in \mathbf{Gl}_n(\mathbb{R})$, $V \in \mathbf{Gl}_m(\mathbb{R})$ and $F \in \mathbb{R}^{m,n}$ it holds

$$[E_1, A_1, B_1] \overset{W,T,V,F}{\sim_{fe}} [E_2, A_2, B_2].$$

We show that the augmented Wong sequences $\mathscr{V}_i^1$, $\mathscr{W}_i^1$ of $[E_1, A_1, B_1]$ and the augmented Wong sequences $\mathscr{V}_i^2$, $\mathscr{W}_i^2$ of $[E_2, A_2, B_2]$ are related by

$$\forall i \in \mathbb{N}_0 : \mathscr{V}_i^1 = T^{-1}\mathscr{V}_i^2 \wedge \mathscr{W}_i^1 = T^{-1}\mathscr{W}_i^2.$$

We proof the statement by induction. It is clear that $\mathscr{V}_0^1 = T^{-1}\mathscr{V}_0^2$. Assuming that $\mathscr{V}_i^1 = T^{-1}\mathscr{V}_i^2$ for some $i \geq 0$ we find that, by (3.1),

$$\begin{aligned}\mathscr{V}_{i+1}^1 &= A_1^{-1}\big(E_1\mathscr{V}_i^1 + \operatorname{im}_{\mathbb{R}} B_1\big)\\ &= \big\{x \in \mathbb{R}^n \,\big|\, \exists y \in \mathscr{V}_i^1\ \exists u \in \mathbb{R}^m : W(A_2T + B_2T)x = WE_2Ty + WB_2Vu\big\}\\ &= \big\{x \in \mathbb{R}^n \,\big|\, \exists z \in \mathscr{V}_i^2\ \exists v \in \mathbb{R}^m : A_2Tx = E_2z + B_2v\big\}\\ &= T^{-1}\big(A_2^{-1}\big(E_2\mathscr{V}_i^1 + \operatorname{im}_{\mathbb{R}} B_2\big)\big) = T^{-1}\mathscr{V}_{i+1}^2.\end{aligned}$$

The statement about $\mathscr{W}_i^1$ and $\mathscr{W}_i^2$ can be proved analogous.

Step 2: By Step 1 we may without loss of generality assume that $[E, A, B]$ is given in feedback form (3.10). We make the convention that if $\alpha \in \mathbb{N}^l$ is some multi-index, then $\alpha - 1 := (\alpha_1 - 1, \ldots, \alpha_l - 1)$. It not follows that

$$\forall i \in \mathbb{N}_0 : \mathscr{V}_i = \mathbb{R}^{|\alpha|} \times \mathbb{R}^{|\beta|} \times \operatorname{im}_{\mathbb{R}} N_{\gamma-1}^i \times \operatorname{im}_{\mathbb{R}}\big(N_{\delta-1}^\top\big)^i \times \operatorname{im}_{\mathbb{R}} N_\kappa^i \times \mathbb{R}^{n_{\overline{c}}}, \quad (6.1)$$

which is immediate from observing that $K_\gamma^\top x = L_\gamma^\top y + E_\gamma u$ for some x, y, u of appropriate dimension yields $x = N_{\gamma-1}y$ and $L_\delta^\top x = K_\delta^\top y$ for some x, y yields $x = N_{\delta-1}^\top y$. Note that in the case $\gamma_i = 1$ or $\delta_i = 1$, i.e., we have a 1×0 block, we find that N_{γ_i-1} and N_{δ_i-1} are absent, so these relations are consistent.

On the other hand we find that

$$\forall i \in \mathbb{N}_0 : \mathscr{W}_i = \ker_{\mathbb{R}} N_\alpha^i \times \ker_{\mathbb{R}} N_\beta^i \times \ker_{\mathbb{R}} N_{\gamma-1}^i \times \{0\}^{|\delta|-\ell(\delta)} \times \ker_{\mathbb{R}} N_\kappa^i \times \{0\}^{n_{\overline{c}}}, \quad (6.2)$$

which indeed needs some more rigorous proof. First observe that $\operatorname{im}_{\mathbb{R}} E_\alpha = \ker_{\mathbb{R}} N_\alpha$, $\ker_{\mathbb{R}} K_\beta = \ker_{\mathbb{R}} N_\beta$ and $(L_\gamma^\top)^{-1}(\operatorname{im}_{\mathbb{R}} E_\gamma) = \operatorname{im}_{\mathbb{R}} E_{\gamma-1} = \ker_{\mathbb{R}} N_{\gamma-1}$. Therefore we have

$$\begin{aligned}\mathscr{W}_1 &= E^{-1}(\operatorname{im}_{\mathbb{R}} B)\\ &= \ker_{\mathbb{R}} N_\alpha \times \ker_{\mathbb{R}} N_\beta \times \ker_{\mathbb{R}} N_{\gamma-1} \times \{0\}^{|\delta|-\ell(\delta)} \times \ker_{\mathbb{R}} N_\kappa \times \{0\}^{n_{\overline{c}}}.\end{aligned}$$

Further observe that $N_\alpha^i N_\alpha^\top = N_\alpha N_\alpha^\top N_\alpha^{i-1}$ for all $i \in \mathbb{N}$ and, hence, if $x = N_\alpha^\top y + E_\alpha u$ for some x, u and $y \in \ker_{\mathbb{R}} N_\alpha^{i-1}$ it follows $x \in \ker_{\mathbb{R}} N_\alpha^i$. Likewise, if $L_\gamma^\top x = K_\gamma^\top y + E_\gamma u$ for some x, u and $y \in \ker_{\mathbb{R}} N_{\gamma-1}^{i-1}$ we find $x = N_{\gamma-1}^\top y + E_{\gamma-1}^\top u$ and hence $x \in \ker_{\mathbb{R}} N_{\gamma-1}^i$. Finally, if $K_\beta x = L_\beta y$ for some x and some $y \in \ker_{\mathbb{R}} N_\beta^{i-1}$ it follows that by adding some zero rows we obtain $N_\beta x = N_\beta N_\beta^\top y$ and hence, as above, $x \in \ker_{\mathbb{R}} N_\beta^i$. This proves (6.2).

Step 3: From (6.1) and (6.2) it follows that

$$\begin{aligned}\mathscr{V}^* &= \mathbb{R}^{|\alpha|} \times \mathbb{R}^{|\beta|} \times \operatorname{im}_{\mathbb{R}}\{0\}^{|\gamma|-\ell(\gamma)} \times \{0\}^{|\delta|-\ell(\delta)} \times \{0\}^{|\kappa|} \times \mathbb{R}^{n_{\overline{c}}},\\ \mathscr{W}^* &= \mathbb{R}^{|\alpha|} \times \mathbb{R}^{|\beta|} \times \operatorname{im}_{\mathbb{R}} \mathbb{R}^{|\gamma|-\ell(\gamma)} \times \{0\}^{|\delta|-\ell(\delta)} \times \mathbb{R}^{|\kappa|} \times \{0\}^{n_{\overline{c}}}.\end{aligned}$$

As by Corollary 3.4(f) the system $[E, A, B]$ is controllable in the behavioral sense if, and only if, $n_{\overline{c}} = 0$ we may immediately deduce that this is the case if, and only if, $\mathscr{V}^* \subseteq \mathscr{W}^*$. This proves the theorem. □

Remark 6.1 (Representation of the reachability space) From Proposition 6.1 and the proof of Theorem 6.4 we may immediately observe that, using the notation from Theorem 3.3, we have

$$\mathscr{R}_{[E,A,B]} = T^{-1}\big(\mathbb{R}^{|\alpha|} \times \mathbb{R}^{|\beta|} \times \operatorname{im}_{\mathbb{R}}\{0\}^{|\gamma|-\ell(\gamma)} \times \{0\}^{|\delta|-\ell(\delta)} \times \{0\}^{|\kappa|} \times \{0\}^{n_{\overline{c}}}\big).$$

7 Kalman Decomposition

Nearly 50 years ago Kalman [82] derived his famous decomposition of linear ODE control systems. This decomposition has later been generalized to regular DAEs by Verghese et al. [155], see also [49]. A Kalman decomposition of general discrete-time DAE systems has been provided by Banaszuk et al. [14] (later generalized to systems with output equation in [15]) in a very nice way using the augmented Wong sequences (cf. Sect. 6). They derive a system

$$\left[\begin{bmatrix} E_{11} & E_{12} \\ 0 & E_{22} \end{bmatrix}, \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \begin{bmatrix} B_1 \\ 0 \end{bmatrix}\right], \tag{7.1}$$

which is system equivalent to given $[E, A, B] \in \Sigma_{k,n,m}$ with the properties that the system $[E_{11}, A_{11}, B_1]$ is completely controllable and the matrix $[E_{11}, A_{11}, B_1]$ has full row rank (strongly $\mathscr{H}$-controllable in the notation of [14]) and, furthermore, $\mathscr{R}_{[E_{22},A_{22},0]} = \{0\}$.

This last condition is very reasonable, as one should wonder what properties a Kalman decomposition of a DAE system should have. In the case of ODEs the decomposition simply is

$$\left[\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \begin{bmatrix} B_1 \\ 0 \end{bmatrix}\right], \quad \text{where } [A_{11}, B_1] \text{ is controllable.}$$

Therefore, an ODE system is decomposed into a controllable and an uncontrollable part, since clearly $[A_{22}, 0]$ is not controllable at all. For DAEs however, the situation is more subtle, since in a decomposition (7.1) with $[E_{11}, A_{11}, B_1]$ completely controllable (and $[E_{11}, A_{11}, B_1]$ full row rank) the conjectural "uncontrollable" part $[E_{22}, A_{22}, 0]$ may still have a controllable subsystem, since systems of the type $[K_\beta, L_\beta, 0]$ are always controllable. To exclude this and ensure that all controllable parts are included in $[E_{11}, A_{11}, B_1]$ we may state the additional condition (as in [14]) that

$$\mathscr{R}_{[E_{22},A_{22},0]} = \{0\}.$$

This then also guarantees certain uniqueness properties of the Kalman decomposition. Hence, any system (7.1) with the above properties which is system equivalent

to $[E, A, B]$ we may call a Kalman decomposition of $[E, A, B]$. We cite the result of [14], but also give some remarks on how the decomposition may be easily derived.

Theorem 7.1 (Kalman decomposition [14]) *For* $[E, A, B] \in \Sigma_{k,n,m}$, *there exist* $W \in \mathbf{Gl}_k(\mathbb{R})$, $T \in \mathbf{Gl}_n(\mathbb{R})$ *such that*

$$[E, A, B] \overset{W,T}{\sim_{se}} \left[\begin{bmatrix} E_{11} & E_{12} \\ 0 & E_{22} \end{bmatrix}, \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \begin{bmatrix} B_1 \\ 0 \end{bmatrix}\right], \tag{7.2}$$

with $E_{11}, A_{11} \in \mathbb{R}^{k_1,n_1}$, $E_{12}, A_{12} \in \mathbb{R}^{k_1,n_2}$, $E_{22}, A_{22} \in \mathbb{R}^{k_2,n_2}$ *and* $B_1 \in \mathbb{R}^{k_1,m}$, *such that* $[E_{11}, A_{11}, B_1] \in \Sigma_{k_1,n_1,m}$ *is completely controllable*, $\operatorname{rk}_{\mathbb{R}}[E_{11}, A_{11}, B_1] = k_1$ *and* $\mathscr{R}_{[E_{22},A_{22},0_{k_2,m}]} = \{0\}$.

Remark 7.1 (Derivation of the Kalman decomposition) Let $[E, A, B] \in \Sigma_{k,n,m}$ be given. The Kalman decomposition (7.2) can be derived using the limits $\mathscr{V}^*$ and $\mathscr{W}^*$ of the augmented Wong sequences presented in Sect. 6. It is clear that these spaces satisfy the following subspace relations:

$$\begin{aligned} E\big(\mathscr{V}^* \cap \mathscr{W}^*\big) &\subseteq \big(E\mathscr{V}^* + \operatorname{im}_{\mathbb{R}} B\big) \cap \big(A\mathscr{W}^* + \operatorname{im}_{\mathbb{R}} B\big), \\ A\big(\mathscr{V}^* \cap \mathscr{W}^*\big) &\subseteq \big(E\mathscr{V}^* + \operatorname{im}_{\mathbb{R}} B\big) \cap \big(A\mathscr{W}^* + \operatorname{im}_{\mathbb{R}} B\big). \end{aligned}$$

Therefore, if we choose any full rank matrices $R_1 \in \mathbb{R}^{n,n_1}$, $P_1 \in \mathbb{R}^{n,n_2}$, $R_2 \in \mathbb{R}^{k,k_1}$, $P_2 \in \mathbb{R}^{k,k_2}$ such that

$$\begin{aligned} \operatorname{im}_{\mathbb{R}} R_1 &= \mathscr{V}^* \cap \mathscr{W}^*, \qquad & \operatorname{im}_{\mathbb{R}} R_2 &= \big(E\mathscr{V}^* + \operatorname{im}_{\mathbb{R}} B\big) \cap \big(A\mathscr{W}^* + \operatorname{im}_{\mathbb{R}} B\big), \\ \operatorname{im}_{\mathbb{R}} R_1 \oplus \operatorname{im}_{\mathbb{R}} P_1 &= \mathbb{R}^n, \qquad & \operatorname{im}_{\mathbb{R}} R_2 \oplus \operatorname{im}_{\mathbb{R}} P_2 &= \mathbb{R}^k, \end{aligned}$$

then $[R_1, P_1] \in \mathbf{Gl}_n(\mathbb{R})$ and $[R_2, P_2] \in \mathbf{Gl}_k(\mathbb{R})$, and, furthermore, there exists matrices $E_{11}, A_{11} \in \mathbb{R}^{k_1,n_1}$, $E_{12}, A_{12} \in \mathbb{R}^{k_1,n_2}$, $E_{22}, A_{22} \in \mathbb{R}^{k_2,n_2}$ such that

$$\begin{aligned} ER_1 &= R_2E_{11}, & AR_1 &= R_2A_{11}, \\ EP_1 &= R_2E_{12} + P_2E_{22}, & AP_1 &= R_2A_{12} + P_2A_{22}. \end{aligned}$$

Since $\operatorname{im}_{\mathbb{R}} B \subseteq (E\mathscr{V}^* + \operatorname{im}_{\mathbb{R}} B) \cap (A\mathscr{W}^* + \operatorname{im}_{\mathbb{R}} B) = \operatorname{im}_{\mathbb{R}} R_2$, there exists $B_1 \in \mathbb{R}^{k_1,m}$ such that $B = R_2B_1$. All these relations together yield the decomposition (7.2) with $W = [R_2, P_2]$ and $T = [R_1, P_1]^{-1}$. The properties of the subsystems essentially rely on the observation that by Proposition 6.1

$$\mathscr{R}_{[E,A,B]} = \mathscr{V}^* \cap \mathscr{W}^* = \operatorname{im}_{\mathbb{R}} R_1 = T^{-1}\big(\mathbb{R}^{n_1} \times \{0\}^{n_2}\big).$$

Remark 7.2 (Kalman decomposition) It is important to note that a trivial reachability space does not necessarily imply that $B = 0$. An intriguing example which illustrates this is the system

$$[E, A, B] = \left[\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right]. \tag{7.3}$$

Another important fact we like to stress by means of this example is that $B \neq 0$ does no necessarily imply $n_1 \neq 0$ in the Kalman decomposition (7.2). In fact, the above system $[E, A, B]$ is already in Kalman decomposition with $k_1 = k_2 = 1, n_1 = 0, n_2 = 1, m = 1$ and $E_{12} = 1$, $A_{12} = 0$, $B_1 = 1$ as well as $E_{22} = 0$, $A_{22} = 1$. Then all the required properties are obtained, in particular $\mathrm{rk}_{\mathbb{R}}[E_{11}, A_{11}, B_1] = \mathrm{rk}_{\mathbb{R}}[1] = 1$ and the system $[E_{11}, A_{11}, B_1]$ is completely controllable as it is in feedback form (3.10) with $\gamma = 1$; complete controllability then follows from Corollary 3.4. However, $[E_{11}, A_{11}, B_1]$ is hard to view as a control system as no equation can be written down. Nevertheless, the space $\mathscr{R}_{[E_{11}, A_{11}, B_1]}$ has dimension zero and obviously every state can be steered to every other state.

We now analyze how two forms of type (7.2) of one system $[E, A, B] \in \Sigma_{k,n,m}$ differ.

Proposition 7.2 (Uniqueness of the Kalman decomposition) *Let* $[E, A, B] \in \Sigma_{k,n,m}$ *be given and assume that, for all* $i \in \{1, 2\}$, *the systems* $[E_i, A_i, B_i] \overset{W_i, T_i}{\sim_{se}} [E, A, B]$ *with*

$$sE_i - A_i = \begin{bmatrix} sE_{11,i} - A_{11,i} & sE_{12,i} - A_{12,i} \\ 0 & sE_{22,i} - A_{22,i} \end{bmatrix}, \qquad B_i = \begin{bmatrix} B_{1,i} \\ 0 \end{bmatrix}$$

where $E_{11,i}, A_{11,i} \in \mathbb{R}^{k_{1,i}, n_{1,i}}$, $E_{12,i}, A_{12,i} \in \mathbb{R}^{k_{1,i}, n_{2,i}}$, $E_{22,i}, A_{22,i} \in \mathbb{R}^{k_{2,i}, n_{2,i}}$, $B_{1,i} \in \mathbb{R}^{k_{1,i}, m}$ *satisfy*

$$\mathrm{rk}_{\mathbb{R}} \begin{bmatrix} E_{11,i} & A_{11,i} & B_{1,i} \end{bmatrix} = k_{1,i}$$

and, in addition, $[E_{11,i}, A_{11,i}, B_{c,i}] \in \Sigma_{k_{1,i}, n_{1,i}, m}$ *is completely controllable and* $\mathscr{R}_{[E_{22,i}, A_{22,i}, 0_{k_{2,i}, m}]} = \{0\}$.

Then $k_{1,1} = k_{1,2}, k_{2,1} = k_{2,2}, n_{1,1} = n_{1,2}, n_{2,1} = n_{2,2}$. *Moreover, for some* $W_{11} \in \mathbf{Gl}_{k_{1,1}}(\mathbb{R})$, $W_{12} \in \mathbb{R}^{k_{1,1}, k_{2,1}}$, $W_{22} \in \mathbf{Gl}_{k_{2,1}}(\mathbb{R})$, $T_{11} \in \mathbf{Gl}_{n_{1,1}}(\mathbb{R})$, $T_{12} \in \mathbb{R}^{n_{1,1}, n_{2,1}}$, $T_{22} \in \mathbf{Gl}_{n_{2,1}}(\mathbb{R})$, *we have*

$$W_2 W_1^{-1} = \begin{bmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{bmatrix}, \qquad T_1^{-1} T_2 = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}.$$

In particular, the systems $[E_{11,1}, A_{11,1}, B_{1,1}]$, $[E_{11,2}, A_{11,2}, B_{1,2}]$ *and, respectively,* $[E_{22,1}, A_{22,1}, 0]$, $[E_{22,2}, A_{22,2}, 0]$ *are system equivalent.*

Proof It is no loss of generality to assume that $W_1 = I_k$, $T_1 = I_n$. Then we obtain

$$\mathbb{R}^{n_{1,1}} \times \{0\} = \mathscr{R}_{[E_1, A_1, B_1]} = T_2 \mathscr{R}_{[E_2, A_2, B_2]} = T_2\big(\mathbb{R}^{n_{1,2}} \times \{0\}\big).$$

This implies $n_{1,1} = n_{1,2}$ and

$$T_2 = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \quad \text{for some } T_{11} \in \mathbf{Gl}_{n_{1,1}},\ T_{12} \in \mathbb{R}^{n_{1,1}, n_{2,1}},\ T_{22} \in \mathbf{Gl}_{n_{2,1}}.$$

Now partitioning

$$W_2 = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix},$$
$$W_{11} \in \mathbb{R}^{k_{1,1},k_{1,2}},\ W_{12} \in \mathbb{R}^{k_{1,1},k_{2,2}},\ W_{21} \in \mathbb{R}^{k_{2,1},k_{c,2}},\ W_{22} \in \mathbb{R}^{k_{2,1},k_{2,2}},$$

the block (2, 1) of the equations $W_1 E_1 T_1 = E_2$, $W_1 A_1 T_1 = A_2$ and $W_1 B_1 = B_2$ give rise to

$$0 = W_{21} \begin{bmatrix} E_{11,2} & A_{11,2} & B_{1,2} \end{bmatrix}.$$

Since the latter matrix is supposed to have full row rank, we obtain $W_{21} = 0$. The assumption of W_2 being invertible then leads to $k_{1,1} \leq k_{1,2}$. Reversing the roles of $[E_1, A_1, B_1]$ and $[E_2, A_2, B_2]$, we further obtain $k_{1,2} \leq k_{1,1}$, whence $k_{1,2} = k_{1,1}$. Using again the invertibility of W, we see that both W_{11} and W_{22} are invertible. □

It is immediate from the form (7.2) that $[E, A, B]$ is completely controllable if, and only if, $n_1 = n$. The following result characterizes the further controllability and stabilizability notions in terms of properties of the submatrices in (7.2).

Corollary 7.3 (Properties induced from the Kalman decomposition) *Consider* $[E, A, B] \in \Sigma_{k,n,m}$ *with*

$$[E, A, B] \overset{W,T}{\sim}_{se} \left[\begin{bmatrix} E_{11} & E_{12} \\ 0 & E_{22} \end{bmatrix}, \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \begin{bmatrix} B_1 \\ 0 \end{bmatrix} \right]$$

such that $[E_{11}, A_{11}, B_1] \in \Sigma_{k_1,n_1,m}$ *is completely controllable*, $\mathrm{rk}_{\mathbb{R}}[E_{11}, A_{11}, B_1] = k_1$ *and* $\mathscr{R}_{[E_{22},A_{22},0_{k_2,m}]} = \{0\}$. *Then the following statements hold true*:

(a) $\mathrm{rk}_{\mathbb{R}(s)}(sE_{22} - A_{22}) = n_2$.
(b) *If* $sE - A$ *is regular, then both pencils* $sE_{11} - A_{11}$ *and* $sE_{22} - A_{22}$ *are regular. In particular, it holds* $k_1 = n_1$ *and* $k_2 = n_2$.
(c) *If* $[E, A, B]$ *is impulse controllable, then the index of the pencil* $sE_{22} - A_{22}$ *is at most one.*
(d) $[E, A, B]$ *is controllable at infinity if, and only if,* $\mathrm{im}_{\mathbb{R}} A_{22} \subseteq \mathrm{im}_{\mathbb{R}} E_{22}$.
(e) $[E, A, B]$ *is controllable in the behavioral sense if, and only if,* $\mathrm{rk}_{\mathbb{R}(s)}(sE_{22} - A_{22}) = \mathrm{rk}_{\mathbb{C}}(\lambda E_{22} - A_{22})$ *for all* $\lambda \in \mathbb{C}$.
(f) $[E, A, B]$ *is stabilizable in the behavioral sense if, and only if,* $\mathrm{rk}_{\mathbb{R}(s)}(sE_{22} - A_{22}) = \mathrm{rk}_{\mathbb{C}}(\lambda E_{22} - A_{22})$ *for all* $\lambda \in \overline{\mathbb{C}}_+$.

Proof (a) Assuming that $\mathrm{rk}_{\mathbb{R}(s)}(sE_{22} - A_{22}) < n_2$, then, in a quasi-Kronecker (3.3) form of $sE_{22} - A_{22}$, it holds $\ell(\beta) > 0$ by (3.6). By the findings of Remark 3.7(ii), we can conclude $\mathscr{R}_{[E_{22},A_{22},0_{k_2,m}]} \neq \{0\}$, a contradiction.

(b) We can infer from (a) that $n_2 \leq k_2$. We can further infer from the regularity of $sE - A$ that $n_2 \geq k_2$. The regularity of $sE_{11} - A_{11}$ and $sE_{22} - A_{22}$ then follows immediately from $\det(sE - A) = \det(W \cdot T) \cdot \det(sE_{11} - A_{11}) \cdot \det(sE_{22} - A_{22})$.

(c) Assume that $[E, A, B]$ is impulse controllable. By Corollary 4.3 and the invariance of impulse controllability under system equivalence this implies that

$$\operatorname{im}_{\mathbb{R}} \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \subseteq \operatorname{im}_{\mathbb{R}} \begin{bmatrix} E_{11} & E_{12} & B_1 & A_{11}Z_1 + A_{12}Z_2 \\ 0 & E_{22} & 0 & A_{22}Z_2 \end{bmatrix},$$

where $Z = [Z_1^\top, Z_2^\top]^\top$ is a real matrix such that $\operatorname{im}_{\mathbb{R}} Z = \ker_{\mathbb{R}} [\begin{smallmatrix} E_{11} & E_{12} \\ 0 & E_{22} \end{smallmatrix}]$. The last condition in particular implies that $\operatorname{im}_{\mathbb{R}} Z_2 \subseteq \ker_{\mathbb{R}} E_{22}$ and therefore we obtain

$$\operatorname{im}_{\mathbb{R}} A_{22} \subseteq \operatorname{im}_{\mathbb{R}} E_{22} + A_{22} \cdot \ker_{\mathbb{R}} E_{22},$$

which is, by (3.4), equivalent to the index of $sE_{22} - A_{22}$ being at most one.

(d) Since $\operatorname{rk}_{\mathbb{R}}[E_{11}, A_{11}, B_1] = k_1$ and the system $[E_{11}, A_{11}, B_1]$ is controllable at infinity, Corollary 4.3 leads to $\operatorname{rk}_{\mathbb{R}}[E_{11}, B_1] = k_1$. Therefore, we have

$$\operatorname{im}_{\mathbb{R}} \begin{bmatrix} E_{11} & E_{12} & B_1 \\ 0 & E_{22} & 0 \end{bmatrix} = \mathbb{R}^{k_1} \times \operatorname{im}_{\mathbb{R}} E_{22}.$$

Analogously, we obtain

$$\operatorname{im}_{\mathbb{R}} \begin{bmatrix} E_{11} & E_{12} & A_{11} & A_{12} & B_1 \\ 0 & E_{22} & 0 & A_{22} & 0 \end{bmatrix} = \mathbb{R}^{k_1} \times (\operatorname{im}_{\mathbb{R}} E_{22} + \operatorname{im}_{\mathbb{R}} A_{22}).$$

Again using Corollary 4.3 and the invariance of controllability at infinity under system equivalence, we see that $[E, A, B]$ is controllable at infinity if, and only if,

$$\mathbb{R}^{k_1} \times (\operatorname{im}_{\mathbb{R}} E_{22} + \operatorname{im}_{\mathbb{R}} A_{22}) = \mathbb{R}^{k_1} \times \operatorname{im}_{\mathbb{R}} E_{22},$$

which is equivalent to $\operatorname{im}_{\mathbb{R}} A_{22} \subseteq \operatorname{im}_{\mathbb{R}} E_{22}$.

(e) Since $\operatorname{rk}_{\mathbb{R}}[E_{11}, A_{11}, B_1] = k_1$ and $[E_{11}, A_{11}, B_1] \in \Sigma_{k_1, n_1, m}$ is completely controllable it holds

$$\operatorname{rk}_{\mathbb{C}}[\lambda E_{11} - A_{11}, B_1] = k_1 \quad \text{for all } \lambda \in \mathbb{C}.$$

Therefore, we have

$$\operatorname{rk}_{\mathbb{C}}[\lambda E - A, B] = \operatorname{rk}_{\mathbb{C}} \begin{bmatrix} \lambda E_{11} - A_{11} & \lambda E_{12} - A_{12} & B_1 \\ 0 & \lambda E_{22} - A_{22} & 0 \end{bmatrix} = k_1 + \operatorname{rk}_{\mathbb{C}}(\lambda E_{22} - A_{22}),$$

and, analogously, $\operatorname{rk}_{\mathbb{R}(s)}[sE - A, B] = k_1 + \operatorname{rk}_{\mathbb{R}(s)}(sE_{22} - A_{22})$. Now applying Corollary 4.3 we find that $[E, A, B]$ is controllable in the behavioral sense if, and only if, $\operatorname{rk}_{\mathbb{R}(s)}(sE_{22} - A_{22}) = \operatorname{rk}_{\mathbb{C}}(\lambda E_{22} - A_{22})$ for all $\lambda \in \mathbb{C}$.

(f) The proof of this statement is analogous to (e). □

Remark 7.3 (Kalman decomposition and controllability) Note that the condition of the index of $sE_{22} - A_{22}$ being at most one in Corollary 7.3(c) is equivalent to the system $[E_{22}, A_{22}, 0_{k_2,m}]$ being impulse controllable. Likewise, the condition

$\mathrm{im}_{\mathbb{R}} A_{22} \subseteq \mathrm{im}_{\mathbb{R}} E_{22}$ in (d) is equivalent to $[E_{22}, A_{22}, 0_{k_2,m}]$ being controllable at infinity. Obviously, the conditions in (e) and (f) are equivalent to behavioral controllability and stabilizability of $[E_{22}, A_{22}, 0_{k_2,m}]$, resp.

Furthermore, the converse implication in (b) does not hold true. That is, the index of $sE_{22} - A_{22}$ being at most one is in general not sufficient for $[E, A, B]$ being impulse controllable. For instance, reconsider system (7.3) which is not impulse controllable, but $sE_{22} - A_{22} = -1$ is of index one. Even in the case where $sE - A$ is regular, the property of the index of $sE_{22} - A_{22}$ being zero or one is not enough to infer impulse controllability of $sE - A$. As a counterexample, consider

$$[E, A, B] = \left[\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right].$$

Acknowledgements We are indebted to Harry L. Trentelman (University of Groningen) for providing helpful comments on the behavioral approach.

References

1. Adams, R.A.: Sobolev Spaces. Pure and Applied Mathematics, vol. 65. Academic Press, New York (1975)
2. Anderson, B.D.O., Vongpanitlerd, S.: Network Analysis and Synthesis—A Modern Systems Theory Approach. Prentice-Hall, Englewood Cliffs (1973)
3. Aplevich, J.D.: Minimal representations of implicit linear systems. Automatica **21**(3), 259–269 (1985)
4. Aplevich, J.D.: Implicit Linear Systems. Lecture Notes in Control and Information Sciences, vol. 152. Springer, Berlin (1991)
5. Armentano, V.A.: Eigenvalue placement for generalized linear systems. Syst. Control Lett. **4**, 199–202 (1984)
6. Armentano, V.A.: The pencil $(sE - A)$ and controllability-observability for generalized linear systems: a geometric approach. SIAM J. Control Optim. **24**, 616–638 (1986)
7. Ascher, U.M., Petzold, L.R.: Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations. SIAM, Philadelphia (1998)
8. Aubin, J.P., Cellina, A.: Differential Inclusions: Set-Valued Maps and Viability Theory. Grundlehren der mathematischen Wissenschaften, vol. 264. Springer, Berlin (1984)
9. Aubin, J.P., Frankowska, H.: Set Valued Analysis. Birkhäuser, Boston (1990)
10. Augustin, F., Rentrop, P.: Numerical methods and codes for differential algebraic equations. In: Surveys in Differential-Algebraic Equations I. Differential-Algebraic Equations Forum, vol. 2. Springer, Berlin (2012)
11. Banaszuk, A., Przyłuski, K.M.: On perturbations of controllable implicit linear systems. IMA J. Math. Control Inf. **16**, 91–102 (1999)
12. Banaszuk, A., Kocięcki, M., Przyłuski, K.M.: On Hautus-type conditions for controllability of implicit linear discrete-time systems. Circuits Syst. Signal Process. **8**(3), 289–298 (1989)
13. Banaszuk, A., Kocięcki, M., Przyłuski, K.M.: Implicit linear discrete-time systems. Math. Control Signals Syst. **3**(3), 271–297 (1990)
14. Banaszuk, A., Kocięcki, M., Przyłuski, K.M.: Kalman-type decomposition for implicit linear discrete-time systems, and its applications. Int. J. Control **52**(5), 1263–1271 (1990)
15. Banaszuk, A., Kocięcki, M., Lewis, F.L.: Kalman decomposition for implicit linear systems. IEEE Trans. Autom. Control **37**(10), 1509–1514 (1992)
16. Basile, G., Marro, G.: Controlled and Conditioned Invariants in Linear System Theory. Prentice-Hall, Englewood Cliffs (1992)

17. Belevitch, V.: Classical Network Theory. Holden-Day, San Francisco (1968)
18. Belur, M., Trentelman, H.: Stabilization, pole placement and regular implementability. IEEE Trans. Autom. Control **47**(5), 735–744 (2002)
19. Bender, D.J., Laub, A.J.: Controllability and observability at infinity of multivariable linear second-order models. IEEE Trans. Autom. Control **AC-30**, 1234–1237 (1985)
20. Bender, D.J., Laub, A.J.: The linear-quadratic optimal regulator for descriptor systems. In: Proc. 24th IEEE Conf. Decis. Control, Ft. Lauderdale, FL, pp. 957–962 (1985)
21. Bender, D., Laub, A.: The linear quadratic optimal regulator problem for descriptor systems. IEEE Trans. Autom. Control **32**, 672–688 (1987)
22. Berger, T., Trenn, S.: The quasi-Kronecker form for matrix pencils. SIAM J. Matrix Anal. Appl. **33**(2), 336–368 (2012)
23. Berger, T., Trenn, S.: Addition to: "The quasi-Kronecker form for matrix pencils". SIAM J. Matrix Anal. Appl. **34**(1), 94–101 (2013). doi:10.1137/120883244
24. Berger, T., Ilchmann, A., Reis, T.: Normal forms, high-gain, and funnel control for linear differential-algebraic systems. In: Biegler, L.T., Campbell, S.L., Mehrmann, V. (eds.) Control and Optimization with Differential-Algebraic Constraints. Advances in Design and Control, vol. 23, pp. 127–164. SIAM, Philadelphia (2012)
25. Berger, T., Ilchmann, A., Reis, T.: Zero dynamics and funnel control of linear differential-algebraic systems. Math. Control Signals Syst. **24**(3), 219–263 (2012)
26. Berger, T., Ilchmann, A., Trenn, S.: The quasi-Weierstraß form for regular matrix pencils. Linear Algebra Appl. **436**(10), 4052–4069 (2012)
27. Bernhard, P.: On singular implicit linear dynamical systems. SIAM J. Control Optim. **20**(5), 612–633 (1982)
28. Birkhoff, G., MacLane, S.: A Survey of Modern Algebra, 4th edn. Macmillan Publishing Co., New York (1977)
29. Bonilla Estrada, M., Malabre, M.: On the control of linear systems having internal variations. Automatica **39**, 1989–1996 (2003)
30. Bonilla, M., Malabre, M., Loiseau, J.J.: Implicit systems reachability: a geometric point of view. In: Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, Shanghai, P.R. China, pp. 4270–4275 (2009)
31. Brenan, K.E., Campbell, S.L., Petzold, L.R.: Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations. North-Holland, Amsterdam (1989)
32. Brunovský, P.: A classification of linear controllable systems. Kybernetika **3**, 137–187 (1970)
33. Bunse-Gerstner, A., Mehrmann, V., Nichols, N.K.: On derivative and proportional feedback design for descriptor systems. In: Kaashoek, M.A., et al. (eds.) Proceedings of the International Symposium on the Mathematical Theory of Networks and Systems, Amsterdam, Netherlands (1989)
34. Bunse-Gerstner, A., Mehrmann, V., Nichols, N.K.: Regularization of descriptor systems by derivative and proportional state feedback. Report, University of Reading, Dept. of Math., Numerical Analysis Group, Reading, UK (1991)
35. Byers, R., Kunkel, P., Mehrmann, V.: Regularization of linear descriptor systems with variable coefficients. SIAM J. Control Optim. **35**, 117–133 (1997)
36. Calahan, D.A.: Computer-Aided Network Design. McGraw-Hill, New York (1972). Rev. edn
37. Campbell, S.L.: Singular Systems of Differential Equations I. Pitman, New York (1980)
38. Campbell, S.L.: Singular Systems of Differential Equations II. Pitman, New York (1982)
39. Campbell, S.L., Carl, D., Meyer, J., Rose, N.J.: Applications of the Drazin inverse to linear systems of differential equations with singular constant coefficients. SIAM J. Appl. Math. **31**(3), 411–425 (1976). http://link.aip.org/link/?SMM/31/411/1. doi:10.1137/0131035
40. Campbell, S.L., Nichols, N.K., Terrell, W.J.: Duality, observability, and controllability for linear time-varying descriptor systems. Circuits Syst. Signal Process. **10**(4), 455–470 (1991)
41. Christodoulou, M.A., Paraskevopoulos, P.N.: Solvability, controllability, and observability of singular systems. J. Optim. Theory Appl. **45**, 53–72 (1985)

42. Cobb, J.D.: Descriptor Variable and Generalized Singularly Perturbed Systems: A Geometric Approach. Univ. of Illinois, Dept. of Electrical Engineering, Urbana-Champaign (1980)
43. Cobb, J.D.: Feedback and pole placement in descriptor variable systems. Int. J. Control **33**(6), 1135–1146 (1981)
44. Cobb, J.D.: On the solution of linear differential equations with singular coefficients. J. Differ. Equ. **46**, 310–323 (1982)
45. Cobb, J.D.: Descriptor variable systems and optimal state regulation. IEEE Trans. Autom. Control **AC-28**, 601–611 (1983)
46. Cobb, J.D.: Controllability, observability and duality in singular systems. IEEE Trans. Autom. Control **AC-29**, 1076–1082 (1984)
47. Crouch, P.E., van der Schaft, A.J.: Variational and Hamiltonian Control Systems. Lecture Notes in Control and Information Sciences, vol. 101. Springer, Berlin (1986)
48. Cuthrell, J.E., Biegler, L.T.: On the optimization of differential-algebraic process systems. AIChE J. **33**(8), 1257–1270 (1987)
49. Dai, L.: Singular Control Systems. Lecture Notes in Control and Information Sciences, vol. 118. Springer, Berlin (1989)
50. Daoutidis, P.: DAEs in chemical engineering: a survey. In: Surveys in Differential-Algebraic Equations I. Differential-Algebraic Equations Forum, vol. 2. Springer, Berlin (2012)
51. Diehla, M., Uslu, I., Findeisen, R., Schwarzkopf, S., Allgöwer, F., Bock, H.G., Bürner, T., Gilles, E.D., Kienle, A., Schlöder, J.P., Stein, E.: Real-time optimization for large scale processes: nonlinear model predictive control of a high purity distillation column. In: Grötschel, M., Krumke, S.O., Rambau, J. (eds.) Online Optimization of Large Scale Systems: State of the Art, pp. 363–384. Springer, Berlin (2001)
52. Diehla, M., Bock, H.G., Schlöder, J.P., Findeisen, R., Nagyc, Z., Allgöwer, F.: Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations. J. Process Control **12**, 577–585 (2002)
53. Dieudonné, J.: Sur la réduction canonique des couples des matrices. Bull. Soc. Math. Fr. **74**, 130–146 (1946)
54. Dziurla, B., Newcomb, R.W.: Nonregular Semistate Systems: Examples and Input-Output Pairing. IEEE Press, New York (1987)
55. Eich-Soellner, E., Führer, C.: Numerical Methods in Multibody Dynamics. Teubner, Stuttgart (1998)
56. Eliopoulou, H., Karcanias, N.: Properties of reachability and almost reachability subspaces of implicit systems: the extension problem. Kybernetika **31**(6), 530–540 (1995)
57. Fletcher, L.R., Kautsky, J., Nichols, N.K.: Eigenstructure assignment in descriptor systems. IEEE Trans. Autom. Control **AC-31**, 1138–1141 (1986)
58. Frankowska, H.: On controllability and observability of implicit systems. Syst. Control Lett. **14**, 219–225 (1990)
59. Führer, C., Leimkuhler, B.J.: Numerical solution of differential-algebraic equations for constrained mechanical motion. Numer. Math. **59**, 55–69 (1991)
60. Gantmacher, F.R.: The Theory of Matrices, vols. I & II. Chelsea, New York (1959)
61. Geerts, A.H.W.T.: Solvability conditions, consistency and weak consistency for linear differential-algebraic equations and time-invariant linear systems: the general case. Linear Algebra Appl. **181**, 111–130 (1993)
62. Geerts, A.H.W.T., Mehrmann, V.: Linear differential equations with constant coefficients: a distributional approach. Tech. Rep. SFB 343 90-073, Bielefeld University, Germany (1990)
63. Glüsing-Lüerßen, H.: Feedback canonical form for singular systems. Int. J. Control **52**(2), 347–376 (1990)
64. Glüsing-Lüerßen, H., Hinrichsen, D.: A Jordan control canonical form for singular systems. Int. J. Control **48**(5), 1769–1785 (1988)
65. Gresho, P.M.: Incompressible fluid dynamics: some fundamental formulation issues. Annu. Rev. Fluid Mech. **23**, 413–453 (1991)
66. Griepentrog, E., März, R.: Differential-Algebraic Equations and Their Numerical Treatment. Teubner-Texte zur Mathematik, vol. 88. Teubner, Leipzig (1986)

67. Haug, E.J.: Computer-Aided Kinematics and Dynamics of Mechanical Systems. Allyn and Bacon, Boston (1989)
68. Hautus, M.L.J.: Controllability and observability condition for linear autonomous systems. Proc. Ned. Akad. Wet., Ser. A **72**, 443–448 (1969)
69. Helmke, U., Shayman, M.A.: A canonical form for controllable singular systems. Syst. Control Lett. **12**(2), 111–122 (1989)
70. Hinrichsen, D., Pritchard, A.J.: Mathematical Systems Theory I. Modelling, State Space Analysis, Stability and Robustness. Texts in Applied Mathematics, vol. 48. Springer, Berlin (2005)
71. Hou, M.: Controllability and elimination of impulsive modes in descriptor systems. IEEE Trans. Autom. Control **49**(10), 1723–1727 (2004)
72. Ilchmann, A., Mehrmann, V.: A behavioural approach to time-varying linear systems, Part 1: general theory. SIAM J. Control Optim. **44**(5), 1725–1747 (2005)
73. Ilchmann, A., Mehrmann, V.: A behavioural approach to time-varying linear systems, Part 2: descriptor systems. SIAM J. Control Optim. **44**(5), 1748–1765 (2005)
74. Ilchmann, A., Nürnberger, I., Schmale, W.: Time-varying polynomial matrix systems. Int. J. Control **40**(2), 329–362 (1984)
75. Ishihara, J.Y., Terra, M.H.: Impulse controllability and observability of rectangular descriptor systems. IEEE Trans. Autom. Control **46**(6), 991–994 (2001)
76. Isidori, A.: Nonlinear Control Systems, 3rd edn. Communications and Control Engineering Series. Springer, Berlin (1995)
77. Isidori, A.: Nonlinear Control Systems II. Communications and Control Engineering Series. Springer, London (1999)
78. Jaffe, S., Karcanias, N.: Matrix pencil characterization of almost (A, B)-invariant subspaces: a classification of geometric concepts. Int. J. Control **33**(1), 51–93 (1981)
79. Julius, A., van der Schaft, A.: Compatibility of behavioral interconnections. In: Proc. 7th European Control Conf. 2003, Cambridge, UK (2003)
80. Kailath, T.: Linear Systems. Prentice-Hall, Englewood Cliffs (1980)
81. Kalman, R.E.: On the general theory of control systems. In: Proceedings of the First International Congress on Automatic Control, Moscow, 1960, pp. 481–493. Butterworth's, London (1961)
82. Kalman, R.E.: Canonical structure of linear dynamical systems. Proc. Natl. Acad. Sci. USA **48**(4), 596–600 (1962)
83. Kalman, R.E.: Mathematical description of linear dynamical systems. SIAM J. Control Optim. **1**, 152–192 (1963)
84. Karcanias, N.: Regular state-space realizations of singular system control problems. In: Proc. 26th IEEE Conf. Decis. Control, Los Angeles, CA, pp. 1144–1146 (1987)
85. Karcanias, N., Hayton, G.E.: Generalised autonomous dynamical systems, algebraic duality and geometric theory. In: Proc. 8th IFAC World Congress, Kyoto, 1981, vol. III, pp. 13–18 (1981)
86. Karcanias, N., Kalogeropoulos, G.: A matrix pencil approach to the study of singular systems: algebraic and geometric aspects. In: Proc. Int. Symp. on Singular Systems, Atlanta, GA, pp. 29–33 (1987)
87. Karcanias, N., Kalogeropoulos, G.: Geometric theory and feedback invariants of generalized linear systems: a matrix pencil approach. Circuits Syst. Signal Process. **8**(3), 375–397 (1989)
88. Karcanias, N., Kouvaritakis, B.: The output zeroing problem and its relationship to the invariant zero structure: a matrix pencil approach. Int. J. Control **30**(3), 395–415 (1979)
89. Knobloch, H.W., Kwakernaak, H.: Lineare Kontrolltheorie. Springer, Berlin (1985)
90. Koumboulis, F.N., Mertzios, B.G.: On Kalman's controllability and observability criteria for singular systems. Circuits Syst. Signal Process. **18**(3), 269–290 (1999)
91. Kouvaritakis, B., MacFarlane, A.G.J.: Geometric approach to analysis and synthesis of system zeros Part 1. Square systems. Int. J. Control **23**(2), 149–166 (1976)
92. Kouvaritakis, B., MacFarlane, A.G.J.: Geometric approach to analysis and synthesis of system zeros Part 2. Non-square systems. Int. J. Control **23**(2), 167–181 (1976)

93. Kronecker, L.: Algebraische Reduction der Schaaren Bilinearer Formen. Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin, pp. 1225–1237 (1890)
94. Kučera, V., Zagalak, P.: Fundamental theorem of state feedback for singular systems. Automatica **24**(5), 653–658 (1988)
95. Kuijper, M.: First-Order Representations of Linear Systems. Birkhäuser, Boston (1994)
96. Kunkel, P., Mehrmann, V.: Differential-Algebraic Equations. Analysis and Numerical Solution. EMS Publishing House, Zürich (2006)
97. Kunkel, P., Mehrmann, V., Rath, W.: Analysis and numerical solution of control problems in descriptor form. Math. Control Signals Syst. **14**, 29–61 (2001)
98. Lamour, R., März, R., Tischendorf, C.: Differential Algebraic Equations: A Projector Based Analysis. Differential-Algebraic Equations Forum, vol. 1. Springer, Heidelberg (2012)
99. Lewis, F.L.: A survey of linear singular systems. IEEE Proc., Circuits Syst. Signal Process. **5**(1), 3–36 (1986)
100. Lewis, F.L.: A tutorial on the geometric analysis of linear time-invariant implicit systems. Automatica **28**(1), 119–137 (1992)
101. Lewis, F.L., Özçaldiran, K.: Reachability and controllability for descriptor systems. In: Proc. 27, Midwest Symp. on Circ. Syst., Morgantown, WV (1984)
102. Lewis, F.L., Özçaldiran, K.: On the Eigenstructure Assignment of Singular Systems. IEEE Press, New York (1985)
103. Lewis, F.L., Özçaldiran, K.: Geometric structure and feedback in singular systems. IEEE Trans. Autom. Control **AC-34**(4), 450–455 (1989)
104. Loiseau, J.: Some geometric considerations about the Kronecker normal form. Int. J. Control **42**(6), 1411–1431 (1985)
105. Loiseau, J., Özçaldiran, K., Malabre, M., Karcanias, N.: Feedback canonical forms of singular systems. Kybernetika **27**(4), 289–305 (1991)
106. Lötstedt, P., Petzold, L.R.: Numerical solution of nonlinear differential equations with algebraic constraints I: convergence results for backward differentiation formulas. Math. Comput. **46**(174), 491–516 (1986)
107. Luenberger, D.G.: Dynamic equations in descriptor form. EEE Trans. Autom. Control **AC-22**, 312–321 (1977)
108. Luenberger, D.G.: Time-invariant descriptor systems. Automatica **14**, 473–480 (1978)
109. Luenberger, D.G.: Introduction to Dynamic Systems: Theory, Models and Applications. Wiley, New York (1979)
110. Luenberger, D.G.: Nonlinear descriptor systems. J. Econ. Dyn. Control **1**, 219–242 (1979)
111. Luenberger, D.G., Arbel, A.: Singular dynamic Leontief systems. Econometrica **45**, 991–995 (1977)
112. Malabre, M.: More Geometry About Singular Systems. IEEE Press, New York (1987)
113. Malabre, M.: Generalized linear systems: geometric and structural approaches. Linear Algebra Appl. **122–124**, 591–621 (1989)
114. Masubuchi, I.: Stability and stabilization of implicit systems. In: Proc. 39th IEEE Conf. Decis. Control, Sydney, Australia, vol. 12, pp. 3636–3641 (2000)
115. Mertzios, B.G., Christodoulou, M.A., Syrmos, B.L., Lewis, F.L.: Direct controllability and observability time domain conditions of singular systems. IEEE Trans. Autom. Control **33**(8), 788–791 (1988)
116. Müller, P.C.: Remark on the solution of linear time-invariant descriptor systems. In: PAMM—Proc. Appl. Math. Mech., GAMM Annual Meeting 2005, Luxemburg, vol. 5, pp. 175–176. Wiley-VCH Verlag GmbH, Weinheim (2005). doi:10.1002/pamm.200510066
117. Newcomb, R.W.: The semistate description of nonlinear time-variable circuits. IEEE Trans. Circuits Syst. **CAS-28**, 62–71 (1981)
118. Özçaldiran, K.: Control of descriptor systems. Ph.D. thesis, Georgia Institute of Technology (1985)
119. Özçaldiran, K.: A geometric characterization of the reachable and controllable subspaces of descriptor systems. IEEE Proc., Circuits Syst. Signal Process. **5**, 37–48 (1986)

120. Özçaldiran, K., Haliloĉlu, L.: Structural properties of singular systems. Kybernetika **29**(6), 518–546 (1993)
121. Özçaldiran, K., Lewis, F.L.: A geometric approach to eigenstructure assignment for singular systems. IEEE Trans. Autom. Control **AC-32**(7), 629–632 (1987)
122. Özçaldiran, K., Lewis, F.L.: Generalized reachability subspaces for singular systems. SIAM J. Control Optim. **27**, 495–510 (1989)
123. Özçaldiran, K., Lewis, F.L.: On the regularizability of singular systems. IEEE Trans. Autom. Control **35**(10), 1156 (1990)
124. Pandolfi, L.: Controllability and stabilization for linear systems of algebraic and differential equations. J. Optim. Theory Appl. **30**, 601–620 (1980)
125. Pandolfi, L.: On the regulator problem for linear degenerate control systems. J. Optim. Theory Appl. **33**, 241–254 (1981)
126. Pantelides, C.C.: The consistent initialization of differential-algebraic systems. SIAM J. Sci. Stat. Comput. **9**, 213–231 (1988)
127. Petzold, L.R.: Numerical solution of differential-algebraic equations in mechanical systems simulation. Physica D **60**, 269–279 (1992)
128. Polderman, J.W., Willems, J.C.: Introduction to Mathematical Systems Theory. A Behavioral Approach. Springer, New York (1997)
129. Popov, V.M.: Hyperstability of Control Systems. Springer, Berlin (1973). Translation based on a revised text prepared shortly after the publication of the Romanian ed., 1966
130. Przyłuski, K.M., Sosnowski, A.M.: Remarks on the theory of implicit linear continuous-time systems. Kybernetika **30**(5), 507–515 (1994)
131. Pugh, A.C., Ratcliffe, P.A.: On the zeros and poles of a rational matrix. Int. J. Control **30**, 213–226 (1979)
132. Rabier, P.J., Rheinboldt, W.C.: Classical and generalized solutions of time-dependent linear differential-algebraic equations. Linear Algebra Appl. **245**, 259–293 (1996)
133. Rath, W.: Feedback design and regularization for linear descriptor systems with variable coefficients. Dissertation, TU Chemnitz, Chemnitz, Germany (1997)
134. Riaza, R.: Differential-Algebraic Systems. Analytical Aspects and Circuit Applications. World Scientific Publishing, Basel (2008)
135. Riaza, R.: DAEs in circuit modelling: a survey. In: Surveys in Differential-Algebraic Equations I. Differential-Algebraic Equations Forum, vol. 2. Springer, Berlin (2012)
136. Rosenbrock, H.H.: State Space and Multivariable Theory. Wiley, New York (1970)
137. Rosenbrock, H.H.: Structural properties of linear dynamical systems. Int. J. Control **20**, 191–202 (1974)
138. Rugh, W.J.: Linear System Theory, 2nd edn. Information and System Sciences Series. Prentice-Hall, New York (1996)
139. Schiehlen, W.: Multibody system dynamics: roots and perspectives. Multibody Syst. Dyn. **1**, 149–188 (1997)
140. Shayman, M.A., Zhou, Z.: Feedback control and classification of generalized linear systems. IEEE Trans. Autom. Control **32**(6), 483–490 (1987)
141. Simeon, B., Führer, C., Rentrop, P.: Differential-algebraic equations in vehicle system dynamics. Surv. Math. Ind. **1**, 1–37 (1991)
142. Sontag, E.D.: Mathematical Control Theory: Deterministic Finite Dimensional Systems, 2nd edn. Springer, New York (1998)
143. Trenn, S.: Distributional differential algebraic equations. Ph.D. thesis, Institut für Mathematik, Technische Universität Ilmenau, Universitätsverlag Ilmenau, Ilmenau, Germany (2009). http://www.db-thueringen.de/servlets/DocumentServlet?id=13581
144. Trenn, S.: Regularity of distributional differential algebraic equations. Math. Control Signals Syst. **21**(3), 229–264 (2009). doi:10.1007/s00498-009-0045-4
145. Trenn, S.: Solution concepts for linear DAEs: a survey. In: Surveys in Differential-Algebraic Equations I. Differential-Algebraic Equations Forum, vol. 2. Springer, Berlin (2013)
146. Trentelman, H., Willems, J.: The behavioral approach as a paradigm for modelling interconnected systems. Eur. J. Control **9**(2–3), 296–306 (2003)

147. Trentelman, H.L., Stoorvogel, A.A., Hautus, M.: Control Theory for Linear Systems. Communications and Control Engineering. Springer, London (2001)
148. van der Schaft, A.J.: System Theoretic Descriptions of Physical Systems. CWI Tract, No. 3. CWI, Amsterdam (1984)
149. van der Schaft, A.J.: Port-Hamiltonian differential-algebraic systems. In: Surveys in Differential-Algebraic Equations I. Differential-Algebraic Equations Forum, vol. 2. Springer, Berlin (2012)
150. van der Schaft, A.J., Schumacher, J.M.H.: The complementary-slackness class of hybrid systems. Math. Control Signals Syst. **9**, 266–301 (1996). doi:10.1007/BF02551330
151. Verghese, G.C.: Infinite-frequency behavio in generalized dynamical systems. Ph.D. thesis, Stanford University (1978)
152. Verghese, G.C.: Further notes on singular systems. In: Proc. Joint American Contr. Conf. (1981). Paper TA-4B
153. Verghese, G.C., Kailath, T.: Eigenvector chains for finite and infinite zeros of rational matrices. In: Proc. 18th Conf. Dec. and Control, Ft. Lauderdale, FL, pp. 31–32 (1979)
154. Verghese, G.C., Kailath, T.: Impulsive behavior in dynamical systems: structure and significance. In: Dewilde, P. (ed.) Proc. 4th MTNS, pp. 162–168 (1979)
155. Verghese, G.C., Levy, B.C., Kailath, T.: A generalized state-space for singular systems. IEEE Trans. Autom. Control **AC-26**(4), 811–831 (1981)
156. Wang, C.J.: Controllability and observability of linear time-varying singular systems. IEEE Trans. Autom. Control **44**(10), 1901–1905 (1999)
157. Wang, C.J., Liao, H.E.: Impulse observability and impulse controllability of linear time-varying singular systems. Automatica **2001**(37), 1867–1872 (2001
158. Weierstraß, K.: Zur Theorie der bilinearen und quadratischen Formen. Berl. Monatsb., pp. 310–338 (1868)
159. Wilkinson, J.H.: Linear differential equations and Kronecker's canonical form. In: de Boor, C., Golub, G.H. (eds.) Recent Advances in Numerical Analysis, pp. 231–265. Academic Press, New York (1978)
160. Willems, J.C.: System theoretic models for the analysis of physical systems. Ric. Autom. **10**, 71–106 (1979)
161. Willems, J.C.: Paradigms and puzzles in the theory of dynamical systems. IEEE Trans. Autom. Control **AC-36**(3), 259–294 (1991)
162. Willems, J.C.: On interconnections, control, and feedback. IEEE Trans. Autom. Control **42**, 326–339 (1997)
163. Willems, J.C.: The behavioral approach to open and interconnected systems. IEEE Control Syst. Mag. **27**(6), 46–99 (2007)
164. Wong, K.T.: The eigenvalue problem $\lambda Tx + Sx$. J. Differ. Equ. **16**, 270–280 (1974)
165. Wonham, W.M.: On pole assignment in multi–input controllable linear systems. IEEE Trans. Autom. Control **AC-12**, 660–665 (1967)
166. Wonham, W.M.: Linear Multivariable Control: A Geometric Approach, 3rd edn. Springer, New York (1985)
167. Wood, J., Zerz, E.: Notes on the definition of behavioural controllability. Syst. Control Lett. **37**, 31–37 (1999)
168. Yamada, T., Luenberger, D.G.: Generic controllability theorems for descriptor systems. IEEE Trans. Autom. Control **30**(2), 144–152 (1985)
169. Yip, E.L., Sincovec, R.F.: Solvability, controllability and observability of continuous descriptor systems. IEEE Trans. Autom. Control **AC-26**, 702–707 (1981)
170. Zhou, Z., Shayman, M.A., Tarn, T.J.: Singular systems: a new approach in the time domain. IEEE Trans. Autom. Control **32**(1), 42–50 (1987)
171. Zubova, S.P.: On full controllability criteria of a descriptor system. The polynomial solution of a control problem with checkpoints. Autom. Remote Control **72**(1), 23–37 (2011)

Robust Stability of Differential-Algebraic Equations

Nguyen Huu Du, Vu Hoang Linh, and Volker Mehrmann

Abstract This paper presents a survey of recent results on the robust stability analysis and the distance to instability for linear time-invariant and time-varying differential-algebraic equations (DAEs). Different stability concepts such as exponential and asymptotic stability are studied and their robustness is analyzed under general as well as restricted sets of real or complex perturbations. Formulas for the distances are presented whenever these are available and the continuity of the distances in terms of the data is discussed. Some open problems and challenges are indicated.

Keywords Differential-algebraic equation · Restricted perturbation · Robust stability · Stability radius · Spectrum · Index

Mathematics Subject Classification 93B35 · 93D09 · 34A09 · 34D10

1 Introduction

In many areas of science and engineering one uses mathematical models to simulate, control or optimize a system or process. These mathematical models, however, are typically inexact or contain uncertainties and thus, the following question is of major importance.

How robust is a specific property of a given system described by differential or difference equations under perturbations to the data?

N.H. Du · V.H. Linh (✉)
Faculty of Mathematics, Mechanics and Informatics, Vietnam National University,
334, Nguyen Trai Str., Thanh Xuan, Hanoi, Vietnam
e-mail: linhvh@vnu.edu.vn

N.H. Du
e-mail: dunh@vnu.edu.vn

V. Mehrmann
Institut für Mathematik, MA 4-5, Technische Universität Berlin, 10623 Berlin,
Fed. Rep. Germany
e-mail: mehrmann@math.tu-berlin.de

A. Ilchmann, T. Reis (eds.), *Surveys in Differential-Algebraic Equations I*,
Differential-Algebraic Equations Forum,
DOI 10.1007/978-3-642-34928-7_2,

Here, we say that a certain property of a system is *robust* if it is preserved when an arbitrary (but sufficiently small) perturbation affects the system. An important quantity in this respect is then the distance (measured by an appropriate metric) between the nominal system and the closest perturbed system that does not possess the mentioned property, this is typically called the *radius* of the system property.

In this paper, we deal with robustness and distance problems for differential-algebraic equations (DAEs), with a focus on robust stability and stability radii. Systems of DAEs, which are also called descriptor systems in the control literature, are a very convenient modeling concept in various real-life applications such as mechanical multibody systems, electrical circuit simulation, chemical reactions, semi-discretized partial differential equations, and in general for automatically generated coupled systems, see [12, 39, 47, 63, 68, 84, 85] and the references therein.

DAEs are generalizations of ordinary differential equations (ODEs) in that certain algebraic equations constrain the dynamical behavior. Since the dynamics of DAEs is constrained to a set which often is only given implicitly, many theoretical and numerical difficulties arise, which may lead to a sensitive behavior of the solution of DAEs to perturbation in the data. The difficulties are characterized by fundamental notions for DAEs such as regularity, index, solution subspace, or hidden constraints, which do not arise for ODEs. These properties may be easily lost when the data are subject to arbitrarily small perturbations. As a consequence, usually restrictions to the allowed perturbations have to be made, leading to robustness questions for DAEs that are very different from those for ODEs.

This paper surveys robustness results for linear DAEs with time-invariant or time-varying coefficients of the form

$$E(t)\dot{x}(t) = A(t)x(t) + f(t), \tag{1.1}$$

on the half-line $\mathbb{I} = [0, \infty)$, together with an initial condition

$$x(t_0) = x_0, \quad t_0 \in \mathbb{I}. \tag{1.2}$$

Here we assume that $E, A \in C(\mathbb{I}, \mathbb{K}^{n\times n})$, and $f \in C(\mathbb{I}, \mathbb{K}^n)$ are sufficiently smooth. We use the notation $C(\mathbb{I}, \mathbb{K}^{n\times n})$ to denote the space of continuous functions from $\mathbb{I}$ to $\mathbb{K}^{n\times n}$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$.

Linear systems of the form (1.1) arise directly in many applications and via linearization around solution trajectories [22]. They describe the local behavior in the neighborhood of a solution for a general implicit nonlinear system of DAEs

$$F\big(t, x(t), \dot{x}(t)\big) = 0, \tag{1.3}$$

the constant coefficient case arising in the case of linearization around stationary solutions.

Definition 1.1 A function $x : \mathbb{I} \to \mathbb{R}^n$ is called a *solution* of (1.1) if $x \in C^1(\mathbb{I}, \mathbb{R}^n)$ and x satisfies (1.1) pointwise. It is called a *solution of the initial value problem* (1.1)–(1.2) if x is a solution of (1.1) and satisfies (1.2). An initial condition (1.2) is called *consistent* if the corresponding initial value problem has at least one solution.

We note that, by the tractability index approach [43, 68], the condition on the smoothness of solutions may be relaxed, namely, only a part of x is required to be continuously differentiable.

Recall the following classical stability concepts for ordinary differential equations:

$$\dot{x}(t) = f\big(t, x(t)\big), \quad t \in \mathbb{I}, \tag{1.4}$$

with initial condition (1.2), see e.g. [55].

Definition 1.2 A solution $x : t \mapsto x(t; t_0, x_0)$ of (1.4) with initial condition (1.2) is called

1. *stable* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that
   (a) the initial value problem (1.4) with initial condition $x(t_0) = \hat{x}_0$ is solvable on $\mathbb{I}$ for all $\hat{x}_0 \in \mathbb{K}^n$ with $\|\hat{x}_0 - x_0\| < \delta$;
   (b) the solution $x(t; t_0, \hat{x}_0)$ satisfies $\|x(t; t_0, \hat{x}_0) - x(t; t_0, x_0)\| < \varepsilon$ on $\mathbb{I}$.
2. *asymptotically stable* if it is stable and there exists $\rho > 0$ such that
   (a) the initial value problem (1.4) with initial condition $x(t_0) = \hat{x}_0$ is solvable on $\mathbb{I}$ for all $\hat{x}_0 \in \mathbb{K}^n$ with $\|\hat{x}_0 - x_0\| < \rho$;
   (b) the solution $x(t; t_0, \hat{x}_0)$ satisfies $\lim_{t\to\infty} \|x(t; t_0, \hat{x}_0) - x(t; t_0, x_0)\| = 0$.
3. *exponentially stable* if it is stable and *exponentially attractive*, i.e., if there exist $\delta > 0$, $L > 0$, and $\gamma > 0$ such that
   (a) the initial value problem (1.4) with initial condition $x(t_0) = \hat{x}_0$ is solvable on $\mathbb{I}$ for all $\hat{x}_0 \in \mathbb{K}^n$ with $\|\hat{x}_0 - x_0\| < \delta$;
   (b) the solution satisfies the estimate $\|x(t; t_0, \hat{x}_0) - x(t; t_0, x_0)\| < Le^{-\gamma(t-t_0)}$ on $\mathbb{I}$.

If δ does not depend on t_0, then we say the solution is uniformly (exponentially) stable.

Note that one can transform the ODE (1.4) in such a way that a given solution $x(t; t_0, x_0)$ is mapped to the trivial solution by simply shifting the arguments. When studying the stability of a selected solution, one may therefore assume without loss of generality that the selected solution is the trivial solution, and also that $t_0 = 0$.

In this paper, we restrict the discussion to *regular* DAEs, i.e., we require that (1.1) (or (1.3) locally) have a unique solution for sufficiently smooth E, A, f (F) and appropriately chosen (consistent) initial conditions. One can immediately extend Definition 1.2 verbatim to regular DAEs. However, one has to be careful with the initial conditions and the inhomogeneities, since they are restricted due to the algebraic constraints in the system. This is, in particular, true if one considers the robustness of the stability concepts under perturbations to the system.

The following examples give an illustration for the possible difficulties in the robustness of the stability concepts for DAEs under small perturbations.

Example 1.1 Consider the homogeneous linear time-invariant DAE

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \tag{1.5}$$

which can be written as $\dot{x}_1 = x_2$, $0 = x_1$ and has only the trivial solution $x_1 = x_2 = 0$.

If we perturb (1.5) by a small ε as

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & \varepsilon \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \tag{1.6}$$

then solving the second equation of (1.6) for x_2 and substituting into the first equation, we obtain

$$\dot{x}_1 = -(1/\varepsilon)x_1. \tag{1.7}$$

Clearly, if $\varepsilon < 0$, then the perturbed DAE (1.6) is unstable. If $\varepsilon > 0$, then the system is asymptotically stable, but it qualitatively differs from the solution of the original system (1.5). For an arbitrarily prescribed initial value $x_1(0) \neq 0$, the initial value problem for (1.7) has a unique solution. Furthermore, the value of $x_2(0)$ is not required and is uniquely determined by $x_1(0)$. In fact, this small perturbation has changed the index of the DAE (1.5), see Definition 2.2 below.

If we add an inhomogeneity to these DAEs, then more essential differences appear.

In Example 1.1 the perturbation is affecting only the coefficient A. The situation is even more complicated if we allow perturbations in the coefficient of $\dot{x}$.

Example 1.2 Consider the well-known singularly perturbed system

$$\begin{bmatrix} I_{n_1} & 0 \\ 0 & \varepsilon I_{n_2} \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \tag{1.8}$$

where A_{ij}, $i, j \in \{1, 2\}$, are constant matrices of appropriate sizes, and $\varepsilon > 0$ is a small parameter. Let us assume that A_{22} is invertible. If ε is set to 0, then the leading matrix becomes singular, i.e., we have a DAE, and we can solve the second equation for x_2, and obtain the so-called *underlying ODE*

$$\dot{x}_1 = \big(A_{11} - A_{12}A_{22}^{-1}A_{21}\big)x_1.$$

It is well known that for sufficiently small ε, the (asymptotic) stability of (1.8) depends not only on the stability of the so-called *slow subsystem* associated with the underlying ODE, but also on that of the so called *fast subsystem* $\dot{x}_2 = A_{22}x_2$, see [32, 34, 61], associated with the algebraic equation.

In this example, the rank of the leading matrix is changed, when ε moves from zero to a nonzero value. In the case $\varepsilon = 0$, the initial condition must be consistent to ensure the existence of a solution, but obviously this is not required in the case of a nonzero ε. The difficulties increase if A_{22} is singular and/or the leading matrix involves a singular perturbation of a more general structure.

The presented examples already indicate some of the possible difficulties which will be discussed in this paper. We present the analysis of robust exponential and asymptotic stability for linear DAEs with time-invariant or time-varying coefficients. This is a relatively young topic, starting with the work in [20, 82], which generalized results on the distance to instability and the concept of stability radius for ODEs in [50] and [92].

The outline of the paper is as follows. In Sect. 2 we summarize recent results on the robust stability and stability radii for linear time-invariant DAEs. In Sect. 3 we study the robust stability of linear time-varying DAEs. Stability radii, their dependence on the data, and the robustness of stability spectra are analyzed. Some discussions and topics for future research close the paper.

2 Robust Stability of Linear Time-Invariant DAEs

In this section we study homogeneous linear time-invariant DAEs of the form

$$E\dot{x}(t) = Ax(t), \quad t \in \mathbb{I}, \tag{2.1}$$

where E and A are given constant matrices in $\mathbb{K}^{n\times n}$, $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$.

Definition 2.1 A matrix pair (E, A), $E, A \in \mathbb{K}^{n\times n}$ is called *regular* if there exists $\lambda \in \mathbb{C}$ such that the determinant of $(\lambda E - A)$, denoted by $\det(\lambda E - A)$, is different from zero. Otherwise, if $\det(\lambda E - A) = 0$ for all $\lambda \in \mathbb{C}$, then we say that (E, A) is *singular*.

If (E, A) is regular, then a complex number λ is called a (*generalized finite*) *eigenvalue* of (E, A) if $\det(\lambda E - A) = 0$. The set of all (finite) eigenvalues of (E, A) is called the (*finite*) *spectrum of the pencil* (E, A) and denoted by $\sigma(E, A)$. If E is singular and the pair is regular, then we say that (E, A) has the eigenvalue ∞.

In the following we only consider regular pairs (E, A). Such pairs can be transformed to *Weierstraß–Kronecker canonical form*, see [12, 41, 43], i.e., there exist nonsingular matrices $W, T \in \mathbb{C}^{n\times n}$ such that

$$E = W \begin{bmatrix} I_r & 0 \\ 0 & N \end{bmatrix} T^{-1}, \qquad A = W \begin{bmatrix} J & 0 \\ 0 & I_{n-r} \end{bmatrix} T^{-1}, \tag{2.2}$$

where I_r, I_{n-r} are identity matrices of indicated size, $J \in \mathbb{C}^{r\times r}$, and $N \in \mathbb{C}^{(n-r)\times(n-r)}$ are matrices in Jordan canonical form and N is nilpotent. If E is invertible, then $r = n$, i.e., the second diagonal block does not occur.

Definition 2.2 Consider a regular pair (E, A) with E, $A \in \mathbb{K}^{n\times n}$ in Weierstraß–Kronecker form (2.2). If $r < n$ and N has nilpotency index $\nu \in \{1, 2, \ldots\}$, i.e., $N^\nu = 0$, $N^i \neq 0$ for $i = 1, 2, \ldots, \nu - 1$, then ν is called the *index of the pair* (E, A) *and the associated DAE* (2.1) and we write $\mathrm{ind}(E, A) = \nu$. If $r = n$ then the DAE has index $\nu = 0$.

The finite spectrum $\sigma(E, A)$ is given by the spectrum of $\sigma(J)$ in the Weierstraß–Kronecker form (2.2) and it is easy to verify that for the *degree of the characteristic polynomial* $\deg\det(\lambda E - A) = \operatorname{rank} E = r$ holds if and only if $\operatorname{ind}(E, A) \leq 1$.

For the regular DAE (2.1) with initial condition (1.2), there always exists a projection matrix $P \in \mathbb{K}^{n\times r}$ so that the projected initial condition $P(x(t_0) - x_0) = 0$ is *consistent*, i.e., the DAE (2.1) with this initial condition has a unique solution, see [12, 43]. Shifting again the solution to the trivial solution $x = 0$, for a DAE of the form (2.1) with regular pair (E, A), E, $A \in \mathbb{K}^{n\times n}$ we say that this solution is *exponentially stable* if there exist $L > 0$, and $\gamma > 0$ such that the initial value problem

$$E\dot{x} = Ax, \qquad P\big(x(t_0) - \hat{x}_0\big) = 0,$$

is solvable on $\mathbb{I}$ for all $\hat{x}_0 \in \mathbb{K}^n$, and the solution satisfies the estimate $\|x(t; t_0, \hat{x}_0)\| < Le^{-\gamma(t-t_0)}\|P\hat{x}_0\|$ for all $t \geq t_0$. If the trivial solution is exponentially stable, then we say that (2.1) *is exponentially stable*.

We remark that the property of exponential stability is independent of the choice of projection P. Furthermore, for linear time-invariant systems, the concept of exponential stability is equivalent to that of asymptotic stability and hence we do not have to distinguish these concepts, and discuss only asymptotic stability as is usually done in the literature.

Using the canonical form (2.2), one can easily verify the following statement, see e.g., [20].

Proposition 2.1 *Consider a DAE of the form* (2.1) *with regular pair* (E, A), E, $A \in \mathbb{K}^{n\times n}$. *System* (2.1) *is asymptotically stable if and only if the pair* (E, A) *is* asymptotically stable, *i.e., the finite spectrum satisfies* $\sigma(E, A) \subset \mathbb{C}^-$, *where* $\mathbb{C}^-$ *denotes the* open left-half complex plane.

After introducing the basic notation, in the next subsection we discuss the stability radius of a DAE.

2.1 Stability Radii for Linear Time-Invariant DAEs

In this section we study the behavior of the finite spectrum of a regular pair (E, A) under structured perturbations in the matrices E, A. Suppose that the system (2.1) is asymptotically stable and consider a perturbed system

$$(E + B_1\Delta_1 C_1)\dot{x} = (A + B_2\Delta_2 C_2)x, \tag{2.3}$$

where $\Delta_i \in \mathbb{K}^{m_i\times q_i}$ $(i = 1, 2)$ are perturbations and $B_i \in \mathbb{K}^{n\times m_i}$, $C_i \in \mathbb{K}^{q_i\times n}$ are given matrix pairs that restrict the structure of the perturbations. The matrix pair $(B_1\Delta_1 C_1, B_2\Delta_2 C_2)$ is called a *structured perturbation*. For simplicity and for the sake of appropriate perturbation structures, let us consider the case that the restricting matrices satisfy $C_1 = C_2 = C$. Alternatively, the other simplifying case $B_1 = B_2$

can be treated as well, see [38]. However, the simplification does not fit to the framework of so-called admissible perturbations characterized by Proposition 2.6 given below. Set

$$\Delta = \begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix}, \qquad B = \begin{bmatrix} B_1 & B_2 \end{bmatrix},$$

and introduce $m = m_1 + m_2$ and $q = q_1 = q_2$. Then we consider the set of destabilizing perturbations

$$\mathscr{V}_{\mathbb{K}}(E, A; B, C) = \left\{ \Delta \in \mathbb{K}^{m \times q}, \text{(2.3) is singular or not asymptotically stable} \right\}$$

and have the following definition.

Definition 2.3 The *structured stability radius* of pair (E, A) subject to structured perturbations as in (2.3) is defined by

$$r_{\mathbb{K}}^{\mathrm{sp}}(E, A; B, C) = \inf\left\{ \|\Delta\|, \Delta \in \mathscr{V}_{\mathbb{K}}(E, A; B, C) \right\},$$

where $\|\cdot\|$ is a matrix norm induced by a vector norm. Depending on $\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$, we talk about the *complex or the real structured stability radius*, respectively.

Note that other properties of pair (E, A) such as the index may still change under a perturbation which is not in $\mathscr{V}_{\mathbb{K}}(E, A; B, C)$. Obviously, we have the estimate

$$r_{\mathbb{C}}^{\mathrm{sp}}(E, A; B, C) \le r_{\mathbb{R}}^{\mathrm{sp}}(E, A; B, C).$$

To obtain a computable formula for the complex stability radius, let us introduce the matrix functions

$$\mathscr{G}_1(s) = -sC(sE - A)^{-1}B_1, \quad \mathscr{G}_2(s) = C(sE - A)^{-1}B_2, \quad \mathscr{G}(s) = \begin{bmatrix} \mathscr{G}_1(s) & \mathscr{G}_2(s) \end{bmatrix},$$

for $s \in \mathbb{C}, \operatorname{Re} s \ge 0$. Denoting by $i\mathbb{R}$ the imaginary axis of the complex plane, the following result is analogous to that for linear time-invariant ODEs of [51].

Theorem 2.2 *Suppose that the matrix pencil (E, A) is regular and asymptotically stable. Then with respect to any matrix norm induced by a vector norm, the complex stability radius of pair (E, A) has the representation*

$$r_{\mathbb{C}}^{\mathrm{sp}}(E, A; B, C) = \left\{ \sup_{s \in i\mathbb{R}} \left\| \mathscr{G}(s) \right\| \right\}^{-1}. \tag{2.4}$$

Proof The proof can be obtained by using the same techniques as in [33, 37, 38]. □

Note that in [33, 37] the complex structured stability radius is considered with respect to perturbations either in E or in A, i.e., either $B_1 = 0$ or $B_2 = 0$. In [38], formula (2.4) has been proven for perturbations in both E, A. Note further that these

papers discuss both the continuous-time and the discrete-time case. Furthermore, it has been shown in [38] that it is always possible to construct a rank 1 destabilizing perturbation Δ, with a norm that approximates the value of $r_{\mathbb{C}}^{\mathrm{sp}}$ within an arbitrarily prescribed accuracy.

Remark 2.1 The concept of stability radius can be extended to more general sets. Suppose that all the eigenvalues of the unperturbed matrix pencil lie in a prescribed open subset $\mathbb{C}_g$ of the complex plane. Then we want to determine the largest perturbations that the system can tolerate so that its spectrum remains in $\mathbb{C}_g$. In the asymptotic stability analysis of differential equations, the open subset $\mathbb{C}_g$ is chosen to be $\mathbb{C}^-$. As in the other cases, it is trivial to obtain a formula of a $\mathbb{C}_g$-stability radius analogously to (2.4). In fact, we simply replace $i\mathbb{R}$ by the boundary set of $\mathbb{C}_b = \mathbb{C}\backslash\mathbb{C}_g$. As a consequence of the definition, the strict positivity of a $\mathbb{C}_g$-stability radius with a relevant subset $\mathbb{C}_g$ implies the continuity of the spectrum with respect to the data.

Unlike for the complex stability radius, a general formula for the real stability radius measured by an arbitrary matrix norm is not available. However, if we consider as vector norm the Euclidean norm, then a computable formula has been obtained in [83]. This formula is based on the notion of *real/complex structured singular values*, which, for a given $M \in \mathbb{K}^{p\times m}$, are defined by

$$\mu_{\mathbb{K}}(M) = \big[\inf\big\{\sigma_1(\Delta),\, \Delta \in \mathbb{K}^{m\times p},\ \text{and } \det(I - \Delta M) = 0\big\}\big]^{-1},$$

respectively, depending on $\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$. Here $\sigma_1(\Delta)$ denotes the largest *singular value*, see [42], of the matrix Δ.

Clearly, if M is real, then the complex and the real structured singular values coincide. While for the complex structured singular values, the formula $\mu_{\mathbb{C}}(M) = \sigma_1(M)$ follows trivially, for the real case the formula is more sophisticated.

Proposition 2.3 ([83]) *The real structured singular value of $M \in \mathbb{K}^{p\times m}$ is given by*

$$\mu_{\mathbb{R}}(M) = \inf_{\gamma\in(0,1]} \sigma_2 \begin{bmatrix} \operatorname{Re} M & -\gamma \operatorname{Im} M \\ \frac{1}{\gamma}\operatorname{Im} M & \operatorname{Re} M \end{bmatrix},$$

where $\sigma_2(A)$ denotes the second largest singular value of A.

With respect to the Euclidean norm and using a similar argument as in [83], we thus have

$$\begin{aligned} r_{\mathbb{K}}^{\mathrm{sp}}(E, A; B, C) &= \inf\big\{\sigma_1(\Delta),\, \Delta \in \mathscr{V}_{\mathbb{K}}(E, A; B, C)\big\} \\ &= \inf_{\operatorname{Re} s\geq 0} \inf\big\{\sigma_1(\Delta),\, \Delta \in \mathbb{K}^{m\times q} \text{ and} \\ &\qquad\qquad \det\big(s(E + B_1\Delta_1 C) - (A + B_2\Delta_2 C)\big) = 0\big\} \end{aligned}$$

$$
\begin{aligned}
&= \inf_{\operatorname{Re} s \geq 0} \inf\{\sigma_1(\Delta),\ \Delta \in \mathbb{K}^{m\times q} \text{ and } \\
&\qquad\qquad \det\big(I + (sE - A)^{-1}(sB_1\Delta_1 C - B_2\Delta_2 C)\big) = 0\} \\
&= \inf_{\operatorname{Re} s \geq 0} \inf\{\sigma_1(\Delta),\ \Delta \in \mathbb{K}^{m\times q} \text{ and } \det\big(I - \Delta\mathcal{G}(s)\big) = 0\} \\
&= \Big\{\sup_{\operatorname{Re} s \geq 0} \mu_{\mathbb{K}}\big(\mathcal{G}(s)\big)\Big\}^{-1},
\end{aligned}
$$

and hence we obtain the following theorem.

Theorem 2.4 *Suppose that the matrix pair (E, A) is regular and asymptotically stable. Then the structured stability radii of pair (E, A), measured in Euclidean norm, are given by*

$$
r_{\mathbb{C}}^{\mathrm{sp}}(E, A; B, C) = \Big\{\sup_{\operatorname{Re} s \geq 0} \sigma_1\big(\mathcal{G}(s)\big)\Big\}^{-1} \tag{2.5}
$$

and

$$
r_{\mathbb{R}}^{\mathrm{sp}}(E, A; B, C) = \left\{\sup_{\operatorname{Re} s \geq 0} \inf_{\gamma\in(0,1]} \sigma_2 \begin{bmatrix} \operatorname{Re}\mathcal{G}(s) & -\gamma \operatorname{Im}\mathcal{G}(s) \\ \frac{1}{\gamma}\operatorname{Im}\mathcal{G}(s) & \operatorname{Re}\mathcal{G}(s) \end{bmatrix}\right\}^{-1}, \tag{2.6}
$$

respectively.

For the case $\mathbb{K} = \mathbb{C}$, due to the maximum principle, it suffices to take the supremum over the imaginary axis instead of the right-half complex plane. The same does not hold for the case $\mathbb{K} = \mathbb{R}$, see Example 2.1 below. Further, unlike the case of ODEs, here one cannot replace sup by max since the supremum may be attained only at infinity. It is important to note that the presented results on the structured stability radii do not reflect the fact that an eigenvalue at ∞ may become finite or conversely a finite eigenvalue may move to ∞, i.e., it may happen that the index or the number of finite eigenvalues of the pair (E, A) changes or that the pair becomes singular.

In the case of inhomogeneous systems, particularly an increase of the index may lead to a loss of solvability of the equation due to inconsistent initial values or a lack of smoothness for the inhomogeneity. As we have demonstrated in Examples 1.1 and 1.2, this can even happen with infinitesimally small perturbations. Furthermore, while the stability radii of an ODE are always strictly positive, those of a DAE may be zero. To see this, considering a pair in canonical form (2.2), we have

$$
(sE - A)^{-1} = T \begin{bmatrix} (sI_r - J)^{-1} & 0 \\ 0 & -\sum_{i=0}^{k-1}(sN)^i \end{bmatrix} W^{-1}.
$$

Obviously, if $N \neq 0$, then $\|\mathcal{G}_1(s)\|$ and $\|\mathcal{G}_2(s)\|$ may tend to ∞ as $|s| \to \infty$, which implies that $r_{\mathbb{C}}^{\mathrm{sp}} = 0$. Hence, the perturbations in (2.3) must be further restricted so that the stability radii are strictly positive.

Partitioning the restriction matrices C, B (after transformation to Weierstraß–Kronecker form) as

$$CT = [C_1 \quad C_2], \qquad W^{-1}B_1 = \begin{bmatrix} B_{11} \\ B_{12} \end{bmatrix}, \qquad W^{-1}B_2 = \begin{bmatrix} B_{21} \\ B_{22} \end{bmatrix}, \tag{2.7}$$

according to the structure of (2.2), it is easy to see that if $\operatorname{ind}(E, A) = 1$ then

$$\sup_{s \in i\mathbb{R}} \left\| \mathscr{G}_1(s) \right\| < \infty \quad \text{if and only if} \quad C_2 B_{12} = 0$$

and if $\operatorname{ind}(E, A) > 1$ then

$$\sup_{s \in i\mathbb{R}} \left\| \mathscr{G}_2(s) \right\| < \infty \quad \text{if and only if} \quad C_2 N^i B_{12} = 0 \quad \text{for } i = 0, 1, \ldots, k-1, \quad \text{and}$$

$$C_2 N^i B_{22} = 0 \quad \text{for } i = 1, 2, \ldots, k-1.$$

These observations are summarized in the following result.

Proposition 2.5 *Consider a regular pair (E, A) and the associated DAE of the form (2.1). If $\operatorname{ind}(E, A) = 1$, then the structured stability radii of (2.1) are strictly positive if and only if $C_2 B_{12} = 0$. If $\operatorname{ind}(E, A) > 1$, then the structured stability radii of (2.1) are strictly positive if and only if $C_2 N^i B_{12} = 0$ for $i = 0, 1, \ldots, k-1$, and $C_2 N^i B_{22} = 0$ for $i = 1, 2, \ldots, k-1$, where the transformed structure matrices are defined by (2.7).*

Moreover, if C_2 is of full rank, then $r^{\mathrm{sp}}_{\mathbb{K}}(E, A, B, C) > 0$ if and only if $B_{12} = 0$ for the case $\operatorname{ind}(E, A) = 1$ and $B_{12} = 0$, $N B_{22} = 0$ for the case $\operatorname{ind}(E, A) > 1$.

According to the characterizations given in Proposition 2.5 and for the sake of simplicity, now the perturbations are further restricted by choosing

$$B_1 = W \begin{bmatrix} B_{11} \\ 0 \end{bmatrix}, \tag{2.8}$$

if $\operatorname{ind}(E, A) = 1$ and

$$B_1 = W \begin{bmatrix} B_{11} \\ 0 \end{bmatrix}, \qquad B_2 = W \begin{bmatrix} B_{21} \\ 0 \end{bmatrix} \tag{2.9}$$

if $\operatorname{ind}(E, A) > 1$.

Definition 2.4 ([20]) A structured perturbation as in (2.3) is called *admissible* if it does not alter the nilpotency structure of the Weierstraß–Kronecker form (2.2) of (E, A), i.e., the nilpotent matrix N and the corresponding left invariant subspace associated with the eigenvalue ∞ are preserved.

In the case that $\operatorname{ind}(E, A) = 1$, one has the following characterization of admissible perturbations, which is in agreement with (2.8).

Proposition 2.6 ([20]) *Consider the regular DAE* (2.1) *with* $\operatorname{ind}(E, A) = 1$, *subject to a general unstructured perturbation*,

$$(E + F)\dot{x} = (A + H)x.$$

Then there exist an orthogonal matrix P *and a permutation matrix* Q *such that*

$$PEQ = \begin{bmatrix} E_{11} & E_{12} \\ 0 & 0 \end{bmatrix}, \qquad PAQ = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where $E_{11} \in \mathbb{K}^{r\times r}$, $E_{12} \in \mathbb{K}^{r\times(n-r)}$, A_{ij} $(i, j = 1, 2)$ *are of corresponding sizes*, $\operatorname{rank}[E_{11}, E_{12}] = \operatorname{rank} E = r$, *and* $\operatorname{rank} A_{22} = n - r$. *Furthermore, if* (F, H) *is an admissible perturbation, then*

$$PFQ = \begin{bmatrix} F_{11} & F_{12} \\ 0 & 0 \end{bmatrix}, \qquad PHQ = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}.$$

Note that in Proposition 2.6, the transformation by the matrices P, Q does not change the structure, the stability, and consequently, the stability radii of (2.1). Note further that Proposition 2.5 can also be used to characterize admissible perturbations for the case $\operatorname{ind}(E, A) > 1$.

After these observations we can introduce the *distance to the nearest pair with a different nilpotency structure:*

$$\begin{aligned} &d_{\mathbb{K}}(E, A; B, C) \\ &\quad = \inf\big\{\|\Delta\|, \Delta \in \mathbb{K}^{m\times q} \text{ and (2.3) does not preserve the nilpotency structure}\big\}, \end{aligned}$$

and obtain the following result, see [11].

Theorem 2.7 *Consider a regular DAE with Weierstraß–Kronecker form* (2.2), *subject to transformed perturbations satisfying* (2.8) *for* $\operatorname{ind}(E, A) = 1$ *and* (2.9) *for* $\operatorname{ind}(E, A) > 1$, *respectively. Then the distance to the nearest system with a different nilpotency structure is given by*

$$d_{\mathbb{K}}(E, A; B, C) = \big\{\mu_{\mathbb{K}}\begin{bmatrix} C_1 B_{11} & C_2 B_{22} \end{bmatrix}\big\}^{-1}$$

if $\operatorname{ind}(E, A) = 1$ *and*

$$d_{\mathbb{K}}(E, A; B, C) = \big\{\mu_{\mathbb{K}}(C_1 B_{11})\big\}^{-1},$$

if $\operatorname{ind}(E, A) > 1$. *Moreover, if the data set is real, then* $d_{\mathbb{C}}(E, A; B, C) = d_{\mathbb{R}}(E, A; B, C)$.

Proof The proof is similar to that for the stability radii, see Remark 2.1. The nilpotency structure of the perturbed system (2.3) is preserved if and only if the perturbed

matrix

$$\begin{bmatrix} I + B_{11}\Delta_1 C_1 & B_{11}\Delta_1 C_2 \\ B_{22}\Delta_2 C_1 & I + B_{22}\Delta_2 C_2 \end{bmatrix}$$

is nonsingular in the case $\mathrm{ind}(E, A) = 1$ and if $I + B_{11}\Delta_1 C_1$ is nonsingular in the case $\mathrm{ind}(E, A) > 1$. Thus, in both cases we have a problem of characterizing the distance of a matrix to singularity, which we obtain by taking $\mathbb{C}_g = \mathbb{C}\backslash\{0\}$. For the equality of the complex and the real stability radii, we note that if the data are real, then the smallest perturbation that makes a matrix singular can always be chosen to be real [42]. □

Now we are in position to define the structured stability radii for the DAE (2.1) and state their formulas.

Definition 2.5 Consider a regular and asymptotically stable DAE (2.1). Then, the structured stability radii of system (2.1) with respect to structured perturbation (2.3) is defined by

$$\begin{aligned} &r_{\mathbb{K}}(E, A; B, C) \\ &\quad = \inf\big\{\|\Delta\|, \Delta \in \mathscr{V}(E, A; B, C) \text{ or (2.3) has different nilpotency structure}\big\}. \end{aligned}$$

It is obvious that $r_{\mathbb{K}}(E, A; B, C) = \min\{r^{\mathrm{sp}}_{\mathbb{K}}(E, A; B, C), d_{\mathbb{K}}(E, A; B, C)\}$, which follows directly from the definition of the stability radii.

Theorem 2.8 *Consider a regular and asymptotically stable DAE* (2.1) *with Weierstraß–Kronecker form* (2.2). *Let the perturbation structure satisfy* (2.8) *for index* $\mathrm{ind}(E, A) = 1$ *and* (2.9) *for* $\mathrm{ind}(E, A) > 1$, *respectively. Then, the complex structured stability radius of system* (2.1) *and that of pair* (E, A) *coincide, i.e.,* $r_{\mathbb{C}}(E, A; B, C) = r^{\mathrm{sp}}_{\mathbb{C}}(E, A; B, C)$. *Furthermore, if the data set is real, then the real structured stability radius of system* (2.1) *and that of pair* (E, A) *are the same, too, that is,* $r_{\mathbb{R}}(E, A; B, C) = r^{\mathrm{sp}}_{\mathbb{R}}(E, A; B, C)$.

Proof Taking into consideration (2.8) for index $\mathrm{ind}(E, A) = 1$ and (2.9) for $\mathrm{ind}(E, A) > 1$, respectively, it is easy to check that

$$\lim_{s\to\infty} \mathscr{G}_1(s) = -C_1 B_{11}, \quad \text{and} \quad \lim_{s\to\infty} \mathscr{G}_2(s) = -C_2 B_{22}.$$

Since $\sup_{s\in i\mathbb{R}} \|\mathscr{G}(s)\| \geq \|\mathscr{G}(\infty)\|$ (due to the continuity of the norm), the statement for the case $\mathbb{K} = \mathbb{C}$ is obtained from Theorems 2.2 and 2.7.

If the data set is real then $\mathscr{G}(s)$ is real along the non-negative real semi-axis, and then due to the continuity of $\mu_{\mathbb{R}}(\mathscr{G}(s))$ on $\mathbb{R}^+$, we have $\sup_{\mathrm{Re}\, s\geq 0} \mu_{\mathbb{R}}(\mathscr{G}(s)) \geq \sup_{s\in\mathbb{R}^+} \mu_{\mathbb{R}}(\mathscr{G}(s)) \geq \mu_{\mathbb{R}}(\mathscr{G}(\infty))$. Invoking again Theorems 2.2 and 2.7, the statement for the real stability radii immediately follows. □

Theorem 2.8 shows that if $\|\Delta\| < r^{\mathrm{sp}}_{\mathbb{K}}(E, A; B, C)$ then the perturbed DAE (2.3) preserves not only the stability, but also the nilpotency structure. Thus, the complex

structured stability radius of system (2.1) are given by (2.4) or (2.5), while if the data set is real, then the real structured stability radius can be computed by (2.6). For ODEs, the perturbed system becomes unstable if and only if at least one of its eigenvalues touches the imaginary axis, while it may happen with DAEs that under the effect of larger and larger perturbations, a finite eigenvalue moves to ∞ (which alters the nilpotency structure) and then appears again as a finite eigenvalue on the right-half complex plane (not necessarily on the imaginary axis).

Example 2.1 Consider the linear constant coefficient DAE with structured perturbation to only the right-hand side $E\dot{x} = (A + B\Delta C)x$, where

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad A = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 1 \end{bmatrix}.$$

This system is of index 1 and has only one finite eigenvalue $\lambda = -2$. Then, a simple calculation yields $\mathscr{G}(s) = 1 - 1/(2 + s)$. Thus, $\|\mathscr{G}(s)\| = \sqrt{1 - 3/(4 + |s|^2)}$ for $s \in i\mathbb{R}$, which implies $\sup_{s\in i\mathbb{R}} \|\mathscr{G}(s)\| = 1$ and $r_{\mathbb{C}}(E, A, B, C) = 1$. We also have

$$\mu_{\mathbb{R}}\big(\mathscr{G}(s)\big) = \begin{cases} 1 - \frac{1}{2+s}, & s \in \mathbb{R},\ s \geq 0, \\ 0, & \text{otherwise}, \end{cases}$$

from which $\sup_{\mathrm{Re}\, s\geq 0} \mu_{\mathbb{R}}(\mathscr{G}(s)) = 1$ and $r_{\mathbb{R}}(E, A, B, C) = 1$ follow. However, we remark that $\sup_{\mathrm{Re}\, s\geq 0} \mu_{\mathbb{R}}(\mathscr{G}(s))$ is attainable only at $+\infty$, but not on the imaginary axis as in the case $\mathbb{K} = \mathbb{C}$. This is also reflected by the quite different effects of complex and real perturbations as the following simple calculations show. The perturbed DAE (2.3) reads

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -2 - \Delta & \Delta \\ -\Delta & -1 + \Delta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{2.10}$$

For $\Delta = 1$, the system (2.10) is of index 2 and the associated finite spectrum is empty (the only finite eigenvalue moves to infinity), i.e., the pair $(E, A + B\Delta C)$ is stable, but the index is changed. With the perturbation $\Delta = 1 + 1/(1 + s)$, $\mathrm{Re}\, s \geq 0$, the pair $(E, A + B\Delta C)$ is again of index 1 and the only eigenvalue is s. This means that choosing $s \in i\mathbb{R}$, $|s| \gg 1$, the norm of the complex perturbation approximates the value of $r_{\mathbb{C}}(E, A, B, C)$ within arbitrary accuracy and the only finite eigenvalue appears on the imaginary axis. If we consider only real perturbations, which happens if and only if s is real, then by taking $s \gg 1$, the norm of the real perturbation approximates the value of $r_{\mathbb{R}}(E, A, B, C)$ within arbitrary accuracy and the finite eigenvalue is located on the positive real semi-axis.

Remark 2.2 The first results for stability radii for linear time-invariant DAEs of index 1 were given in [82] and [20]. In [82] only unstructured perturbations in A were considered and the formula for the unstructured complex stability radius measured in Euclidean norm is exactly a special case of (2.5) with restriction matrices $B_1 = 0$ and $B_2 = C = I$. A more general result was obtained in [20], where the

authors considered admissible perturbations as in Proposition 2.6 and formulated the complex stability radius using the Frobenius norm which is not a matrix norm induced by a vector norm. However, using the fact that the Frobenius norm gives an upper bound for the Euclidean norm, the stability radii in Frobenius norm are upper bounds for those in Euclidean norm. On the other hand, in the proof of Theorem 2.2, a rank one destabilizing perturbation can be constructed whose (Euclidean) norm approximates the true value of the stability radius with an arbitrarily small accuracy. Since the Euclidean and the Frobenius norm of a rank one matrix are equal, the formula of the complex stability radius given in [20] and (2.5) yield the same value, i.e., the Frobenius norm case can be considered as a special case of (2.4) and (2.5).

Remark 2.3 A somewhat more general and extensive analysis of stability radii for DAEs is given in [11], where the robustness of the structure and the spectrum are treated separately. The quantities $r_{\mathbb{K}}^{\mathrm{sp}}(E, A; B, C)$ and $d_{\mathbb{K}}(E, A; B, C)$ are called the *spectral* and the *structure-preserving* stability radii, respectively. This approach makes the characterization of stability radii for higher index DAEs possible as well as that for the special uncertainty structure of affine perturbations. The latter one also extends the result in [69] to the case when both coefficient matrices are perturbed with a one-parameter family. In addition, the work in [11] is partially devoted to robust stability of second order DAEs with applications in electrical networks.

Remark 2.4 Since in general the real stability radius is more complicated than the complex one, the question when they are equal is of great practical interest. It has been shown in [87] that for the class of positive systems the complex and the real stability radii coincide and they are easily computable. Attempts to extend this result to DAEs and to other implicit dynamic equations are presented in [33] and [38], respectively.

Remark 2.5 In many applications, one does not only have static perturbations as in (2.3), but also linear time-varying, nonlinear or even nonlinear dynamic perturbations. In the context of regular DAEs (2.1) of index at most one, if perturbations are admitted in A only, then it is possible to extend the concept of structured stability radii with respect to linear time-varying, nonlinear, or nonlinear dynamic perturbations. It can be shown that all the complex structured stability radii with respect to different classes of perturbations are equal, as is well known in the ODE case [53], see also Corollary 3.4 below.

Remark 2.6 Numerical algorithms for computing the stability radii for ODEs are proposed in a number of works, e.g., see [10, 13, 18, 40, 44–46, 48, 57, 79–81, 89]. Some extensions to DAEs are discussed in [2–4, 11]. Since the robust stability of a linear time-invariant system is closely related to the sensitivity of the spectrum, the characterization and the computation of stability radii is also very closely related to the topics of spectral value sets [49, 54] and pseudospectra [17, 91].

Remark 2.7 Robustness questions can also be discussed for other fundamental concepts of control theory such as controllability, observability, stabilizability, or detectability. These concepts have been extended to DAEs in many different publications, see e.g., [14–16, 23, 27, 76, 78] and a recent survey [8]. It is natural to analyze the robustness of these properties when the control systems are subject to uncertain perturbations, which leads to similar distance problems as for robust stability. For ODEs, such distance problems are extensively investigated in a number of works, e. g., see [58, 79] and the references therein. Results on the controllability radius for linear time-invariant descriptor systems are given in [19, 66, 67, 88].

It is well known that solutions of DAEs are more sensitive to data than those of ODEs. This topic has been discussed in [75] for a perturbed index two DAE in semi-explicit form and in [30, 90] for general singularly perturbation problems of DAEs. But a general perturbation theory for linear time-invariant DAEs is *still open*. This is partly due to the fact that no complete characterization of the distance to the nearest singular pencil is available [21].

2.2 Dependence of Stability Radii on the Data

In view of the numerical computation of the stability radii, a natural question is whether the structured stability radii $r_{\mathbb{K}}(E, A; B, C)$ depend continuously on the data E, A, B, C. In the ODE case, i.e., $E = I$ and if only A is perturbed, it was shown in [52, 83] that the complex structured stability radius depends continuously on data, but the real one does not. This is due to the continuity (discontinuity) of the complex (real) structured singular value, see [83]. Extending these results to DAEs, it follows that the complex structured stability radius $r_{\mathbb{C}}(E, A; B, C)$ depends continuously on the data, provided that the nilpotency structure is preserved, i.e., we are restricted to the set of DAEs (2.1) that have the same nilpotency structure and to the set of admissible perturbations. In [34, 35] the robust stability of the parameterized DAE system

$$(E + \varepsilon F)\dot{x} = Ax, \tag{2.11}$$

is considered, where $\varepsilon > 0$ is a small parameter and the unperturbed DAE ($\varepsilon = 0$) is assumed to be regular, of index at most one and asymptotically stable. The classical singularly perturbed system (1.8) is a special case of this more general system. If εF belongs to the class of admissible perturbations characterized by Proposition 2.6, then it is easily shown that the complex structured stability radius depends continuously on the parameter ε. This, however, is not the case when the appearance of εF changes the index and/or the number of finite eigenvalues, i.e., the nilpotency structure of (E, A).

Sufficient conditions can be given to ensure that (2.11) is asymptotically stable for all sufficiently small and positive ε. Namely, if the unperturbed DAE ($\varepsilon = 0$) and the fast subsystem (which is associated with the algebraic part of the DAE) are simultaneously asymptotically stable, then for all sufficiently small ε, so is the

parameterized system (2.11). Furthermore, the complex structured stability radius of (2.11) converges to the minimum of the stability radius of the reduced system and that of the fast subsystem. This implies that the stability radius of (2.11) may be discontinuous at $\varepsilon = 0$, when the nilpotency structure is no longer invariant. As a special case, the result for the robust stability of (1.8), investigated in [31] by a different approach, then follows immediately. The asymptotic behavior of the real structured stability radius for (2.11) is of interest as well, but it is still an open problem.

Remark 2.8 In [69], the robust stability of a DAE subject to perturbations of the form $E\dot{x} = (A + \varepsilon H)x$ is considered, where E, A are given as in (2.1), H is a given matrix, and ε is an uncertain parameter. Assuming that the unperturbed system is regular, of index at most one and asymptotically stable, a computable formula for the maximal stability interval $(\varepsilon_1, \varepsilon_2)$ is derived, i.e., the perturbed system retains the index and is asymptotically stable for all $\varepsilon \in (\varepsilon_1, \varepsilon_2)$.

In [37], complex structured stability radii for the discrete-time analog of DAEs, i.e., singular difference equations are analyzed. In particular, again due to the continuity, the complex structured stability radius of the discretized system (using the implicit Euler method) of (2.1) is shown to converge to the corresponding one of the continuous-time system as the stepsize tends to zero. The analogous question concerning the real structured stability radii is **still open**.

3 Robust Stability of Linear Time-Varying DAEs

In this section, we investigate the exponential and asymptotic stability and its robustness for linear time-varying DAEs of the form

$$E(t)\dot{x}(t) = A(t)x(t), \quad t \in \mathbb{I}, \tag{3.1}$$

with matrix functions $E, A \in C(\mathbb{I}, \mathbb{K}^{n\times n})$, $\mathbb{K} \in \{\mathbb{C}, \mathbb{R}\}$.

Analyzing the different stability concepts for (3.1) is, however, much more complicated than for linear time-invariant systems. Even if for all $t \in \mathbb{I}$, the finite eigenvalues of $(E(t), A(t))$ have negative real part, system (3.1) may be unstable, as many well-known examples demonstrate for the ODE case, see e.g., [55].

Example 3.1 For all $t \in \mathbb{R}$

$$A(t) = \begin{bmatrix} \cos^2(3t) - 5\sin^2(3t) & -6\cos(3t)\sin(3t) + 3 \\ -6\cos(3t)\sin(3t) + 3 & \sin^2(3t) - 5\cos^2(3t) \end{bmatrix}$$

has a double eigenvalue at -2 but the solution of $\dot{x} = Ax$, with $x(0) = \left[\begin{smallmatrix} c_1 \\ 0 \end{smallmatrix}\right]$ is given by $x(t) = \left[\begin{smallmatrix} c_1 e^t \cos(3t) \\ -c_1 e^t \sin(3t) \end{smallmatrix}\right]$, which is obviously unstable.

We use the following standard notation as in [36, 60]. Let X, Y be finite dimensional vector spaces. For every $p, 1 \le p < \infty$ and $s, t, t_0 \le s < t < \infty$, we denote by $L_p(s,t;X)$ the space of measurable functions f with values in X and norm $\|f\|_p := (\int_s^t \|f(\rho)\|^p \, d\rho)^{1/p} < \infty$ and by $L_\infty(s,t;X)$ the space of measurable and essentially bounded functions f with $\|f\|_\infty := \operatorname{ess\,sup}_{\rho\in[s,t]} \|f(\rho)\|$. We also consider the spaces $L_p^{\mathrm{loc}}(t_0,\infty;X)$ and $L_\infty^{\mathrm{loc}}(t_0,\infty;X)$, which contain all functions $f \in L_p(s,t;X)$ and $f \in L_\infty(s,t;X)$ for some $s, t, t_0 \le s < t < \infty$, respectively. We, furthermore, use the notation $\mathscr{L}(L_p(t_0,\infty;X), L_p(t_0,\infty;Y))$ to denote the Banach space of linear bounded operators $\mathbb{P}$ from $L_p(t_0,\infty;X)$ to $L_p(t_0,\infty;Y)$ supplied with the norm

$$\|\mathbb{P}\| := \sup_{x\in L_p(t_0,\infty;X),\,\|x\|=1} \|\mathbb{P}x\|_{L_p(t_0,\infty;Y)}.$$

In the following we assume that (3.1) is of index 1, in the sense of the tractability index [43] or that it is strangeness-free in the sense of the strangeness-index [63]. Let us briefly introduce these index concepts. First, to introduce the tractability index and the projector chain approach, see e.g., [43, 68], we assume that $E(t)$ is singular for all t and $S := \ker E$ is absolutely continuous. Then, there exists an absolutely continuous projector Q onto S, i.e., $Q \in C(0,\infty;\mathbb{K}^{n\times n})$, Q is differentiable almost everywhere, $Q^2 = Q$, and $\operatorname{Im} Q = S$ for all $t \in \mathbb{I}$. If we assume in addition that $\dot{Q} \in L_\infty^{\mathrm{loc}}(0,\infty;\mathbb{K}^{n\times n})$, then $P = I - Q$ is a projector along S and system (3.1) can be rewritten in the form

$$E\frac{d}{dt}(Px) = \widehat{A}x, \tag{3.2}$$

where $\widehat{A} := A + E\dot{P} \in L_\infty^{\mathrm{loc}}(0,\infty;\mathbb{K}^{n\times n})$. Setting

$$G := E - \widehat{A}Q. \tag{3.3}$$

Definition 3.1 The linear DAE (3.1) is said to be *tractable of index* 1, if $G(t)$ defined by (3.3) is invertible for almost every $t \in [0,\infty)$ and $G^{-1} \in L_\infty^{\mathrm{loc}}(0,\infty;\mathbb{K}^{n\times n})$.

Multiplying both sides of (3.2) by PG^{-1}, QG^{-1}, and taking into account the identities $G^{-1}E = P$, $G^{-1}\widehat{A} = -Q + G^{-1}\widehat{A}P$, we obtain

$$\frac{d}{dt}(Px) = \left(\frac{d}{dt}P + PG^{-1}\widehat{A}\right)Px,$$
$$Qx = QG^{-1}\widehat{A}Px,$$

which decomposes the DAE into a differential part and an algebraic part. With $z = Px$, the dynamics of the system is given by the *inherent ODE*

$$\dot{z} = \left(\dot{P} + PG^{-1}\widehat{A}\right)z \tag{3.4}$$

of (3.1). Let $\Phi_0(t,s)$ denote the *fundamental solution matrix* associated with the inherent ODE (3.4), i.e., the matrix function satisfying

$$\frac{d}{dt}\Phi_0(t,s) = \big(\dot{P} + PG^{-1}\widehat{A}\big)\Phi_0(t,s), \qquad \Phi_0(s,s) = I,$$

for $t > s \geq 0$, then the fundamental solution operator generated by (3.1) is defined by

$$E\frac{d}{dt}\Phi(t,s) = A\Phi(t,s), \qquad P(s)\big(\Phi(s,s) - I\big) = 0$$

and Φ can be expressed as $\Phi(t,s) = (I + QG^{-1}\widehat{A}(t))\Phi_0(t,s)P(s)$. It can be shown by direct calculations that, despite its construction, the fundamental solution operator is in fact independent of the choice of projector Q.

The concept of the strangeness index uses the DAE and its derivatives to construct a so-called strangeness-free DAE with the same solution [63]. In the homogeneous case this *strangeness-free* system has the form (3.1), where

$$E = \begin{bmatrix} E_1 \\ 0 \end{bmatrix}, \qquad A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \tag{3.5}$$

with $E_1 \in C(\mathbb{I}, \mathbb{K}^{d\times n})$ and $A_2 \in C(\mathbb{I}, \mathbb{K}^{(n-d)\times n})$ such that the matrix $\hat{E}(t) := \left[\begin{smallmatrix} E_1(t) \\ A_2(t) \end{smallmatrix}\right]$ is invertible for all $t \in \mathbb{I}$.

By using a global kinematic equivalence transformation, see [70, Remark 13], (3.1) can be transformed to the special form,

$$\begin{bmatrix} \tilde{E}_{11} & \tilde{E}_{12} \\ 0 & 0 \end{bmatrix}\dot{\tilde{x}} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}\tilde{x}, \tag{3.6}$$

so that the *essential underlying ODE* is readily given by $\tilde{E}_{11}\dot{x}_1 = \tilde{A}_{11}x_1$ with nonsingular $\tilde{E}_{11}$.

A matrix function $\Phi \in C^1(\mathbb{I}, \mathbb{R}^{n\times d})$ is called *minimal fundamental solution matrix of the strangeness-free DAE* (3.1) if each of its columns is a solution to (3.1) and $\operatorname{rank}\Phi(t) = d$ for all $t \in \mathbb{I}$.

In the following we assume that the DAE is of index at most one or alternatively strangeness-free. These conditions are equivalent if the coefficients are sufficiently smooth [77].

The characterization of the different stability concepts for linear variable coefficient ODEs is well established via the concepts of *Bohl and Lyapunov exponents* [1, 9, 28] and Sacker–Sell spectra [29, 86]. These concepts were extended from ODEs to DAEs in [24, 70, 71].

To analyze exponential stability, we introduce first the Bohl exponent.

Definition 3.2 The *Bohl exponent* for an index 1 system of the form (3.1) with fundamental solution Φ is given by

$$k_B(E, A) = \inf\big\{-\alpha \in \mathbb{R};\ \text{there exists } L_\alpha > 0 \text{ such that for all } t \geq t_0 \geq 0 : \big\|\Phi(t, t_0)\big\| \leq L_\alpha e^{-\alpha(t-t_0)}\big\}.$$

It follows that (3.1) is exponentially stable if and only if its Bohl exponent is negative (including the case $k_B(E, A) = -\infty$).

Analogous to the ODE case (see [28]), using the fundamental solution operator Φ, it follows that the Bohl exponent of (3.1) is bounded from above, i.e., $k_B(E, A) < \infty$ if and only if $\sup_{0\leq|t-s|\leq 1} \|\Phi(t, s)\| < \infty$. Furthermore, the Bohl exponent of (3.1) can be determined by

$$k_B(E, A) = \limsup_{s,t-s\to\infty} \frac{\ln \|\Phi(t, s)\|}{t - s}.$$

In [24], various properties of the Bohl exponent, as well as the connection between the exponential stability of (3.1) and the boundedness of solutions to nonhomogeneous DAE with bounded inhomogeneity are investigated.

Remark 3.1 In the ODE case, the boundedness of the coefficient function A ensures the finiteness of Bohl exponent. This is not true for DAEs (3.1) even with both bounded coefficient functions E and A, where the Bohl exponent may be $+\infty$ or $-\infty$. We note also that, by assuming that (3.1) is of index 1, we exclude degenerate cases such as non-uniqueness or finite escape time of solutions, which may happen with nonregular DAEs and are discussed in [6, 7].

For ODEs, the asymptotic stability of solutions can be characterized by the Lyapunov exponents, see [74]. The extension of the theory of Lyapunov exponents to linear time-varying DAEs has been given in [25, 26, 70–72], using either the projector-based tractability index or the derivative array-based strangeness index approach.

Definition 3.3 For a given minimal fundamental solution matrix Φ of a strangeness-free DAE system of the form (3.1), and for $1 \leq i \leq d$, we introduce

$$\lambda_i^u = \limsup_{t\to\infty} \frac{1}{t} \ln\big\|\Phi(t)e_i\big\| \quad \text{and} \quad \lambda_i^\ell = \liminf_{t\to\infty} \frac{1}{t} \ln\big\|\Phi(t)e_i\big\|,$$

where e_i denotes the ith unit vector and $\|\cdot\|$ denotes the Euclidean norm. Let all the Lyapunov exponents be finite, then the columns of a minimal fundamental solution matrix form a *normal basis* if $\sum_{i=1}^d \lambda_i^u$ is minimal. The λ_i^u, $i = 1, 2, \ldots, d$ belonging to a normal basis are called (*upper*) *Lyapunov exponents* and the intervals $[\lambda_i^\ell, \lambda_i^u]$, $i = 1, 2, \ldots, d$, are called *Lyapunov spectral intervals*.

The strangeness-free DAE system (3.1) then is asymptotically stable if and only if the largest upper Lyapunov exponent is negative. Note that for linear time-invariant DAEs (2.1), the Lyapunov exponents are exactly the real parts of the finite eigenvalues of pencil (E, A).

This brings us to another major difference between linear time-invariant and linear time-varying DAEs. The following example shows that an infinitesimally small time-varying perturbation applied to both coefficient matrices may change the asymptotic stability, even if the perturbation does not change the index.

Example 3.2 ([71]) Consider the system $\dot{x}_1 = -x_1$, $0 = x_2$, which is strangeness-free and asymptotically stable. For the perturbed DAE

$$\big(1 + \varepsilon^2 \sin(2nt)\big)\dot{x}_1 - \varepsilon \cos(nt)\dot{x}_2 = -x_1, \qquad 0 = -2\varepsilon \sin(nt)x_1 + x_2, \tag{3.7}$$

where ε is a small parameter and n is a given number, from the second equation of (3.7), we obtain $x_2 = 2\varepsilon \sin nt\, x_1$. Differentiating this expression for x_2 and inserting the result into the first equation, after some elementary calculations, we obtain $\dot{x}_1 = (-1 + n\varepsilon^2 + n\varepsilon^2 \cos(2nt))x_1$. Explicit integration yields $x_1(t) = e^{(-1+n\varepsilon^2)t+\varepsilon^2 \sin(2nt)/2}x_1(0)$. Clearly, even if ε is arbitrarily small (hence the perturbation in the coefficient matrices is arbitrarily small in the sup-norm), (3.7) may become unstable if n is sufficiently large.

3.1 Stability Radii for Linear Time-Varying DAEs

In this section we discuss the stability radii for linear DAEs with variable coefficients. Formulas for the stability radii of exponential stability were derived in [24, 36] extending the results for ODEs in [56, 60].

We assume that the DAE is of index at most one and discuss perturbed systems

$$E(t)\dot{x}(t) = \big(A(t) + H(t)\big)x(t), \quad t \in \mathbb{I}, \tag{3.8}$$

with a perturbation function $H \in L_\infty(0, \infty; \mathbb{K}^{n\times n})$ as well as structured perturbations of the form

$$E(t)\dot{x}(t) = A(t)x(t) + B(t)\Delta\big(C(\cdot)x(\cdot)\big)(t), \quad t \in \mathbb{I}, \tag{3.9}$$

where $B \in L_\infty(0, \infty; \mathbb{K}^{n\times m})$ and $C \in L_\infty(0, \infty; \mathbb{K}^{q\times n})$ are given matrix functions, restricting the structure of the perturbation and $\Delta : L_p(0, \infty; \mathbb{K}^q) \to L_p(0, \infty; \mathbb{K}^m)$ is an unknown perturbation operator.

To obtain formulas for the exponential stability radius, we assume that (3.1) is exponentially stable, i.e., there exist constants $L > 0$, $\gamma > 0$ such that $\|\Phi_0(t, s)P(s)\| \le Le^{-\gamma(t-s)}$, for $t \ge s \ge 0$, and that it is robustly index 1 in the following sense.

Definition 3.4 Consider the DAE (3.1) in the form (3.2) and let G be as in (3.3). Then the DAE is said to be *robustly index* 1 if, supplied with a bounded projection Q, the matrix functions G^{-1} and $Q_s := -QG^{-1}\widehat{A}$ are essentially bounded on $\mathbb{I}$.

To extend the tractability index concept to the perturbed system (3.9), we assume that the perturbation operator $\Delta \in \mathscr{L}(L_p(0,\infty;\mathbb{K}^q), L_p(0,\infty;\mathbb{K}^m))$ is causal which is defined as follows. For $T \in \mathbb{I}$, the *truncation operator* π_T at T on $L_p(0,\infty;X)$ is defined via

$$\pi_T u(t) := \begin{cases} u(t), & t \in [0,T], \\ 0, & t > T. \end{cases}$$

An operator $\mathbb{P} \in \mathscr{L}(L_p(0,\infty;X), L_p(0,\infty;Y))$ then is said to be *causal*, if $\pi_T \mathbb{P} \pi_T = \pi_T \mathbb{P}$ for every $T \in \mathbb{I}$. Let the linear operator $\widetilde{G} \in \mathscr{L}(L_p^{\mathrm{loc}}(0,\infty;\mathbb{K}^n), L_p^{\mathrm{loc}}(0,\infty;\mathbb{K}^n))$ be defined via

$$(\widetilde{G}z) = (E - \widehat{A}Q)z - B\Delta(CQz).$$

If for every $T > 0$, the operator $\widetilde{G}$ restricted to $L_p(0,T;\mathbb{K}^n)$ is invertible and the inverse operator $\widetilde{G}^{-1}$ is bounded, then we say the perturbed DAE (3.9) is of *index* 1 (*in a generalized sense*). The structured perturbation in (3.9) is called dynamic perturbation, which is different from static perturbations considered in Sect. 2. Then we can employ the concept of *mild solution*.

Definition 3.5 We say that the initial value problem for the perturbed system (3.9) with initial condition

$$P(t_0)\big(x(t_0) - x_0\big) = 0, \tag{3.10}$$

admits a *mild solution* if there exists $x \in L_p^{\mathrm{loc}}(t_0,\infty;\mathbb{K}^n)$ satisfying

$$\begin{aligned} x(t) &= \Phi(t,t_0)P(t_0)x_0 \\ &\quad + \int_{t_0}^{t} \Phi(t,\rho)PG^{-1}B\Delta\big(\big[Cx(\cdot)\big]_{t_0}\big)(\rho)\,d\rho + QG^{-1}B\Delta\big(\big[Cx(\cdot)\big]_{t_0}\big)(t) \end{aligned}$$

for $t \geq t_0$, where

$$\big[Cx(\cdot)\big]_{t_0} = \begin{cases} 0, & t \in [0,t_0), \\ C(t)x(t), & t \in [t_0,\infty). \end{cases}$$

The mild solution plays an important role in the characterization of robust stability, since solutions of the perturbed DAE (3.9) in the classical sense do not exist, in general. The existence and uniqueness of mild solutions is given by the following result.

Theorem 3.1 ([36]) *Consider the initial value problem* (3.9)–(3.10). *If* (3.9) *is of index at most one, then it admits a unique mild solution* $x \in L_p^{\mathrm{loc}}(t_0,\infty;\mathbb{K}^n)$ *with absolutely continuous* $z = Px$ *for all* $t_0 \in \mathbb{I}$, $x_0 \in \mathbb{K}^n$. *Furthermore, for an arbitrary* $T > 0$, *there exists a constant* L_1 *such that pointwise*

$$\big\|P(t)x(t)\big\| \leq L_1 \big\|P(t_0)x_0\big\| \quad \textit{for all } t \in [t_0,T].$$

With (3.9), we associate the *perturbation operator* (in [35, 60] it is called the input–output operator)

$$\mathbb{L}_{t_0} z = C \int_{t_0}^{t} \Phi(t,\rho) P G^{-1} B(\rho) z(\rho)\, d\rho + C Q G^{-1} B z, \tag{3.11}$$

which is defined for all $t \geq t_0 \geq 0$, $z \in L_p(0,\infty;\mathbb{K}^m)$. It is important to note that, by direct algebraic verifications, the mild solution as well as the perturbation operator are independent of the choice of projector Q. It is shown that, analogously to the ODE case [56, 60], there exists a close relationship between the perturbation operator associated with (3.9) and the robust stability of (3.1) with respect to structured and dynamic perturbations given by (3.9). For exponentially stable robustly index 1 systems this operator is linear, bounded, and monotonically non-increasing with respect to t, i.e.,

$$\|\mathbb{L}_{t_0}\| \geq \|\mathbb{L}_{t_1}\| \quad \text{for all } t_1 \geq t_0 \geq 0,$$

and for all $t \in \mathbb{I}$ the bound

$$\|\mathbb{L}_t\| \leq \frac{L}{\gamma} \left\| P G^{-1} \right\|_\infty \|B\|_\infty \|C\|_\infty + \left\| C Q G^{-1} B \right\|_\infty$$

holds.

Definition 3.6 ([36]) Consider a DAE of the form (3.9) of index 1 and denote by $x(t;t_0,x_0)$ the solution satisfying initial condition (3.10). The trivial solution of (3.9) is said to be *globally L_p-stable*, if there exist positive constant L_2 and L_3 such that

$$\begin{aligned} &\left\| P(t) x(t;t_0,x_0) \right\|_{\mathbb{K}^n} \leq L_2 \left\| P(t_0) x_0 \right\|_{\mathbb{K}^n}, \\ &\left\| x(\cdot;t_0,x_0) \right\|_{L_p(t_0,\infty;\mathbb{K}^n)} \leq L_3 \left\| P(t_0) x_0 \right\|_{\mathbb{K}^n}, \end{aligned} \tag{3.12}$$

for all $t \geq t_0$, $x_0 \in \mathbb{K}^n$.

Note that this kind of stability notion is equivalent to the concept of output stability, see [59] for various stability concepts in the ODE case.

We then have the following definition of stability radii for time-varying DAEs.

Definition 3.7 If system (3.1) is exponentially stable and robustly index 1, then the *complex/real structured stability radii* of (3.1) subject to dynamic structured perturbation as in (3.9), are defined by

$$\begin{aligned} &r_{\mathbb{K}}(E,A;B,C) \\ &\quad = \inf \left\{ \|\Delta\|, \text{ the trivial solution of (3.9) is not globally } L_p\text{-stable} \right. \\ &\qquad\qquad \left. \text{or (3.9) is not of index 1} \right\}, \end{aligned}$$

where $\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$, respectively.

In [36] the following formulas for the stability radii were derived.

Theorem 3.2 ([36]) *If system* (3.1) *is exponentially stable and robustly index* 1, *then*

$$r_{\mathbb{K}}(E, A; B, C) = \min\left\{ \lim_{t_0 \to \infty} \|\mathbb{L}_{t_0}\|^{-1}, \left(\operatorname{ess\,sup}_{t \in \mathbb{I}} \left\| C Q G^{-1} B(t) \right\| \right)^{-1} \right\}.$$

This implies the following corollary.

Corollary 3.3 ([36]) *If system* (3.1) *is exponentially stable and robustly index* 1 *and the data* $(E, A; B, C)$ *are real, then*

$$r_{\mathbb{C}}(E, A; B, C) = r_{\mathbb{R}}(E, A; B, C).$$

As special case we obtain the formula (2.4) for the complex stability radius of time-invariant systems (with respect to dynamic perturbations).

Corollary 3.4 ([36]) *Let* E, A, B, C *be time-invariant, let the system* (3.1) *be index* 1 *and exponentially stable. If* $p = 2$, *i.e., for the space* L_2 *of square integrable functions, then*

$$r_{\mathbb{C}}(E, A; B, C) = \|\mathbb{L}_0\|^{-1} = \left(\sup_{\omega \in i\mathbb{R}} \left\| C(\omega E - A)^{-1} B \right\| \right)^{-1}.$$

Comparing with a special case of Theorem 2.2 when only A is subject to perturbations, Corollary 3.4 states that the complex stability radius with respect to dynamic perturbations and that to static perturbations are equal. This statement in fact generalizes a previous result for the ODE case in [53].

Example 3.3 ([36, Sect. 5.1]) Consider a DAE in semi-explicit form

$$E = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}, \qquad A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{3.13}$$

with appropriate partitioning. The index 1 assumption means that $A_{22}(t)$ is invertible almost everywhere in $\mathbb{I}$. One gets $Q = \operatorname{diag}(0, I_{n-r})$,

$$G = \begin{bmatrix} I_r & -A_{12} \\ 0 & -A_{22} \end{bmatrix}, \qquad \Phi(t, s) = \begin{bmatrix} \widehat{\Phi}(t, s) \\ -A_{22}^{-1} A_{21} \widehat{\Phi}(t, s) \end{bmatrix},$$

where $\widehat{\Phi}(t, s)$ is the fundamental solution operator of the underlying ordinary differential equation $\dot{y} = (A_{11} - A_{12} A_{22}^{-1} A_{21}) y$, which is assumed to be exponentially stable. The assumption that the system is robustly index 1 means the essential boundedness of A_{22}^{-1}, $A_{22}^{-1} A_{21}$, and $A_{12} A_{22}^{-1}$. Partitioning the restriction matrices B, C as

$B = [B_1^T \ B_2^T]^T$, $C = [C_1 \ C_2]$, analogously, we obtain

$$(\mathbb{L}_{t_0} u)(t) = \big(C_1 - C_2 A_{22}^{-1} A_{21}\big) \int_{t_0}^{t} \widehat{\Phi}(t,\rho)\big(B_1(\rho) - A_{12} A_{22}^{-1} B_2(\rho)\big)u(\rho)\, d\rho$$
$$- C_2 A_{22}^{-1} B_2 u(t),$$

and by Theorem 3.2 we have

$$r_{\mathbb{K}}(E, A; B, C) = \min\Big\{ \lim_{t_0 \to \infty} \|\mathbb{L}_{t_0}\|^{-1}, \Big(\operatorname*{ess\,sup}_{t\in\mathbb{I}} \big\| C_2 A_{22}^{-1} B_2(t) \big\|\Big)^{-1} \Big\}.$$

In summary, we have seen in this section that in the index 1 case the exponential stability radii in the linear time-varying case are similar to the ones in the linear time-invariant case.

A similar analysis can be performed in principle for the asymptotic stability, by studying the Lyapunov exponents under perturbations, see the next section. Unfortunately, however, in contrast to the Bohl exponents, the Lyapunov exponents themselves may be very sensitive to small changes in the data.

3.2 Dependence of Stability Radii on the Data

The robustness analysis of the Bohl exponent was extended from ODEs in [56] to DAEs in [24]. In the context of deriving numerical methods for the numerical computation of Bohl and Lyapunov exponents for linear time-varying DAEs, see [70, 71, 73] the robustness of these exponents under perturbations was studied and the concept of admissible perturbations was extended to the variable coefficient case. In this section we summarize these results and study how the stability radii depend on the data.

To see that the Bohl exponent is robust under sufficiently small index-1-preserving perturbations, we consider the perturbed equation

$$E\frac{d}{dt}(Px) = (\widehat{A} + H)x, \tag{3.14}$$

where $\widehat{A}$ is the same as in (3.2). Here H is assumed to be piecewise continuous (just for sake of simplicity) and to satisfy

$$\sup_{t\in\mathbb{I}} \big\| H(t) \big\| < \Big(\sup_{t\in\mathbb{I}} \big\| QG^{-1}(t) \big\|\Big)^{-1}. \tag{3.15}$$

Note that this condition implies the inequality $\sup_{t\in\mathbb{I}} \|QG^{-1}H(t)\| < 1$, which is essential in the analysis.

Theorem 3.5 ([24]) *Suppose that system* (3.1) *is robustly index* 1, *the perturbation* H *satisfies* (3.15), *and suppose further that for any* $\varepsilon > 0$ *there exists* $\delta = \delta(\varepsilon) > 0$ *such that for all* H *satisfying*

$$\limsup_{s,t-s\to\infty} \frac{1}{t-s} \int_s^t \left\| PG^{-1}H(\tau) \right\| d\tau < \delta,$$

then

$$k_B(E, A+H) < k_B(E, A) + \varepsilon.$$

As a consequence of Theorem 3.5, the exponential stability is preserved under all sufficiently small perturbations H.

Remark 3.2 The robustness analysis of the Bohl exponent is extendable to the case of general perturbations arising in both coefficients of (3.1). It is clear that additional assumptions on the perturbation structure and/or the smoothness of the admissible perturbations are necessary in this case. The same can be said for the analysis of Bohl exponents for general higher-index DAEs and for nonlinear perturbations, see [6].

We now discuss how the structured stability radius of (3.8) with respect to the same structured perturbation as in (3.9) depends on the perturbation H and the restriction matrices B, C. To this end, we first establish the asymptotic behavior of the norm of the input-output operator defined in (3.11).

Theorem 3.6 [24] *Suppose that system* (3.1) *is exponentially stable*, *robustly index* 1 *and satisfies* (3.15) *and suppose*, *in addition*, *that the perturbation function* H *satisfies*

$$\lim_{t\to\infty} \left\| H(t) \right\| = 0.$$

Then the operator $\mathbb{L}_t$ *defined in* (3.11) *and the corresponding operator associated with* (3.14), *denoted by* $\widetilde{\mathbb{L}}_t$, *satisfy*

$$\lim_{t\to\infty} \|\mathbb{L}_t\| = \lim_{t\to\infty} \|\widetilde{\mathbb{L}}_t\|.$$

By invoking Theorem 3.2, we obtain sufficient conditions for a sequence of perturbations H_k under which the structured stability radius of the perturbed systems converges to that of the unperturbed system.

Theorem 3.7 ([24]) *Suppose that system* (3.1) *is exponentially stable*, *robustly index* 1 *and satisfies* (3.15). *Let* $\{H_k(\cdot)\}_{k=1}^{\infty}$ *be a sequence of measurable matrix functions and suppose that* $\sup_{t\in\mathbb{I}} \|H_k(t)\| < (\sup_{t\in\mathbb{I}} \|QG^{-1}(t)\|)^{-1}$ *for all* $k = 1, 2, \ldots$ *and* $\lim_{k\to\infty} \sup_{t\in\mathbb{I}} \|H_k(t)\| = 0$. *Then*

$$\lim_{k\to\infty} r_{\mathbb{C}}(E, A+H_k; B, C) = r_{\mathbb{C}}(E, A; B, C).$$

Theorem 3.7 implies that the stability radius for the system (3.1) depends continuously on the coefficient matrix function A. As a consequence of Theorem 3.7 and Corollary 3.4, we get the following result.

Corollary 3.8 *Let E, A, B, C be constant matrices, let system (3.1) be of index 1 and exponentially stable. Furthermore, assume that the sequence of time-varying perturbation $\{H_k\}_{k=1}^{\infty}$ fulfills the conditions of Theorem 3.7. Then, for the Euclidean norm, one has*

$$\lim_{k\to\infty} r_{\mathbb{C}}(E, A + H_k; B, C) = \Big(\sup_{w\in i\mathbb{R}} \big\|C(wE - A)^{-1}B\big\|\Big)^{-1}.$$

For illustration, the following numerical example shows that the complex stability radius of a linear time-invariant system under time-varying perturbations can be well approximated by that of the corresponding time-invariant DAE.

Example 3.4 ([24, Example 5.13]) Consider the simple example of a linear constant coefficient DAE with data

$$E = \begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix}, \qquad A = \begin{bmatrix} -2 & -1 \\ 2 & 2 \end{bmatrix}, \qquad B = I, \qquad C = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}. \tag{3.16}$$

Let a sequence of time-varying perturbations be defined by

$$F_k(t) = \begin{bmatrix} -\frac{1}{3+t^2} & \frac{1}{4\sqrt{1+t}} \\ \frac{e^{-2t}}{k+1} & -\frac{e^{-t}}{2k} \end{bmatrix}, \quad k = 1, 2, \ldots. \tag{3.17}$$

Here, we choose

$$Q = \begin{bmatrix} -1 & -1 \\ 2 & 2 \end{bmatrix}, \qquad G = E - AQ = \begin{bmatrix} 2 & 1 \\ -2 & -2 \end{bmatrix}.$$

Then it is easy to check that $\lim_{t\to\infty} \|F_k(t)\| = 0$ and

$$\sup_{t\geq 0} \big\|F_k(t)\big\| < \left\| \begin{bmatrix} 1/3 & 1/4 \\ 1/2 & 1/2 \end{bmatrix} \right\| \approx 0.8192 < \big\|QG^{-1}\big\|^{-1} \approx 0.8945.$$

Furthermore, we have

$$QG^{-1}F_k(t) = \begin{bmatrix} \frac{e^{-2t}}{2(k+1)} & -\frac{e^{-t}}{4k} \\ -\frac{e^{-2t}}{k+1} & \frac{e^{-t}}{2k} \end{bmatrix},$$

thus $\lim_{k\to\infty} \sup_{t\geq 0} \|QG^{-1}F_k(t)\| = 0$. On the other hand, by elementary calculations, we obtain

$$\Big(\sup_{w\in i\mathbb{R}} \big\|C(wE - A)^{-1}B\big\|\Big)^{-1} = 1.$$

Invoking Corollary 3.8, we have

$$\lim_{k\to\infty} r_{\mathbb{C}}(E, A + F_k; B, C) = 1.$$

Finally, one also obtains that the complex structured stability radius of (3.1) depends continuously on the restriction matrices B and C as in the time-invariant case.

Theorem 3.9 ([24]) *Suppose that system* (3.1) *is exponentially stable, robustly index* 1 *and satisfies* (3.15). *Let* B_k *and* C_k *be two sequences of measurable and essentially bounded matrix functions satisfying*

$$\lim_{k\to\infty} \operatorname{ess\,sup}_{t\in\mathbb{I}} \|B_k(t) - B(t)\| = 0, \quad \lim_{k\to\infty} \operatorname{ess\,sup}_{t\in\mathbb{I}} \|C_k(t) - C(t)\| = 0$$

then,

$$\lim_{k\to\infty} r_{\mathbb{C}}(E, A; B_k, C_k) = r_{\mathbb{C}}(E, A; B, C).$$

We stress once more that, since the dynamics of DAEs is constrained and the index-1 property should be preserved, only weaker results hold for the continuity of the stability radius and more restrictive assumptions are required than in the ODE case. Furthermore, for time-varying DAEs, we have to restrict the analysis to perturbations in A only, see (3.8) and (3.9), simply because the study of perturbations associated with the leading term E is a long-standing *open problem*. Just very recently, in [5], a result on robust stability with respect to perturbations associated with the leading term E has been derived within the framework of DAEs of tractability-index 1.

In order to study the robustness of Lyapunov exponents, we consider the specially perturbed system for (3.1) given in the form (3.5)

$$[E + F]\dot{x} = [A + H]x, \quad t \in \mathbb{I}, \tag{3.18}$$

where $F = \left[\begin{smallmatrix} F_1 \\ 0 \end{smallmatrix}\right]$ and $H = \left[\begin{smallmatrix} H_1 \\ H_2 \end{smallmatrix}\right]$, and where F_1 and H_1, H_2 are assumed to have the same order of smoothness as E_1 and A_1, A_2, respectively. Perturbations of this structure are called *admissible*, generalizing the concept for the constant coefficient DAEs studied in [20].

The DAE (3.1) is said to be *robustly strangeness-free* if it stays strangeness-free under all sufficiently small admissible perturbations and it is easy to see that this property holds if and only if the matrix function $\hat{E}$ is boundedly invertible.

If we assume that (3.1) is already given in the form (3.6), then the perturbed DAE has the form

$$\begin{aligned}(E_{11} + F_{11})\frac{d}{dt}\tilde{x}_1 + (E_{12} + F_{12})\frac{d}{dt}\tilde{x}_2 &= (A_{11} + H_{11})\tilde{x}_1 + (A_{12} + H_{12})\tilde{x}_2,\\ 0 &= H_{21}\tilde{x}_1 + (A_{22} + H_{22})\tilde{x}_2.\end{aligned}$$

In the following we restrict ourselves to robustly strangeness-free DAE systems under admissible perturbations.

Definition 3.8 The upper Lyapunov exponents $\lambda_1^u \geq \cdots \geq \lambda_d^u$ of (3.1) are said to be *stable* if for any $\varepsilon > 0$, there exists $\delta > 0$ such that the conditions $\sup_t \|F(t)\| < \delta$, $\sup_t \|H(t)\| < \delta$, and $\sup_t \|\dot{H}_2(t)\| < \delta$ on the perturbations imply that the perturbed DAE system (3.18) is strangeness-free and

$$\left|\lambda_i^u - \gamma_i^u\right| < \varepsilon, \quad \text{for all } i = 1, 2, \ldots, d,$$

where the γ_i^u are the ordered upper Lyapunov exponents of (3.18).

The boundedness condition on $\dot{H}_2$, which is obviously satisfied in the time-invariant setting [20], is an extra condition and it seems to be somehow unusual. However, the DAE (3.7) shows the necessity.

As in the ODE case, see [1, 29], to have stability of the Lyapunov spectrum, one needs the property of *integral separation*, i.e., for the columns of the minimal fundamental solution matrix Φ of (3.1) there exist constants $c_1 > 0$ and $c_2 > 0$ such that

$$\frac{\|X(t)e_i\|}{\|X(s)e_i\|} \cdot \frac{\|X(s)e_{i+1}\|}{\|X(t)e_{i+1}\|} \geq c_2 e^{c_1(t-s)},$$

for all t, s with $t \geq s \geq 0$ and $i = 1, \ldots, d-1$. Then we have the following sufficient conditions for the stability of the upper Lyapunov exponents of (3.1).

Theorem 3.10 ([70]) *Consider the DAE* (3.1) *in the form* (3.6). *Suppose that the matrix* $\hat{E}$ *is boundedly invertible and that* $E_{11}^{-1}A_{11}$, $A_{12}A_{22}^{-1}$ *and the derivative of* A_{22} *are bounded on* $\mathbb{I}$. *Then, the upper Lyapunov exponents of* (3.1) *are distinct and stable if and only if the system has the integral separation property.*

This shows that if perturbations are performed in E as well, then the perturbation analysis of time-varying DAEs requires more restrictive conditions than in the time-invariant case. However, for some classes of structured problems and/or structured perturbation, parts of these conditions can be relaxed.

If the perturbation block H_{21} disappears, i.e., if H and A have the same block triangular structure, then for example the restrictive conditions on the derivatives in Definition 3.8 and Theorem 3.10 can be omitted. Similar situations happen in the case that $E_{12} = F_{12} = 0$ as discussed in [70, Sect. 3.2] and the case of perturbations in A only as in (3.8).

In [70] another stability concept, the *Sacker–Sell spectrum* has been extended to linear DAEs with variable coefficients. It is also shown that unlike the Lyapunov spectrum, the Sacker–Sell spectrum is robust in the sense that it is stable without requiring integral separation.

This means that for general strangeness-free time-varying systems, the exponential stability is robust, but the asymptotic stability is not. Note that this remark does not apply to time-invariant systems, for which the two stability notions are equivalent.

Remark 3.3 In [56], the robust stability of time-varying ODEs with respect to time-varying perturbations was characterized in terms of differential Riccati equations. In particular, the relation between the perturbation operator and the existence of solutions to the Riccati equations was pointed out. A similar analysis for time-varying DAEs would be of interest as well. However, this gives rise to differential-algebraic Riccati equations, which may have a very complicated solution structure, see e.g. [62, 65].

Remark 3.4 The robustness analysis for linear DAEs of index higher than one and under general perturbations is essentially an *open problem*. The same is true for the distances to the other important control properties such as controllability and observability. The robustness of these concepts for linear DAEs with variable coefficients presents a major challenge.

4 Discussion

In this paper we have surveyed recent results on the robustness of asymptotic and exponential stability for linear time-invariant and time-varying DAEs. We have analyzed robust stability and its distance measures, the real or complex structured stability radius and presented formulas and various properties of the stability radii. We have seen that the robustness analysis for DAEs is much more complicated than that for ODEs. In general, results already known for ODEs now hold for DAEs only under extra assumptions, mainly restricting the set of admissible perturbations. DAE aspects also give rise to new robustness and distance problems. While for time-invariant DAEs, most of the robustness and distance problems are well understood, many problems for time-varying DAEs are still open. These and robustness analysis for general nonlinear and/or high-index DAEs and time-delay DAEs are interesting and challenging topics for future work.

Acknowledgements V.H. Linh supported by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 101.01-2011.14. V. Mehrmann supported by *Deutsche Forschungsgemeinschaft*, through project A02 within Collaborative Research Center 910 *Control of self-organizing nonlinear systems*.

We thank the anonymous referees for their useful suggestions that led to improvements of the paper.

References

1. Adrianova, L.Ya.: Introduction to Linear Systems of Differential Equations. Trans. Math. Monographs, vol. 146. AMS, Providence (1995)
2. Benner, P., Voigt, M.: Numerical computation of structured complex stability radii of large-scale matrices and pencils. In: Proceedings of the 51st IEEE Conference on Decision and Control, Maui, HI, USA, December 10–13, 2012, pp. 6560–6565. IEEE Publications, New York (2012)

3. Benner, P., Voigt, M.: A structured pseudospectral method for H_∞-norm computation of large-scale descriptor systems. MPI Magdeburg Preprint MPIMD/12-10 (2012)
4. Benner, P., Sima, V., Voigt, M.: L_∞-norm computation for continuous-time descriptor systems using structured matrix pencils. IEEE Trans. Autom. Control **57**(1), 233–238 (2012)
5. Berger, T.: Robustness of stability of time-varying index-1 DAEs. Institute for Mathematics, Ilmenau University of Technology, Preprint 12-10 (2012)
6. Berger, T.: Bohl exponent for time-varying linear differential-algebraic equations. Int. J. Control **85**(10), 1433–1451 (2012)
7. Berger, T., Ilchmann, A.: On stability of time-varying linear differential-algebraic equations. Institute for Mathematics, Ilmenau University of Technology, Preprint 10-12 (2012)
8. Berger, T., Reis, T.: Controllability of linear differential-algebraic systems—a survey. Institute for Mathematics, Ilmenau University of Technology, Preprint 12-02 (2012)
9. Bohl, P.: Über differentialungleichungen. J. Reine Angew. Math. **144**, 284–313 (1913)
10. Boyd, S., Balakrishnan, V.: A regularity result for the singular values of a transfer matrix and a quadratically convergent algorithm for computing its L_∞-norm. Syst. Control Lett. **15**, 1–7 (1990)
11. Bracke, M.: On stability radii of parametrized linear differential-algebraic systems. Ph.D. Thesis, University of Kaiserslautern (2000)
12. Brenan, K.E., Campbell, S.L., Petzold, L.R.: Numerical Solution of Initial-Value Problems in Differential Algebraic Equations, 2nd edn. SIAM, Philadelphia (1996)
13. Bruinsma, N.A., Steinbuch, M.: A fast algorithm to compute the H_∞ of a transfer function matrix. Syst. Control Lett. **14**, 287–293 (1990)
14. Bunse-Gerstner, A., Mehrmann, V., Nichols, N.K.: Regularization of descriptor systems by derivative and proportional state feedback. SIAM J. Matrix Anal. Appl. **13**, 46–67 (1992)
15. Bunse-Gerstner, A., Mehrmann, V., Nichols, N.K.: Regularization of descriptor systems by output feedback. IEEE Trans. Autom. Control **39**, 1742–1748 (1994)
16. Bunse-Gerstner, A., Byers, R., Mehrmann, V., Nichols, N.K.: Feedback design for regularizing descriptor systems. Linear Algebra Appl. **299**, 119–151 (1999)
17. Burke, J., Lewis, A.S., Overton, M.L.: Optimization and pseudospectra, with applications to robust stability. SIAM J. Matrix Anal. Appl. **25**, 80–104 (2003)
18. Byers, R.: A bisection method for measuring the distance of a stable matrix to the unstable matrices. SIAM J. Sci. Stat. Comput. **9**, 875–881 (1988)
19. Byers, R.: The descriptor controllability radius. In: Systems and Networks: Mathematical Theory and Application, Proceedings of MTNS'93, pp. 85–88. Akademie Verlag, Berlin (1994)
20. Byers, R., Nichols, N.K.: On the stability radius of a generalized state-space system. Linear Algebra Appl. **188–189**, 113–134 (1993)
21. Byers, R., He, C., Mehrmann, V.: Where is the nearest non-regular pencil? Linear Algebra Appl. **285**, 81–105 (1998)
22. Campbell, S.L.: Linearization of DAE's along trajectories. Z. Angew. Math. Phys. **46**, 70–84 (1995)
23. Campbell, S.L., Nichols, N.K., Terrell, W.J.: Duality, observability, and controllability for linear time-varying descriptor systems. Circuits Syst. Signal Process. **10**, 455–470 (1991)
24. Chyan, C.J., Du, N.H., Linh, V.H.: On data-dependence of exponential stability and the stability radii for linear time-varying differential-algebraic systems. J. Differ. Equ. **245**, 2078–2102 (2008)
25. Cong, N.D., Nam, H.: Lyapunov's inequality for linear differential algebraic equation. Acta Math. Vietnam. **28**, 73–88 (2003)
26. Cong, N.D., Nam, H.: Lyapunov regularity of linear differential algebraic equations of index 1. Acta Math. Vietnam. **29**, 1–21 (2004)
27. Dai, L.: Singular Control Systems. Lecture Notes in Control and Information Science. Springer, Berlin (1989)
28. Daleckii, J.L., Krein, M.G.: Stability of Solutions of Differential Equations in Banach Spaces. American Mathematical Society, Providence (1974)

29. Dieci, L., Van Vleck, E.S.: Lyapunov spectral intervals: theory and computation. SIAM J. Numer. Anal. **40**, 516–542 (2002)
30. Dmitriev, M., Kurina, G.: Singular perturbation in control problems. Autom. Remote Control **67**, 1–43 (2006)
31. Dragan, V.: The asymptotic behavior of the stability radius for a singularly perturbed linear system. Int. J. Robust Nonlinear Control **8**, 817–829 (1998)
32. Dragan, V., Halanay, A.: Stabilization of Linear Systems. Birkhäuser, Boston (1999)
33. Du, N.H.: Stability radii of differential-algebraic equations with structured perturbations. Syst. Control Lett. **57**, 546–553 (2008)
34. Du, N.H., Linh, V.H.: Implicit-system approach to the robust stability for a class of singularly perturbed linear systems. Syst. Control Lett. **54**, 33–41 (2005)
35. Du, N.H., Linh, V.H.: Robust stability of implicit linear systems containing a small parameter in the leading term. IMA J. Math. Control Inf. **23**, 67–84 (2006)
36. Du, N.H., Linh, V.H.: Stability radii for linear time-varying differential-algebraic equations with respect to dynamic perturbations. J. Differ. Equ. **230**, 579–599 (2006)
37. Du, N.H., Lien, D.T., Linh, V.H.: Complex stability radii for implicit discrete-time systems. Vietnam J. Math. **31**, 475–488 (2003)
38. Du, N.H., Thuan, D.D., Liem, N.C.: Stability radius of implicit dynamic equations with constant coefficients on time scales. Syst. Control Lett. **60**, 596–603 (2011)
39. Eich-Soellner, E., Führer, C.: Numerical Methods in Multibody Systems. Teubner Verlag, Stuttgart (1998)
40. Freitag, M.A., Spence, A.: A Newton-based method for the calculation of the distance to instability. Linear Algebra Appl. **435**(12), 3189–3205 (2011)
41. Gantmacher, F.R.: Theory of Matrices, vol. 2. Chelsea, New York (1959)
42. Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. The Johns Hopkins University Press, Baltimore (1996)
43. Griepentrog, E., März, R.: Differential-Algebraic Equations and Their Numerical Treatment. Teubner Verlag, Leipzig (1986)
44. Gu, M., Mengi, E., Overton, M.L., Xia, J., Zhu, J.: Fast methods for estimating the distance to uncontrollability. SIAM J. Matrix Anal. Appl. **28**, 477–502 (2006)
45. Guglielmi, N., Overton, M.L.: Fast algorithms for the approximation of the pseudospectral abscissa and pseudospectral radius of a matrix. SIAM J. Matrix Anal. Appl. **32**, 1166–1192 (2011)
46. Gürbüzbalaban, M., Overton, M.L.: Some regularity results for the pseudospectral abscissa and pseudospectral radius of a matrix. SIAM J. Optim. **22**, 281–285 (2012)
47. Hairer, E., Wanner, G.: Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems, 2nd edn. Springer, Berlin (1996)
48. He, C., Watson, G.A.: An algorithm for computing the distance to instability. SIAM J. Matrix Anal. Appl. **20**, 101–116 (1999)
49. Hinrichsen, D., Kelb, B.: Spectral value sets: a graphical tool for robustness analysis. Syst. Control Lett. **21**(2), 127–136 (1993)
50. Hinrichsen, D., Pritchard, A.J.: Stability radii of linear systems. Syst. Control Lett. **7**, 1–10 (1986)
51. Hinrichsen, D., Pritchard, A.J.: Stability radii for structured perturbations and the algebraic Riccati equations. Syst. Control Lett. **8**, 105–113 (1986)
52. Hinrichsen, D., Pritchard, A.J.: A note on some difference between real and complex stability radii. Syst. Control Lett. **14**, 401–408 (1990)
53. Hinrichsen, D., Pritchard, A.J.: Destabilization by output feedback. Differ. Integral Equ. **5**, 357–386 (1992)
54. Hinrichsen, D., Pritchard, A.J.: On spectral variations under bounded real matrix perturbations. Numer. Math. **60**, 509–524 (1992)
55. Hinrichsen, D., Pritchard, A.J.: Mathematical Systems Theory I. Modelling, State Space Analysis, Stability and Robustness. Springer, New York (2005)

56. Hinrichsen, D., Ilchmann, A., Pritchard, A.J.: Robustness of stability of time-varying linear systems. J. Differ. Equ. **82**, 219–250 (1989)
57. Hinrichsen, D., Kelb, B., Linnemann, A.: An algorithm for the computation of the complex stability radius. Automatica **25**, 771–775 (1989)
58. Hu, G.: Robustness measures for linear time-invariant time-delay systems. Ph.D. Thesis, University of Toronto (2001)
59. Ilchmann, A., Mareels, I.M.Y.: On stability radii of slowly time-varying systems. In: Advances in Mathematical System Theory, pp. 55–75. Birkhäuser, Boston (2001)
60. Jacob, B.: A formula for the stability radius of time-varying systems. J. Differ. Equ. **142**, 167–187 (1998)
61. Kokotovic, P., Khalil, H.K., O'Reilly, J.: Singular Perturbation Method in Control: Analysis and Design. Academic Press, New York (1986)
62. Kunkel, P., Mehrmann, V.: Numerical solution of differential algebraic Riccati equations. Linear Algebra Appl. **137/138**, 39–66 (1990)
63. Kunkel, P., Mehrmann, V.: Differential-Algebraic Equations. Analysis and Numerical Solution. EMS Publishing House, Zürich (2006)
64. Kunkel, P., Mehrmann, V., Rath, W., Weickert, J.: A new software package for linear differential–algebraic equations. SIAM J. Sci. Comput. **18**, 115–138 (1997)
65. Kurina, G.A., März, R.: Feedback solutions of optimal control problems with DAE constraints. In: Proceedings of the 47th IEEE Conf. on Decision and Control, Cancun, Mexico, Dec. 9–11 (2008)
66. Lam, S.: Real robustness radii and performance limitations of LTI control systems. Ph.D. Thesis, University of Toronto (2011)
67. Lam, S., Davison, E.J.: Real controllability radius of high-order, descriptor, and time-delay LTI systems. In: Proceedings of the 18th IFAC World Congress, Milano, Italy, August 28–September 2 (2011). 6 pages
68. Lamour, R., März, R., Tischendorf, C.: Differential-Algebraic Equations: A Projector Based Analysis. Springer, Berlin (2013)
69. Lee, L., Fang, C.-H., Hsieh, J.-G.: Exact unidirectional perturbation bounds for robustness of uncertain generalized state-space systems: continuous-time cases. Automatica **33**, 1923–1927 (1997)
70. Linh, V.H., Mehrmann, V.: Lyapunov, Bohl and Sacker–Sell spectral intervals for differential-algebraic equations. J. Dyn. Differ. Equ. **21**, 153–194 (2009)
71. Linh, V.H., Mehrmann, V.: Approximation of spectral intervals and associated leading directions for linear differential-algebraic systems via smooth singular value decompositions. SIAM J. Numer. Anal. **49**, 1810–1835 (2011)
72. Linh, V.H., Mehrmann, V.: Spectral analysis for linear differential-algebraic systems. In: Proceedings of the 8th AIMS International Conference on Dynamical Systems, Differential Equations and Applications, Dresden, May 24–28, pp. 991–1000 (2011). DCDS Supplement
73. Linh, V.H., Mehrmann, V., Van Vleck, E.: QR methods and error analysis for computing Lyapunov and Sacker–Sell spectral intervals for linear differential-algebraic equations. Adv. Comput. Math. **35**, 281–322 (2011)
74. Lyapunov, A.M.: The general problem of the stability of motion. Translated by A.T. Fuller from E. Davaux's French translation (1907) of the 1892 Russian original. Int. J. Control, 521–790 (1992)
75. Mattheij, R.M.M., Wijckmans, P.M.E.J.: Sensitivity of solutions of linear DAE to perturbations of the system matrices. Numer. Algorithms **19**, 159–171 (1998)
76. Mehrmann, V.: The Autonomous Linear Quadratic Control Problem: Theory and Numerical Solution. Lecture Notes in Control and Information Sciences, vol. 163. Springer, Heidelberg (1991)
77. Mehrmann, V.: Index concepts for differential-algebraic equations. Preprint 2012-03, Institut für Mathematik, TU Berlin (2012). http://www.math.tu-berlin.de/preprints/
78. Mehrmann, V., Stykel, T.: Descriptor systems: a general mathematical framework for modelling, simulation and control. Automatisierungstechnik **54**(8), 405–415 (2006)

79. Mengi, E.: Measures for robust stability and controllability. Ph.D. Thesis, University of New York (2006)
80. Mengi, E., Overton, M.L.: Algorithms for the computation of the pseudospectral radius and the numerical radius of a matrix. IMA J. Numer. Anal. **25**, 48–669 (2005)
81. Motscha, M.: Algorithm to compute the complex stability radius. Int. J. Control **48**, 2417–2428 (1988)
82. Qiu, L., Davison, E.J.: The stability robustness of generalized eigenvalues. IEEE Trans. Autom. Control **37**, 886–891 (1992)
83. Qiu, L., Bernhardsson, B., Rantzer, A., Davison, E.J., Young, P.M., Doyle, J.C.: A formula for computation of the real stability radius. Automatica **31**, 879–890 (1995)
84. Rabier, P.J., Rheinboldt, W.C.: Theoretical and Numerical Analysis of Differential-Algebraic Equations. Handbook of Num. Analysis, vol. VIII. Elsevier, Amsterdam (2002)
85. Riaza, R.: Differential-Algebraic Systems. Analytical Aspects and Circuit Applications. World Scientific Publishing Co. Pte. Ltd., Hackensack (2008)
86. Sacker, R.J., Sell, G.R.: A spectral theory for linear differential systems. J. Differ. Equ. **27**, 320–358 (1978)
87. Son, N.K., Hinrichsen, D.: Robust stability of positive continuous time systems. Numer. Funct. Anal. Optim. **17**, 649–659 (1996)
88. Son, N.K., Thuan, D.D.: The structured distance to non-surjectivity and its application to calculating the controllability radius of descriptor systems. J. Math. Anal. Appl. **388**, 272–281 (2012)
89. Sreedhar, J., Van Dooren, P., Tits, A.: A fast algorithm to compute the real structured stability radius. Int. Ser. Numer. Math. **121**, 219–230 (1996)
90. Tidefelt, H.: Differential-algebraic equations and matrix-valued singular perturbation. Ph.D. Thesis, Linköping University (2009)
91. Trefethen, L.N., Embree, M.: Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators. Princeton University Press, Princeton (2005)
92. Van Loan, C.F.: How near is a stable matrix to an unstable matrix? Contemp. Math. **47**, 465–477 (1985)

DAEs in Circuit Modelling: A Survey

Ricardo Riaza

Abstract This paper surveys different analytical aspects of differential-algebraic models of electrical and electronic circuits. The use of DAEs in circuit modelling has increased in the last two decades, and differential-algebraic (or *semistate*) models play nowadays a key role in circuit simulation programs and also in the analysis of several aspects of nonlinear circuit dynamics. We discuss not only nodal systems, including MNA, but also branch-oriented and hybrid ones, as well as the models arising in other approaches to circuit analysis. Different results characterizing the index of DAE models, for both passive and active circuits, are reviewed in detail. We also present a detailed discussion of memristive devices (memristors, memcapacitors and meminductors), displaying a great potential impact in electronics in the near future, and address how to accommodate them in differential-algebraic models. Some dynamical aspects and other topics in circuit theory in which DAEs play a role, regarding e.g. model reduction, coupled problems or fault diagnosis, are discussed in less detail.

Keywords Differential-algebraic equation · Electrical circuit · Electronic circuit · Semistate model · State equation · Index · Nodal analysis · Hybrid analysis · Loop analysis · Equilibrium · Bifurcation · Memristor · Memcapacitor · Meminductor

Mathematics Subject Classification 05C50 · 34A09 · 94C05 · 94C15

1 Introduction

Circuit modelling, analysis and simulation in the nonlinear setting have greatly benefited from the use of the DAE formalism in the last three decades. The ubiquitous presence of nonlinear devices in modern electronic circuits naturally leads to time-domain models; the differential-algebraic form of circuit equations emanates

R. Riaza (✉)
Departamento de Matemática Aplicada a las Tecnologías de la Información, Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain
e-mail: ricardo.riaza@upm.es

A. Ilchmann, T. Reis (eds.), *Surveys in Differential-Algebraic Equations I*, Differential-Algebraic Equations Forum,
DOI 10.1007/978-3-642-34928-7_3,

from the combination of differential equations coming from reactive elements with algebraic (non-differential) relations modelling Kirchhoff laws and device characteristics. In the opposite direction, a considerable amount of research on analytical and numerical aspects of differential-algebraic equations has been motivated by applications in circuit theory.

The term 'algebraic-differential system' was already used in the circuit context by Brown in 1963 [25]. Actually, the work of Bashkow, Brown, Bryant and others in the late 1950s and in the 1960s [10, 15, 29, 30, 48, 49, 105, 128, 181] on the formulation of state space models defines an important precedent of the use of DAEs in circuit modelling. A nice compilation of the state-of-the-art of state space modelling of nonlinear circuits up to 1980 can be found in Chua's paper [35]. The state space approach to circuit modelling displays, however, some important limitations. For several circuit configurations an explicit state space equation may not exist, not even locally. Additionally, when state space descriptions do exist, their formulation may be hardly automatable. The latter is extremely important from the computational point of view, specially in very large scale integration systems.

These limitations led, in the 1970s and 1980s, to the systematic formulation of *semistate* models, which use larger sets of network variables allowing some redundancy between them. Semistate models are currently framed in the differential-algebraic context. A milestone is the 1971 paper of Gear [71], addressing numerical aspects and which many consider the first paper on DAEs. Less known but also relevant from an analytical perspective is the paper of Takens [191]. The term "semistate", which was proposed by Dziurla (see p. 31 in [94]), appeared for the first time in the joint work of Dziurla and Newcomb [53]; an important reference in this regard is Newcomb's paper [126]. Other early contributions in this direction can be found in [31, 32, 172]. Later in the 1980s are worth recalling the 1986 and 1989 special issues of *Circuits, Systems, and Signal Processing* [108, 109], the book [92] and the papers [36, 37, 42, 43, 86–88, 127].

Since the 1990s, DAEs have been pervasive in nonlinear circuit analysis and design, specially because of their appearance in nodal analysis methods (cf. [57, 61, 79–83, 116, 117, 143, 156, 169, 193, 194]) used to set up automatically network equations in circuit simulation. As detailed later, this is the case of Modified Nodal Analysis (MNA), whose origin can be traced back to [93] and which is used in different circuit simulation programs, notably in SPICE and its commercial variants. Other techniques, not based on the use of node potentials, have also been the object of attention in the DAE context [69, 101, 149, 156, 190]. Different analytical and numerical aspects have been examined in the last decade within the context of properly stated DAE models: see [115, 117, 194] and the forthcoming title [107]. Section 2 will present a detailed discussion of different families of differential-algebraic circuit models.

In the numerical simulation of circuit dynamics, a key aspect is the computation and monitorization of the index of the differential-algebraic models, a problem which has attracted much attention; find in [61, 80, 81, 100, 101, 149, 156, 190, 193, 194] several results concerning the index of different circuit models under passivity assumptions; extensions to the non-passive context are discussed in [55, 69]. The main results in this direction will be compiled in Sect. 3.

A large amount of recent research has been motivated by the introduction in nonlinear circuit theory of so-called *mem-devices*. The memory-resistor or *memristor* is an electronic device defined by a nonlinear relation between the charge and the flux, and its existence was predicted by Leon Chua in 1971 for symmetry reasons [34]. The memristor would be the fourth basic circuit element, in addition to resistors, capacitors and inductors, whose characteristics relate voltage and current, voltage and charge, and flux and current, respectively. The report in 2008 of a nanoscale device with a memristive characteristic [184] had a great impact in electrical and electronic engineering, making the memristor a topic of active research (cf. [11, 45, 70, 97–99, 103, 121–124, 131–133, 157, 159–161, 167, 203] and references therein), which has been further motivated by the announcement of HP that commercial memory chips based on the memristor will be released in 2013 [1]. In 2009 the idea of a device with memory was extended to reactive elements by Di Ventra, Pershin and Chua [51]. In Sect. 4 we will discuss some features of DAE models of circuits with mem-devices. In this context, circuits without mem-devices are often referred to as "classical circuits".

The analysis of different theoretical aspects of circuit dynamics benefits from the use of differential-algebraic models. These include the state formulation problem or the study of different qualitative issues of nonlinear circuits, including circuits with memristors. The DAE formalism is mandatory in the analysis of impasse behaviour and phenomena related to singularities, which cannot be displayed in the setting of explicit ODEs. These aspects are addressed in Sect. 5; cf. [36, 37, 46, 63, 92, 148, 151, 153–156, 165, 166, 168]. Other aspects of DAE-based circuit modelling, related to model reduction, coupled problems or fault diagnosis, will be discussed more briefly in Sect. 6.

This survey will be focussed on analytical aspects of DAEs in circuit modelling. For different aspects of numerics in circuit simulation, the reader is referred to the references compiled on this topic in Sect. 6. For the sake of brevity no proofs are included, but the papers where these proofs can be found are cited after each result. Also for the sake of simplicity, controlled sources are not included in the models, although in most cases the results apply also to circuits with controlled sources under certain topological restrictions (cf. [61, 101, 156]). Other surveys published on this topic or on closely related ones in the last 15 years are [59, 62, 80, 81, 84, 169]; related aspects are also addressed in the papers on PDAEs and port-Hamiltonian systems within the present volume.

2 Model Families for Classical Circuits

The semistate or differential-algebraic framework provides a valuable tool for circuit modelling since it makes it possible to accommodate, in a comprehensive manner, different families of circuit models. A difference between model families is made by the set of semistate variables that they use. In particular, *nodal* methods are characterized by the use (in addition to some branch variables) of node potentials e as the fundamental model variables, by expressing Kirchhoff's voltage law

as $v = A^T e$. This approach comprises Modified Nodal Analysis (MNA) techniques, but also other methods such as Node Tableau Analysis (NTA) [41, 44, 80, 81, 85] or Augmented Nodal Analysis (ANA) [55, 110, 169]. After providing some preliminary material on graphs and elementary aspects of circuit theory in Sects. 2.1 and 2.2, nodal analysis models are discussed in Sect. 2.3.

Branch-oriented models (also known as "standard", "mixed" or "hybrid" models; note that the latter is used below with a different meaning) are characterized by the statement of Kirchhoff laws in the form $Ai = 0$ (or $Qi = 0$) and $Bv = 0$. Here A, B and Q are incidence, loop and cutset matrices (cf. Sects. 2.1 and 2.2). This will make it possible to formulate the circuit equations just in terms of the branch currents i and voltages v, which explains the 'branch-oriented' label for these methods. See [92, 149, 156]. So-called *hybrid models* (cf. [4, 20, 21, 33, 91, 104, 171]) can be obtained as a reduction of branch-oriented equations and have received very recent attention because their index does not exceed one, either in a passive or a non-passive context [69, 101, 190]. Branch-oriented, tree-based and hybrid models are presented in Sect. 2.4. We will also briefly consider other DAE models in Sects. 2.5 and 2.6.

2.1 Graph-Theoretic Results

The formulation of nodal models for electrical circuits will based on the description of the underlying digraph in terms of the so-called *reduced incidence matrix* $A = (a_{ij}) \in \mathbb{R}^{(n-1)\times b}$, where

$$a_{ij} = \begin{cases} 1 & \text{if branch } j \text{ leaves node } i, \\ -1 & \text{if branch } j \text{ enters node } i, \\ 0 & \text{if branch } j \text{ is not incident with node } i. \end{cases}$$

We are assuming the digraph to be connected and to have n nodes and b branches. Note that the reference node does not have an associated row in A.

An alternative description of the circuit's underlying digraph can be given in terms of the so-called reduced loop and cutset matrices. This description will be used in the formulation of branch-oriented and related models. We will use both the term *loop* and *cycle* to mean the branches in a closed path without self-intersections. A *reduced loop matrix* is defined, for a connected digraph with n nodes, b branches and after having chosen $b - n + 1$ independent loops, as $B = (b_{ij}) \in \mathbb{R}^{(b-n+1)\times b}$, with

$$b_{ij} = \begin{cases} 1 & \text{if branch } j \text{ is in loop } i \text{ with the same orientation}, \\ -1 & \text{if branch } j \text{ is in loop } i \text{ with the opposite orientation}, \\ 0 & \text{if branch } j \text{ is not in loop } i. \end{cases}$$

Similarly, the entries of a *reduced cutset matrix* $Q = (q_{ij}) \in \mathbb{R}^{(n-1)\times b}$ are given by

$$q_{ij} = \begin{cases} 1 & \text{if branch } j \text{ is in cutset } i \text{ with the same orientation,} \\ -1 & \text{if branch } j \text{ is in cutset } i \text{ with the opposite orientation,} \\ 0 & \text{if branch } j \text{ is not in cutset } i. \end{cases}$$

Recall that a subset S of the set of branches of a connected digraph is a *cutset* if the removal of S results in a disconnected graph, and it is minimal with respect to this property, that is, the removal of any proper subset of S does not disconnect the graph; a cutset may be oriented just by directing it from one of the connected components resulting from the cutset deletion towards the other.

The rows of any reduced loop matrix span the so-called *cycle space* $\operatorname{im} B^T$. Those of either a reduced incidence matrix or a reduced cutset matrix span the *cut space* $\operatorname{im} Q^T$. Details can be found in [19]. Both spaces are orthogonal to each other, so that $\operatorname{im} B^T = \ker Q$ and $\operatorname{im} Q^T = \ker B$.

Let K be a set of branches of a given digraph $\mathcal{G}$. The presence or absence of loops and cutsets within K can be described in terms of the aforementioned matrices as indicated in Lemmas 2.1 and 2.2 below. Find details in [5, 6, 64, 156]. By A_K (resp. $A_{\mathcal{G}-K}$) we mean the submatrix of A defined by the columns which correspond to branches in (resp. not in) K. The same applies to B and Q.

Lemma 2.1 *A subset K of the set of branches of a connected digraph $\mathcal{G}$ does not contain loops if and only if A_K has full column rank. In particular, if K has $n-1$ branches, A_K is nonsingular if and only if K defines a spanning tree; in this case,* $\det A_K = \pm 1$.

Lemma 2.2 *Consider a connected digraph $\mathcal{G}$. The following statements are equivalent*:

1. *K does not contain cutsets*;
2. *$A_{\mathcal{G}-K}$ has full row rank*;
3. *$Q_{\mathcal{G}-K}$ has full row rank*;
4. *B_K has full column rank.*

The absence of loops can be also characterized in terms of the loop and cutset matrices by requiring $B_{\mathcal{G}-K}$ to have full row rank and Q_K to have full column rank, respectively.

In a connected digraph, the choice of a spanning tree (that is, a connected subgraph including all nodes and having no loops) yields two sets of so-called fundamental cutsets and fundamental loops; fundamental cutsets are defined by a tree branch (or *twig*) together with some cotree branches (or *links*) and, analogously, fundamental loops are defined by a link together with some twigs. Choosing the orientation of the cutsets and loops coherently with that of the corresponding twigs or links, and using the orthogonality property $QB^T = 0$, the corresponding reduced

cutset and loop matrices have the form

$$Q = \begin{pmatrix} I & -K^T \end{pmatrix}, \tag{2.1a}$$

$$B = \begin{pmatrix} K & I \end{pmatrix}, \tag{2.1b}$$

for a certain matrix K. The first columns of both matrices correspond to twigs, whereas the last ones are associated with the links. These matrices will be used in the formulation of tree-based models. The reader can find at the end of Sect. 2.2 an example illustrating all the notions introduced above.

2.2 Some Preliminaries from Circuit Theory

Kirchhoff Laws Possibly the most fundamental properties of electrical circuits are expressed by Kirchhoff laws. Kirchhoff's current law (KCL) states that *the sum of the currents leaving any circuit node is zero*. This must be understood as follows: if the current in branch k is denoted by i_k and this branch is directed away from (resp. towards) a given node, then the current leaving this node is i_k (resp. $-i_k$). In terms of the reduced incidence and cutset matrices A, Q, this law can be expressed as $Ai = 0$ or $Qi = 0$, respectively.

In turn, Kirchhoff's voltage law (KVL) states that *the sum of the voltage drops along the branches of any loop is zero*. In this statement, provided that an orientation is defined in every loop and denoting by v_k the voltage in branch k, the corresponding voltage drop must be understood as v_k if branch k has the same orientation as the loop, and $-v_k$ otherwise. In terms of the loop matrix, KVL can be expressed as $Bv = 0$. It is also possible to express Kirchhoff's voltage law in terms of the *node potentials* e as $v = A^T e$; this will be one of the key features of nodal models.

Component Relations The basic components of electrical circuits are resistors, capacitors, inductors, and voltage and current sources (find in Sect. 4 the recently introduced *memristor* and other mem-devices). In a nonlinear setting, resistors can be either current-controlled, being governed by a relation of the form $v_r = \gamma_r(i_r)$, or voltage-controlled, this relation having the form $i_r = \gamma_g(v_r)$. For the sake of simplicity, capacitors and inductors will be supposed to be voltage-controlled and current-controlled (respectively) by relations of the form $q_c = \gamma_c(v_c)$ and $\varphi_l = \gamma_l(i_l)$; here q and φ are the charge and the magnetic flux. Note that in some cases, these descriptions may exist only locally: this is the case, for instance, in the Josephson junction [44], a device composed of two superconductors separated by an oxide barrier and which is governed by a current-flux characteristic of the form $i_l = I_0 \sin(k\varphi_l)$ for certain physical constants I_0, k. Points where $\cos(k\varphi_l) = 0$ do not admit any current-controlled description and, away from them, such a current-controlled description is only locally defined.

Also for simplicity, voltage and current sources will be assumed to be independent, their voltage and current being, respectively, defined by explicit functions of time, to be denoted by $v_s(t)$ and $i_s(t)$.

When the characteristics defined above are differentiable, the *incremental capacitance matrix* of the set of capacitors is defined as $C(v_c) = \gamma_c'(v_c)$, whereas the *incremental inductance matrix* of inductors is $L(i_l) = \gamma_l'(i_l)$. Analogously, the *incremental conductance matrix* for voltage-controlled resistors is defined as $G(v_r) = \gamma_g'(v_r)$. In a current-controlled setting, the *incremental resistance matrix* is $R(i_r) = \gamma_r'(i_r)$. All these matrices will be diagonal in the absence of coupling effects. When they are positive definite (resp. semidefinite) in a given region, the corresponding devices are said to be *strictly locally passive* (resp. *locally passive*) in that region [35, 44]; recall that an $m \times m$ matrix M is positive definite (resp. semidefinite) if $u^T Mu > 0$ (resp. ≥ 0) for any $u \in \mathbb{R}^m - \{0\}$; we do not require M to be symmetric. In the absence of coupling effects, these passivity (resp. strict passivity) conditions amount to requiring that the individual capacitances, inductances, conductances or resistances are non-negative (resp. positive).

Topologically Degenerate Configurations An important role in the index analysis will be played by so-called *topologically degenerate configurations,* namely, VC-loops (loops defined by capacitors and/or voltage sources), and IL-cutsets (cutsets composed of inductors and/or current sources). The former restrict the set of admissible values for the branch voltages of the capacitors within the loop, whereas the latter restrict the values for the currents of the inductors belonging to the cutset. In most models these configurations lead to a higher index DAE, as detailed in Sect. 3.

Note that, using Lemmas 2.1 and 2.2, the presence of such configurations can be characterized in terms of the incidence, loop and cutset matrices. For instance, the existence of a VC-loop makes the columns of the submatrix $(A_c \ A_u)$ (defined by the columns of the reduced incidence matrix which correspond to capacitors and voltage sources) linearly dependent; analogously, the presence of an IL-cutset makes the rows of $(A_r \ A_c \ A_u)$ linearly dependent. Equations (3.2a)–(3.2c), (3.3) and (3.4) in Sect. 3.2.1 illustrate how to use this type of results.

Example We will use the circuit depicted in Fig. 1 as a running example for the non-expert reader, aimed to illustrate different concepts and results introduced throughout the document. It includes a voltage and a current source (labelled as V_s and I_s, respectively), two linear inductors with inductances L_1 and L_2, a linear capacitor with capacitance C, and a diode (labelled as D). The latter is an example of a voltage-controlled nonlinear resistor. We are not concerned with the specific form of its characteristic; it is enough to write it as $i_d = \gamma_g(v_d)$ for a certain function γ_g. We will denote its incremental conductance $\gamma_g'(v_d)$ as $G(v_d)$ (or simply as G). At certain points in the index analysis we will consider cases in which G may vanish or even become negative for certain values of v_d; in practice, this may happen for instance if the diode exhibits a tunnelling effect which makes it locally active at a certain operating region [44, 187].

We may already use this example to show the form that the reduced incidence, loop and cutset matrices take. Let us begin with the reduced incidence matrix A, using the ground terminal as reference node. We will order the branches as V_s, C,

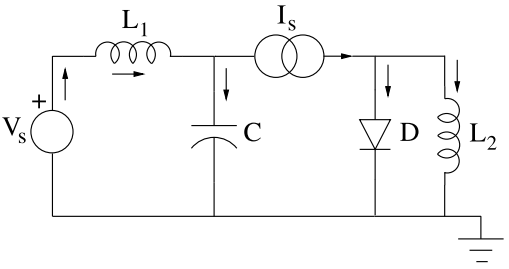


Fig. 1 Example

D, L_1, I_s, L_2. The nodes on top of the voltage source, the capacitor and the diode will have numbers 1, 2 and 3, respectively. With these conventions and the branch directions defined by the arrows in Fig. 1, the reduced incidence matrix reads

$$A = \begin{pmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{pmatrix}.$$

For instance, the branch accommodating the voltage source leaves the reference node and enters node 1, hence the entries -1, 0, 0 in the first column of A. Analogously, the capacitor and the diode leave nodes 2 and 3, respectively, both entering the reference node. The L_1-inductor (cf. the fourth column of A) leaves node 1 and enters node 2, etc.

A reduced loop matrix B is easily constructed by considering the meshes defined by V_s, L_1 and C; by C, I_s and D; and by D and L_2. The corresponding loops are assumed to be oriented clockwise. This yields

$$B = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{pmatrix}. \tag{2.2}$$

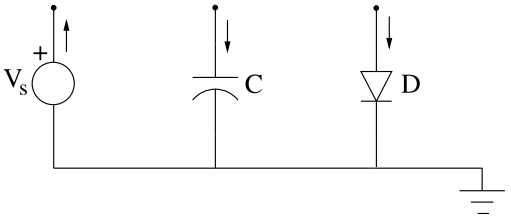
Here, the first row reflects that the branches accommodating V_s, L_1 and C are oriented as the loop itself; in the second one, the -1 in the second entry is due to the fact that the C-branch has the opposite direction to the loop, contrary to I_s and D. In the last row, the L_2-branch has the same direction as the loop but the D-branch has not.

A reduced cutset matrix Q can be built analogously. Let us consider the cutsets defined by V_s and L_1; by L_1, C and I_s; and by I_s, D and L_2. Note that the removal of each one of these sets of branches leaves nodes 1, 2 and 3 (respectively) isolated. By orientating these cutsets coherently with V_s, C and D, respectively, we get

$$Q = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{pmatrix}. \tag{2.3}$$

Indeed, the first row corresponds to the cutset defined by V_s and L_1; note that the removal of these two branches leaves node 1 isolated. Since the orientation of the

Fig. 2 Spanning tree



cutset is defined by that of the V_s-branch (which enters node 1), there is a $+1$ in the first entry; now, the fact that the L_1-branch leaves node 1 yields a -1 in the fourth column. The second and third rows are constructed analogously.

It is worth mentioning that B and Q in (2.2) and (2.3) are actually the fundamental matrices defined by the spanning tree depicted in Fig. 2.

Indeed, the fundamental cutsets are defined by the twigs V_s, C and D, whereas the fundamental loops correspond to the links L_1, I_s and L_2. Note in particular that the matrix K arising in (2.1a), (2.1b) reads

$$K = \begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}.$$

Mind the identity blocks in B and Q and also the appearance of $-K^T$ within the last three columns of Q in (2.3).

2.3 Nodal Analysis. MNA

By using the subscripts r, c, l, u and j to denote resistors, capacitors, inductors, voltage sources and current sources, respectively, we may express the *Node Tableau Analysis* (NTA) equations [85] as

$$\big(\gamma_c(v_c)\big)' = i_c, \tag{2.4a}$$

$$\big(\gamma_l(i_l)\big)' = v_l, \tag{2.4b}$$

$$0 = i_r - \gamma_g(v_r), \tag{2.4c}$$

$$0 = A_r i_r + A_l i_l + A_c i_c + A_u i_u + A_j i_s(t), \tag{2.4d}$$

$$0 = v_r - A_r^T e, \tag{2.4e}$$

$$0 = v_l - A_l^T e, \tag{2.4f}$$

$$0 = v_c - A_c^T e, \tag{2.4g}$$

$$0 = v_s(t) - A_u^T e, \tag{2.4h}$$

$$0 = v_j - A_j^T e, \tag{2.4i}$$

where γ_c, γ_l, γ_g describe the characteristics of capacitors, inductors and resistors, as introduced above; note that resistors are assumed here to be voltage-controlled. For the sake of brevity we will not address cases in which capacitors are charge-controlled or inductors are flux-controlled, which lead to so-called charge-oriented nodal models (see e.g. [61, 169, 193]). Note that Kirchhoff laws are expressed as $Ai = 0$ in (2.4d) and $v = A^T e$ in (2.4e)–(2.4i). We are splitting the reduced incidence matrix A as $(A_r \ A_l \ A_c \ A_u \ A_j)$; the subscript s (from "source") is used in v_s and i_s to indicate that in voltage and current sources the voltage and the current (respectively) are explicitly given functions of time.

Provided that γ_c and γ_l are differentiable and as far as one is not concerned with optimal smoothness descriptions of the solutions, the differential relations for reactive elements (2.4a)–(2.4b) can be recast in terms of the incremental capacitance and inductance matrices $C(v_c) = \gamma_c'(v_c)$, $L(i_l) = \gamma_l'(i_l)$ as

$$C(v_c)v_c' = i_c, \tag{2.5a}$$

$$L(i_l)i_l' = v_l. \tag{2.5b}$$

We will use this form in later models, but the reader should keep in mind that the proper formulation (2.4a)–(2.4b) (cf. [107, 115, 194]) may be used in all of them.

Augmented Nodal Analysis (ANA) models [110, 169] arise as a reduction of the NTA model, retaining its semiexplicit form and (as detailed in Sect. 3) its index. ANA models are defined by a system of the form

$$C(v_c)v_c' = i_c, \tag{2.6a}$$

$$L(i_l)i_l' = A_l^T e, \tag{2.6b}$$

$$0 = A_r\gamma_g\left(A_r^T e\right) + A_l i_l + A_c i_c + A_u i_u + A_j i_s(t), \tag{2.6c}$$

$$0 = v_c - A_c^T e, \tag{2.6d}$$

$$0 = v_s(t) - A_u^T e. \tag{2.6e}$$

The insertion of (2.6d) into (2.6a) and this in turn into (2.6c) yields the Modified Nodal Analysis (MNA) model

$$A_c C\left(A_c^T e\right)A_c^T e' = -A_r\gamma_g\left(A_r^T e\right) - A_l i_l - A_u i_u - A_j i_s(t), \tag{2.7a}$$

$$L(i_l)i_l' = A_l^T e, \tag{2.7b}$$

$$0 = v_s(t) - A_u^T e. \tag{2.7c}$$

Among all nodal techniques, MNA models are by far the most widely used, mainly because of its compact form; indeed, reducing the number of variables while at the same time being easy to set up in an automatic way makes them well-suited for computational purposes [61, 80, 81, 193, 194]. On the other hand it loses the semiexplicit form of NTA and ANA and this makes the index analysis a bit more intricate; note, however, that the index of MNA models never exceeds that of NTA or ANA, being strictly lower in some cases (cf. Sect. 3).

2.4 Branch-Oriented Models, Tree-Based Formulations and Hybrid Analysis

Branch-oriented models avoid the use of node potentials and express the circuit equations in terms of branch currents and voltages in the form

$$C(v_c)v_c' = i_c, \tag{2.8a}$$

$$L(i_l)i_l' = v_l, \tag{2.8b}$$

$$0 = A_r i_r + A_l i_l + A_c i_c + A_u i_u + A_j i_s(t), \tag{2.8c}$$

$$0 = B_r v_r + B_l v_l + B_c v_c + B_u v_s(t) + B_j v_j, \tag{2.8d}$$

$$0 = g_r(v_r, i_r). \tag{2.8e}$$

Because of the symmetric nature of the model, we do not make any assumptions on the controlling variables for resistors in (2.8e). Kirchhoff's current and voltage laws are expressed in (2.8c) and (2.8d) as $Ai = 0$, $Bv = 0$. A similar formulation results from recasting the former in terms of a reduced cutset matrix as $Qi = 0$.

Tree-Based Models The use of fundamental matrices makes it possible to rewrite Kirchhoff laws $Qi = 0$ and $Bv = 0$ as

$$i_{\mathrm{tr}} = K^T i_{\mathrm{co}}, \tag{2.9a}$$

$$v_{\mathrm{co}} = -K v_{\mathrm{tr}}. \tag{2.9b}$$

Throughout the document the subscripts $_{\mathrm{tr}}$ and $_{\mathrm{co}}$ refer to tree and cotree elements, respectively, that is, twigs and links. For instance, the vector of tree voltages v_{tr} can be split in terms of twig capacitors, inductors, resistors and voltage sources as $v_{\mathrm{tr}} = (v_{c_{\mathrm{tr}}}, v_{l_{\mathrm{tr}}}, v_{r_{\mathrm{tr}}}, v_u)$; analogously, the vector of cotree voltages v_{co} can be written in terms of link capacitors, inductors, resistors and current sources as $v_{\mathrm{co}} = (v_{c_{\mathrm{co}}}, v_{l_{\mathrm{co}}}, v_{r_{\mathrm{co}}}, v_j)$. Note that we assume that all voltage (resp. current) sources are located in the tree (resp. in the cotree); this is possible because all circuits are assumed to be well-posed, namely, that they display neither loops of voltage sources nor cutsets of current sources. An analogous splitting holds for the tree and cotree current vectors i_{tr} and i_{co}. The relations depicted in (2.9a), (2.9b) for Kirchhoff laws express the widely used fact that tree currents can be written in terms of cotree currents, whereas cotree voltages can be expressed in terms of tree voltages, a result which can be traced back at least to [186].

Once a tree has been chosen, the model (2.8a)–(2.8e) can be rewritten in the form

$$C(v_c)v_c' = i_c, \tag{2.10a}$$

$$L(i_l)i_l' = v_l, \tag{2.10b}$$

$$0 = i_{\mathrm{tr}} - K^T i_{\mathrm{co}}, \tag{2.10c}$$

$$0 = K v_{\mathrm{tr}} + v_{\mathrm{co}}, \tag{2.10d}$$

$$0 = g_r(v_r, i_r). \tag{2.10e}$$

Hybrid Analysis Recent research [100, 101, 190] has framed the hybrid analysis of Kron [104] (cf. also [4, 20, 21, 33, 91, 171]) in a differential-algebraic context. Among other advantages, hybrid models display an index which never exceeds that of MNA (cf. Sect. 3).

These hybrid models can be understood as a reduction of the tree-based model (2.10a)–(2.10e), when the latter is based on a *normal reference tree*, that is, a spanning tree with all voltage sources and no current source, as many twig capacitors as possible, among those satisfying the previous requirements one with as many twig voltage-controlled resistors as possible, and then one with as many twig current-controlled resistors as possible. As a byproduct, a normal reference tree has as few twig inductors as possible. This is an extension of Bryant's original notion of a *normal tree*, arising in connection with the state formulation problem [29, 30] (see also [25, 105]); this concept has been revisited since then in connection to different aspects of circuit analysis [101, 150, 164, 190].

For the sake of notational simplicity, hybrid models will be presented below for connected circuits including only capacitors, voltage- and current-controlled resistors (their branch variables being labelled with the subscripts g and r, respectively) and inductors. Independent sources offer no difficulties and are only omitted to simplify the discussion. In this context, a normal reference tree is a spanning tree which verifies the following requirements: it includes as many capacitors as possible; among the ones satisfying the previous condition, it includes as many voltage-controlled resistors as possible and, among these, it includes as many current-controlled resistors as possible. This yields for the matrix K in (2.1a), (2.1b) the structure

$$\begin{pmatrix} K_{11} & 0 & 0 & 0 \\ K_{21} & K_{22} & 0 & 0 \\ K_{31} & K_{32} & K_{33} & 0 \\ K_{41} & K_{42} & K_{43} & K_{44} \end{pmatrix}$$

and Kirchhoff laws read

$$\begin{pmatrix} v_{c_{\mathrm{co}}} \\ v_{g_{\mathrm{co}}} \\ v_{r_{\mathrm{co}}} \\ v_{l_{\mathrm{co}}} \end{pmatrix} = - \begin{pmatrix} K_{11} & 0 & 0 & 0 \\ K_{21} & K_{22} & 0 & 0 \\ K_{31} & K_{32} & K_{33} & 0 \\ K_{41} & K_{42} & K_{43} & K_{44} \end{pmatrix} \begin{pmatrix} v_{c_{\mathrm{tr}}} \\ v_{g_{\mathrm{tr}}} \\ v_{r_{\mathrm{tr}}} \\ v_{l_{\mathrm{tr}}} \end{pmatrix}$$

and

$$\begin{pmatrix} i_{c_{\mathrm{tr}}} \\ i_{g_{\mathrm{tr}}} \\ i_{r_{\mathrm{tr}}} \\ i_{l_{\mathrm{tr}}} \end{pmatrix} = \begin{pmatrix} K_{11}^T & K_{21}^T & K_{31}^T & K_{41}^T \\ 0 & K_{22}^T & K_{32}^T & K_{42}^T \\ 0 & 0 & K_{33}^T & K_{43}^T \\ 0 & 0 & 0 & K_{44}^T \end{pmatrix} \begin{pmatrix} i_{c_{\mathrm{co}}} \\ i_{g_{\mathrm{co}}} \\ i_{r_{\mathrm{co}}} \\ i_{l_{\mathrm{co}}} \end{pmatrix},$$

respectively; as above, the subscripts $_{\mathrm{tr}}$ and $_{\mathrm{co}}$ specify tree and cotree elements. By splitting all circuit variables and characteristic maps into tree and cotree ones, hybrid models are expressed in terms of the variables $v_{c_{\mathrm{tr}}}$, $v_{g_{\mathrm{tr}}}$, $i_{r_{\mathrm{co}}}$ and $i_{l_{\mathrm{co}}}$, as follows:

$$\begin{aligned}
&\big(C_{\mathrm{tr}}(v_{c_{\mathrm{tr}}}) + K_{11}^T C_{\mathrm{co}}(-K_{11} v_{c_{\mathrm{tr}}}) K_{11}\big) v'_{c_{\mathrm{tr}}} \\
&\quad = K_{21}^T \gamma_{g_{\mathrm{co}}}(-K_{21} v_{c_{\mathrm{tr}}} - K_{22} v_{g_{\mathrm{tr}}}) + K_{31}^T i_{r_{\mathrm{co}}} + K_{41}^T i_{l_{\mathrm{co}}}, \qquad (2.11a) \\
&\big(L_{\mathrm{co}}(i_{l_{\mathrm{co}}}) + K_{44} L_{\mathrm{tr}}\big(K_{44}^T i_{l_{\mathrm{co}}}\big) K_{44}^T\big) i'_{l_{\mathrm{co}}} \\
&\quad = -K_{41} v_{c_{\mathrm{tr}}} - K_{42} v_{g_{\mathrm{tr}}} - K_{43} \gamma_{r_{\mathrm{tr}}}\big(K_{33}^T i_{r_{\mathrm{co}}} + K_{43}^T i_{l_{\mathrm{co}}}\big), \qquad (2.11b) \\
&\gamma_{g_{\mathrm{tr}}}(v_{g_{\mathrm{tr}}}) = K_{22}^T \gamma_{g_{\mathrm{co}}}(-K_{21} v_{c_{\mathrm{tr}}} - K_{22} v_{g_{\mathrm{tr}}}) + K_{32}^T i_{r_{\mathrm{co}}} + K_{42}^T i_{l_{\mathrm{co}}}, \qquad (2.11c) \\
&\gamma_{r_{\mathrm{co}}}(i_{r_{\mathrm{co}}}) = -K_{31} v_{c_{\mathrm{tr}}} - K_{32} v_{g_{\mathrm{tr}}} - K_{33} \gamma_{r_{\mathrm{tr}}}\big(K_{33}^T i_{r_{\mathrm{co}}} + K_{43}^T i_{l_{\mathrm{co}}}\big). \qquad (2.11d)
\end{aligned}$$

Detailed derivations of this model can be found in [69, 101, 190].

2.5 Multiport Model and Hessenberg Form

As a starting point in the analysis of dynamic circuits (cf. for instance [10, 35, 154, 172, 179]) it is often assumed that resistive variables as well as i_u, v_j can be eliminated from (2.8a)–(2.8e). This leads to a model of the form

$$\begin{aligned}
C(v_c) v'_c &= i_c, \qquad (2.12a) \\
L(i_l) i'_l &= v_l, \qquad (2.12b) \\
0 &= \Psi(v_c, i_c, v_l, i_l, t) \qquad (2.12c)
\end{aligned}$$

for a certain map Ψ. System (2.12a)–(2.12c) is called a *multiport model*. Indeed, letting m stand for the number of reactive elements (capacitors and inductors), the map Ψ within (2.12c) can be understood to define m abstract relations involving the currents and voltages of certain m ports of a (possibly nonlinear) subnetwork which includes all resistors and sources, with a time dependence coming from the sources. The connection of reactances at those ports leads to the dynamical system modelled by the DAE (2.12a)–(2.12c). The map Ψ comprises implicitly the topology of the network. A detailed discussion of conditions which allow for such a reduction can be found in [156]. Additionally, as detailed there, in problems with loops defined by capacitors and/or voltage sources, or with cutsets composed of inductors and/or current sources, the choice of a normal tree (where "normal" is used in the sense specified in Sect. 3.2.2) allows one to eliminate twig capacitor currents and link inductor voltages, making it possible to express the circuit equations in Hessenberg form (cf. Sect. 2.7 below).

2.6 Loop Analysis

Another classical technique for setting up the circuit equation is the so-called *loop analysis* [33, 44], which in a nonlinear context leads to a differential-algebraic model (cf. for instance [162]), as detailed in the sequel.

Assume that the circuit is connected. Recalling that b and n stand for the number of branches and nodes, respectively, fix $b-n+1$ linearly independent loops, and let B stand for the associated reduced loop matrix. Assign a *loop current* j_k, $k=1,\dots,b-n+1$, to each one of these loops. When the circuit is planar, these loop currents can be taken as the ones defined by the *meshes*, that is, the loops encircling the different faces in a planar description of the circuit; note, however, that the circuit needs not be planar for the loop analysis to be feasible. We will denote by $\mathbf{j}$ the vector of loop currents.

The branch currents i can be computed from $\mathbf{j}$ simply as $i=B^T\mathbf{j}$. The loop analysis begins with the description of Kirchhoff's voltage law in the form $Bv=0$, and then proceeds by replacing as far as possible the branch voltages of current-controlled devices in terms of branch currents and, eventually, of loop currents. Voltage-controlled devices, such as capacitors or current sources, will introduce additional equations in the circuit model.

Assuming (for simplicity) the resistors to be current-controlled by a C^1 map of the form $v_r=\gamma_r(i_r)$, and splitting the cycle matrix B as $(B_r\ B_l\ B_c\ B_u\ B_j)$, as we did in the formulation of the branch-oriented model (2.8a)–(2.8e), the loop analysis equations then read

$$C(v_c)v_c' = B_c^T\mathbf{j}, \tag{2.13a}$$

$$L(i_l)i_l' = v_l, \tag{2.13b}$$

$$0 = B_r\gamma_r\left(B_r^T\mathbf{j}\right) + B_l v_l + B_c v_c + B_u v_s(t) + B_j v_j, \tag{2.13c}$$

$$0 = i_l - B_l^T\mathbf{j}, \tag{2.13d}$$

$$0 = i_s(t) - B_j^T\mathbf{j}. \tag{2.13e}$$

Note that this is the dual model of (2.6a)–(2.6e).

2.7 DAE Form of the Models

All the models discussed above have the form of a quasilinear DAE, that is,

$$A(u)u' = F(u,t) \tag{2.14}$$

where A and F are matrix-valued and vector-valued maps. More precisely (cf. (2.6a)–(2.6e), (2.7a)–(2.7c), (2.8a)–(2.8e), (2.10a)–(2.10e), (2.11a)–(2.11d), (2.12a)–(2.12c), (2.13a)–(2.13e)) these models take the form of a semiimplicit DAE

$$M(x)x' = f(x, y, t), \tag{2.15a}$$

$$0 = g(x, y, t), \tag{2.15b}$$

for different sets of vectors x, y, matrices $M(x)$ and maps f, g. For instance, in the hybrid model (2.11a)–(2.11d) we have $x = (v_{c_{\text{tr}}}, i_{l_{\text{co}}})$, $y = (v_{g_{\text{tr}}}, i_{r_{\text{co}}})$,

$$M = \begin{pmatrix} C_{\text{tr}}(v_{c_{\text{tr}}}) + K_{11}^T C_{\text{co}}(-K_{11} v_{c_{\text{tr}}}) K_{11} & 0 \\ 0 & L_{\text{co}}(i_{l_{\text{co}}}) + K_{44} L_{\text{tr}}(K_{44}^T i_{l_{\text{co}}}) K_{44}^T \end{pmatrix}, \tag{2.16}$$

and f, g capture the right-hand side of (2.11a)–(2.11d).

For all but the MNA and hybrid models (2.7a)–(2.7c), (2.11a)–(2.11d), the matrix $M(x)$ is nonsingular if and only if the inductance and capacitance matrices $L(i_l)$ and $C(v_c)$ are nonsingular, and under this assumption the corresponding DAE can be obviously rewritten in a semiexplicit form

$$x' = h(x, y, t), \tag{2.17a}$$

$$0 = g(x, y, t). \tag{2.17b}$$

Generally speaking, such semiexplicit reduction is not feasible for the MNA and hybrid models (2.7a)–(2.7c), (2.11a)–(2.11d); this is due to the fact that more variables are eliminated in these models, which offers a computational advantage, but as a drawback the index analysis (cf. Sect. 3) is a bit more cumbersome for MNA and hybrid systems.

It is worth mentioning that in some cases, notably after a reduction to the multiport form (2.12a)–(2.12c), in the presence of VC-loops (loops defined by capacitors and (maybe) voltage sources), and/or IL-cutsets (cutsets composed of inductors and (maybe) current sources), the semiexplicit system (2.17a) can be recast in Hessenberg form, namely

$$x' = h(x, y, t), \tag{2.18a}$$

$$0 = g(x, t). \tag{2.18b}$$

Specifically, when coming from a tree-based multiport model of the form (2.12a)–(2.12c), x in (2.18a), (2.18b) stands for capacitor voltages and inductor currents, whereas y comprises link capacitor currents and twig inductor voltages. The absence of y in (2.18b) confers (2.18a), (2.18b) a Hessenberg structure, which displays several advantages and can be understood to define a standard form for higher index DAEs. Find details in [156].

3 The Index of DAE Circuit Models

The characterization of the index of the above-introduced models is a central problem in circuit simulation [61, 80, 81, 117, 156, 193]. Not only the index of nodal

models but also that of branch-oriented and hybrid ones has been carefully examined [55, 61, 69, 100, 101, 156, 190, 193]. This interest stems from the fact that the index has a major impact in the numerical techniques to be used in the simulation of circuit dynamics. The index is also relevant with regard to other analytical properties of nonlinear circuits, related e.g. to the state formulation problem or to different qualitative issues [36, 37, 63, 92, 179, 180, 198]. The reader is referred to [24, 75, 89, 90, 106, 107, 140, 156] for detailed introductions to the different index notions in DAE theory.

As indicated above, a great deal of research in this direction has been focussed on the characterization of the index of nodal models. This has been mainly motivated by the use of nodal techniques such as Modified Nodal Analysis (MNA) in circuit simulators, notably in SPICE and its commercial variants [80, 81, 143]. Under passivity assumptions, the index of nodal models is known to be not greater than two, according to the results in [61, 193]. A key feature of this approach is its *topological* emphasis, aiming at the characterization of different properties in terms of the underlying digraph and the electric nature of every branch, disregarding the specific characteristic equations of each device. Some recent results extend the analysis to low index configurations in non-passive nodal models [55].

On the other hand, recent research has been focussed on the *hybrid models* presented in Sect. 2.4 above. The recent use of a DAE formalism to accommodate hybrid models has made it possible to show that their index does not exceed one in passive contexts [101, 190], in contrast to MNA and other nodal techniques, for which certain configurations yield index two systems. This result is of great interest from the computational point of view, since index two DAEs are known to be more involved and to pose more difficulties than lower index problems, specially when they do not admit a Hessenberg form. This characterization is extended to branch-oriented and hybrid models of non-passive circuits in [68, 69]. The present section summarizes these results.

3.1 On the Index Notion

The Tractability Index Most papers on index analysis for DAE circuit models are based on the *tractability index* notion (cf. [75, 107, 112, 114, 156, 194]). The characterization of the tractability index of a given DAE model paves the way for an appropriate numerical treatment in simulation and, from an analytical standpoint, reduces the description of the dynamical behaviour to that of an inherent explicit ODE; this is performed by means of a decoupling of the different solution components [114, 156, 194]. We present below a brief introduction to this index concept. The discussion will be restricted to cases with index not greater than two; this way we avoid certain technical difficulties arising in problems with arbitrary index. Note that the index of DAEs modelling a very large class of electrical and electronic circuits does not exceed two [61, 149, 156, 193].

Consider the quasilinear DAE (2.14), and assume that the kernel of $A(u)$ is constant. This holds in particular for (2.15a), (2.15b) if $M(x)$ is nonsingular (a condition whose characterization is itself of interest for models such as (2.7a)–(2.7c) or (2.11a)–(2.11d); note that for other models this requirement just relies on the nonsingularity of the capacitance and inductance matrices $C(v_c)$, $L(i_l)$). Assume, as happens in all the circuit models above, that the right-hand side $F(u,t)$ of (2.14) can be written as $F_1(u) + F_2(t)$. Let $B(u)$ stand for the matrix of partial derivatives $-F_u(u,t)$; note that B does not depend on t because of the splitting of F. Denoting by Q a constant projector onto $\ker A(u)$ (so that $Q^2 = Q$ with $\operatorname{im} Q = \ker A(u)$), the DAE (2.14) has tractability index one if the matrix $A_1(u) = A(u) + B(u)Q$ is nonsingular.

For the DAE (2.15a), (2.15b), provided that $M(x)$ is nonsingular, this index one notion can be checked to amount to the nonsingularity of the matrix of partial derivatives g_y (this index one notion being the same in other index concepts, cf. [24, 89, 90, 106, 140]). In this situation, a straightforward application of the implicit function theorem makes it possible it describe the local system dynamics in the form $M(x)x' = f(x, \varphi(x,t), t)$, where $y = \varphi(x,t)$ comes from (2.15b).

The notion of an index two DAE is more cumbersome and the different index notions make a difference in this regard (cf. [24, 75, 89, 90, 106, 107, 140, 156]). In the tractability index framework, consider a setting in which $A_1(u)$ is rank-deficient everywhere, in such a way that there exists a continuous projector $Q_1(u)$ onto $\ker A_1(u)$ (forcing $A_1(u)$ to have constant rank). Let $B_1(u)$ stand for the product $B(u)(I - Q)$. Basing on the special form of the circuit equations (cf. [194, Remark A.18]), the DAE (2.14) will be said to have tractability index two if $A_2(u) = A_1(u) + B_1(u)Q_1(u)$ is nonsingular. This definition of the index is simpler than the one for general nonlinear DAEs [114, 194].

These notions are a bit simpler for linear time-invariant DAEs, a context in which the analysis can be performed in terms of the associated *matrix pencil*; cf. [67]. For time-varying and/or linearly implicit problems the relation with matrix pencil theory is more involved [24, 140, 156]. In particular, the relation between the tractability index of (2.14) and the Weierstraß–Kronecker index of the matrix pencil $\{A(u^*),\ B(u^*)\}$ arising from the linearized problem were thoroughly examined in [75, 76, 111] (recent related results can be found in [113, 114]). In particular, a matrix pencil $\{A,\ B\}$ with singular A is regular with Weierstraß–Kronecker index one if and only if the matrix $A_1 = A + BQ$ is nonsingular, Q being any projector onto $\ker A$. Additionally, if A_1 is singular, the pencil can be shown to be regular with Weierstraß–Kronecker index two if and only if $A_2 = A_1 + B_1Q_1$ is nonsingular, where Q_1 is any projector onto $\ker A_1$ and $B_1 = B(I - Q)$.

Other Index Notions Other index concepts, including the differentiation index and the geometric index, have also been used in the analysis of DAE circuit models (cf. [61, 156]). Broadly speaking, the idea supporting the *differentiation index* is to compute the number of differentiations needed to recast a DAE as an explicit ODE; more precisely, the constraints are differentiated in order to realize an explicit underlying ODE for which the solution set of the DAE is an invariant manifold. Find a detailed discussion in the book [24].

In contrast, the *geometric index* notion and the reduction methods supported on it describe the behaviour of a given DAE in terms of a vector field defined on the so-called solution manifold. Using local parametrizations, this vector field locally leads to a reduced ODE on an open subset of $\mathbb{R}^r$. In the original problem coordinates, the reduction process can be roughly described as the elimination of certain variables by solving the constraints. The solution manifold and the reduced ODE are computed in an iterative manner, and the number of iteration steps needed for the algorithm to stabilize defines the index. A detailed introduction to the geometric index framework can be found in [136, 137, 140–142]. See also [152].

When applied in particular to a linear time-invariant DAE with a regular matrix pencil, all these notions amount to the Weierstraß–Kronecker index of the pencil [67].

Solvability Solvability is an important concept in DAE theory (see e.g. [24, 107]). Broadly speaking, when an index is well-defined (at least in a local sense), then a (local) flow is defined on a lower-dimensional solution set which accommodates the solutions of the DAE. This is sometimes known as the set of consistent initial values, for which a unique solution is well-defined. Find details in the above-mentioned references and in [89, 106, 140, 156]. Things in this regard are different (and more complicated) in so-called singular DAEs, since solutions may bifurcate or even collapse in finite time at singular points; cf. Sect. 5.2 below and [36, 37, 135, 138–140, 148, 151, 153, 155, 156, 170].

In the circuit context, any index characterization guarantees the solvability of the circuit equations; that is, provided that an index is well-defined and that an initial point satisfying the constraints (including the possibly hidden ones) is given, then a unique solution emanates from that point. From this point of view, the index results reported in the remainder of this section can be understood to comprise implicitly the solvability of the circuit equations. Notably, this requires a well-posedness assumption on the circuit; this means that neither loops of voltage sources nor cutsets of current sources are present. This is a standard working hypothesis, which is sometimes assumed in circuit analysis without explicit mention.

3.2 Nodal Models

3.2.1 Passive Problems

We present in Theorem 3.1 below a slightly modified form of the topological characterizations of index one and index two configurations in MNA discussed in [193], accommodating also the index zero analysis of [194]. With respect to [193, 194], the statement below relaxes some passivity requirements, since the original statements in the aforementioned references assume the positive definiteness of the incremental conductance, capacitance and inductance matrices.

Theorem 3.1 *Consider a well-posed, connected circuit with nonsingular inductance matrix L and positive definite capacitance matrix C.*

(1) *The MNA system* (2.7a)–(2.7c) *is index zero if and only if*
   (a) *there are no voltage sources*; *and*
   (b) *there exists a capacitive tree.*
   Assume in the sequel that at least one of the conditions (a) *or* (b) *fails.*
(2) *Suppose additionally that the conductance matrix G is positive definite. Then the MNA model* (2.7a)–(2.7c) *has tractability index one if and only if the network contains neither VC-loops* (*except for C-loops*) *nor IL-cutsets.*
(3) *Let also L be positive definite. If the network contains VC-loops* (*with at least one voltage source*) *and/or IL-cutsets, then* (2.7a)–(2.7c) *has tractability index two.*

In order to avoid cumbersome computations, the reader is referred to [193, 194] for detailed proofs of these claims; the proofs there apply identically in this slightly broader setting. However, the reader can have a glimpse of some of the main ideas involved in these index analyses by considering the (simpler) index one case for the ANA model (2.6a)–(2.6e). Indeed, provided that C and L are nonsingular, the index one condition for (2.6a)–(2.6e) relies on the nonsingularity of

$$J = \begin{pmatrix} A_r G A_r^T & A_c & A_u \\ A_c^T & 0 & 0 \\ A_u^T & 0 & 0 \end{pmatrix}. \tag{3.1}$$

If the conductance matrix $G = \gamma_g'$ is positive definite, and the circuit displays neither VC-loops nor IL-cutsets, then J is actually nonsingular (and therefore the model (2.6a)–(2.6e) is index one). Note that J in (3.1) is nonsingular if and only if the unique solution to

$$A_r G A_r^T x + A_c y + A_u z = 0, \tag{3.2a}$$

$$A_c^T x = 0, \tag{3.2b}$$

$$A_u^T x = 0 \tag{3.2c}$$

is the trivial one. We show below that a non-trivial solution to (3.2a)–(3.2c) is in contradiction with the absence of VC-loops and IL-cutsets. Premultiplying (3.2a) by x^T, and using (3.2b) and (3.2c) to derive $x^T A_r G A_r^T x = 0$, one gets

$$A_r^T x = 0, \tag{3.3}$$

because of the positive definiteness assumption on G. The non-vanishing of x would indicate the presence of an IL-cutset, in the light of (3.2b), (3.2c), (3.3) and according to Lemma 2.2. In the absence of these configurations, the vanishing of x reduces (3.2a) to

$$A_c y + A_u z = 0, \tag{3.4}$$

but from Lemma 2.1, the non-vanishing of (y, z) necessarily describes the existence of a VC-loop.

The index two case, for both ANA and MNA, is more complicated, and as mentioned above details can be found in [156, 193, 194]. It is also worth indicating that these results can be extended to include a broad family of controlled sources; cf. [61, 156] in this regard. Finally, as detailed in [156], a similar characterization also holds for NTA models.

3.2.2 Low Index Configurations in the Non-passive Context

Many devices in electrical and electronic circuit theory do not meet the passivity requirements arising in Theorem 3.1. Examples range from the locally active resistors in Van der Pol or Chua circuits to tunnel diodes or Josephson junctions (see e.g. [44, 187]). For this reason it is of interest to extend the index characterization of DAE models to the non-passive context, a task which can be accomplished by means of tree-based methods. In this direction, the results compiled below can be found in [55]; the techniques used in the proof of these properties are based on the use of determinantal expansions and the Cauchy–Binet formula [95], extending the analysis of nodal admittance matrices originally performed by Maxwell. A *proper tree* is a spanning tree which includes all voltage sources and capacitors, and neither current sources nor inductors; in this context a *normal tree* is a spanning tree which contains all voltage sources, no current sources, as many capacitors as possible and as few inductors as possible [25, 29, 30, 105]. Note that the notion of a *normal reference tree* used in Sect. 2.4, which makes a difference between voltage- and current-controlled resistors, is an extension of this one.

Theorem 3.2 *Consider a well-posed, connected circuit with nonsingular inductance matrix L and no coupling among capacitors.*

(1) *The MNA system* (2.7a)–(2.7c) *is index zero if and only if*
   (a) *there are no voltage sources*;
   (b) *there exists at least one capacitive spanning tree*; *and*
   (c) *the sum of products* $\sum_{T\in\mathcal{T}_c}\prod_{C_i\in T} C_i$, *extended over the set* $\mathcal{T}_c$ *of capacitive spanning trees in the circuit, does not vanish.*

   Assume below that at least one of the conditions (a) *or* (b) *fails.*

(2) *Suppose that the sum in item* (c) *above does not vanish, and that there is no coupling among resistors. Then the MNA model* (2.7a)–(2.7c) *has tractability index one if and only if either there exists a proper V-tree or*
   (d) *there are neither VC-loops* (*except for C-loops*) *nor IL-cutsets*; *and*
   (e) *the sum of conductance products* $\sum_{T\in\mathcal{T}_n}\prod_{G_i\in T} G_i$, *extended over the set* $\mathcal{T}_n$ *of normal trees, does not vanish.*

A similar characterization of index one configurations is also discussed for the ANA model (2.6a)–(2.6e) in [55]. As in the passive context, the simpler structure of Augmented Nodal Analysis models makes it easier to give some hints on the proof

of this result (note additionally that ANA models are never index zero). Indeed, the nonsingularity of the matrix J in (3.1), without a positive definite assumption on G, can be addressed via the factorization

$$J = \begin{pmatrix} A_r & A_{cu} & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} G & 0 & 0 \\ 0 & 0 & I \\ 0 & I & 0 \end{pmatrix} \begin{pmatrix} A_r^T & 0 \\ A_{cu}^T & 0 \\ 0 & I \end{pmatrix}, \tag{3.5}$$

where A_{cu} joins together the columns of A_c and A_u. The Cauchy–Binet formula [95] allows one to express the determinant of J as the sum of determinantal products of maximal, nonsingular square submatrices. These necessarily correspond to spanning trees including all voltage sources and capacitors and neither current sources nor inductors; the determinants of the submatrices coming from the second factor of (3.5) then yield the conductance products referred to above. Find details in [55]. In contrast, the description of index two nodal models in non-passive settings remains open.

3.3 Branch-Oriented and Hybrid Models

3.3.1 Branch-Oriented Models

The index of branch-oriented circuit models in a passive context parallelizes the MNA case, except for the fact that C-loops lead to index two DAEs, as happens for ANA models. Find in [149] an analysis of the tractability index of these models, and in [156] a characterization of the geometric index (cf. Sect. 3.1). Non-passive circuits are addressed in [68], where both index one and index two configurations are fully characterized for circuits including both voltage-controlled and current-controlled resistors. Details on the index analysis of branch-oriented models can be found in the above-mentioned references; the remainder of this section is focussed on the hybrid model (2.11a)–(2.11d).

3.3.2 Hybrid Models of Passive Circuits

As detailed in [69, 100, 101, 190], a key advantage of hybrid circuit models is the fact that their index is not greater than one under very mild assumptions. Broadly speaking, this is a consequence of the elimination of, say, higher index variables in the formulation of the model.

The analysis carried out in [100, 101, 190] shows that under the assumption that the circuit is strictly locally passive, the hybrid model (2.11a)–(2.11d) is either index zero or one. For it to be index zero, an obvious requirement is the absence of the algebraic restrictions (2.11c)–(2.11d). This will happen if all voltage-controlled resistors are located in the cotree and every current-controlled one is in the tree. Recall that the construction of the hybrid model is based on choosing a tree with as

many twig capacitors and (subsequently) as many twig voltage-controlled resistors as possible, and as many link inductors and (subsequently) as many link current-controlled resistors as possible. This means that the algebraic restrictions are absent if and only if there is no chance to have voltage-controlled resistors in the tree or current-controlled ones in the cotree. This is equivalent to requiring the so-called *resistor-acyclic condition*, according to which (in a circuit without sources) every voltage-controlled resistor defines a loop together with some capacitors, and every current-controlled resistor defines a cutset together with some inductors. When the resistor-acyclic condition is not met, the hybrid model (2.11a)–(2.11d) of a strictly locally passive circuit is index one; find details in the references [100, 101, 190] mentioned above.

3.3.3 Hybrid Models of Non-passive Circuits

The index characterization of hybrid models of non-passive, uncoupled circuits is addressed in [69]. As in the nodal case, such a characterization is obtained in terms of certain trees. We compile below the main results in this regard, which make use of certain circuit minors; allowing for the presence of sources, the minor $\mathcal{G}_1$ is the one obtained after short-circuiting voltage sources and open-circuiting all other devices (except for capacitors); $\mathcal{G}_{23}$ is the minor which results from short-circuiting voltage sources and capacitors, and open-circuiting current sources and inductors. Finally, the minor $\mathcal{G}_4$ is obtained after short-circuiting all elements except for inductors and current sources, and open-circuiting current sources. In the statement of Theorems 3.3 and 3.4 below, note that a forest is implicitly understood to be a spanning forest.

Theorem 3.3 *Assume that a given circuit does not display capacitive or inductive coupling. Then the hybrid model* (2.11a)–(2.11d) *is index zero if and only if the resistor-acyclic condition is met, and neither the sum of capacitance products in the forests of* $\mathcal{G}_1$ *nor the sum of inductance products in the coforests of* $\mathcal{G}_4$ *do vanish.*

The idea supporting the result above is to factorize the blocks $C_{\text{tr}}(v_{c_{\text{tr}}}) + K_{11}^T C_{\text{co}}(-K_{11} v_{c_{\text{tr}}}) K_{11}$ and $L_{\text{co}}(i_{l_{\text{co}}}) + K_{44} L_{\text{tr}}(K_{44}^T i_{l_{\text{co}}}) K_{44}^T$ within the matrix M depicted in (2.16) as

$$\begin{pmatrix} I & -K_{11}^T \end{pmatrix} \begin{pmatrix} C_{\text{tr}} & 0 \\ 0 & C_{\text{co}} \end{pmatrix} \begin{pmatrix} I \\ -K_{11} \end{pmatrix}$$

and

$$\begin{pmatrix} I & -K_{44} \end{pmatrix} \begin{pmatrix} L_{\text{co}} & 0 \\ 0 & L_{\text{tr}} \end{pmatrix} \begin{pmatrix} I \\ -K_{44}^T \end{pmatrix},$$

respectively, in order to perform a Cauchy–Binet expansion. Find details in [69].

Under very broad assumptions, the failing of the resistor-acyclic condition leads to an index one model, as stated below.

Theorem 3.4 *Assume that a given uncoupled circuit does not meet the resistor-acyclic condition, and that the sums of capacitance and inductance products arising in Theorem 3.3 do not vanish.*

Then the hybrid model (2.11a)–(2.11d) is index one if and only if the sum of products of the conductances of voltage-controlled twig resistors and the resistances of current-controlled link resistors, extended over the forests of $\mathcal{G}_{23}$, does not vanish.

Again, this is a consequence of the Cauchy–Binet formula, but in this setting the analysis is more involved. The reader is referred to [69] for a detailed proof of this result.

3.4 Example

Let us go back to the circuit depicted in Fig. 1. We will use this example to show the form that the main models discussed in Sect. 2 (namely, MNA and hybrid ones) take, and also to illustrate the index results reported above.

MNA Denoting by e_1, e_2 and e_3 the potentials at the nodes on top of the voltage source, the capacitor and the diode, respectively, the MNA model (2.7a)–(2.7c) for the circuit in Fig. 1 can be checked to read

$$Ce_2' = i_{l_1} - i_s(t), \tag{3.6a}$$

$$L_1 i_{l_1}' = e_1 - e_2, \tag{3.6b}$$

$$L_2 i_{l_2}' = e_3, \tag{3.6c}$$

$$0 = i_u - i_{l_1}, \tag{3.6d}$$

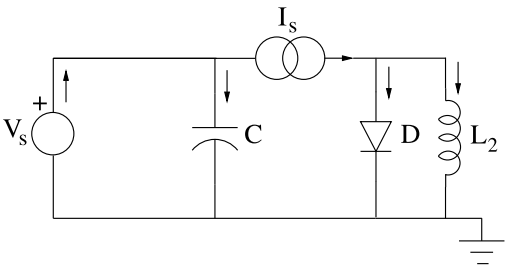
$$0 = -\gamma_g(e_3) - i_{l_2} + i_s(t), \tag{3.6e}$$

$$0 = v_s(t) - e_1. \tag{3.6f}$$

Provided that neither the capacitance nor the inductances vanish, this model is easily checked to be index one if and only if the incremental conductance at the diode, that is, $G(e_3) = \gamma_g'(e_3)$, is not zero. In particular, this is (obviously) the case if the diode is strictly locally passive, that is, if G is positive. This index one configuration reflects the fact that the circuit has neither VC-loops nor IL-cutsets; cf. Theorem 3.1. The condition $G \neq 0$ can be understood to express the fact that open-circuiting the diode would result in an IL-cutset.

This simple example also offers a glance at the index analysis in the non-passive context reported in Sect. 3.2.2. Note that G being positive is not a mandatory requirement for (3.6a)–(3.6f) to be index one. The accurate requirement is that G does not vanish, and this follows from the fact that the unique normal tree is the one depicted in Fig. 2. The sum of conductance products arising in Theorem 3.2 amounts in this case to the single conductance G, and therefore the condition $G \neq 0$

Fig. 3 Short-circuiting L_1



fully characterizes index one configurations in this model. Other examples in this direction can be found in [55, 157].

Let us analyze what happens if L_1 vanishes, that is, if we short-circuit the first inductor; cf. Fig. 3. Now a VC-loop shows up and in light of Theorem 3.1 this should make the MNA model index two. Both e_1 and i_{l_1} disappear from the model, which in this setting takes the form

$$Ce_2' = i_u - i_s(t), \tag{3.7a}$$

$$L_2 i_{l_2}' = e_3, \tag{3.7b}$$

$$0 = -\gamma_g(e_3) - i_{l_2} + i_s(t), \tag{3.7c}$$

$$0 = v_s(t) - e_2. \tag{3.7d}$$

Omitting technical details, it can be checked that the VC-loop indeed makes this model index two. This follows from the fact that the algebraic variable i_u does not enter equations (3.7c)–(3.7d), in contrast to its appearance in (3.6d).

Hybrid Analysis Using the tree depicted in Fig. 2, hybrid equations for the circuit in Fig. 1 read

$$Cv_c' = i_{l_1} - i_s(t), \tag{3.8a}$$

$$L_1 i_{l_1}' = v_s(t) - v_c, \tag{3.8b}$$

$$L_2 i_{l_2}' = v_d, \tag{3.8c}$$

$$0 = -\gamma_g(v_d) - i_{l_2} + i_s(t). \tag{3.8d}$$

Note that this system is formulated in terms of tree capacitor voltages (which amount in this case to v_c), link inductor currents (i_{l_1}, i_{l_2}) and the voltages of tree voltage-controlled resistors (that is, v_d). The diode prevents the resistor-acyclic condition from being satisfied and, as in MNA, the index is one provided that the incremental conductance $G(v_d) = \gamma_g'(v_d)$ does not vanish.

A difference with nodal analysis is made in the case $L_1 = 0$ which, as indicated above, leads to an index two system in MNA. Within the hybrid approach, in this situation the capacitor must be driven to the cotree (since it is now in parallel with the voltage source; cf. Fig. 3) and therefore both (3.8a) and (3.8b) disappear from the

model, so that the resulting equations are still index one. A more detailed discussion of the computational advantages which result from the low index configurations arising in hybrid analysis can be found in [100, 101, 190].

4 Memristors and Mem-Devices

The report in 2008 of a nanometre-scale device displaying a memristive characteristic [184] has had a great impact in the electrical and electronic engineering communities; cf. [11, 17, 45, 51, 70, 97–99, 102, 103, 121–124, 131–133, 157, 159–161, 163, 167, 203] and references therein. The existence of the memory-resistor or *memristor* was already postulated by Leon Chua in 1971 [34], and the actual appearance of memristors in nanoscale electronics has raised a renewed interest in these devices. The memristor, which is defined by a nonlinear charge-flux characteristic, is considered as the fourth basic circuit element, besides the resistor, the inductor and the capacitor which relate the voltage–current, current-flux and voltage–charge pairs, respectively. This device is likely to play a relevant role in electronics in the near future, especially at the nanometre scale. Many applications are already reported, e.g. in pattern recognition, design of associative memories, signal processing, adaptive systems, etc. (see [11, 45, 51, 97–99, 123, 132, 133]). HP has announced the release by 2013 of commercial memory chips based on the memristor [1]. The notion of a device with memory was extended to the reactive setting by Di Ventra, Pershin and Chua [51] to define *memcapacitors* and *meminductors*.

4.1 Memristors

The characteristic of a memristor may have either a charge-controlled or a flux-controlled form. In a charge-controlled setting, the characteristic reads

$$\varphi = \phi(q), \tag{4.1}$$

for some C^1 map ϕ. The incremental *memristance* is

$$M(q) = \phi'(q).$$

Using the relations $\varphi' = v$, $q' = i$ we get the voltage–current characteristic

$$v = M(q)i. \tag{4.2}$$

This relation shows that the device behaves as a resistor in which the resistance depends on $q(t) = \int_{-\infty}^{t} i(\tau)\,d\tau$, hence the *memory-resistor* name. This is the key feature of the device. In greater generality, one may consider (4.2) as a particular case of a fully nonlinear characteristic of the form

$$v = \eta(q, i),$$

as proposed in [160]. We refer the reader to [40] for a discussion of the more general family of *memristive systems*.

In turn, a flux-controlled memristor has a characteristic of the form

$$q = \xi(\varphi), \tag{4.3}$$

and the incremental *memductance* is

$$W(\varphi) = \xi'(\varphi).$$

The voltage-current relation has in this case the form

$$i = W(\varphi)v \tag{4.4}$$

or, in a fully nonlinear context,

$$i = \zeta(\varphi, v).$$

A memristor governed by (4.2) or (4.4) is said to be *strictly locally passive* if $M(q) > 0$ or $W(\varphi) > 0$ for all q or φ, respectively. In the presence of coupling effects (if eventually displayed), this requirement must be restated by asking the memristance or memductance matrices to be positive definite.

4.2 Memcapacitors, Meminductors and Higher Order Devices

Di Ventra, Pershin and Chua extended in [51] the idea of a device with memory to reactive elements. A (voltage-controlled) *memcapacitor* has a characteristic of the form

$$q = C_{\mathrm{m}}(\varphi)v. \tag{4.5}$$

Here C_{m} is the *memcapacitance*. The distinct feature of this device is that the memcapacitance depends on the state variable $\varphi(t) = \int_{-\infty}^{t} v(\tau)\,d\tau$, so that the relation $q(t) = C_{\mathrm{m}}(\int_{-\infty}^{t} v(\tau)\,d\tau)v(t)$ reflects the device history. Analogously, a (current-controlled) *meminductor* is governed by

$$\varphi = L_{\mathrm{m}}(q)i, \tag{4.6}$$

and $L_{\mathrm{m}}(q)$ is the *meminductance*, which reflects the device history via the variable q.

Note that both (4.5) and (4.6) come from differentiating a two-variable relation, namely $\sigma = \alpha(\varphi)$ for voltage-controlled memcapacitors and $\rho = \beta(q)$ for current-controlled meminductors; here σ and ρ arise as the time-integrals of q and φ, respectively. By using the differentiated relations (4.5) and (4.6) we get rid of these second order variables. This is not the case for so-called *second-order devices*, for

which either σ or ρ appear explicitly in the memcapacitance or the meminductance. Specifically, a *charge-controlled memcapacitor* is a device defined by the relations

$$\sigma' = q, \tag{4.7a}$$
$$q' = i, \tag{4.7b}$$
$$v = C^{-1}(\sigma)q, \tag{4.7c}$$

whereas a *flux-controlled meminductor* is characterized by

$$\rho' = \varphi, \tag{4.8a}$$
$$\varphi' = v, \tag{4.8b}$$
$$i = L^{-1}(\rho)\varphi. \tag{4.8c}$$

We refer the reader to [17, 51] for an introduction to these and other related devices.

4.3 DAE Models of Circuits with Mem-Devices

The nodal, branch-oriented and hybrid models discussed in Sect. 2 can be easily extended to accommodate memristors. In particular, the ANA model of a circuit including charge-controlled memristors can be written as (cf. [167])

$$C(v_c)v_c' = i_c, \tag{4.9a}$$
$$L(i_l)i_l' = A_l^T e, \tag{4.9b}$$
$$q_m' = i_m, \tag{4.9c}$$
$$0 = A_r\gamma_g\left(A_r^T e\right) + A_l i_l + A_c i_c + A_m i_m + A_u i_u + A_j i_s(t), \tag{4.9d}$$
$$0 = v_c - A_c^T e, \tag{4.9e}$$
$$0 = v_s(t) - A_u^T e, \tag{4.9f}$$
$$0 = M(q_m)i_m - A_m^T e \tag{4.9g}$$

or, under a flux-control assumption on memristors, as

$$C(v_c)v_c' = i_c, \tag{4.10a}$$
$$L(i_l)i_l' = A_l^T e, \tag{4.10b}$$
$$\varphi_m' = A_m^T e, \tag{4.10c}$$
$$0 = A_r\gamma_g\left(A_r^T e\right) + A_l i_l + A_c i_c + A_m i_m + A_u i_u + A_j i_s(t), \tag{4.10d}$$
$$0 = v_c - A_c^T e, \tag{4.10e}$$

$$0 = v_s(t) - A_u^T e, \tag{4.10f}$$

$$0 = i_m - W(\varphi_m) A_m^T e. \tag{4.10g}$$

Branch-oriented models of circuits with charge-controlled memristors and current-controlled resistors are discussed in [70]. These models can be written in the form

$$C(v_c)v_c' = i_c, \tag{4.11a}$$

$$L(i_l)i_l' = v_l, \tag{4.11b}$$

$$q_m' = i_m, \tag{4.11c}$$

$$0 = v_m - M(q_m)i_m, \tag{4.11d}$$

$$0 = v_r - \gamma_r(i_r), \tag{4.11e}$$

$$0 = A_c i_c + A_l i_l + A_m i_m + A_r i_r + A_u i_u + A_j i_s(t), \tag{4.11f}$$

$$0 = B_c v_c + B_l v_l + B_m v_m + B_r v_r + B_u v_s(t) + B_j v_j. \tag{4.11g}$$

Kirchhoff's current law (4.11f) can be written, alternatively, in terms of a reduced cutset matrix Q. Flux-controlled memristors and voltage-controlled resistors can be easily accommodated in this framework.

Hybrid models can be also extended to accommodate both charge-controlled and flux-controlled memristors; find details in [69], where a full index characterization is carried out without passivity assumptions. Previous index results directed to the above-mentioned nodal and branch-oriented models in a strictly locally passive setting can be found in [70, 167]. In contrast, a general characterization of the index of circuits with other mem-devices remains open.

5 Dynamical Aspects

5.1 The State Formulation Problem

The differential-algebraic formalism makes it possible to revisit the problem of formulating a state space model a given nonlinear circuit. The problem is how to obtain an explicit ODE modelling the dynamics, either in a local or a global sense. Note that this is a theoretical problem which is important in the study of different analytical features of nonlinear circuits (concerning e.g. stability aspects, oscillations, bifurcations or chaotic phenomena), although from the computational point of view it is often preferred to simulate the circuit behaviour using directly a differential-algebraic model and appropriate numerical tools. In the state formulation problem the goal is often to formulate state space models not for individual problems but for general circuit families, in terms of their topology. In our framework, the state formulation problem can be naturally addressed as a *reduction* of a semistate model.

This reduction can be discussed in terms of different DAE models, although the best-suited ones in this regard are the branch-oriented and hybrid models of Sect. 2.4. Note e.g. that the problem can be stated as the elimination of resistive variables and i_u, v_j, i_c, v_l in the branch-oriented model (2.8a)–(2.8e), or just i_c, v_l in the multiport model (2.12a)–(2.12c); it is worth noticing that this model can be seen as an intermediate step between the branch-oriented one and a state space model formulated in terms of v_c, i_l. In index two settings, cotree capacitor voltages and tree inductor currents must also be removed; this can be alternatively performed by eliminating $v_{g_{\text{tr}}}$ and $i_{r_{\text{co}}}$ in the hybrid model (2.11a)–(2.11d).

The state formulation problem as a reduction of semistate models is discussed in detail in [156] and, for memristive circuits, in [159]. For the sake of completeness, it is important to mention that the state space formulation problem has been also tackled without using a differential-algebraic formalism, stemming from the original work of Bashkow and Bryant [15, 29, 30]: some recent results in this direction can be found in [110, 125, 150, 177–180].

5.2 Singularities and Impasse Phenomena

Some readers might conjecture at this point that, since DAEs may be eventually reduced to an explicit ODE on a manifold, no new local dynamic phenomena should be expected in the differential-algebraic setting. This should apply in particular to the local dynamics of nonlinear circuits. This point of view, which essentially looks at DAEs just as a modelling tool, is incomplete. Besides the fact that in practice this reduction may not be feasible, even from a theoretical point of view such a reduction to an explicit ODE is possible only when the index is well-defined, but this is not the case in the presence of so-called *singularities*.

It is interesting to note that research on singular DAEs was motivated to a large extent by certain phenomena observed in circuit theory, namely, the so-called *jump phenomenon* (cf. [172] and references therein). This was formalized as *impasse phenomena* roughly at the same time by Chua and Deng [36, 37] and Rabier [135]. An impasse point is defined as a point where a pair of trajectories collapse in (either backward or forward) finite time with infinite speed; more precisely, denoting by Ω the semistate space of the DAE, x^* is a *forward impasse point* if there exists a $\delta > 0$ and two distinct solutions in $C^1((-\delta, 0), \Omega) \cap C^0((-\delta, 0], \Omega)$ with $x(0) = x^*$, whose derivatives blow up at $t = 0$. A *backward impasse point* is defined analogously, just requiring the solutions to be defined in $C^1((0, \delta), \Omega) \cap C^0([0, \delta), \Omega)$.

Noteworthy, Chua and Deng addressed the problem in the setting of semiexplicit DAEs, whereas Rabier tackled it for quasilinear ODEs amounting to an explicit equation except on a hypersurface of singular points. Only later the connection between both settings would be made, once it was realized that quasilinear ODEs with a singular hypersurface provide the natural reduction of DAEs with arbitrary index near singular points (cf. [138–140, 156]), much as explicit ODEs arise as the reduction of DAEs near points with a well-defined index. Singular DAEs in the circuit context have been specifically discussed in [148, 155, 158]. Normal forms and closely related topics are addressed in [151, 191].

5.3 Qualitative Properties in the Semistate Context

Qualitative aspects play a key role in dynamical systems theory, and in particular in nonlinear circuit theory. The local or global behaviour of a system may often be characterized in terms of some of its invariants (equilibria, periodic solutions, chaotic attractors); their stability and their dependence on certain parameters are often the key ingredients to understand the system behaviour.

In the nonlinear circuit context, qualitative properties are often addressed via state space models; see e.g. [35, 38, 39, 63, 72–74, 92, 118, 120, 188, 189] and references therein. However, performing a state space reduction may be unnecessary from the qualitative point of view; the DAE formalism provides different tools which make it possible to analyze stability features and bifurcations in the semistate context. This is specially relevant in large scale integration circuits. In contrast to low scale circuits (even if they have rich dynamics) such as Van der Pol's or Chua's circuits, where modelling is not an issue, in large scale circuits it is often difficult to perform a state reduction, which is not actually necessary for a qualitative analysis.

Indeed, the linearization about equilibria can be characterized in the DAE context using matrix pencils; in this direction, find a characterization of different stability properties of equilibria of nonlinear circuits in the paper [165]. In particular, the role of so-called non-hyperbolic configurations (yielding purely imaginary eigenvalues and oscillations) were further studied in [166, 168]. In these papers, the role of certain configurations such as VCL-loops and ICL-cutsets were carefully examined, extending several properties involving VL-loops and IC-cutsets which were already known to yield null eigenvalues [59, 86, 118]. Beyond the linear time-invariant context, oscillations were analyzed using a DAE formulation in [46]. Different aspects concerning DC operating points are addressed in [72–74]. Several qualitative results based on Lyapunov function methods can be found in [38, 63, 92, 189], although most of these references do not make specific use of a DAE formalism. It is also worth mentioning that qualitative properties of nonlinear circuits can be studied via the geometric approach stemming from the work of Brayton and Moser [22, 23]; later results in this direction can be found in [50, 86, 87, 176, 200, 201].

Qualitative properties of circuits with memristors have been studied in recent years [97, 121–124, 160]. Peculiar dynamics arise from the presence of non-isolated equilibria, which had been observed in particular instances of memristive circuits e.g. in [121]. This issue has been recently addressed in broader generality [161], where graph-theoretic conditions guaranteeing the normal hyperbolicity of such manifolds of equilibria are provided. The failing of certain passivity assumptions along the manifold of equilibria motivates the analysis of certain bifurcations without parameters in memristive circuits; find details in [161]. Chaotic phenomena in memristive circuits have been studied by several authors; see e.g. [11, 122–124]. The differential-algebraic formalism seems to be promising regarding further qualitative analyses of circuits with mem-devices.

6 Other Topics in DAE-Based Circuit Modelling

More briefly, we compile in this section some references addressing other aspects of circuit theory in which DAE models play a role.

Model Reduction Model order reduction techniques aim at a substantial reduction of the number of variables involved in the description of a given dynamical system, while at the same time retaining the essential features of the original system. In circuit analysis, this is especially relevant when using distributed models; the space discretization of these yields an ordinary differential equation or a DAE with a very large number of state variables. Model reduction approaches in this context are discussed in [65, 66, 144–146, 185, 195, 199].

Coupled Problems Recent research has also been directed to *coupled problems*, in which electrical circuits interact with other systems of different nature. Electrical circuits with semiconductor devices are often modelled combining DAEs for the lumped elements and PDEs for the distributed semiconductor devices, leading to a PDAE formalism: see [2, 3, 14, 18, 77, 78, 119, 147, 175, 194] and references therein, as well as the survey on PDAEs within this volume. Semiconductors and electrical circuits including thermal effects are considered in [26]. In [28], a model accommodating electrical circuits, semiconductors and thermal networks modelling heat evolution in lumped elements is analyzed and simulated. A discussion of the coupling of electrical circuits with optoelectronic devices (semiconductor lasers) can be found in [27]. For field/circuit coupling the reader is referred to [13, 14, 174] and references therein.

Numerics in Circuit Simulation via DAE Models This survey is focussed on analytical aspects of DAE-based circuit modelling. Certainly, numerical simulation is of major importance in nonlinear circuit theory and we compile here some references in this direction, without any attempt at being exhaustive. With regard to the problem of consistent initialization in circuit simulation, cf. [16, 56–58, 60] and references therein. Numerical aspects of properly stated DAEs modelling nonlinear circuits are addressed in [107, 115, 194]. Multirate methods are discussed in [12, 174, 183, 192, 197]. For the use of dynamic iteration methods the reader is referred to [7, 14, 54] and the references compiled therein. Numerical aspects involving stochastic DAEs in circuit simulation are discussed in [130, 173, 202]. Other numerical aspects in circuit simulation can be found in [47, 80, 81, 116, 129].

Other Topics DAEs have been also used in connection to circuit synthesis of passive semistate systems [143]. Fault diagnosis problems are discussed using a DAE formalism in [52, 59, 182] and via a sensitivity analysis in [96]. Index reduction methods are addressed in [8, 9]. For a discussion of port-Hamiltonian formulations the reader is referred to [102, 134, 196] and to the corresponding survey in this volume.

7 Concluding Remarks

Differential-algebraic equations play nowadays a key role in nonlinear circuit modelling, specially because of the chance to set up automatically the circuit equations in semistate (differential-algebraic) form. In this survey we have presented a detailed introduction to the main families of DAE circuit models, emanating from different methods of circuit analysis; these include nodal analysis, as well as branch-oriented and hybrid modelling. We include a detailed compilation of index characterizations for different models, applying also in non-passive settings. Some results in this direction remain open; for instance, a full index-two characterization of nodal models is not yet known in a non-passive context. The models and the index analysis can be extended in a natural manner to circuits including memristors; in this regard, the index of circuits with reactive and higher-order mem-devices has not been addressed in the literature. We have discussed more briefly other analytical issues, regarding the dynamics of nonlinear circuits as well as other aspects related e.g. to model reduction, coupled problems or numerics. Finally, the reader can find several related results in the papers on PDAEs and on port-Hamiltonian systems within this volume.

Acknowledgements Research supported by Research Project MTM2010-15102 of Ministerio de Ciencia e Innovación, Spain.

References

1. Adee, S.: Memristor inside. IEEE Spectrum, Sept. (2010)
2. Alí, G., Bartel, A., Günther, M., Tischendorf, C.: Elliptic partial differential-algebraic multiphysics models in electrical network design. Math. Models Methods Appl. Sci. **13**, 1261–1278 (2003)
3. Alí, G., Bartel, A., Günther, M.: Parabolic differential-algebraic models in electrical network design. Multiscale Model. Simul. **4**, 813–838 (2005)
4. Amari, S.: Topological foundations of Kron's tearing of electrical networks. RAAG Mem. **3**, 322–350 (1962)
5. Andrásfai, B.: Introductory Graph Theory. Akadémiai Kiadó, Budapest (1977)
6. Andrásfai, B.: Graph Theory: Flows, Matrices. Adam Hilger, Bristol (1991)
7. Arnold, M., Günther, M.: Preconditioned dynamic iteration for coupled differential-algebraic systems. BIT Numer. Math. **41**, 1–25 (2001)
8. Bächle, S., Ebert, F.: A structure preserving index reduction method for MNA. Proc. Appl. Math. Mech. **6**, 727–728 (2006)
9. Bächle, S., Ebert, F.: Index reduction by element-replacement for electrical circuits. In: Proc. SCEE 2006, pp. 191–198. Springer, Berlin (2007)
10. Balabanian, N., Bickart, T.A.: Electrical Network Theory. Wiley, New York (1969)
11. Bao, B., Ma, Z., Xu, J., Liu, Z., Xu, Q.: A simple memristor chaotic circuit with complex dynamics. Int. J. Bifurc. Chaos **21**, 2629–2645 (2011)
12. Bartel, A., Günther, M.: A multirate W-method for electrical networks in state-space formulation. J. Comput. Appl. Math. **147**, 411–425 (2002)
13. Bartel, A., Baumanns, S., Schöps, S.: Structural analysis of electrical circuits including magnetoquasistatic devices. Appl. Numer. Math. **61**, 1257–1270 (2011)

14. Bartel, A., Brunk, M., Günther, M., Schöps, S.: Dynamic iteration for coupled problems of electric circuits and distributed devices. Preprint BUW-IMACM 11/16, University of Wuppertal (2011)
15. Bashkow, T.R.: The *A* matrix, new network description. IRE Trans. Circuit Theory **4**, 117–119 (1957)
16. Baumanns, S., Selva, M., Tischendorf, C.: Consistent initialization for coupled circuit-device simulation. In: Proc. SCEE 2008, pp. 297–304. Springer, Berlin (2010)
17. Biolek, D., Biolek, Z., Biolkova, V.: SPICE modeling of memristive, memcapacitive and meminductive systems. In: Proc. Eur. Conf. Circuit Theory and Design 2009, pp. 249–252 (2009)
18. Bodestedt, M., Tischendorf, C.: PDAE models of integrated circuits and index analysis. Math. Comput. Model. Dyn. Syst. **13**, 1–17 (2007)
19. Bollobás, B.: Modern Graph Theory. Springer, Berlin (1998)
20. Branin, F.H.: The relation between Kron's method and the classical methods of network analysis. Matrix Tensor Q. **12**, 69–115 (1962)
21. Branin, F.H.: Computer methods of network analysis. Proc. IEEE **55**, 1787–1801 (1967)
22. Brayton, R.K., Moser, J.K.: A theory of nonlinear networks, I. Q. Appl. Math. **22**, 1–33 (1964)
23. Brayton, R.K., Moser, J.K.: A theory of nonlinear networks, II. Q. Appl. Math. **22**, 81–104 (1964)
24. Brenan, K.E., Campbell, S.L., Petzold, L.R.: Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations. SIAM, Philadelphia (1996)
25. Brown, D.P.: Derivative-explicit differential equations for RLC graphs. J. Franklin Inst. **275**, 503–514 (1963)
26. Brunk, M., Jüngel, A.: Numerical coupling of electrical circuit equations and energy-transport models for semiconductors. SIAM J. Sci. Comput. **30**, 873–894 (2008)
27. Brunk, M., Jüngel, A.: Simulation of thermal effects in optoelectronic devices using coupled energy-transport and circuit models. Math. Models Methods Appl. Sci. **18**, 1–26 (2008)
28. Brunk, M., Jüngel, A.: Self-heating in a coupled thermo-electric circuit-device model. J. Comput. Electron. **10**, 163–178 (2011)
29. Bryant, P.R.: The order of complexity of electrical networks. IEE Proc. Part C **106**, 174–188 (1959)
30. Bryant, P.R.: The explicit form of Bashkow's A matrix. IRE Trans. Circuit Theory **9**, 303–306 (1962)
31. Campbell, S.L.: Singular Systems of Differential Equations. Pitman, New York (1980)
32. Campbell, S.L.: Singular Systems of Differential Equations II. Pitman, New York (1982)
33. Chen, W.K.: Net Theory and Its Applications. World Scientific, Singapore (2003)
34. Chua, L.O.: Memristor—the missing circuit element. IEEE Trans. Circuit Theory **18**, 507–519 (1971)
35. Chua, L.O.: Dynamic nonlinear networks: state-of-the-art. IEEE Trans. Circuits Syst. **27**, 1059–1087 (1980)
36. Chua, L.O., Deng, A.D.: Impasse points, I: numerical aspects. Int. J. Circuit Theory Appl. **17**, 213–235 (1989)
37. Chua, L.O., Deng, A.D.: Impasse points, II: analytical aspects. Int. J. Circuit Theory Appl. **17**, 271–282 (1989)
38. Chua, L.O., Green, D.N.: A qualitative analysis of the behavior of dynamic nonlinear networks: stability of autonomous networks. IEEE Trans. Circuits Syst. **23**, 355–379 (1976)
39. Chua, L.O., Green, D.N.: Graph-theoretic properties of dynamic nonlinear networks. IEEE Trans. Circuits Syst. **23**, 292–312 (1976)
40. Chua, L.O., Kang, S.M.: Memristive devices and systems. Proc. IEEE **64**, 209–223 (1976)
41. Chua, L.O., Lin, P.M.: Computer-Aided Analysis of Electronic Circuits: Algorithms and Computational Techniques. Prentice-Hall, New York (1975)
42. Chua, L.O., Oka, H.: Normal forms for constrained nonlinear differential equations, Part I: theory. IEEE Trans. Circuits Syst. **35**, 881–901 (1988)

43. Chua, L.O., Oka, H.: Normal forms for constrained nonlinear differential equations, Part II: bifurcation. IEEE Trans. Circuits Syst. **36**, 71–88 (1989)
44. Chua, L.O., Desoer, C.A., Kuh, E.S.: Linear and Nonlinear Circuits. McGraw-Hill, New York (1987)
45. Corinto, F., Ascoli, A., Gilli, M.: Analysis of current-voltage characteristics for memristive elements in pattern recognition systems. Int. J. Circuit Theory Appl. **40**, 1277–1320 (2012)
46. Demir, A.: Floquet theory and non-linear perturbation analysis for oscillators with differential-algebraic equations. Int. J. Circuit Theory Appl. **28**, 163–185 (2000)
47. Denk, G., Penski, C.: Integration schemes for highly oscillatory DAEs with applications to circuit simulation. J. Comput. Appl. Math. **82**, 19–91 (1997)
48. Dervisoglu, A.: Bashkow's A-matrix for active RLC networks. IEEE Trans. Circuit Theory **11**, 404–406 (1964)
49. Dervisoglu, A.: The realization of the A-matrix of a certain class of RLC networks. IEEE Trans. Circuit Theory **13**, 164–170 (1966)
50. Desoer, Ch.A., Wu, F.F.: Trajectories of nonlinear RLC networks: a geometric approach. IEEE Trans. Circuit Theory **19**, 562–571 (1972)
51. Di Ventra, M., Pershin, Y.V., Chua, L.O.: Circuit elements with memory: memristors, memcapacitors and meminductors. Proc. IEEE **97**, 1717–1724 (2009)
52. Domínguez-García, A.D., Trenn, S.: Detection of impulsive effects in switched DAEs with applications to power electronics reliability analysis. In: Proc. 49th IEEE Conf. Dec. Control, pp. 5662–5667 (2010)
53. Dziurla, B., Newcomb, R.: The Drazin inverse and semi-state equations. In: Proc. Intl. Symp. Math. Theory of Networks and Systems, pp. 283–289 (1979)
54. Ebert, F.: On partitioned simulation of electrical circuits using dynamic iteration methods. PhD Thesis, Technical University of Berlin (2008)
55. Encinas, A., Riaza, R.: Tree-based characterization of low index circuit configurations without passivity restrictions. Int. J. Circuit Theory Appl. **36**, 135–160 (2008)
56. Estévez-Schwarz, D.: Consistent initialization for index-2 differential algebraic equations and its application to circuit simulation. PhD Thesis, Humboldt University of Berlin (2000)
57. Estévez-Schwarz, D.: A step-by-step approach to compute a consistent initialization for the MNA. Int. J. Circuit Theory Appl. **30**, 1–16 (2002)
58. Estévez-Schwarz, D.: Consistent initialization for DAEs in Hessenberg form. Numer. Algorithms **52**, 629–648 (2009)
59. Estévez-Schwarz, D., Feldmann, U.: Actual problems of circuit simulation in industry. In: Modeling, Simulation, and Optimization of Integrated Circuits, Oberwolfach, 2001. Int. Ser. Numer. Math., vol. 146, pp. 83–99 (2003)
60. Estévez-Schwarz, D., Lamour, R.: The computation of consistent initial values for nonlinear index-2 differential-algebraic equations. Numer. Algorithms **26**, 49–75 (2001)
61. Estévez-Schwarz, D., Tischendorf, C.: Structural analysis of electric circuits and consequences for MNA. Int. J. Circuit Theory Appl. **28**, 131–162 (2000)
62. Estévez-Schwarz, D., Tischendorf, C.: Mathematical problems in circuit simulation. Math. Comput. Model. Dyn. Syst. **7**, 215–223 (2001)
63. Fosséprez, M.: Non-Linear Circuits: Qualitative Analysis of Non-linear, Non-reciprocal Circuits. Wiley, New York (1992)
64. Foulds, L.R.: Graph Theory Applications. Springer, Berlin (1992)
65. Freund, R.W.: Krylov-subspace methods for reduced-order modeling in circuit simulation. J. Comput. Appl. Math. **123**, 395–421 (2000)
66. Freund, R.W.: The SPRIM algorithm for structure-preserving order reduction of general RCL circuits. In: Model Reduction in Circuit Simulation. Lecture Notes in Electrical Engineering, vol. 74, pp. 25–52. Springer, Berlin (2011). Part I
67. Gantmacher, F.R.: The Theory of Matrices, vols. 1 & 2. Chelsea, New York (1959)
68. García de la Vega, I., Riaza, R.: Mixed determinantal expansions in circuit theory. Preprint (2012)

69. García de la Vega, I., Riaza, R.: Hybrid analysis of nonlinear circuits: DAE models with indices zero and one. Circuits Syst. Signal Process. (2013, in press)
70. García-Redondo, F., Riaza, R.: The tractability index of memristive circuits: branch-oriented and tree-based models. Math. Methods Appl. Sci. **35**, 1659–1699 (2012)
71. Gear, C.W.: The simultaneous numerical solution of differential-algebraic equations. IEEE Trans. Circuit Theory **18**, 89–95 (1971)
72. Green, M.M., Willson, A.N. Jr.: How to identify unstable dc operating points. IEEE Trans. Circuits Syst. I **39**, 820–832 (1992)
73. Green, M.M., Willson, A.N. Jr.: (Almost) half on any circuit's operating points are unstable. IEEE Trans. Circuits Syst. I **41**, 286–293 (1994)
74. Green, M.M., Willson, A.N. Jr.: An algorithm for identifying unstable operating points using SPICE. IEEE Trans. Comput.-Aided Des. Circuits Syst. **14**, 360–370 (1995)
75. Griepentrog, E., März, R.: Differential-Algebraic Equations and Their Numerical Treatment. Teubner-Texte zur Mathematik, vol. 88. Teubner, Leipzig (1986)
76. Griepentrog, E., März, R.: Basic properties of some differential-algebraic equations. Z. Anal. Anwend. **8**, 25–40 (1989)
77. Günther, M.: A joint DAE/PDE model for interconnected electrical networks. Math. Comput. Model. Dyn. Syst. **6**, 114–128 (2000)
78. Günther, M.: A PDAE model for interconnected linear RLC networks. Math. Comput. Model. Dyn. Syst. **7**, 189–203 (2001)
79. Günther, M., Feldmann, U.: The DAE-index in electric circuit simulation. Math. Comput. Simul. **39**, 573–582 (1995)
80. Günther, M., Feldmann, U.: CAD-based electric-circuit modeling in industry. I: mathematical structure and index of network equations. Surv. Math. Ind. **8**, 97–129 (1999)
81. Günther, M., Feldmann, U.: CAD-based electric-circuit modeling in industry. II: impact of circuit configurations and parameters. Surv. Math. Ind. **8**, 131–157 (1999)
82. Günther, M., Rentrop, P.: The differential-algebraic index concept in electric circuit simulation. Z. Angew. Math. Mech. **76**(S. 1), 91–94 (1996)
83. Günther, M., Rentrop, P.: Numerical simulation of electrical circuits. Mitt. Ges. Angew. Math. Mech. **23**, 51–77 (2000)
84. Günther, M., Feldmann, U., ter Maten, J.: Modelling and discretization of circuit problems. In: Handbook of Numerical Analysis, vol. 13, pp. 523–659. Elsevier, Amsterdam (2005)
85. Hachtel, G.D., Brayton, R.K., Gustafson, F.G.: The sparse tableau approach to network analysis and design. IEEE Trans. Circuits Syst. **18**, 101–113 (1971)
86. Haggman, B.C., Bryant, P.R.: Solutions of singular constrained differential equations: a generalization of circuits containing capacitor-only loops and inductor-only cutsets. IEEE Trans. Circuits Syst. **31**, 1015–1029 (1984)
87. Haggman, B.C., Bryant, P.R.: Geometric properties of nonlinear networks containing capacitor-only cutsets and/or inductor-only loops. Part I: conservation laws. Circuits Syst. Signal Process. **5**, 279–319 (1986)
88. Haggman, B.C., Bryant, P.R.: Geometric properties of nonlinear networks containing capacitor-only cutsets and/or inductor-only loops. Part II: symmetries. Circuits Syst. Signal Process. **5**, 435–448 (1986)
89. Hairer, E., Wanner, G.: Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems. Springer, Berlin (1996)
90. Hairer, E., Lubich, C., Roche, M.: The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods. Lect. Notes Maths., vol. 1409. Springer, Berlin (1989)
91. Harrison, B.K.: A discussion of some mathematical techniques used in Kron's method of tearing. J. Soc. Ind. Appl. Math. **11**, 258–280 (1963)
92. Hasler, M., Neirynck, J.: Nonlinear Circuits. Artech House, Norwood (1986)
93. Ho, C.W., Ruehli, A.E., Brennan, P.A.: The modified nodal approach to network analysis. IEEE Trans. Circuits Syst. **22**, 504–509 (1975)
94. Hodge, A.M., Newcomb, R.W.: Semistate theory and analog VLSI design. IEEE Circuits Syst. Mag. **2**, 30–51 (2002)

95. Horn, R.A., Johnson, Ch.R.: Matrix Analysis. Cambridge Univ. Press, Cambridge (1985)
96. Ilievski, Z., Xu, H., Verhoeven, A., ter Maten, E.J.W., Schilders, W.H.A., Mattheij, R.M.M.: Adjoint transient sensitivity analysis in circuit simulation. In: Proc. SCEE 2006, pp. 183–189. Springer, Berlin (2006)
97. Itoh, M., Chua, L.O.: Memristor oscillators. Int. J. Bifurc. Chaos **18**, 3183–3206 (2008)
98. Itoh, M., Chua, L.O.: Memristor cellular automata and memristor discrete-time cellular neural networks. Int. J. Bifurc. Chaos **19**, 3605–3656 (2009)
99. Itoh, M., Chua, L.O.: Memristor Hamiltonian circuits. Int. J. Bifurc. Chaos **21**, 2395–2425 (2011)
100. Iwata, S., Takamatsu, M.: Index minimization of differential-algebraic equations in hybrid analysis for circuit simulation. Math. Program., Ser. A **121**, 105–121 (2010)
101. Iwata, S., Takamatsu, M., Tischendorf, C.: Tractability index of hybrid equations for circuit simulation. Math. Comput. **81**, 923–939 (2012)
102. Jeltsema, D., van der Schaft, A.J.: Memristive port-Hamiltonian systems. Math. Comput. Model. Dyn. Syst. **16**, 75–93 (2010)
103. Kavehei, O., Iqbal, A., Kim, Y.S., Eshraghian, K., Al-Sarawi, S.F., Abbott, D.: The fourth element: characteristics, modelling and electromagnetic theory of the memristor. Proc. R. Soc. A **466**, 2175–2202 (2010)
104. Kron, G.: Tensor Analysis of Networks. Wiley, New York (1939)
105. Kuh, E.S., Rohrer, R.A.: The state-variable approach to network analysis. Proc. IEEE **53**, 672–686 (1965)
106. Kunkel, P., Mehrmann, V.: Differential-Algebraic Equations. Analysis and Numerical Solution. EMS, Zürich (2006)
107. Lamour, R., März, R., Tischendorf, C.: Differential-Algebraic Equations: A Projector Based Analysis. Springer, Berlin (2013)
108. Lewis, F.L. (ed.): Special Issue on Semistate Systems. Circuits Syst. Signal Process. **5**(1) (1986)
109. Lewis, F.L., Mertzios, V.G. (eds.): Special Issue on Recent Advances in Singular Systems. Circuits Syst. Signal Process. **8**(3) (1989)
110. Li, F., Woo, P.Y.: A new method for establishing state equations: the branch replacement and augmented node-voltage equation approach. Circuits Syst. Signal Process. **21**, 149–161 (2002)
111. März, R.: A matrix chain for analyzing differential algebraic equations. Preprint 162, Inst. Math., Humboldt University, Berlin (1987)
112. März, R.: Numerical methods for differential algebraic equations. Acta Numer. **1992**, 141–198 (1992)
113. März, R.: The index of linear differential algebraic equations with properly stated leading terms. Results Math. **42**, 308–338 (2002)
114. März, R.: Differential algebraic equations anew. Appl. Numer. Math. **42**, 315–335 (2002)
115. März, R.: Differential algebraic systems with properly stated leading term and MNA equations. In: Modeling, Simulation, and Optimization of Integrated Circuits, Oberwolfach, 2001. Int. Ser. Numer. Math., vol. 146, pp. 135–151 (2003)
116. März, R., Tischendorf, C.: Recent results in solving index-2 differential-algebraic equations in circuit simulation. SIAM J. Sci. Comput. **18**, 139–159 (1997)
117. März, R., Estévez-Schwarz, D., Feldmann, U., Sturtzel, S., Tischendorf, C.: Finding beneficial DAE structures in circuit simulation. In: Jäger, W., et al. (eds.) Mathematics—Key Technology for the Future, pp. 413–428. Springer, Berlin (2003)
118. Matsumoto, T., Chua, L.O., Makino, A.: On the implications of capacitor-only cutsets and inductor-only loops in nonlinear networks. IEEE Trans. Circuits Syst. **26**, 828–845 (1979)
119. Matthes, M., Tischendorf, C.: Convergence analysis of a partial differential algebraic system from coupling a semiconductor model to a circuit model. Appl. Numer. Math. **61**, 382–394 (2011)

120. Mees, A.I., Chua, L.O.: The Hopf bifurcation theorem and its applications to nonlinear oscillations in circuits and systems. IEEE Trans. Circuits Syst. **26**, 235–254 (1979)
121. Messias, M., Nespoli, C., Botta, V.A.: Hopf bifurcation from lines of equilibria without parameters in memristors oscillators. Int. J. Bifurc. Chaos **20**, 437–450 (2010)
122. Muthuswamy, B.: Implementing memristor based chaotic circuits. Int. J. Bifurc. Chaos **20**, 1335–1350 (2010)
123. Muthuswamy, B., Chua, L.O.: Simplest chaotic circuit. Int. J. Bifurc. Chaos **20**, 1567–1580 (2010)
124. Muthuswamy, B., Kokate, P.P.: Memristor-based chaotic circuits. IETE Tech. Rev. **26**, 417–429 (2009)
125. Natarajan, S.: A systematic method for obtaining state equations using MNA. IEE Proc. G **138**, 341–346 (1991)
126. Newcomb, R.W.: The semistate description of nonlinear time-variable circuits. IEEE Trans. Circuits Syst. **28**, 62–71 (1981)
127. Newcomb, R.W., Dziurla, B.: Some circuits and systems applications of semistate theory. Circuits Syst. Signal Process. **8**, 235–260 (1989)
128. Ohtsuki, T., Watanabe, H.: State-variable analysis of RLC networks containing nonlinear coupling elements. IEEE Trans. Circuit Theory **16**, 26–38 (1969)
129. Penski, C.: A numerical scheme for highly oscillatory DAEs and its application to circuit simulation. Numer. Algorithms **19**, 173–181 (1998)
130. Penski, C.: A new numerical method for SDEs and its application in circuit simulation. J. Comput. Appl. Math. **115**, 461–470 (2000)
131. Pershin, Y.V., Di Ventra, M.: Practical approach to programmable analog circuits with memristors. IEEE Trans. Circuits Syst. I **57**, 1857–1864 (2010)
132. Pershin, Y.V., Di Ventra, M.: Memory effects in complex materials and nanoscale systems. Adv. Phys. **60**, 145–227 (2011)
133. Pershin, Y.V., Di Ventra, M.: Neuromorphic, digital and quantum computation with memory circuit elements. Proc. IEEE **100**, 2071–2080 (2012)
134. Polyuga, R.V., van der Schaft, A.J.: Structure preserving port-Hamiltonian model reduction of electrical circuits. In: Model Reduction for Circuit Simulation. Lecture Notes in Electrical Engineering, vol. 74, II, pp. 241–260 (2011)
135. Rabier, P.J.: Implicit differential equations near a singular point. J. Math. Anal. Appl. **144**, 425–449 (1989)
136. Rabier, P.J., Rheinboldt, W.C.: A general existence and uniqueness theory for implicit differential-algebraic equations. Differ. Integral Equ. **4**, 563–582 (1991)
137. Rabier, P.J., Rheinboldt, W.C.: A geometric treatment of implicit differential-algebraic equations. J. Differ. Equ. **109**, 110–146 (1994)
138. Rabier, P.J., Rheinboldt, W.C.: On impasse points of quasi-linear differential-algebraic equations. J. Math. Anal. Appl. **181**, 429–454 (1994)
139. Rabier, P.J., Rheinboldt, W.C.: On the computation of impasse points of quasi-linear differential-algebraic equations. Math. Comput. **62**, 133–154 (1994)
140. Rabier, P.J., Rheinboldt, W.C.: Theoretical and numerical analysis of differential-algebraic equations. In: Handbook of Numerical Analysis, vol. VIII, pp. 183–540. North-Holland, Amsterdam (2002)
141. Reich, S.: On a geometrical interpretation of differential-algebraic equations. Circuits Syst. Signal Process. **9**, 367–382 (1990)
142. Reich, S.: On an existence and uniqueness theory for nonlinear differential-algebraic equations. Circuits Syst. Signal Process. **10**, 343–359 (1991)
143. Reis, T.: Circuit synthesis of passive descriptor systems: a modified nodal approach. Int. J. Circuit Theory Appl. **38**, 44–68 (2010)
144. Reis, T., Heinkenschloss, M.: Model reduction for a class of nonlinear electrical circuits by reduction of linear subcircuits. Matheon Preprint 702-2010 (2010)
145. Reis, T., Stykel, T.: PABTEC: passivity-preserving balanced truncation for electrical circuits. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **29**, 1354–1367 (2010)

146. Reis, T., Stykel, T.: Lyapunov balancing for passivity-preserving model reduction of RC circuits. SIAM J. Appl. Dyn. Syst. **10**, 1–34 (2011)
147. Reis, T., Tischendorf, C.: Frequency domain methods and decoupling of linear infinite dimensional differential algebraic systems. J. Evol. Equ. **5**, 357–385 (2005)
148. Reiszig, G.: Differential-algebraic equations and impasse points. IEEE Trans. Circuits Syst. I **43**, 122–133 (1996)
149. Reiszig, G.: The index of the standard circuit equations of passive RLCTG-networks does not exceed 2. In: Proc. ISCAS'98, vol. 3, pp. 419–422 (1998)
150. Reiszig, G.: Extension of the normal tree method. Int. J. Circuit Theory Appl. **27**, 241–265 (1999)
151. Reiszig, G., Boche, H.: On singularities of autonomous implicit ordinary differential equations. IEEE Trans. Circuits Syst. I **50**, 922–931 (2003)
152. Rheinboldt, W.C.: Differential-algebraic systems as differential equations on manifolds. Math. Comput. **43**, 473–482 (1984)
153. Riaza, R.: On the singularity-induced bifurcation theorem. IEEE Trans. Autom. Control **47**, 1520–1523 (2002)
154. Riaza, R.: A matrix pencil approach to the local stability analysis of nonlinear circuits. Int. J. Circuit Theory Appl. **32**, 23–46 (2004)
155. Riaza, R.: Singularity-induced bifurcations in lumped circuits. IEEE Trans. Circuits Syst. I **52**, 1442–1450 (2005)
156. Riaza, R.: Differential-Algebraic Systems. Analytical Aspects and Circuit Applications. World Scientific, Singapore (2008)
157. Riaza, R.: Nondegeneracy conditions for active memristive circuits. IEEE Trans. Circuits Syst. II **57**, 223–227 (2010)
158. Riaza, R.: Graph-theoretic characterization of bifurcation phenomena in electrical circuit dynamics. Int. J. Bifurc. Chaos **20**, 451–465 (2010)
159. Riaza, R.: Explicit ODE reduction of memristive systems. Int. J. Bifurc. Chaos **21**, 917–930 (2011)
160. Riaza, R.: Dynamical properties of electrical circuits with fully nonlinear memristors. Nonlinear Anal., Real World Appl. **12**, 3674–3686 (2011)
161. Riaza, R.: Manifolds of equilibria and bifurcations without parameters in memristive circuits. SIAM J. Appl. Math. **72**, 877–896 (2012)
162. Riaza, R.: Cyclic matrices of weighted digraphs. Discrete Appl. Math. **160**, 280–290 (2012)
163. Riaza, R.: First order mem-circuits: modeling, nonlinear oscillations and bifurcations. IEEE Trans. Circuits Syst. I (2013, in press)
164. Riaza, R., Encinas, A.: Augmented nodal matrices and normal trees. Discrete Appl. Math. **158**, 44–61 (2010)
165. Riaza, R., Tischendorf, C.: Qualitative features of matrix pencils and DAEs arising in circuit dynamics. Dyn. Syst. **22**, 107–131 (2007)
166. Riaza, R., Tischendorf, C.: The hyperbolicity problem in electrical circuit theory. Math. Methods Appl. Sci. **33**, 2037–2049 (2010)
167. Riaza, R., Tischendorf, C.: Semistate models of electrical circuits including memristors. Int. J. Circuit Theory Appl. **39**, 607–627 (2011)
168. Riaza, R., Tischendorf, C.: Structural characterization of classical and memristive circuits with purely imaginary eigenvalues. Int. J. Circuit Theory Appl. (2013, in press)
169. Riaza, R., Torres-Ramírez, J.: Nonlinear circuit modeling via nodal methods. Int. J. Circuit Theory Appl. **33**, 281–305 (2005)
170. Riaza, R., Campbell, S.L., Marszalek, W.: On singular equilibria of index-1 DAEs. Circuits Syst. Signal Process. **19**, 131–157 (2000)
171. Roth, J.P.: An application of algebraic topology: Kron's method of tearing. Q. Appl. Math. **17**, 1–24 (1959)
172. Sastry, S.S., Desoer, C.A.: Jump behavior of circuits and systems. IEEE Trans. Circuits Syst. **28**, 1109–1124 (1981)

173. Schein, O., Denk, G.: Numerical solution of stochastic differential-algebraic equations with applications to transient noise simulation of microelectronic circuits. J. Comput. Appl. Math. **100**, 77–92 (1998)
174. Schöps, S., De Gersem, H., Bartel, A.: A cosimulation framework for multirate time integration of field/circuit coupled problems. IEEE Trans. Magn. **46**, 3233–3236 (2010)
175. Selva Soto, M., Tischendorf, C.: Numerical analysis of DAEs from coupled circuit and semiconductor simulation. Appl. Numer. Math. **53**, 471–488 (2005)
176. Smale, S.: On the mathematical foundations of electrical circuit theory. J. Differ. Geom. **7**, 193–210 (1972)
177. Sommariva, A.M.: On a specific substitution theorem. Int. J. Circuit Theory Appl. **26**, 509–512 (1998)
178. Sommariva, A.M.: On a specific substitution theorem: further results. Int. J. Circuit Theory Appl. **27**, 277–281 (1999)
179. Sommariva, A.M.: State-space equations of regular and strictly topologically degenerate linear lumped time-invariant networks: the multiport method. Int. J. Circuit Theory Appl. **29**, 435–453 (2001)
180. Sommariva, A.M.: State-space equations of regular and strictly topologically degenerate linear lumped time-invariant networks: the implicit tree-tableau method. IEEE Int. Symp. Circuits Syst. Proc. **8**, 1139–1141 (2001)
181. Stern, T.E.: On the equations of nonlinear networks. IEEE Trans. Circuit Theory **13**, 74–81 (1966)
182. Straube, B., Reinschke, K., Vermeiren, W., Röbenack, K., Müller, B., Clauß, C.: DAE-index increase in analogue fault simulation. In: Workshop on System Design Automation 2000, Dresden, pp. 99–104 (2000)
183. Striebel, M., Günther, M.: A charge oriented mixed multirate method for a special class of index-1 network equations in chip design. Appl. Numer. Math. **53**, 489–507 (2005)
184. Strukov, D.B., Snider, G.S., Stewart, D.R., Williams, R.S.: The missing memristor found. Nature **453**, 80–83 (2008)
185. Stykel, T.: Balancing-related model reduction of circuit equations using topological structure. In: Model Reduction in Circuit Simulation. Lecture Notes in Electrical Engineering, vol. 74, I, pp. 53–83. Springer, Berlin (2011)
186. Synge, J.L.: The fundamental theorem of electrical networks. Q. Appl. Math. **9**, 113–127 (1951)
187. Sze, S.M., Ng, K.K.: Physics of Semiconductor Devices. Wiley, New York (2007)
188. Tadeusiewicz, M.: A method for identification of asymptotically stable equilibrium points of a certain class of dynamic circuits. IEEE Trans. Circuits Syst. I **46**, 1101–1109 (1999)
189. Tadeusiewicz, M.: Global and local stability of circuits containing MOS transistors. IEEE Trans. Circuits Syst. I **48**, 957–966 (2001)
190. Takamatsu, M., Iwata, S.: Index characterization of differential-algebraic equations in hybrid analysis for circuit simulation. Int. J. Circuit Theory Appl. **38**, 419–440 (2010)
191. Takens, F.: Constrained equations; a study of implicit differential equations and their discontinuous solutions. In: Lect. Notes Maths., vol. 525, pp. 143–234. Springer, Berlin (1976)
192. Tasic, B., Verhoeven, A., ter Maten, E.J.W., Beelen, T.G.J.: Compound BDF multirate transient analysis applied to circuit simulation. In: Proc. ICNAAM'06, pp. 480–483 (2006)
193. Tischendorf, C.: Topological index calculation of DAEs in circuit simulation. Surv. Math. Ind. **8**, 187–199 (1999)
194. Tischendorf, C.: Coupled systems of differential algebraic and partial differential equations in circuit and device simulation. Modeling and numerical analysis. Habilitationsschrift, Humboldt-Univ. Berlin (2003)
195. van der Meijs, N.P.: Model order reduction of large RC circuits. In: Schilders, W.H.A., et al. (eds.) Model Order Reduction: Theory, Research Aspects and Applications, pp. 421–446. Springer, Berlin (2008)

196. van der Schaft, A.J.: Port-Hamiltonian systems: network modeling and control of nonlinear physical systems. In: Irschik, H., Schlacher, K. (eds.) Advanced Dynamics and Control of Structures and Machines, pp. 127–168. Springer, Berlin (2004)
197. Verhoeven, A., Beelen, T.G.J., El Guennouni, A., ter Maten, E.J.W., Mattheij, R.M.M., Tasic, B.: Error analysis of BDF compound-fast multirate method for differential-algebraic equations. In: 9th Copper Mountain Conf. Iterative Methods (2006). http://www.mgnet.org/mgnet-conferences.html#VirtualProceedings
198. Vlach, J., Singhal, K.: Computer Methods for Circuits Analysis and Design. Van Nostrand Reinhold ITP, New York (1994)
199. Voss, T., Verhoeven, A., Bechtold, T., ter Maten, J.: Model order reduction for nonlinear differential algebraic equations in circuit simulation. In: Proc. ECMI 2006, pp. 518–523. Springer, Berlin (2007)
200. Weiss, L., Mathis, W.: A Hamiltonian formulation for complete nonlinear RLC-networks. IEEE Trans. Circuits Syst. I **44**, 843–846 (1997)
201. Weiss, L., Mahtis, W., Trajkovic, L.: A generalization of Brayton-Moser's mixed potential function. IEEE Trans. Circuits Syst. I **45**, 423–427 (1998)
202. Winkler, R.: Stochastic differential algebraic equations of index 1 and applications in circuit simulation. J. Comput. Appl. Math. **163**, 435–463 (2004)
203. Yang, J.J., Pickett, M.D., Li, X., Ohlberg, D.A.A., Stewart, D.R., Williams, R.S.: Memristive switching mechanism for metal/oxide/metal nanodevices. Nat. Nanotechnol. **3**, 429–433 (2008)

Solution Concepts for Linear DAEs: A Survey

Stephan Trenn

Abstract This survey aims at giving a comprehensive overview of the solution theory of linear differential-algebraic equations (DAEs). For classical solutions a complete solution characterization is presented including explicit solution formulas similar to the ones known for linear ordinary differential equations (ODEs). The problem of inconsistent initial values is treated and different approaches are discussed. In particular, the common Laplace-transform approach is discussed in the light of more recent distributional solution frameworks.

Keywords Differential algebraic equations · Descriptor systems · Distributional solution theory · Laplace transform

Mathematics Subject Classification (2010) 34A09 · 34A12 · 34A05 · 34A25

1 Introduction

Modeling physical phenomena relates physical variables via differential equations as well as algebraic equations leading in general to a system description of the form

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

a *differential-algebraic equation* (DAE). However, this survey will not treat this most general system description but it will consider its linear counterpart

$$E\dot{x} = Ax + f, \tag{1.1}$$

where $E, A \in \mathbb{R}^{m\times n}$, $m, n \in \mathbb{N}$, are constant matrices and $f : \mathbb{R} \to \mathbb{R}^m$ is some inhomogeneity. If the matrix E is square and invertible, the DAE is equivalent to an ordinary differential equation (ODE) of the form

$$\dot{x} = Ax + f. \tag{1.2}$$

S. Trenn (✉)
University of Kaiserslautern, 67663 Kaiserslautern, Germany
e-mail: trenn@mathematik.uni-kl.de

A. Ilchmann, T. Reis (eds.), *Surveys in Differential-Algebraic Equations I*,
Differential-Algebraic Equations Forum,
DOI 10.1007/978-3-642-34928-7_4, © Springer-Verlag Berlin Heidelberg 2013

For this ODE the solution theory is well understood and there have been no disputes or different viewpoints on it in the last five or more decades. In fact, the solution formula can concisely be expressed with the matrix exponential:

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)} f(\tau)\,\mathrm{d}\tau, \quad x_0 \in \mathbb{R}^n; \tag{1.3}$$

although the Jordan canonical form of A is essential to grasp the whole of the possibilities of solution behaviors. Some features of the solutions of an ODE are highlighted:

Existence. For every initial condition $x(0) = x_0$, $x_0 \in \mathbb{R}^n$, and each (locally integrable) inhomogeneity f there exists a solution.

Uniqueness. For any fixed inhomogeneity f the initial value $x(0)$ uniquely determines the whole solution; in fact each single value $x(t)$, $t \in \mathbb{R}$, determines the solution on the whole time axis.

Inhomogeneity. The solution is always one degree "smoother" then the inhomogeneity, i.e. if f is differentiable then x is at least twice differentiable, in particular, non-smoothness of f does not prevent the ODE of having a solution (at least in the sense of Carathéodory).

In Sects. 2.4 and 2.5 solution formulas similar to (1.3) will be presented for *regular* DAEs; however, for general DAEs none of these three properties have to hold anymore as the following example shows.

Example 1.1 Consider the DAE

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \dot{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} x + f$$

which implies $x_2 = -f_2$, $x_1 = \dot{x}_2 - f_1 = -\dot{f}_2 - f_1$ and $f_3 = 0$. In particular, not for all initial values or all inhomogeneities there exists a solution. Furthermore, x_3 is not restricted at all, hence uniqueness of solutions is not present. Finally, x_1 contains the derivative of the inhomogeneity so that the solution is "less smooth" than the inhomogeneity which could lead to non-existence of solutions if the inhomogeneities is not sufficiently smooth.

The aim of this survey is twofold: (1) to present a fairly complete classical solution theory for the DAE (1.1) also for the singular case; (2) to discuss the approaches to treat inconsistent initial values and the corresponding distributional solution concepts. In particular, a rigorous discussion of the so-called Laplace-transform approach to treat inconsistent initial values and its connection to distributional solution concepts is carried out. This is a major difference with the already available survey by Lewis [32], which is not so much concerned with distributional solutions. The focus of Lewis' survey is more on system theoretic topics like controllability, observability, stability and feedback, which are not treated here.

This survey is structured as follows. In Sect. 2 classical (i.e. differentiable) solutions of (1.1) are studied. It is shown how the Weierstraß and Kronecker canonical form of the matrix pencil $sE - A \in \mathbb{R}^{m\times n}[s]$ can be used to fully characterize the solutions. Solution formulas which do not need the complete knowledge of the canonical forms will be presented, too. A short overview over the situation for time-varying DAEs is given as well. Inconsistent initial values are the most discussed topics concerning DAEs and different arguments how to treat them have been proposed. One common approach to treat inconsistent values is the application of the Laplace transform to (1.1); the details are explained in Sect. 4. However, the latter approach led to much confusion and therefore a time-domain approach based on distributional solutions was developed and studied by a number of authors, see Sect. 5.

2 Classical Solutions

In this section classical solutions of the DAE (1.1) are considered:

Definition 2.1 (Classical solution) A classical solution of the DAE (1.1) is any differential function $x \in \mathscr{C}^1(\mathbb{R} \to \mathbb{R}^n)$ such that $E\dot{x}(t) = Ax(t) + f(t)$ holds for all $t \in \mathbb{R}$.

It will turn out that existence of a classical solution in general also depends on the smoothness properties of the inhomogeneity; if not mentioned otherwise it will be assumed therefore in the following that the inhomogeneity f is sufficiently smooth, e.g. by assuming that f is in fact smooth (i.e. arbitrarily often differentiable).

2.1 The Kronecker and Weierstraß Canonical Forms

The first appearance of DAEs (1.1) with a complete solution discussion seems to be the one by Gantmacher [21] (Russian original 1953), where he considered classical solutions. His analysis is based on the following notion of equivalence of matrix pairs (called strict equivalence by him):

$$(E_1, A_1) \cong (E_2, A_2)$$
$$:\Leftrightarrow \ \exists S \in \mathbb{R}^{m\times m}, T \in \mathbb{R}^{n\times n} \text{ both invertible: } (E_1, A_1) = (SE_2T, SA_2T).$$

It is clear that for equivalent matrix pairs (E_1, A_1) and (E_2, A_2) (via the transformation matrices S and T) the following equivalence holds:

$$x \text{ solves } E_1\dot{x} = A_1x + f \quad \Leftrightarrow \quad z = T^{-1}x \text{ solves } E_2\dot{z} = Az + S^{-1}f.$$

Gantmacher's focus is actually on matrix pencils $sE - A \in \mathbb{R}^{m\times n}[s]$ and the derivation of a canonical form corresponding to the above equivalence—the *Kronecker canonical form* (KCF). The solution theory of the DAE (1.1) is a mere application of the KCF. In particular, he does not consider inconsistent initial values or non-smooth inhomogeneities. The existence and representation of the KCF is formulated with the following result.

Theorem 2.1 (Kronecker canonical form [21, 28]) *For every matrix pencil $sE - A \in \mathbb{R}^{m\times n}[s]$ there exist invertible matrices $S \in \mathbb{C}^{m\times m}$ and $T \in \mathbb{C}^{n\times n}$ such that, for $a, b, c, d \in \mathbb{N}$ and $\varepsilon_1, \dots, \varepsilon_a, \rho_1, \dots, \rho_b, \sigma_1, \dots, \sigma_c, \eta_1, \dots, \eta_d \in \mathbb{N}$,*

$$S(sE - A)T = \operatorname{diag}\big(\mathscr{P}_{\varepsilon_1}(s), \dots, \mathscr{P}_{\varepsilon_a}(s), \mathscr{J}_{\rho_1}(s), \dots, \mathscr{J}_{\rho_b}(s), \mathscr{N}_{\sigma_1}(s), \dots, \mathscr{N}_{\sigma_c}(s), \mathscr{Q}_{\eta_1}(s), \dots, \mathscr{Q}_{\eta_d}(s)\big), \tag{2.1}$$

where

$$\mathscr{P}_\varepsilon(s) = s\begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix} \in \mathbb{R}^{\varepsilon\times(\varepsilon+1)}[s], \quad \varepsilon \in \mathbb{N},$$

$$\mathscr{J}_\rho(s) = sI - \begin{bmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix} \in \mathbb{C}^{\rho\times\rho}[s], \quad \rho \in \mathbb{N}, \lambda \in \mathbb{C},$$

$$\mathscr{N}_\sigma(s) = s\begin{bmatrix} 0 & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix} - I \in \mathbb{R}^{\sigma\times\sigma}[s], \quad \sigma \in \mathbb{N},$$

$$\mathscr{Q}_\eta(s) = s\begin{bmatrix} 0 & & \\ 1 & \ddots & \\ & \ddots & 0 \\ & & 1 \end{bmatrix} - \begin{bmatrix} 1 & & \\ 0 & \ddots & \\ & \ddots & 1 \\ & & 0 \end{bmatrix} \in \mathbb{R}^{(\eta+1)\times\eta}[s], \quad \eta \in \mathbb{N}.$$

The block-diagonal form (2.1) *is unique up to reordering of the blocks and is called Kronecker canonical form* (*KCF*) *of the matrix pencil* $(sE - A)$.

Note that in the KCF $\mathscr{P}_\varepsilon(s)$-blocks with $\varepsilon = 0$ and $\mathscr{Q}_\eta(s)$-blocks with $\eta = 0$ are possible, which results in zero columns (for $\varepsilon = 0$) and/or zero rows (for $\eta = 0$) in the KCF, see the following example.

Example 2.1 (KCF of Example 1.1) By a simple interchanging of rows and columns the KCF is obtained from Example 1.1 and has the following form

$$\left[\begin{array}{c|cc} 0 & -1 & 0 \\ 0 & s & -1 \\ \hline 0 & 0 & 0 \end{array}\right]$$

i.e. the KCF consists of one $\mathscr{P}_0(s)$-block, one $\mathscr{N}_2(s)$-block and one $\mathscr{Q}_0(s)$-block.

In the canonical coordinates the solution analysis is now rather straightforward because each block on the diagonal and the associated variables can be considered separately. The four different block types lead to the following solution characterizations:

$\mathscr{P}_\varepsilon(s)$-**block** If $\varepsilon = 0$ then this simply means that the corresponding variable does not appear in the equations and is therefore free and can be chosen arbitrarily. For $\varepsilon > 0$ consider the differential equation $\mathscr{P}_\varepsilon(\frac{\mathrm{d}}{\mathrm{d}t})(x) = f$ which equivalently can be written as the ODE

$$\begin{pmatrix} \dot{x}_2 \\ \dot{x}_3 \\ \vdots \\ \dot{x}_{\varepsilon+1} \end{pmatrix} = \begin{bmatrix} 0 & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} x_2 \\ x_3 \\ \vdots \\ x_{\varepsilon+1} \end{pmatrix} + \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_\varepsilon \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{pmatrix} x_1.$$

Hence for any x_1 and any inhomogeneity f there exist solutions for $x_2, x_3, \ldots, x_{\varepsilon+1}$ uniquely determined by the initial values $x_2(0), \ldots, x_3(0)$. In particular, for all initial values and all inhomogeneities there exist solutions which are not unique because x_1 can freely be chosen.

$\mathscr{J}_\rho(s)$-**block** The differential equation $\mathscr{J}_\rho(\frac{\mathrm{d}}{\mathrm{d}t})(x) = f$ is a standard linear ODE, i.e. it holds that for all initial values and all inhomogeneities a unique solution.

$\mathscr{N}_\rho(s)$-**block** Write $\mathscr{N}_\rho(s) = sN - I$, then it is easily seen that the differential operator $\mathscr{N}_\rho(\frac{\mathrm{d}}{\mathrm{d}t}) : \mathscr{C}^\infty \to \mathscr{C}^\infty$ is invertible with inverse

$$\left(N\frac{\mathrm{d}}{\mathrm{d}t} - I\right)^{-1} = -\sum_{i=0}^{\rho-1} N^i \frac{\mathrm{d}}{\mathrm{d}t}^i. \tag{2.2}$$

In particular for any smooth inhomogeneity the solution of the differential equation $\mathscr{N}(\frac{\mathrm{d}}{\mathrm{d}t})(x) = f$ is uniquely given by

$$x = -\sum_{i=0}^{\rho-1} N^i f^{(i)} = \begin{pmatrix} -f_1 \\ -f_2 - \dot{f}_1 \\ \vdots \\ -f_\rho - \dot{f}_{\rho-1} - \cdots - f_1^{(\rho-1)} \end{pmatrix}. \tag{2.3}$$

In particular it is not possible to specify the initial values arbitrarily—they are completely determined by the inhomogeneity.

$\mathscr{Q}_\eta(s)$-**block** If $\eta = 0$ then no variable is present and the equation reads $0 = f$, hence not for all inhomogeneities the overall DAE is solvable. If $\eta > 0$ then the solution of the differential equation $\mathscr{Q}_\eta(\frac{\mathrm{d}}{\mathrm{d}t})(x) = f$ is given by (2.3) with ρ replaced by η but only if the inhomogeneity fulfills

$$f_{\eta+1} = \dot{x}_\eta = -\dot{f}_\eta - \ddot{f}_\eta - \cdots - f_1^{(\eta)}.$$

In particular not for all inhomogeneities and not for all initial values solutions exist. However, when solutions exist they are uniquely given by (2.3).

A consequence of the above blockwise analysis is the following result.

Corollary 2.2 (Existence and uniqueness of solutions) *The DAE (1.1) has a smooth solution x for all smooth inhomogeneities f if, and only if, in the KCF the $\mathscr{Q}_\eta(s)$-blocks are not present. Any solution x of (1.1) with fixed inhomogeneity f is uniquely determined by the initial value $x(0)$ if, and only if, in the KCF the $\mathscr{P}_\varepsilon(s)$-blocks are not present.*

The KCF without the $\mathscr{P}_\varepsilon(s)$ and $\mathscr{Q}_\eta(s)$ blocks is also called the *Weierstraß canonical form* (WCF) and can be characterized directly in terms of the original matrices. For this the notion of regularity is needed.

Definition 2.2 (Regularity) The matrix pencil $sE - A \in \mathbb{R}^{m\times n}[s]$ is called regular if, and only if, $n = m$ and $\det(sE - A)$ is not the zero polynomial. The matrix pair (E, A) and the corresponding DAE (1.1) is called regular whenever $sE - A$ is regular.

Theorem 2.3 (Weierstraß canonical form [49]) *The matrix pencil $sE - A \in \mathbb{R}^{n\times n}[s]$ is regular if, and only if, there exist invertible matrices $S, T \in \mathbb{C}^{n\times n}$ such that $sE - A$ is transformed into the Weierstraß canonical form (WCF)*

$$S(sE - A)T = s\begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix} - \begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix},$$

where $J \in \mathbb{C}^{n_1\times n_1}$, $N \in \mathbb{C}^{n_2\times n_2}$, $n_1 + n_2 = n$, are matrices in Jordan canonical form and N is nilpotent.

In conclusion, if one aims at similar solution properties as for classical linear ODEs the class of regular DAEs is exactly the one to consider, see also Sects. 2.4 and 2.5. In the classical solution framework there is still a gap between ODEs and regular DAEs because (1.1) does not have solutions for all initial values and not for insufficiently smooth inhomogeneities. However, in a distributional solution framework these two missing properties can also be recaptured, see Sect. 5.

2.2 Solution Formulas Based on the Wong Sequences: General Case

For practical problems the above solution characterization is not so useful as the determination of the KCF is numerically ill posed. Therefore, solution formulas which do not need the complete KCF are of interest. One of the first work in this direction is the one by Wilkonson [50], who presents an iterative algorithm to obtain the solutions. More geometrical approaches can be traced back to Dieudonné [15] and Wong [51]; the latter introduced the two important subspace sequences for a matrix pair $(E, A) \in (\mathbb{R}^{m\times n})^2$:

$$\begin{aligned} \mathscr{V}_0 &= \mathbb{R}^n, \qquad \mathscr{V}_{i+1} = A^{-1}(E\mathscr{V}_i), \quad i = 0, 1, 2, \ldots, \\ \mathscr{W}_0 &= \{0\}, \qquad \mathscr{W}_{j+1} = E^{-1}(A\mathscr{W}_j), \quad j = 0, 1, 2, \ldots, \end{aligned} \tag{2.4}$$

which therefore will be called *Wong sequences* in the following. It is easily seen that the Wong sequences are nested and terminate after finitely many steps, i.e.

$$\begin{aligned} \exists i^* \in \{0, 1, \ldots, n\} : \qquad \mathscr{V}^* &:= \bigcap_{i\in\mathbb{N}} \mathscr{V}_i = \mathscr{V}_{i^*}, \\ \exists j^* \in \{0, 1, \ldots, n\} : \qquad \mathscr{W}^* &:= \bigcup_{j\in\mathbb{N}} \mathscr{W}_j = \mathscr{W}_{j^*}. \end{aligned}$$

Bernhard [6] used the first Wong sequence in his geometrical analysis of (1.1) where the inhomogeneity has the special form $f = Bu$ for some suitable matrix B. Utilizing both Wong sequences Armentano [2] was able to obtain a Kronecker like form. However, his arguments are purely geometrical and it is not apparent how to characterize the solutions of (1.1) because the necessary transformation matrices are not given explicitly. This problem was resolved recently in [4], where the following connection between the Wong sequences and a quasi-Kronecker form was established.

Theorem 2.4 (Quasi Kronecker form (QKF) [4]) *Consider the DAE* (1.1) *and the corresponding limits* $\mathscr{V}^*$ *and* $\mathscr{W}^*$ *of the Wong sequences* (2.4). *Choose any invertible matrices* $[P_1, R_1, Q_1] \in \mathbb{R}^{n\times n}$ *and* $[P_2, R_2, Q_2] \in \mathbb{R}^{m\times m}$ *such that*

$$\begin{aligned} \operatorname{im} P_1 &= \mathscr{V}^* \cap \mathscr{W}^*, & \operatorname{im}[P_1, R_1] &= \mathscr{V}^* + \mathscr{W}^*, \\ \operatorname{im} P_2 &= E\mathscr{V}^* \cap A\mathscr{W}^*, & \operatorname{im}[P_2, R_2] &= E\mathscr{V}^* + A\mathscr{W}^*, \end{aligned}$$

then $T = [P_1, R_1, Q_1]$, $S = [P_2, R_2, Q_2]^{-1}$ *put the matrix pencil* $sE - A$ *into quasi-Kronecker triangular form* (*QKTF*):

$$S(sE - A)T = \begin{bmatrix} sE_P - A_P & sE_{PR} - A_{PR} & sE_{PQ} - A_{PQ} \\ 0 & sE_R - A_R & sE_{RQ} - A_{RQ} \\ 0 & 0 & sE_Q - A_Q \end{bmatrix}, \tag{2.5}$$

where $\lambda E_P - A_P$ has full row rank for all $\lambda \in \mathbb{C} \cup \{\infty\}$, $sE_R - A_R$ is regular, and $\lambda E_Q - A_Q$ has full column rank for all $\lambda \in \mathbb{C} \cup \{\infty\}$. Furthermore, the following generalized Sylvester equations are solvable:

$$\begin{aligned} 0 &= E_{RQ} + E_R F_1 + F_2 E_Q, & 0 &= A_{RQ} + A_R F_1 + F_2 A_Q, \\ 0 &= E_{PR} + E_P G_1 + G_2 E_R, & 0 &= A_{PR} + A_P G_1 + G_2 A_R, \\ 0 &= (E_{PQ} + E_{PR} F_1) + E_P H_1 + H_2 E_Q, \\ 0 &= (A_{PQ} + A_{PR} F_1) + A_P H_1 + H_2 A_Q, \end{aligned}$$

and any solutions $F_1, F_2, G_1, G_2, H_1, H_2$ yield a quasi-Kronecker form (QKF) via

$$\begin{aligned} &\begin{bmatrix} I & -G_2 & -H_2 \\ 0 & I & -F_2 \\ 0 & 0 & I \end{bmatrix}^{-1} S(sE - A)T \begin{bmatrix} I & G_1 & H_1 \\ 0 & I & F_1 \\ 0 & 0 & I \end{bmatrix} \\ &= \begin{bmatrix} sE_P - A_P & 0 & 0 \\ 0 & sE_R - A_R & 0 \\ 0 & 0 & sE_Q - A_Q \end{bmatrix}, \end{aligned} \tag{2.6}$$

where the diagonal block entries are the same as in (2.5).

The solution analysis can now be carried out via analyzing the blocks in the QKF (2.6) individually:

- $sE_P - A_P$: Due to the full rank assumption there exists a unimodular[1] matrix $[M_P(s), K_P(s)]$ such that

$$(sE_P - A_P)\big[M_P(s), K_P(s)\big] = [I, 0], \tag{2.7}$$

  see e.g. [4, Lem. 3.1]. The solutions x_P of the DAE $E_P \dot{x}_P = A_P x_P + f_P$ are given by

$$x_P = M_P\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)(f_P) + K_P\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)(u)$$

  where $u : \mathbb{R} \to \mathbb{R}^{n_P - m_P}$ is an arbitrary (sufficiently smooth) function and where $m_P \times n_P$ with $m_P < n_P$ is the size of the matrix pencil $sE_P - A_P$. Furthermore, each initial condition $x_P(0) = x_P^0$ can be achieved by an appropriate choice of u.
- $sE_R - A_R$: The solution behavior for a regular DAE was already discussed at the end of Sect. 2.1, a further discussion is carried out in Sects. 2.4 and 2.5.

[1] A polynomial matrix is called unimodular if it is invertible and its inverse is again a polynomial matrix.

- $sE_Q - A_Q$: Analogous to the $sE_P - A_P$ block there exists a unimodular matrix $\left[\begin{smallmatrix} M_Q(s) \\ K_Q(s) \end{smallmatrix}\right]$ such that

$$\begin{bmatrix} M_Q(s) \\ K_Q(s) \end{bmatrix} (sE_Q - A_Q) = \begin{bmatrix} I \\ 0 \end{bmatrix}. \tag{2.8}$$

  Then $E_Q \dot{x}_Q = A_Q x_Q + f_Q$ is solvable if, and only if,

$$K_Q\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)(f_Q) = 0$$

  and the solution is uniquely determined by

$$x_Q = M_Q\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)(f_Q).$$

  In particular, the initial values cannot be specified as they are already fixed by $x_Q(0) = M_Q(\frac{\mathrm{d}}{\mathrm{d}t})(f_Q)(0)$.

In summary, the QKF decouples the corresponding DAE into the underdetermined part (existence but non-uniqueness of solutions), the regular part (existence and uniqueness of solutions) and the overdetermined part (uniqueness of solution but possible non-existence). Furthermore, the above solution characterization can also be carried out directly with the QKTF (2.5), where the analysis for the $sE_Q - A_Q$ block remains unchanged, for the regular block the inhomogeneity f_R is replaced by $f_R + (E_{RQ}\frac{\mathrm{d}}{\mathrm{d}t} - A_{RQ})(x_Q)$ and for the $sE_P - A_P$ block the inhomogeneity f_P is replaced by $f_P + (E_{PR}\frac{\mathrm{d}}{\mathrm{d}t} - A_{PR})(x_R) + (E_{PQ}\frac{\mathrm{d}}{\mathrm{d}t} - A_{PQ})(x_Q)$.

Remark 2.1 (Refinement of QKF [3]) If R_1 and R_2 in Theorem 2.4 are chosen in the special way $R_1 = [R_1^J, R_1^N]$ and $R_2 = [R_2^J, R_2^N]$ where

$$\operatorname{im}\left[P_1, R_1^J\right] = \mathscr{V}^*, \qquad \operatorname{im}\left[P_2, R_2^J\right] = E\mathscr{V}^*,$$

then a decoupling of the regular part in (2.5) corresponding the WCF is obtained as well. In particular, applying the Wong sequences again to the regular part (see next section) is not necessary for a further analysis.

2.3 Existence and Uniqueness of Solutions with Respect to In- and Outputs

In practical application the inhomogeneity f in the DAE (1.1) is often generated by a lower dimensional input u, i.e. $f = Bu$ for some suitable matrix B; furthermore, an output $y = Cx + Du$ is introduced to represent the signals of the systems which are available for measurement and/or are of interest. The resulting DAE is then

often called *descriptor system* [52] (other common names are singular systems [8] or generalized state-space system [48])

$$\begin{aligned} E\dot{x} &= Ax + Bu, \\ y &= Cx + Du. \end{aligned} \tag{2.9}$$

Clearly, a solution theory for general DAEs (1.1) is also applicable to descriptor systems (2.9). In particular, regularity of the matrix pair (E, A) guarantees existence and uniqueness of solutions for any sufficiently smooth input. However, existence and uniqueness of solutions with respect to the input and output might hold for descriptor systems even when the matrix pair (E, A) is not regular as the following example shows.

Example 2.2 Consider the following descriptor system:

$$\begin{aligned} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \dot{x} &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u, \\ y &= \begin{bmatrix} 1 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \end{bmatrix} u, \end{aligned}$$

which has for any input u the unique output $y = -u$. However, the corresponding matrix pair $(E, A) = \left(\left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]\right)$ is not regular.

It is therefore useful to define the notion of *external regularity*.

Definition 2.3 (External regularity) The descriptor system (2.9) and the corresponding matrix tuple (E, A, B, C, D) are called *externally regular* if, only if, for all sufficiently smooth inputs u there exist (classical) solutions x of (2.9) and the output y is uniquely determined by u and $x(0)$.

With the help of the quasi-Kronecker form it is now possible to prove the following characterization of external regularity.

Theorem 2.5 (Characterization of external regularity) *The descriptor system* (2.9) *is externally regular if, and only if,*

$$\operatorname{rk}[sE - A, B] = \operatorname{rk}[sE - A] = \operatorname{rk} \begin{bmatrix} sE - A \\ C \end{bmatrix} \tag{2.10}$$

for infinitely many $s \in \mathbb{C}$.

Proof The rank of a matrix does not change when multiplied with invertible matrices (from the left and the right), hence it can be assumed that the matrix pair (E, A) is already in QKF (2.6) with corresponding transformation matrices S and T. According to the block size in (2.6) let $SB = [B_P^\top, B_R^\top, B_Q^\top]^\top$ and $CT =$

$[C_P, C_R, C_Q]$. Then (2.10) is equivalent to

$$\operatorname{rk}[sE_Q - A_Q, B_Q] = \operatorname{rk}[sE_Q - A_Q] \quad \text{and} \quad \operatorname{rk}\begin{bmatrix} sE_P - A_P \\ C_P \end{bmatrix} = \operatorname{rk}[sE_P - A_P]$$

for infinitely many $s \in \mathbb{C}$. The rank is also invariant under multiplication with unimodular polynomial matrices, hence (2.10) is also equivalent to, invoking (2.8) and (2.7),

$$\operatorname{rk}\begin{bmatrix} I & M_Q(s)B_Q \\ 0 & K_Q(s)B_Q \end{bmatrix} = \operatorname{rk}\begin{bmatrix} I \\ 0 \end{bmatrix} \quad \text{and} \quad \operatorname{rk}\begin{bmatrix} I & 0 \\ C_P M_P(s) & C_P K_P(s) \end{bmatrix} = \operatorname{rk}\begin{bmatrix} I & 0 \end{bmatrix}.$$

Because a polynomial matrix is zero if and only if it is zero at infinitely values it follows that (2.10) is equivalent to the condition $K_Q(s)B_Q \equiv 0$ and $C_P K_p(s) \equiv 0$. Taking into account the solution characterization given in conclusion to Theorem 2.4 the characterization of external regularity is shown. □

Note that condition (2.10) already appears in the survey paper by Lewis [32] based on arguments in the frequency domain.

2.4 Solution Formulas Based on the Wong Sequences: Regular Case

If the Wong sequences (2.4) are applied to a regular matrix pencil $sE - A \in \mathbb{R}^{n\times n}[s]$ then the limits $\mathscr{V}^*$ and $\mathscr{W}^*$ fulfill (see [2, 5, 51])

$$\mathscr{V}^* \cap \mathscr{W}^* = \{0\}, \qquad \mathscr{V}^* + \mathscr{W}^* = \mathbb{R}^n,$$
$$E\mathscr{V}^* \cap A\mathscr{W}^* = \{0\}, \qquad E\mathscr{V}^* + A\mathscr{W}^* = \mathbb{R}^n.$$

In particular $[V, W]$ and $[EV, AW]$ are invertible matrices for all basis matrices V and W of $\mathscr{V}^*$ and $\mathscr{W}^*$. In fact, any of these invertible matrices yield a transformation which put the matrix pencil $sE - A$ into a quasi-Weierstraß form (QWF):

Theorem 2.6 (Quasi Weierstraß form (QWF) [2, 5]) *Consider a regular matrix pencil $sE - A \in \mathbb{R}^{n\times n}[s]$ and the corresponding Wong sequences with limits $\mathscr{V}^*$ and $\mathscr{W}^*$. For any full rank matrices V, W with $\operatorname{im} V = \mathscr{V}^*$ and $\operatorname{im} W = \mathscr{W}^*$ let $T = [V, W]$ and $S = [EV, AW]^{-1}$. Then*

$$S(sE - A)T = s\begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix} - \begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix}, \tag{2.11}$$

where $J \in \mathbb{R}^{n_1\times n_1}$, $n_1 \in \mathbb{N}$, is some matrix and $N \in \mathbb{R}^{n_2\times n_2}$, $n_2 = n - n_1$, is nilpotent. In particular, $\mathscr{V}^*$ is exactly the space of consistent initial values, i.e. for all $x_0 \in \mathscr{V}^*$ there exists a unique (classical) solution x of $E\dot{x} = Ax$ with $x(0) = x_0$.

The difference to the WCF from Theorem 2.3 is that J and N are not assumed to be in Jordan canonical form. Furthermore, the transformation matrices for the QWF can be chosen easily; it is only necessary to calculate the Wong sequences.

The knowledge of the two limiting spaces $\mathscr{V}^*$ and $\mathscr{W}^*$ is enough to obtain an explicit solution formula similar to the solution formula (1.3) for ODEs as the next result shows. To formulate the explicit solution formula it is necessary to define certain projectors as follows.

Definition 2.4 (Consistency, differential and impulse projector[43]) Consider a regular matrix pair (E, A) and use the same notation as in Theorem 2.6. The *consistency projector* is given by

$$\Pi_{(E,A)} := T \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} T^{-1},$$

the *differential projector* is given by

$$\Pi^{\text{diff}}_{(E,A)} := T \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} S,$$

and the *impulse projector* is given by

$$\Pi^{\text{imp}}_{(E,A)} := T \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} S,$$

where the block structure is as in the QWF (2.11). Furthermore, let

$$A^{\text{diff}} := \Pi^{\text{diff}}_{(E,A)} A \quad \text{and} \quad E^{\text{imp}} = \Pi^{\text{imp}}_{(E,A)} E.$$

Note that the above defined matrices do not depend on the specific choice of the matrices V and W, because when choosing different basis matrices $\widetilde{V}$ and $\widetilde{W}$ it must hold that $V = \tilde{V} Q$ and $W = \tilde{W} P$ for some invertible P and Q. Hence

$$\widetilde{T} = [\widetilde{V}, \widetilde{W}] = T \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \quad \text{and} \quad \widetilde{S} = [E\widetilde{V}, A\widetilde{W}]^{-1} = \begin{bmatrix} P^{-1} & 0 \\ 0 & Q^{-1} \end{bmatrix} S$$

and the invariance of the above definitions with respect to the choice of V and W is obvious. Furthermore, the differential and impulse projectors are not projectors in the usual sense because they are in general not idempotent.

Theorem 2.7 (Explicit solution formula based on Wong sequences [47]) *Let (E, A) be a regular matrix pair and use the notation from Definition 2.4. Then all solutions of (1.1) are given by, for $c \in \mathbb{R}^n$,*

$$x(t) = e^{A^{\text{diff}} t} \Pi_{(E,A)} c + \int_0^t e^{A^{\text{diff}}(t-\tau)} \Pi^{\text{diff}}_{(E,A)} f(\tau)\, \mathrm{d}\tau - \sum_{i=0}^{n-1} \left(E^{\text{imp}}\right)^i \Pi^{\text{imp}}_{(E,A)} f^{(i)}(t). \tag{2.12}$$

In particular,

$$x(0) = \Pi_{(E,A)} c - \sum_{i=0}^{n-1} \left(E^{\mathrm{imp}}\right)^i \Pi^{\mathrm{imp}}_{(E,A)} f^{(i)}(0)$$

i.e. $c \in \mathbb{R}^n$ *implicitly specifies the initial value* (*but in general* $x(0) \neq c$ *even when* $c \in \mathscr{V}^*$).

In the homogeneous case the following equivalence holds [43]:

$$E\dot{x} = Ax \quad \Leftrightarrow \quad \dot{x} = A^{\mathrm{diff}} x \wedge x(0) \in \mathscr{V}^*,$$

which motivates the name differential projector. There is also a motivation for the name of the impulse projector, see the end of Sect. 4 as well as Sect. 5.

The Wong sequences appeared sporadically in the DAE literature: For example, Yip and Sincovec [52] used them to characterize regularity of the matrix pencil, Owens and Debeljkovic [36] characterized the space of consistent initial values via the Wong sequences; they are also included in the text books [1, 29] but not in the text books [7–9, 14, 30]. In general it seems that the connection between the Wong sequences and the (quasi-)Weierstraß/Kronecker form and their role in the solution characterization is not well known or appreciated in the DAE community (especially in the case of singular matrix pencils).

2.5 The Drazin Inverse Solution Formula

Another explicit solution formula was proposed by Campbell et al. [11] already in 1976 and is based on the Drazin inverse.

Definition 2.5 (Drazin inverse [17]) For $M \in \mathbb{R}^{n\times n}$ a matrix $D \in \mathbb{R}^{n\times n}$ is called Drazin inverse if, and only if,

1. $DM = MD$,
2. $D = DMD$,
3. $\exists \nu \in \mathbb{N}: \; M^\nu = M^{\nu+1} D$.

In [17] it is shown that the Drazin inverse is unique and it is easy to see that the Drazin inverse of M is given by

$$M^D = T \begin{bmatrix} J^{-1} & 0 \\ 0 & 0 \end{bmatrix} T^{-1},$$

where the invertible matrix T is such that

$$M = T \begin{bmatrix} J & 0 \\ 0 & N \end{bmatrix} T^{-1},$$

J is invertible and N is nilpotent. In particular, for invertible M the Drazin inverse is just the classical inverse, i.e. $M^{-1} = M^D$.

The following solution formula for the DAE (1.1) based on the Drazin inverse needs commutativity of the matrices E and A, however, as also regularity is assumed the following result shows that this is not a restriction of generality.

Lemma 2.8 (Commutativation of (E, A) [11]) *Assume (E, A) is regular and chose $\lambda \in \mathbb{R}$ such that $\lambda E - A$ is invertible. Then*

$$(\lambda E - A)^{-1} E \quad \text{and} \quad (\lambda E - A)^{-1} A$$

commute, i.e. the whole equation (1.1) can simply be multiplied from the left with $(\lambda E - A)^{-1}$ which will not change the solution properties but will guarantee commutativity of the coefficient matrices.

Theorem 2.9 (Explicit solution formula based on the Drazin inverse [11]) *Consider the regular DAE (1.1) with $EA = AE$. Then all solutions x are given by*

$$x(t) = e^{E^D A t} E^D E c + \int_0^t e^{E^D A (t-\tau)} E^D f(\tau)\, \mathrm{d}\tau - \big(I - E^D E\big) \sum_{i=0}^{n-1} \big(E A^D\big)^i A^D f^{(i)}(t). \tag{2.13}$$

A direct comparison of the solution formula (2.12) based on the Wong sequences and (2.13) indicates that $E^D A$ plays the role of A^{diff}, $E^D E$ plays the role of the consistency projector and E^D plays the role of the differential projector. However, the connection between the impulse projector and E^{imp} to the expressions involving the Drazin inverse of A is not immediately clear. The following result justifies the previous observations.

Lemma 2.10 (Wong sequences and Drazin inverse [5]) *Consider the regular matrix pair (E, A) with $EA = AE$ and use the notation from Theorem 2.6. Then*

$$E^D = T \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} S \quad \text{and} \quad A^D = T \begin{bmatrix} J^D & 0 \\ 0 & I \end{bmatrix} S.$$

In particular, also taking into account $E = S^{-1} \left[\begin{smallmatrix} I & 0 \\ 0 & N \end{smallmatrix}\right] T^{-1}$ and $A = S^{-1} \left[\begin{smallmatrix} J & 0 \\ 0 & I \end{smallmatrix}\right] T^{-1}$,

$$E^D = \Pi^{\mathrm{diff}}_{(E,A)},$$

$$E^D A = \Pi^{\mathrm{diff}}_{(E,A)} A = A^{\mathrm{diff}},$$

$$E^D E = T \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} S S^{-1} \begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix} T^{-1} = \Pi_{(E,A)},$$

$$\big(E A^D\big)^i = \left(S^{-1} \begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix} T^{-1} T \begin{bmatrix} J^D & 0 \\ 0 & I \end{bmatrix} S \right)^i = S^{-1} \begin{bmatrix} (J^D)^i & 0 \\ 0 & N^i \end{bmatrix} S, \quad i \in \mathbb{N},$$

and with some more effort, using $E\mathscr{V}^* = \mathscr{V}^*$ and $A\mathscr{W}^* = \mathscr{W}^*$ in the commuting case (see [5]), it follows that

$$\big(I - E^D E\big)\big(EA^D\big)^i A^D = T \begin{bmatrix} 0 & 0 \\ 0 & N^i \end{bmatrix} S = \big(E^{\mathrm{imp}}\big)^i \Pi^{\mathrm{imp}}_{(E,A)}.$$

This shows that indeed the two solution formulas (2.12) and (2.13) are identical in the commuting case. Note that in the solution formula (2.13) the Drazin inverse A^D appears and one might therefore think that the occurrence of zero eigenvalues in A plays some special role for the solution. However, this is just an artifact and it turns out that in the expression

$$A^D = T \begin{bmatrix} J^D & 0 \\ 0 & I \end{bmatrix} S$$

the matrix J^D can be replaced by an arbitrary matrix without changing the result of the solution formula (2.13). One canonical choice is to replace J^D by the zero matrix which yields the impulse projector and which makes the "correction term" $(I - E^D E)$ superfluous.

2.6 Time-Varying DAEs

In this section the time-varying version of (1.1), i.e.

$$E(t)\dot{x}(t) = A(t)x(t) + f(t),$$

is briefly discussed.

Campbell and Petzold [10] proved that if $E(\cdot)$ and $A(\cdot)$ have real analytical entries then a solution characterization similar to Corollary 2.2 holds. In particular, they showed that unique solvability is equivalent to finding time-varying (analytical) transformation matrices $S(\cdot)$, $T(\cdot)$, such that

$$\big(S(t)E(t)T(t), S(t)A(t)T(t) - S(t)E(t)T'(t)\big) = \left(\begin{bmatrix} I & 0 \\ 0 & N(t) \end{bmatrix}, \begin{bmatrix} J(t) & 0 \\ 0 & I \end{bmatrix}\right),$$

where $N(t)$ is a strictly lower triangular (and hence nilpotent) matrix. In particular, as in the time-invariant case, the DAE decouples into an ODE part and a pure DAE part. It is easily seen that for a strictly lower triangular matrix $N(t)$ also the differential operator $N(\cdot)\frac{\mathrm{d}}{\mathrm{d}t}$ is nilpotent, hence the inverse operator of $(N(\cdot)\frac{\mathrm{d}}{\mathrm{d}t} - I)$ can be calculated nearly identically as in (2.2):

$$\left(N(\cdot)\frac{\mathrm{d}}{\mathrm{d}t} - I\right)^{-1} = -\sum_{i=0}^{\nu-1}\left(N(\cdot)\frac{\mathrm{d}}{\mathrm{d}t}\right)^i,$$

where $\nu \in \mathbb{N}$ is the nilpotency index of the operator $N(\cdot)\frac{\mathrm{d}}{\mathrm{d}t}$.

If the coefficient matrices are not analytical the situation is not so clear anymore and different approaches have been proposed. Most methods have their motivation in numerical simulations and a detailed description and discussion is outside the scope of this survey. The interested reader is referred to the nice survey by Rabier and Rheinboldt [38], and to the text book by Kunkel and Mehrmann [30] as well as the recent monograph by Lamour, März and Tischendorf [31]. However, all these approaches do not allow for discontinuous coefficient matrices. These are studied in [46] and because of the connection to inconsistent initial value problems the problem of discontinuous coefficient matrices is further discussed in Sect. 5.

3 Inconsistent Initial Values and Distributional Solutions

After having presented a rather extensive discussion of classical solutions, this section presents an introductory discussion of the problem of inconsistent initial values. From the above derived solution formulas for (1.1) it becomes apparent that $x(0)$ cannot be chosen arbitrarily, certain parts of $x(0)$ are already fixed by the DAE and the inhomogeneity, cf. Theorem 2.7. In the extreme case that the QWF of (E, A) only consists of the nilpotent part, the initial value $x(0)$ is completely determined by the inhomogeneity and no freedom to choose the initial value is left. However, there are situations where one wants to study the response of a system described by a DAE when an *inconsistent initial value* is given. Examples are electrical circuits which are switched on at a certain time [48]. There have been different approaches to deal with inconsistent initial values, e.g. [12, 18, 35, 37, 39, 42], some of them will be presented in detail in the later sections. All have in common that jumps as well as Dirac impulses may occur in the solutions. The Dirac impulse is a distribution (a generalized function), hence one must enlarge the considered solution space to also include distributions. In fact, also the presence of non-smooth inhomogeneities (or inputs) can lead to distributional solutions. However, the latter do not produce conceptional difficulties as the solution characterization of the previous section basically remains unchanged.

In order to be able to make mathematical precise statements the classical distribution theory [41] is revised first. The space of *test functions* is given by

$$\mathscr{C}_0^\infty := \big\{\varphi : \mathbb{R} \to \mathbb{R} \mid \varphi \in \mathscr{C}^\infty \text{ has compact support}\big\},$$

which is equipped with a certain topology.[2] The space of distributions, denoted by $\mathbb{D}$, is then the dual of the space of test functions, i.e.

$$\mathbb{D} := \big\{D : \mathscr{C}_0^\infty \to \mathbb{R} \mid D \text{ is linear and continuous}\big\}.$$

[2]The topology is such that a sequence $(\varphi_k)_{k\in\mathbb{N}}$ of test functions converges to zero if, and only if, (1) the supports of all φ_k are contained within one common compact set $K \subseteq \mathbb{R}$ and (2) for all $i \in \mathbb{N}$, $\varphi_k^{(i)}$ converges uniformly to zero as $k \to \infty$.

A large class of ordinary functions, namely locally integrable functions, can be embedded into $\mathbb{D}$ via the following injective[3] homomorphism:

$$f \mapsto f_{\mathbb{D}}, \quad \text{with } f_{\mathbb{D}}(\varphi) := \int_{\mathbb{R}} f\varphi.$$

The main feature of distributions is the ability to take derivatives for any distribution $D \in \mathbb{D}$ via

$$D'(\varphi) := -D\big(\varphi'\big).$$

Simple calculations show that this is consistent with the classical derivative, i.e. if f is differentiable, then

$$(f_{\mathbb{D}})' = \big(f'\big)_{\mathbb{D}}.$$

In particular, the Heaviside unit step $\mathbb{1}_{[0,\infty)}$ has a distributional derivative which can easily be calculated to be

$$(\mathbb{1}_{[0,\infty)\mathbb{D}})'(\varphi) = \varphi(0) =: \delta(\varphi),$$

hence it results in the well known Dirac impulse δ (at $t = 0$). In general, the Dirac impulse δ_t at time $t \in \mathbb{R}$ is given by $\delta_t(\varphi) := \varphi(t)$. Furthermore, if g is a piecewise differentiable function with one jump at $t = t_j$, i.e. g is given as

$$g(t) = \begin{cases} g_1(t), & t < t_j, \\ g_2(t), & t \geq t_j, \end{cases}$$

where g_1 and g_2 are differentiable functions and

$$g^1(t) := \begin{cases} g_1'(t), & t < t_j, \\ g_2'(t), & t \geq t_j, \end{cases}$$

then

$$(g_{\mathbb{D}})' = \big(g^1\big)_{\mathbb{D}} + \big(g(t_J+) - g(t_J-)\big)\delta_{t_j}. \tag{3.1}$$

In other words, taking derivatives of a general jump results in a Dirac impulse at the jump position whose amplitude is the height of the jump.

Finally, distributions can be multiplied with smooth functions α:

$$(\alpha D)(\varphi) = D(\alpha\varphi)$$

and it is easily seen that this multiplication is consistent with the pointwise multiplication of functions and that the Leibniz product rule holds:

$$(\alpha D)' = \alpha' D + \alpha D'.$$

[3]Two locally integrable functions which only differ on a set of measure zero are identified with each other.

Now it is no problem to consider the DAE (1.1) in a distributional solution space, instead of x and f being vectors of functions they are now vectors of distributions, i.e. $x \in \mathbb{D}^n$ and $f \in \mathbb{D}^m$ where $m \times n$ is the size of the matrices E and A. The definition of the matrix vector product remains unchanged[4] so that (1.1) reads as m equations in $\mathbb{D}$.

Considering distributional solutions, however, does not help to treat inconsistent initial value; au contraire, distributions cannot be evaluated at a certain time because they are not functions of time, so writing $x(0) = x_0$ makes no sense. Even when assuming that a pointwise evaluation is well defined for certain distributions, the DAE (1.1) will still not exhibit (distributional) solution with arbitrary initial values. This is easily seen when considering the DAE $N\dot{x} = x + f$ with nilpotent N. Then also in the distributional solution framework the operator $N\frac{\mathrm{d}}{\mathrm{d}t} - I : \mathbb{D} \to \mathbb{D}$ is invertible with inverse as in (2.2) and there exists a unique (distributional) solution given by

$$x = -\sum_{i=0}^{n-1} N^i f^{(i)},$$

hence the initial value of x cannot be assigned arbitrarily (i.e. independently of the inhomogeneity).

So what does it then mean to speak of a solution of (1.1) with inconsistent initial value? The motivation for inconsistent initial values is the situation that the system descriptions gets active at the initial time $t = 0$ and before that the system was governed by different (maybe unknown) rules. This viewpoint was also expressed by Doetsch [16, p. 108] in the context of distributional solutions for ODEs:

> The concept of "initial value" in the physical science can be understood only when the past, that is, the interval $t < 0$, has been included in our considerations. This occurs naturally for distributions which, without exception, are defined on the entire t-axis.

So mathematically, there is some given past trajectory x^0 for x up to the initial time and the DAE (1.1) only holds on the interval $[0, \infty)$. This means that a solution of the following *initial trajectory problem* (ITP) is sought:

$$\begin{aligned} x_{(-\infty,0)} &= x^0_{(-\infty,0)}, \\ (E\dot{x})_{[0,\infty)} &= (Ax + f)_{[0,\infty)}, \end{aligned} \tag{3.2}$$

where $x^0 \in \mathbb{D}^n$ is an arbitrary past trajectory and D_I for some interval $I \subseteq \mathbb{R}$ and $D \in \mathbb{D}$ denotes a distributional restriction generalizing the restrictions of functions given by

$$f_I(t) = \begin{cases} f(t), & t \in I, \\ 0, & t \notin I. \end{cases}$$

[4] Some authors [30, 38] use a different definition for the matrix vector product which is due to the different viewpoint of a distributional vector x as a map from $(\mathscr{C}_0^\infty)^n$ to $\mathbb{R}$ instead of a map from $\mathscr{C}_0^\infty$ to $\mathbb{R}^n$. The latter seems the more natural approach in view of applying it to (1.1), but it seems that both approaches are equivalent at least with respect to the solution theory of DAEs.

A fundamental problem is the fact (see Lemma 5.1) that such a distributional restriction does not exist!

This problem was resolved especially in older publication [8, 9, 48] by ignoring it and/or by arguing with the Laplace transform (see the next section). Cobb [13] seems to be the first to be aware of this problem and he resolved it by introducing the space of piecewise-continuous distributions; Geerts [22, 23] was the first to use the space of impulsive-smooth distributions (introduced in [27]) as a solution space for DAEs. Seemingly unaware of these two approaches, Tolsa and Salichs [44] developed a distributional solution framework which can be seen as a mixture between the approaches of Cobb and Geerts. The more comprehensive space of piecewise-smooth distributions was later introduced [45] to combine the advantages of the piecewise-continuous and impulsive-smooth distributional solution spaces. The details are discussed in Sect. 5.

Cobb [12] also presented another approach by justifying the impulsive response due to inconsistent initial values via his notion of *limiting solutions*. The idea is to replace the singular matrix E in (1.1) by a "disturbed" version E_ε which is invertible for all $\varepsilon > 0$ and $E_\varepsilon \to E$ as $\varepsilon \to 0$. If the solutions of the corresponding initial value ODE problem $\dot{x} = E_\varepsilon^{-1} Ax$, $x(0) = x_0$ converges to a distribution, then Cobb calls this the limiting solution. He is then able to show that the limiting solution is unique and equal to the one obtained via the Laplace-transform approach. Campbell [9] extends this result also to the inhomogeneous case.

4 Laplace Transform Approaches

Especially in the signal theory community it is common to study systems like (1.1) or (2.9) in the so called *frequency domain* (in contrast to the *time domain*). In particular, when the input-output mapping is of interest the frequency domain approach significantly simplifies the analysis. The transformation between time and frequency domain is given by the *Laplace transform* defined via the Laplace integral:

$$\hat{g}(s) := \int_0^\infty e^{-st} g(t)\,\mathrm{d}t \tag{4.1}$$

for some function g and $s \in \mathbb{C}$. Note that in general the Laplace integral is not well defined for all $s \in \mathbb{C}$ and a suitable domain for $\hat{g}$ must be chosen [16]. If a suitable domain exists, then $\hat{g} = \mathscr{L}\{g\}$ is called the *Laplace transform* of g and, in general, $\mathscr{L}\{\cdot\}$ denotes the Laplace transform operator. Again note that it is not specified at this point which class of functions have a Laplace transform and which class of functions are obtained as the image of $\mathscr{L}\{\cdot\}$. The main feature of the Laplace transform is the following property, where g is a differentiable function for which g and g' have Laplace transforms:

$$\mathscr{L}\{g'\}(s) = s\mathscr{L}\{g\}(s) - g(0), \tag{4.2}$$

which is a direct consequence of the definition of the Laplace integral invoking partial differentiation. If g is not continuous at $t = 0$ but $g(0+)$ exists and g' denotes the derivative of g on $\mathbb{R} \setminus \{0\}$, then (4.2) still holds in a slightly altered form:

$$\mathscr{L}\{g'\}(s) = s\mathscr{L}\{g\}(s) - g(0+). \tag{4.3}$$

In particular, the Laplace transform does not take into account at all how g behaved for $t < 0$ which is a trivial consequence of the definition of the Laplace integral. This observation will play an important role when studying inconsistent initial values.

Taking into account the linearity of the Laplace transform the descriptor system (2.9) is transformed into

$$\begin{aligned} sE\hat{x}(s) &= A\hat{x}(s) + B\hat{u}(s) + Ex(0+), \\ \hat{y}(s) &= C\hat{x}(s) + D\hat{u}(s). \end{aligned} \tag{4.4}$$

If the matrix pair (E, A) is regular and $x(0+) = 0$, the latter can be solved easily algebraically:

$$\hat{y}(s) = \big(C(sE - A)^{-1}B + D\big)\hat{u}(s) =: G(s)\hat{u}(s), \tag{4.5}$$

where $G(s)$ is a matrix over the field of rational functions and is usually called transfer function. As there are tables of functions and its Laplace transforms it is often possible to find the solutions of descriptor system with given input simply by plugging the Laplace transform of the input in the above formula and lookup the resulting output $\hat{y}(s)$ to obtain the solution $y(t)$ in the time domain. Furthermore, many important system properties can be deduced from properties (like the zeros and poles) of the transfer function directly.

A first systematic treatment of descriptor systems in the frequency domain was carried out by Rosenbrock [40]. He, however, only considered zero initial values and the input-output behavior. In particular, he was not concerned with a solution theory for general DAEs (1.1) with possible inconsistent values. Furthermore, he restricted attention to inputs which are exponentially bounded (guaranteeing existence of the Laplace transform), hence formally his framework could not deal with arbitrary (sufficiently smooth) inputs.

The definition of the Laplace transform can be extended to be well defined for certain distributions as well [16], therefore consider the following class of distributions:

$$\mathbb{D}_{\geq 0,k} := \big\{ D = (g_{\mathbb{D}})^{(k)} \;\big|\; \text{where } g : \mathbb{R} \to \mathbb{R} \text{ is continuous and } g(t) = 0 \text{ on } (-\infty, 0) \big\}.$$

For $D \in \mathbb{D}_{\geq 0,k}$ with $D = (g_{\mathbb{D}})^{(k)}$ the (distributional) Laplace transform is now given by

$$\mathscr{L}_{\mathbb{D}}\{D\}(s) := s^k \mathscr{L}\{g\}(s)$$

on a suitable domain in $\mathbb{C}$. Note that $\delta \in \mathbb{D}_{\geq 0,2}$ and it is easily seen that

$$\mathscr{L}_{\mathbb{D}}\{\delta\} = 1. \tag{4.6}$$

Furthermore, for every locally integrable function g for which $\mathscr{L}\{g\}$ is defined on a suitable domain it holds that

$$\mathscr{L}_{\mathbb{D}}\{g_{\mathbb{D}}\} = s\mathscr{L}\left\{\int_0 g\right\} \overset{[16]}{=} s\frac{1}{s}\mathscr{L}\{g\} = \mathscr{L}\{g\}, \tag{4.7}$$

i.e. the distributional Laplace transform coincides with the classical Laplace transform defined by (4.1).

A direct consequence of the definition of $\mathscr{L}_{\mathbb{D}}$ is the following derivative rule for all $D \in \bigcup_k \mathbb{D}_{\geq 0,k}$:

$$\mathscr{L}_{\mathbb{D}}\{D'\}(s) = s\mathscr{L}_{\mathbb{D}}\{D\} \tag{4.8}$$

which seems to be in contrast to the derivative rule (4.3), because no initial value occurs. The latter can actually not be expected because general distributions do not have a well defined function evaluation at a certain time t. However, the derivative rule (4.8) is consistent with (4.3); to see this let g be a function being zero on $(-\infty, 0)$, differentiable on $(0, \infty)$ with well defined value $g(0+)$. Denote with g' the (classical) derivative of g on $\mathbb{R} \setminus \{0\}$, then (invoking linearity of $\mathscr{L}_{\mathbb{D}}$)

$$\begin{aligned}\mathscr{L}_{\mathbb{D}}\{(g_{\mathbb{D}})'\}(s) &\overset{(3.1)}{=} \mathscr{L}_{\mathbb{D}}\{(g')_{\mathbb{D}} + g(0+)\delta\}(s)\\ &= \mathscr{L}_{\mathbb{D}}\{(g')_{\mathbb{D}}\}(s) + g(0+)\mathscr{L}_{\mathbb{D}}\{\delta\}(s) \overset{(4.6),(4.7)}{=} \mathscr{L}\{g'\} + g(0+),\end{aligned}$$

which shows equivalence of (4.8) and (4.3). The key observation is that the distributional derivative takes into account the jump at $t = 0$ whereas the classical derivative ignores it, i.e. in the above context

$$(g_{\mathbb{D}})' \neq (g')_{\mathbb{D}}.$$

As it is common to identify g with $g_{\mathbb{D}}$ (even in [16]), the above distinction is difficult to grasp, in particular for inexperienced readers. As this problem plays an important role when dealing with inconsistent initial values, it is not surprising that researchers from the DAE community who are simply using the Laplace transform as a tool, struggle with the treatment of inconsistent initial values, cf. [34].

Revisiting the treatment of the descriptor system (2.9) in the frequency domain one has now to decide whether to use the usual Laplace transform resulting in (4.4) or the distributional Laplace transform resulting in

$$\begin{aligned}sE\hat{x}(s) &= A\hat{x}(s) + B\hat{u}(s),\\ \hat{y}(s) &= C\hat{x}(s) + D\hat{u}(s),\end{aligned} \tag{4.9}$$

where the initial value $x(0+)$ does not occur anymore. In particular, if the matrix pair (E, A) is regular, the only solution of (4.9) is given by (4.5) independently of $x(0+)$. In particular, if $u = 0$ the only solution of (4.9) is $\hat{x}(s) = 0$ and $\hat{y}(s) = 0$. Assuming a well defined inverse Laplace transform this implies that the only solution of (2.9) with $u = 0$ is the trivial solution, which is of course not true in general. Altogether the following dilemma occurs.

Dilemma (Discrepancy between time domain and frequency domain) Consider the regular DAE (1.1) or more specifically (2.9) with zero inhomogeneity (input) but non-zero initial value.

- An ad hoc analysis calls for *distributional solutions* in response to inconsistent initial values. For consistent initial value there exist classical (non-zero) solutions.
- Using the *distributional* Laplace transform to analyze the (distributional) solutions of (1.1) or (2.9) reveals that the *only* solution is the trivial one. In particular, no initial values (neither inconsistent nor consistent ones) are taken into account at all.

This problem was already observed in [16, p. 108] and is based on the definition of the distributional Laplace transform which is only defined for distributions vanishing on $(-\infty, 0)$. The following "solution" to this dilemma was suggested [16, p. 129]: Define for $D \in \bigcup_k \mathbb{D}_{\geq 0,k}$ the "past-aware" derivative operator $\frac{\mathrm{d}_-}{\mathrm{d}t}$:

$$\frac{\mathrm{d}_-}{\mathrm{d}t} D := D' - d_0^- \delta, \tag{4.10}$$

where $d_0^- \in \mathbb{R}$ is interpreted as a "virtual" initial value for $D(0-)$. Note, however, that, by definition, $D(0-) = 0$ for every $D \in \bigcup_k \mathbb{D}_{\geq 0,k}$; hence at this stage it is not clear why this definition makes sense. This problem was also pointed out by Cobb [12]. Nevertheless, a motivation for this choice will be given in Sect. 5.

Using now the past-aware derivative in the distributional formulation of (1.1) one obtains

$$\begin{aligned} Ex' &= Ax + Bu + Ex_0^- \delta, \\ y &= Cx + Du, \end{aligned} \tag{4.11}$$

where $x_0^- \in \mathbb{R}^n$ is the virtual (possible inconsistent) initial value for $x(0-)$ and solutions are sought in the space $(\bigcup_k \mathbb{D}_{\geq 0,k})^n$, i.e. x is assumed to be zero on $(-\infty, 0)$. Applying the distributional Laplace transform to (4.11) yields

$$\begin{aligned} sE\hat{x}(s) &= A\hat{x}(s) + B\hat{u}(s) + Ex_0^-, \\ \hat{y}(s) &= C\hat{x}(s) + D\hat{u}(s). \end{aligned} \tag{4.12}$$

In contrast to (4.4), x_0^- is not the initial value for $x(0+)$ but is the virtual initial value for $x(0-)$. If the matrix pair (E, A) is regular, the solution of (4.12) can now be obtained via

$$\hat{x}(s) = (sE - A)^{-1}\big(B\hat{u}(s) + Ex_0^-\big)$$

and using the inverse Laplace transform. Because E is not invertible in general, the rational matrix $(sE - A)^{-1}$ may contain polynomial entries resulting in polynomial parts in $\hat{x}$ corresponding to Dirac impulses in the time domain, for details see the end of this section.

The solution formula for $\hat{x}(s)$ is possible to calculate analytically when the matrices E, A, and B are known and for suitable inputs u the inverse Laplace transform of $\hat{x}(s)$ can also be obtained analytically. This is the main advantage of the Laplace transform approach. There are, however, the following major drawbacks:

1. Within the frequency domain it is not possible to motivate the incorporation of the (inconsistent) initial values as in (4.11); in fact, Doetsch [16] who seems to have introduced this notion, needs to argue with the help of the distributional derivative and (4.10) within the time domain!
2. The Laplace transform ignores everything that was in the past, i.e. on the interval $(-\infty, 0)$; this is true for the classical Laplace transform (by definition of the Laplace integral) as well as for the distributional Laplace transform (by only considering distributions which vanish for $t < 0$). Hence the natural viewpoint of an initial trajectory problem (3.2) as also informally advocated by Doetsch cannot possibly be treated with the Laplace transform approach.
3. A frequency domain analysis gets useless when the original system is time-varying or nonlinear, whereas (linear) time-domain methods may in principle be extended to also treat time-variance and certain non-linearities. In fact, the piecewise-smoothly distributional solution framework as presented in Sect. 5 can be used without modification for linear time-varying DAEs [46] and also for certain non-linear DAEs [33].
4. Making statements about existence and uniqueness of solution with the help of the frequency domain heavily depends on an isomorphism between the time-domain and the frequency domain; there are, however, only a few special isomorphisms between certain special subspaces of the frequency and time domain, no general isomorphism is available, see also the discussion concerning (4.9).

This section on the Laplace domain concludes with the calculation of the re-initialization of the inconsistent initial value as well as the resulting Dirac impulses occurring in the solution. Therefore, consider the "distributional version" (following Doetsch) of (1.1):

$$E\dot{x} = Ax + f_{\mathbb{D}} + Ex_0^-\delta, \tag{4.13}$$

where $x_0^- \in \mathbb{R}^n$, and its corresponding Laplace transformed version in frequency domain

$$sE\hat{x}(s) = A\hat{x}(s) + \hat{f}(s) + Ex_0^-. \tag{4.14}$$

The unique solution of (4.14) in frequency domain is given by

$$\hat{x}(s) = (sE - A)^{-1}\big(\hat{f}(s) + Ex_0^-\big),$$

which needs regularity of the matrix pair (E, A) to be well defined, which will therefore be assumed in the following. Applying a coordinate transformation $x = T\binom{v}{w}$

according to the QWF (2.11), the solution in the new coordinates is given by

$$\begin{pmatrix} \hat{v}(s) \\ \hat{w}(s) \end{pmatrix} = T^{-1}(sE - A)^{-1}\left(\hat{f}(s) + ET\begin{pmatrix} v_0^- \\ w_0^- \end{pmatrix}\right)$$
$$= (sSET - SAT)^{-1}\left(S\hat{f}(s) + SET\begin{pmatrix} v_0^- \\ w_0^- \end{pmatrix}\right),$$

where $x_0^- =: T\left(\begin{smallmatrix} v_0^- \\ w_0^- \end{smallmatrix}\right)$. Hence, invoking the QWF (2.11), the solution formula decouples into

$$\hat{v}(s) = (sI - J)^{-1}\big(\hat{f}_1(s) + v_0^-\big),$$
$$\hat{w}(s) = (sN - I)^{-1}\big(\hat{f}_2(s) + Nw_0^-\big) = -\sum_{i=0}^{\nu-1} N^i s^i \big(\hat{f}_2(s) + Nw_0^-\big),$$

where $Sf =: \left(\begin{smallmatrix} f_1 \\ f_2 \end{smallmatrix}\right)$ and $\nu \in \mathbb{N}$ is the nilpotency index of N. Since $(sI - J)^{-1}$ is a strictly proper rational matrix, the solution for v (resulting from taking the inverse Laplace transform) is the corresponding standard ODE solution (1.3). In particular, $v(0+) = v_0^-$ and no Dirac impulses occur in v. Applying the inverse Laplace transformation on the solution formula for $\hat{w}(s)$, one obtains the solution $w = w_f + w_i$, where w_f is the response with respect to the inhomogeneity given by

$$w_f := -\sum_{i=0}^{\nu-1} N^i (f_{2\mathbb{D}})^{(i)}$$

and w_i consists of Dirac impulses at $t = 0$ produced by the inconsistent initial value:

$$w_i := -\sum_{i=0}^{\nu-1} N^{i+1} w_0^- \delta^{(i)}.$$

Note that in order to obtain w_f by using the correspondence (4.8), the distributional derivatives of f_2 have to be considered. As the (distributional) Laplace transform can only be applied to distributions vanishing on $(-\infty, 0)$, the inhomogeneity f_2 will in general have a jump at $t = 0$, hence w_f will also contain Dirac impulses depending on $f_2^{(i)}(0+)$, $i = 0, 1, \ldots, \nu - 1$. In summary:

Theorem 4.1 (Solution formula obtained via the Laplace transform approach) *Consider the regular DAE (1.1) with its "distributional version" (4.13). Let $\nu \in \mathbb{N}$ be the nilpotency index of N in the QWF (2.11) of the matrix pair (E, A). Assume $f : \mathbb{R} \to \mathbb{R}^n$ is zero on $(-\infty, 0)$ and $\nu - 1$ times differentiable on $(0, \infty)$ with well defined values $f^{(i)}(0+)$, $i = 0, 1, \ldots, \nu - 1$. Use the notation from Definition 2.4. Then $x \in (\bigcup_k \mathbb{D}_{\geq 0,k})^n$ given by (2.12) on $(0, \infty)$ with $c = x_0^-$ and by the impulsive*

part at $t = 0$, *denoted by* $x[0]$,

$$x[0] = -\sum_{i=0}^{\nu-2} \left(E^{\mathrm{imp}}\right)^{i+1} \sum_{j=0}^{i} \Pi_{(E,A)}^{\mathrm{imp}} f^{(i-j)}(0+)\delta^{(j)}$$
$$-\sum_{i=0}^{\nu-2} \left(E^{\mathrm{imp}}\right)^{i+1} (I - \Pi_{(E,A)}) x_0^- \delta^{(i)} \tag{4.15}$$

is the unique solution of (4.13) *obtained via solving* (4.14). *In particular,*

$$x(0+) = \Pi_{(E,A)} x_0^- + \sum_{i=0}^{n-1} \left(E^{\mathrm{imp}}\right)^i \Pi_{(E,A)}^{\mathrm{imp}} f^{(i)}(0+), \tag{4.16}$$

hence if $f \equiv 0$ *then the consistent reinitialization is given by the consistency projector* $\Pi_{(E,A)}$ *via*

$$x(0+) = \Pi_{(E,A)} x_0^-.$$

Proof Invoking (3.1), one obtains

$$(f_{2\mathbb{D}})^{(i)}[0] = \sum_{j=0}^{i-1} f_2^{(i-1-j)}(0+)\delta^{(j)},$$

hence

$$w_f[0] = -\sum_{i=0}^{\nu-2} N^{i+1} \sum_{j=0}^{i} f_2^{(i-j)}(0+)\delta^{(j)}.$$

Now using the identities, cf. [47],

$$A^{\mathrm{diff}} = T \begin{bmatrix} J & 0 \\ 0 & 0 \end{bmatrix} T^{-1}, \qquad E^{\mathrm{imp}} = T \begin{bmatrix} 0 & 0 \\ 0 & N \end{bmatrix} T^{-1},$$
$$T \begin{pmatrix} f_1 \\ 0 \end{pmatrix} = \Pi_{(E,A)}^{\mathrm{diff}} f, \qquad T \begin{pmatrix} 0 \\ f_2 \end{pmatrix} = \Pi_{(E,A)}^{\mathrm{imp}} f, \qquad T \begin{pmatrix} v_0^- \\ 0 \end{pmatrix} = \Pi_{(E,A)} x_0^-,$$
$$T \begin{pmatrix} 0 \\ w_0^- \end{pmatrix} = (I - \Pi_{(E,A)}) x_0^-$$

yields the claimed solution formula. □

5 Distributional Solutions

The previous section introduced distributional solutions in order to treat inconsistent initial values with the help of the Laplace transform. This leads to the consideration

of the distributional space $\bigcup_k \mathbb{D}_{\geq 0,k}$ which contains all distributions which can be written as a (distributional) kth derivative, $k \in \mathbb{N}$, of a continuous function being zero on $(-\infty, 0)$ and of which a Laplace transform exists. This choice is motivated by the applicability of the Laplace transform and is actually not motivated by dealing with inconsistent initial values. In fact, as was pointed out in the previous section, the Laplace transform ignores by definition/design all what has happened before $t < 0$ and is therefore in principle not suitable to treat inconsistent initial values coming from the past. Most researchers in the field agree with the notion that an inconsistent initial is due to a past which was *not* governed by the system description (1.1). One way of formalizing this viewpoint is the ITP (3.2). In general, having a past which obeys different rules then the present means that the overall system description is *time-variant* which gives another reason why the Laplace-transform approach runs into difficulties.

5.1 The Problem of Distributional Restrictions

Treating the ITP (3.2) in a distributional solution framework is, however, also not straightforward, because (as already mentioned above) the distributional restriction used in (3.2) is not well defined.

Lemma 5.1 (Bad distribution [45]) *Let D be the (distributional, i.e. weak*) limit of the distributions*:

$$D_k := \sum_{i=0}^{k} d_i \delta_{d_i}, \quad \text{where } d_i := \frac{(-1)^i}{i+1}, \quad i, k \in \mathbb{N}.$$

Then the restriction (in the sense of [45]) of D to the interval $[0, \infty)$ is not a well-defined distribution.

Proof Clearly,

$$D_{[0,\infty)} = \sum_{j=0}^{\infty} d_{2j} \delta_{d_{2j}},$$

however, applying $D_{[0,\infty)}$ to a test function φ which is identically one on $[0, 1]$ yields

$$D_{[0,\infty)}(\varphi) = \sum_{j=0}^{\infty} d_{2j} \delta_{d_{2j}}(\varphi) = \sum_{j=0}^{\infty} \frac{1}{2j} = \infty,$$

which shows that $D_{[0,\infty)}$ is not a well defined distribution. □

Remark 5.1 (Restriction to open intervals) The above results remain true when considering restriction to *open* intervals. However, it should be mentioned here that

nevertheless the *equation* $F_I = G_I$ makes sense for arbitrary distributions $F, G \in \mathbb{D}$ and any open interval $I \subseteq \mathbb{R}$ by defining:

$$F_I = G_I \quad :\Leftrightarrow \quad \forall \varphi \in \mathscr{C}_0^\infty \text{ with } \operatorname{supp}\varphi \subseteq I : \ F(\varphi) = G(\varphi).$$

In fact, this definition is consistent with the restriction-definition to be established in the following for a special class of distributions [45, Prop. 2.2.10]. Nevertheless, restricting the second equation in the ITP (3.2) to the *closed* interval $[0, \infty)$ is essential. Taking an open restriction in both equations of (3.2) would imply that the past and the present are decoupled so that the initial trajectory would not influence the future trajectory. To be more precise: Any (distributional) solution x of (3.2) will exhibit a jump at $t = 0$ in response to an inconsistent value $x_0(0-)$, but the derivative of this jump appears as a Dirac impulse in the expression $E\dot{x}$. While the restriction to the open interval $(0, \infty)$ would neglect this Dirac impulse, the restriction to the closed interval $[0, \infty)$ keeps the Dirac impulse in the second equation of the ITP (3.2) and hence the past can influence the present.

5.2 Cobb's Space of Piecewise-Continuous Distributions

The need to define a restriction for distributions was already advocated by Cobb [13]; although his motivation was not the ITP (3.2) but a rigorous definition of the impulsive term $D[t]$ of a distribution D at time $t \in \mathbb{R}$ which can be viewed as a restriction to the interval $[t, t]$. To this end, Cobb first defined the space of piecewise-continuous distributions given by

$$\mathbb{D}_{\mathrm{pw}\mathscr{C}^0} := \left\{ D \in \mathbb{D} \;\middle|\; \begin{array}{l} \exists T = \{t_i \in \mathbb{R} | i \in \mathbb{Z}\} \text{ ordered and locally finite} \\ \exists g \in \mathscr{C}^0_{\mathrm{pw}} \forall i \in \mathbb{Z} : \ D_{(t_i, t_{i+1})} = (g_{\mathbb{D}})_{(t_i, t_{i+1})} \\ \text{in the sense of Remark 5.1} \end{array} \right\},$$

where $\mathscr{C}^0_{\mathrm{pw}}$ denotes the space of piecewise-continuous functions, in particular, for any $g \in \mathscr{C}^0_{\mathrm{pw}}$ the values $g(t+)$ and $g(t-)$ are well defined for all $t \in \mathbb{R}$.

Definition 5.1 (Cobb's distributional restriction [13]) Let $D \in \mathbb{D}_{\mathrm{pw}\mathscr{C}^0}$ with $g \in \mathscr{C}^0_{\mathrm{pw}}$ and $T = \{t_i \in \mathbb{R} | i \in \mathbb{Z}\}$ such that D coincides with $g_{\mathbb{D}}$ on each interval (t_i, t_{i+1}), $i \in \mathbb{Z}$. For any $\tau \in \mathbb{R}$ choose $\varepsilon > 0$ such that $(\tau - \varepsilon, \tau) \subseteq (t_i, t_{i+1})$ for some $i \in \mathbb{Z}$. Then the restriction of D to the interval $[\tau, \infty)$ is defined via

$$D_{[\tau,\infty)}(\varphi) = \begin{cases} 0, & \text{if } \operatorname{supp}\varphi \subseteq (-\infty, \tau], \\ D(\varphi) - \int_{\tau-\varepsilon}^{\tau} g(t)\varphi(t)\,\mathrm{d}t, & \text{if } \operatorname{supp}\varphi \subseteq [\tau - \varepsilon, \infty), \\ D_{[\tau,\infty)}(\varphi^\varepsilon), & \text{otherwise,} \end{cases}$$

where $\varphi^\varepsilon \in \mathscr{C}_0^\infty$ is such that $\varphi = \varphi_\tau + \varphi^\varepsilon$ with $\operatorname{supp}\varphi_\tau \subseteq (-\infty, \tau]$ and $\operatorname{supp}\varphi^\varepsilon \subseteq [\tau - \varepsilon, \infty)$.

It is easily seen that this definition does not depend on the specific choice of φ^ε, hence $D_{[\tau,\infty)}$ is a well defined (continuous) operator on $\mathscr{C}_0^\infty$ and therefore a distribution. In fact, $D_{[\tau,\infty)} \in \mathbb{D}_{\mathrm{pw}\mathscr{C}^0}$ with $g_{[\tau,\infty)}$ as the corresponding piecewise-continuous function. The restriction to the closed interval $(-\infty, \tau]$ is defined analogously, and the restriction to arbitrary intervals can be defined as follows, $s, t \in \mathbb{R} \cup \{\infty\}$:

$$\begin{aligned} D_{(s,t)} &= D - D_{[t,\infty)} - D_{(-\infty,s]}, \\ D_{[s,t]} &= D_{[s,\infty)} - D_{(t,\infty)}, \\ D_{[s,t)} &= D_{[s,\infty)} - D_{[t,\infty)}, \\ D_{(s,t]} &= D_{(s,\infty)} - D_{(t,\infty)}. \end{aligned}$$

It is worth noting that it is not difficult to show that

$$\mathbb{D}_{\mathrm{pw}\mathscr{C}^0} = \left\{ D = g_{\mathbb{D}} + \sum_{t\in T} D_t \;\middle|\; \begin{array}{l} g \in \mathscr{C}^0_{\mathrm{pw}},\ T \subseteq \mathbb{R} \text{ is locally finite, } \forall t \in T \\ \exists n_t \in \mathbb{N}, \alpha_1^t, \ldots, \alpha_{n_t}^t \in \mathbb{R} : D_t = \sum_{k=0}^{n_t} \alpha_k^t \delta_t^{(k)} \end{array} \right\}$$

and the restriction of $D \in \mathbb{D}_{\mathrm{pw}\mathscr{C}^0}$ with the above representation $D = g_{\mathbb{D}} + \sum_{t\in T} D_t$ to an interval $I \in \mathbb{R}$ is given by

$$D_I = g_{I\mathbb{D}} + \sum_{t\in T\cap I} D_t.$$

The space of piecewise-continuous distributions also allows a pointwise evaluation in the following three senses, for $t \in \mathbb{R}$ and $D \in \mathbb{D}_{\mathrm{pw}\mathscr{C}^0}$ with corresponding $g \in \mathscr{C}^0_{\mathrm{pw}}$:

- the right sided evaluation: $D(t+) := g(t+)$,
- the left sided evaluation: $D(t-) = g(t-)$,
- the impulsive part: $D[t] := D_{[t,t]}$.

The following relates the restriction with the derivative.

Lemma 5.2 (Derivative of a restriction [13, Prop. 1]) *Let $D \in \mathbb{D}_{\mathrm{pw}\mathscr{C}^0}$ and assume $D' \in \mathbb{D}_{\mathrm{pw}\mathscr{C}^0}$ as well. Then, for any $\tau \in \mathbb{R}$,*

$$(D_{[\tau,\infty)})' = \big(D'\big)_{[\tau,\infty)} + D(\tau-)\delta_\tau.$$

Note that Cobb did not include the assumption $D' \in \mathbb{D}_{\mathrm{pw}\mathscr{C}^0}$ in his result; however, without this assumption the restriction of D' to some interval is not defined, because in general D' is not a piecewise-continuous distributions anymore (actually Cobb claims that the result is "obvious"; this is quite often a hint that there might be something wrong).

Remark 5.2 (A distributional motivation of Doetsch's past-aware derivative) Lemma 5.2 now gives a justification of the past-aware derivative (4.10) as propagated by Doetsch, because $D_{[0,\infty)}$ as well as $(D')_{[0,\infty)}$ are elements of the space $\bigcup_k \mathbb{D}_{\geq 0,k}$, however, D can still be non-zero on $(-\infty, 0)$ and $D(0-) \neq 0$ in general.

A connection between (consistent) distributional solution of (1.1) and the solutions of "distributional" DAEs (4.13) was established in [13, Prop. 2], a clearer connection, also allowing for inconsistent initial values, will be formulated in the context of piecewise-smooth distributions (see Sect. 5.4).

5.3 Impulsive-Smooth Distributions as Solution Space

The space of impulsive-smooth distributions was introduced by Hautus [26] (without denoting them as such) and was first used by this name in the context of optimal control problems [27]. Geerts [22–24] was then the first to use them as a solution space for DAEs. The space of impulsive-smooth distributions $\mathscr{C}_{\text{imp}}$ is defined in this earlier work as follows:

$$\mathscr{C}_{\text{imp}} := \left\{ D = g_{[0,\infty)\mathbb{D}} + D_{\text{imp}} \;\middle|\; g \in \mathscr{C}^\infty, D_{\text{imp}} = \sum_{i=0}^{k} \alpha_i \delta^{(i)}, k \in \mathbb{N}, \alpha_0, \ldots, \alpha_k \in \mathbb{R} \right\}.$$

Similar as in the Laplace transform approach, Geerts considers the distributional version (4.13) instead of (2.9) and he rewrites the (distributional) derivative as the convolution with δ':

$$\delta' * Ex = Ax + f + Ex_0\delta. \tag{5.1}$$

By viewing $\mathscr{C}_{\text{imp}}$ as a commutative algebra with convolution as multiplication, the distributional DAE can now be written as

$$pEx = Ax + f + Ex_0,$$

where $p = \delta'$ and δ is the unit with respect to convolution and hence denoted by one. The (time-domain) equation is now algebraically identically to the one obtained by the Laplace transformation approach without the need to think about problems like the existence of the Laplace transform and domain of convergence. In particular, existence and uniqueness results directly apply because no isomorphism between different solution spaces is needed. Nevertheless, the definition of $\mathscr{C}_{\text{imp}}$ still assumes that all involved variables are identically zero on $(-\infty, 0)$, hence speaking of inconsistent initial values is conceptionally as difficult as for the Laplace transform approach. In summary, viewing x_0 in (5.1) as the initial value for $x(0-)$ cannot be motivated within the impulsive-smooth distributional framework, because, by definition, $x(0-) = 0$.

In fact, there is no reason to consider variables which have to vanish on $(-\infty, 0)$: Rabier and Rheinboldt [37] were the first to use the space of impulsive-smooth distributions which can also be non-zero in the past. The formal definition is

$$\mathscr{C}_{\mathrm{imp}}\big(\mathbb{R}^*\big) := \left\{ D = f^-_{(-\infty,0)\mathbb{D}} + D_{\mathrm{imp}} + f^+_{(0,\infty)\mathbb{D}} \;\middle|\; \begin{array}{l} f^-, f^+ \in \mathscr{C}^\infty,\ D_{\mathrm{imp}} = \sum_{i=0}^{k} \alpha_i \delta^{(i)}, \\ k \in \mathbb{N}, \alpha_0, \ldots, \alpha_k \in \mathbb{R} \end{array} \right\}.$$

Clearly,

$$\mathscr{C}_{\mathrm{imp}} \subset \mathscr{C}_{\mathrm{imp}}\big(\mathbb{R}^*\big) \subset \mathbb{D}_{\mathrm{pw}\mathscr{C}^0} \subset \mathbb{D},$$

in particular, the three types of evaluation defined for piecewise-continuous distributions are also well defined for impulsive-smooth distribution as well as the distributional restriction. The main difference to the space of piecewise-continuous distribution is the fact that the space of impulsive-smooth distribution is closed under differentiation. In particular, impulsive-smooth distributions are arbitrarily often differentiable within the space of impulsive-smooth distributions.

Within the impulsive-smooth distributional framework the ITP (3.2)

$$\begin{aligned} x_{(-\infty,0)} &= x^0_{(-\infty,0)}, \\ (E\dot{x})_{[0,\infty)} &= (Ax + f)_{[0,\infty)} \end{aligned}$$

is well defined for all initial trajectories $x^0 \in \mathscr{C}_{\mathrm{imp}}(\mathbb{R}^*)^n$, all inhomogeneities $f \in \mathscr{C}_{\mathrm{imp}}(\mathbb{R}^*)^m$ and solutions x are sought in $\mathscr{C}_{\mathrm{imp}}(\mathbb{R}^*)^n$. In fact, the following result holds, which finally gives a satisfying and rigorous motivation for the incorporation of the (inconsistent) initial value as in (4.13).

Theorem 5.3 (Equivalent description of the ITP (3.2)) *Consider the ITP (3.2) within the impulsive-smooth distributional solution framework with fixed initial trajectory $x^0 \in \mathscr{C}_{\mathrm{imp}}(\mathbb{R}^*)^n$ and inhomogeneity $f \in \mathscr{C}_{\mathrm{imp}}(\mathbb{R}^*)^m$. Then $x \in \mathscr{C}_{\mathrm{imp}}(\mathbb{R}^*)^n$ solves the ITP (3.2) if, and only if, $z := x - x^0_{(-\infty,0)} = x_{[0,\infty)}$ solves*

$$\begin{aligned} z_{(-\infty,0)} &= 0, \\ (E\dot{z})_{[0,\infty)} &= (Az + f)_{[0,\infty)} + Ex^0(0-)\delta. \end{aligned} \tag{5.2}$$

Proof Let x be a solution of the ITP (3.2) and let $z = x_{[0,\infty)}$. Then, clearly, $z_{(-\infty,0)} = 0$. Furthermore,

$$\begin{aligned} (E\dot{z})_{[0,\infty)} &= (E\dot{x})_{[0,\infty)} - \big(E\big(x^0_{(-\infty,0)}\big)'\big)_{[0,\infty)} = (Ax + f)_{[0,\infty)} + Ex^0(0-)\delta \\ &= (Az + f)_{[0,\infty)} + Ex^0(0-)\delta, \end{aligned}$$

which shows that $z = x_{[0,\infty)}$ is indeed a solution of (5.2). On the other hand, let z be a solution of (5.2) and define $x := z + x^0_{(-\infty,0)}$. Then, clearly, $x_{(-\infty,0)} = x^0_{(-\infty,0)}$. Furthermore,

$$\begin{aligned}(E\dot{x})_{[0,\infty)} &= (E\dot{z})_{[0,\infty)} + \big(E\big(x^0_{(-\infty,0)}\big)'\big)_{[0,\infty)} \\ &= (Az + f)_{[0,\infty)} + Ex^0(0-)\delta - Ex^0(0-)\delta \\ &= (Ax + f)_{[0,\infty)}.\end{aligned}$$

□

Remark 5.3

1. If (5.2) is considered within the one-sided impulsive-smooth distributional framework, i.e. $f \in (\mathscr{C}_{\mathrm{imp}})^m$ and $z \in (\mathscr{C}_{\mathrm{imp}})^n$ then (5.2) simplifies to

$$E\dot{z} = Az + f + Ex^0(0-)\delta. \tag{5.3}$$

2. Comparing the result of Theorem 5.3 with the result of Cobb [13, Prop. 2] reveals three main differences: (1) Cobb only states one direction and not the equivalence, (2) instead of the ITP (3.2) Cobb just considers the original DAE (1.1), hence his result concerns only consistent solutions, (3) Cobb assumes that (5.3) has a unique solution.
3. Regularity of the matrix pair (E, A) is not assumed; in particular, neither is it assumed that for all inhomogeneities f there exist solutions to (3.2) and (5.2), nor is it assumed that solutions of (3.2) and (5.2) are uniquely given for fixed initial trajectory and fixed inhomogeneity. However, due to the established equivalence all existence and uniqueness results obtained for (5.3) carry over to the ITP (3.2).

Although Rabier and Rheinboldt [37] introduced the space of impulsive-smooth distribution which allow a clean treatment of the ITP (3.2), they did not follow this approach. Instead, they redefine the inhomogeneity to make inconsistent initial values consistent. To this end, let $x^0 \in \mathscr{C}_{\mathrm{imp}}(\mathbb{R}^*)^n$ be a given initial trajectory and $f \in \mathscr{C}_{\mathrm{imp}}(\mathbb{R}^*)^m$ a given inhomogeneity and consider the ITP-DAE

$$\begin{aligned}x_{(-\infty,0)} &= x^0_{(-\infty,0)}, \\ E\dot{x} &= Ax + f_{\mathrm{ITP}},\end{aligned} \tag{5.4}$$

where

$$f_{\mathrm{ITP}} := E\dot{x}^0_{(-\infty,0)} - Ax^0_{(-\infty,0)} + f_{[0,\infty)}.$$

Note that $x_{(-\infty,0)} = x^0_{(-\infty,0)}$ already implies, due to the special choice of f_{ITP}, that

$$(E\dot{x})_{(-\infty,0)} = (Ax + f_{\mathrm{ITP}})_{(-\infty,0)},$$

which shows that (5.4) is in fact equivalent to the ITP (3.2). However, the form of (5.4) has certain disadvantages compared to the ITP formulation (3.2):

1. The second equation of (5.4) suggest that the DAE (1.1) is valid globally (just with a different inhomogeneity), which conflicts with the intuition that an inconsistent initial value is due to the fact that the system description (1.1) is only valid on $[0, \infty)$ and not in the past.
2. In (5.4) the past trajectory of x is formally determined by two equations which could in general be conflicting (depending on the choice of f_{ITP}).
3. When studying an autonomous system (i.e. without the presence of an inhomogeneity), the formulation (5.4) formally leaves the class of autonomous systems.

On the other hand, an interesting advantage of the formulation (5.4) is that, due to Remark 5.1, (5.4) makes sense even when x is an arbitrary distribution and f as well as x^0 are such that f_{ITP} is well defined. In fact, Rabier and Rheinboldt [37, Thm. 4.1] do consider arbitrary distributions $x \in \mathbb{D}^n$ and show that under certain regularity assumptions the solutions are in fact impulsive-smooth.

5.4 Piecewise-Smooth Distributions as Solution Space

Comparing Cobb's piecewise-continuous distributional solution framework with the impulsive-smooth distributional solution framework the following differences are apparent:

1. $\mathbb{D}_{\text{pw}\mathscr{C}^0}$ is not closed under differentiation.
2. $\mathscr{C}_{\text{imp}}(\mathbb{R}^*)$ does not allow non-smooth inhomogeneities away from $t = 0$.

Rabier and Rheinboldt [37] seem to be aware of the latter problem as they introduce the space $\mathscr{C}_{\text{imp}}(\mathbb{R} \setminus \mathscr{S})$, where $\mathscr{S} = \{t_i \in \mathbb{R} | i \in \mathbb{Z}\}$ is a strictly ordered set with $t_i \to \pm\infty$ as $i \to \pm\infty$ and $D \in \mathscr{C}_{\text{imp}}(\mathbb{R} \setminus \mathscr{S})$ is such that $D_{(t_i, t_{i+1})}$ is induced by the corresponding restriction of a smooth function. A similar idea is proposed in [25], however, in both cases the resulting distributional space is not studied in detail. A more detailed treatment can be found in [45, 46] where, in the spirit of Cobb's definition, the space of piecewise-smooth distributions is defined as follows:

$$\mathbb{D}_{\text{pw}\mathscr{C}^\infty} := \left\{ D = f_{\mathbb{D}} + \sum_{t \in T} D_t \;\middle|\; \begin{array}{l} f \in \mathscr{C}^\infty_{\text{pw}}, T \subseteq \mathbb{R} \\ \text{locally finite } \forall t \in T : D_t \in \operatorname{span}\{\delta_t, \delta_t', \delta_t'', \ldots\} \end{array} \right\},$$

where $f \in \mathscr{C}^\infty_{\text{pw}}$ is a piecewise-smooth function if, and only if, there exists a strictly ordered locally finite set $\{s_i \in \mathbb{R} | i \in \mathbb{Z}\}$ and $f_i \in \mathscr{C}^\infty$, $i \in \mathbb{Z}$, such that $f = \sum_{i \in \mathbb{Z}} f_i|_{[s_i, s_{i+1})}$. Clearly,

$$\mathscr{C}_{\text{imp}}(\mathbb{R}^*) \subset \mathbb{D}_{\text{pw}\mathscr{C}^\infty} \subset \mathbb{D}_{\text{pw}\mathscr{C}^0},$$

and the space of piecewise-smooth distributions resolves each of the above mentioned drawbacks of the piecewise-continuous and impulsive-smooth distributions.

However, the major advantage of considering the space of piecewise-smooth distributions becomes apparent when considering time-varying DAEs:

$$E(t)\dot{x}(t) = A(t)x(t) + f(t). \tag{5.5}$$

If the coefficient matrices $E(\cdot)$ and $A(\cdot)$ are smooth it is no problem to use any of the above distributional solution concepts because the product of a smooth function with any distribution is well defined so that (5.5) makes sense as an equation of distributions. In the discussion of the drawbacks of the Laplace transform approach it was already mentioned that an inconsistent initial value could be seen as the results from the presence of a time-varying system. In fact, the ITP (3.2) can be reformulated as the following time-varying DAE [45, Thm. 3.1.7]:

$$E_{\mathrm{ITP}}(t)\dot{x}(t) = A_{\mathrm{ITP}}(t)x(t) + f_{\mathrm{ITP}}(t),$$

where

$$E_{\mathrm{ITP}}(t) = \begin{cases} 0, & t < 0, \\ E, & t \geq 0, \end{cases} \qquad A_{\mathrm{ITP}}(t) = \begin{cases} I, & t < 0, \\ A, & t \geq 0, \end{cases}$$

$$f_{\mathrm{ITP}}(t) = \begin{cases} -x^0(t), & t < 0, \\ f(t), & t \geq 0. \end{cases}$$

The problem is now that the time-varying coefficient matrices are not smooth anymore so that the multiplication with a distribution is not well defined. Rabier and Rheinboldt [37] treated already time-varying DAEs (5.5); however, the interpretation of inconsistent initial values as a time-variant DAE with non-smooth coefficients did not occur to them, maybe because they considered (5.4) where formally the original DAE (with a special choice of the inhomogeneity) with smooth coefficient is considered globally (i.e. in the whole of $\mathbb{R}$ and not only on $[0, \infty)$). Another important motivation for studying time-varying DAEs with non-smooth coefficient matrices is switched DAEs [47]:

$$E_\sigma \dot{x} = A_\sigma x + f,$$

where $\sigma : \mathbb{R} \to \{1, 2, \ldots, P\}$, $P \in \mathbb{N}$, and $(E_1, A_1), \ldots, (E_P, A_P)$ are constant matrices.

It turns out that for the space of piecewise-smooth distributions a (non-commutative) multiplication can be defined, named *Fuchssteiner multiplication* after [19, 20], which in particular defines the multiplication of a piecewise-smooth function with a piecewise-smooth distribution. Hence (5.5) makes sense even for coefficient matrices which are only piecewise-smooth.

Remark 5.4 (The square of the Dirac impulse) The multiplication of distributions occurs several times in the context of DAEs. The different approaches can be best illustrated by the different treatments of the square of the Dirac impulse:

1. In the context of impulsive-smooth distributions [22, 23, 27] convolution is viewed as a multiplication and the Dirac impulse is the unit element for that multiplication. Hence $\delta^2 = \delta$ in this framework.
2. The Fuchssteiner multiplication for piecewise-smooth distributions yields
$$\delta^2 = 0.$$
3. It is well known that a commutative and associative multiplication which generalizes the multiplication of functions to distributions is not possible in general, but when enlarging the space of distributions the square of the Dirac impulse is well defined (but not a classical distribution). In the context of DAEs this approach was considered in [44], where the square of the Dirac impulse occurs in the analysis of the connection energy (the product of the voltage and current).

Within the framework of piecewise-smooth distributions it is now possible to show [45] that the ITP (3.2) is uniquely solvable for all initial trajectories and all inhomogeneities if, and only if, the matrix pair (E, A) is regular. In particular, the impulses and jumps derived in this framework [47, Thm. 6.5.1] are identical to (4.15) and (4.16) obtained via the Laplace transform approach.

6 Conclusion

The role of the Wong sequences of the matrix pair (E, A) for characterizing the (classical) solutions was highlighted. In particular, explicit solution formulas where given which are similar to the ones obtained for linear ODEs. The quasi-Kronecker form (QKF) and quasi-Weierstraß form (QWF) play a prominent role. For time-varying DAEs with analytical coefficients a time-varying QWF is available, however, time-varying Wong sequences and their connection to a time-varying QWF (or even QKF) have not been studied yet. The problem of inconsistent initial values was discussed and it was shown how the Laplace transform was used to treat this problem. However, it is argued that the Laplace transform approach cannot justify the notion of an inconsistent initial value. With the help of certain distributional solution spaces the notion of inconsistent initial values can be treated in a satisfying way and it also justifies the Laplace transform approach.

References

1. Aplevich, J.D.: Implicit Linear Systems. Lecture Notes in Control and Information Sciences, vol. 152. Springer, Berlin (1991)
2. Armentano, V.A.: The pencil $(sE - A)$ and controllability-observability for generalized linear systems: a geometric approach. SIAM J. Control Optim. **24**, 616–638 (1986)
3. Berger, T., Trenn, S.: Addition to: "The quasi-Kronecker form for matrix pencils". SIAM. J. Matrix Anal. Appl. **34**(1), 94–101 (2013)

4. Berger, T., Trenn, S.: The quasi-Kronecker form for matrix pencils. SIAM J. Matrix Anal. Appl. **33**(2), 336–368 (2012)
5. Berger, T., Ilchmann, A., Trenn, S.: The quasi-Weierstraß form for regular matrix pencils. Linear Algebra Appl. **436**(10), 4052–4069 (2012). doi:10.1016/j.laa.2009.12.036
6. Bernhard, P.: On singular implicit linear dynamical systems. SIAM J. Control Optim. **20**(5), 612–633 (1982)
7. Brenan, K.E., Campbell, S.L., Petzold, L.R.: Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations. North-Holland, Amsterdam (1989)
8. Campbell, S.L.: Singular Systems of Differential Equations I. Pitman, New York (1980)
9. Campbell, S.L.: Singular Systems of Differential Equations II. Pitman, New York (1982)
10. Campbell, S.L., Petzold, L.R.: Canonical forms and solvable singular systems of differential equations. SIAM J. Algebr. Discrete Methods **4**, 517–521 (1983)
11. Campbell, S.L., Meyer, C.D. Jr., Rose, N.J.: Applications of the Drazin inverse to linear systems of differential equations with singular constant coefficients. SIAM J. Appl. Math. **31**(3), 411–425 (1976). http://link.aip.org/link/?SMM/31/411/1. doi:10.1137/0131035
12. Cobb, J.D.: On the solution of linear differential equations with singular coefficients. J. Differ. Equ. **46**, 310–323 (1982)
13. Cobb, J.D.: Controllability, observability and duality in singular systems. IEEE Trans. Autom. Control **AC-29**, 1076–1082 (1984)
14. Dai, L.: Singular Control Systems. Lecture Notes in Control and Information Sciences, vol. 118. Springer, Berlin (1989)
15. Dieudonné, J.: Sur la réduction canonique des couples des matrices. Bull. Soc. Math. Fr. **74**, 130–146 (1946)
16. Doetsch, G.: Introduction to the Theory and Application of the Laplace Transformation. Springer, Berlin (1974)
17. Drazin, M.P.: Pseudo-inverses in associative rings and semigroups. Am. Math. Mon. **65**(7), 506–514 (1958)
18. Frasca, R., Çamlıbel, M.K., Goknar, I.C., Iannelli, L., Vasca, F.: Linear passive networks with ideal switches: consistent initial conditions and state discontinuities. IEEE Trans. Circuits Syst. I, Fundam. Theory Appl. **57**(12), 3138–3151 (2010)
19. Fuchssteiner, B.: Eine assoziative Algebra über einen Unterraum der Distributionen. Math. Ann. **178**, 302–314 (1968)
20. Fuchssteiner, B.: Algebraic foundation of some distribution algebras. Stud. Math. **76**, 439–453 (1984)
21. Gantmacher, F.R.: The Theory of Matrices, vols. I & II. Chelsea, New York (1959)
22. Geerts, A.H.W.T.: Invariant subspaces and invertibility properties for singular systems: the general case. Linear Algebra Appl. **183**, 61–88 (1993). doi:10.1016/0024-3795(93)90424-M
23. Geerts, A.H.W.T.: Solvability conditions, consistency and weak consistency for linear differential-algebraic equations and time-invariant linear systems: the general case. Linear Algebra Appl. **181**, 111–130 (1993)
24. Geerts, A.H.W.T.: Regularity and singularity in linear-quadratic control subject to implicit continuous-time systems. IEEE Proc. Circuits Syst. Signal Process. **13**, 19–30 (1994)
25. Geerts, A.H.W.T., Schumacher, J.M.H.: Impulsive-smooth behavior in multimode systems. Part I: state-space and polynomial representations. Automatica **32**(5), 747–758 (1996)
26. Hautus, M.L.J.: The formal Laplace transform for smooth linear systems. In: Marchesini, G., Mitter, S.K. (eds.) Mathematical Systems Theory. Lecture Notes in Economics and Mathematical Systems, vol. 131, pp. 29–47. Springer, New York (1976)
27. Hautus, M.L.J., Silverman, L.M.: System structure and singular control. Linear Algebra Appl. **50**, 369–402 (1983)
28. Kronecker, L.: Algebraische Reduction der Schaaren bilinearer Formen. Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin, pp. 1225–1237 (1890)
29. Kuijper, M.: First-Order Representations of Linear Systems. Birkhäuser, Boston (1994)
30. Kunkel, P., Mehrmann, V.: Differential-Algebraic Equations. Analysis and Numerical Solution. EMS Publishing House, Zürich (2006)

31. Lamour, R., März, R., Tischendorf, C.: Differential Algebraic Equations: A Projector Based Analysis. Differential-Algebraic Equations Forum, vol. 1. Springer, Heidelberg (2013)
32. Lewis, F.L.: A survey of linear singular systems. IEEE Proc. Circuits Syst. Signal Process. **5**(1), 3–36 (1986)
33. Liberzon, D., Trenn, S.: Switched nonlinear differential algebraic equations: solution theory, Lyapunov functions, and stability. Automatica **48**(5), 954–963 (2012). doi:10.1016/j.automatica.2012.02.041
34. Lundberg, K.H., Miller, H.R., Trumper, D.L.: Initial conditions, generalized functions, and the Laplace transform. IEEE Control Syst. Mag. **27**(1), 22–35 (2007). doi:10.1109/MCS.2007.284506
35. Opal, A., Vlach, J.: Consistent initial conditions of linear switched networks. IEEE Trans. Circuits Syst. **37**(3), 364–372 (1990)
36. Owens, D.H., Debeljkovic, D.L.: Consistency and Liapunov stability of linear descriptor systems: a geometric analysis. IMA J. Math. Control Inf. **2**, 139–151 (1985)
37. Rabier, P.J., Rheinboldt, W.C.: Time-dependent linear DAEs with discontinuous inputs. Linear Algebra Appl. **247**, 1–29 (1996)
38. Rabier, P.J., Rheinboldt, W.C.: Theoretical and numerical analysis of differential-algebraic equations. In: Ciarlet, P.G., Lions, J.L. (eds.) Handbook of Numerical Analysis, vol. VIII, pp. 183–537. Elsevier, Amsterdam (2002)
39. Reißig, G., Boche, H., Barton, P.I.: On inconsistent initial conditions for linear time-invariant differential-algebraic equations. IEEE Trans. Circuits Syst. I, Fundam. Theory Appl. **49**(11), 1646–1648 (2002)
40. Rosenbrock, H.H.: State Space and Multivariable Theory. Wiley, New York (1970)
41. Schwartz, L.: Théorie des Distributions. Hermann, Paris (1957, 1959)
42. Sincovec, R.F., Erisman, A.M., Yip, E.L., Epton, M.A.: Analysis of descriptor systems using numerical algorithms. IEEE Trans. Autom. Control **AC-26**, 139–147 (1981)
43. Tanwani, A., Trenn, S.: On observability of switched differential-algebraic equations. In: Proc. 49th IEEE Conf. Decis. Control, Atlanta, USA, pp. 5656–5661 (2010)
44. Tolsa, J., Salichs, M.: Analysis of linear networks with inconsistent initial conditions. IEEE Trans. Circuits Syst. **40**(12), 885–894 (1993). doi:10.1109/81.269029
45. Trenn, S.: Distributional differential algebraic equations. Ph.D. thesis, Institut für Mathematik, Technische Universität Ilmenau, Universitätsverlag Ilmenau, Ilmenau, Germany (2009). http://www.db-thueringen.de/servlets/DocumentServlet?id=13581
46. Trenn, S.: A normal form for pure differential algebraic systems. Linear Algebra Appl. **430**(4), 1070–1084 (2009). doi:10.1016/j.laa.2008.10.004
47. Trenn, S.: Switched differential algebraic equations. In: Vasca, F., Iannelli, L. (eds.) Dynamics and Control of Switched Electronic Systems—Advanced Perspectives for Modeling, Simulation and Control of Power Converters, pp. 189–216. Springer, London (2012). Chap. 6
48. Verghese, G.C., Levy, B.C., Kailath, T.: A generalized state-space for singular systems. IEEE Trans. Autom. Control **AC-26**(4), 811–831 (1981)
49. Weierstraß, K.: Zur Theorie der bilinearen und quadratischen Formen. Berl. Monatsb., pp. 310–338 (1868)
50. Wilkinson, J.H.: Linear differential equations and Kronecker's canonical form. In: de Boor, C., Golub, G.H. (eds.) Recent Advances in Numerical Analysis, pp. 231–265. Academic Press, New York (1978)
51. Wong, K.T.: The eigenvalue problem $\lambda Tx + Sx$. J. Differ. Equ. **16**, 270–280 (1974)
52. Yip, E.L., Sincovec, R.F.: Solvability, controllability and observability of continuous descriptor systems. IEEE Trans. Autom. Control **AC-26**, 702–707 (1981)

Port-Hamiltonian Differential-Algebraic Systems

A.J. van der Schaft

Abstract The basic starting point of port-Hamiltonian systems theory is *network modeling*; considering the overall physical system as the *interconnection* of simple subsystems, mutually influencing each other via energy flow. As a result of the interconnections *algebraic constraints* between the state variables commonly arise. This leads to the description of the system by *differential-algebraic equations* (DAEs), i.e., a combination of ordinary differential equations with algebraic constraints. The basic point of view put forward in this survey paper is that the differential-algebraic equations that arise are not just arbitrary, but are endowed with a special mathematical structure; in particular with an underlying geometric structure known as a Dirac structure. It will be discussed how this knowledge can be exploited for analysis and control.

Keywords Port-Hamiltonian systems · Passivity · Algebraic constraints · Kinematic constraints · Casimirs · Switching systems · Dirac structure · Interconnection

Mathematics Subject Classification (2010) 34A09 · 37J05 · 70G45 · 93B10 · 93B27 · 93C10

1 Introduction to Port-Hamiltonian Differential-Algebraic Systems

The framework of port-Hamiltonian systems is intended to provide a systematic approach to the modeling, analysis, simulation and control of, possibly large-scale, multi-physics systems; see [9, 15, 19, 20, 24, 25, 29, 31–34, 38, 39] for some key references. Although the framework includes distributed-parameter systems as well, we will focus in this paper on lumped-parameter, i.e., finite-dimensional, systems.

A.J. van der Schaft (✉)
Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, PO Box 407, 9700 AK Groningen, The Netherlands
e-mail: A.J.van.der.Schaft@rug.nl

A. Ilchmann, T. Reis (eds.), *Surveys in Differential-Algebraic Equations I*, Differential-Algebraic Equations Forum,
DOI 10.1007/978-3-642-34928-7_5,

The basic starting point of port-Hamiltonian systems theory is (power-based) *network modeling*; considering the overall system as the *interconnection* of simple subsystems, mutually influencing each other via energy flow [27]. As a result of the interconnections *algebraic constraints* between the state variables commonly arise. This leads to the description of the system by *differential-algebraic equations* (DAEs), i.e., a combination of ordinary differential equations with algebraic constraints. However, the basic point of view put forward in this paper is that the differential-algebraic equations that arise *are not just arbitrary differential-algebraic equations*, but are endowed with a special mathematical structure, which may be fruitfully used for analysis, simulation and control.

As a motivating and guiding example for the theory surveyed in this paper we will start with the following example.

1.1 A Motivating Example

Consider an LC-circuit consisting of two capacitors and one inductor, all in parallel. Naturally this system can be seen as the interconnection of three subsystems, the two capacitors and the inductor, interconnected by Kirchhoff's current and voltage laws. The capacitors (first assumed to be linear) are described by the following dynamical equations:

$$\begin{aligned} \dot{Q}_i &= I_i, \\ V_i &= \frac{Q_i}{C_i}, \end{aligned} \qquad i = 1, 2. \tag{1.1}$$

Here I_i and V_i are the currents through, respectively the voltages across, the two capacitors, and C_i are their capacitances. Furthermore, Q_i are the *charges* stored at the capacitors; regarded as basic state variables.[1]

Similarly, the linear inductor is described by the dynamical equations

$$\begin{aligned} \dot{\varphi} &= V_L, \\ I_L &= \frac{\varphi}{L}, \end{aligned} \tag{1.2}$$

where I_L is the current through the inductor, and V_L is the voltage across the inductor. Here the (magnetic) *flux* φ is taken as the state variable of the inductor, and L denotes its inductance.

Parallel interconnection of these three subsystems by Kirchhoff's laws amounts to the interconnection equations

$$V_1 = V_2 = V_L, \qquad I_1 + I_2 + I_L = 0, \tag{1.3}$$

[1] In the port-Hamiltonian formulation there is a clear preference for taking the charges to be the state variables instead of the voltages V_i. Although this comes at the expense of the introduction of extra variables, it will turn out to be very advantageous from a geometric point of view.

where the equation $V_1 = V_2$ gives rise to the algebraic constraint

$$\frac{Q_1}{C_1} = \frac{Q_2}{C_2} \tag{1.4}$$

relating the two state variables Q_1, Q_2.

There are multiple ways to describe the total system. One is to regard either I_1 or I_2 as a *Lagrange multiplier* for the constraint $\frac{Q_1}{C_1} - \frac{Q_2}{C_2} = 0$. Indeed, by defining $\lambda = I_1$ one may write the total system as

$$\begin{aligned} \begin{bmatrix} \dot{Q}_1 \\ \dot{Q}_2 \\ \dot{\varphi} \end{bmatrix} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{Q_1}{C_1} \\ \frac{Q_2}{C_2} \\ \frac{\varphi}{L} \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \lambda, \\ 0 &= \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \frac{Q_1}{C_1} \\ \frac{Q_2}{C_2} \\ \frac{\varphi}{L} \end{bmatrix}, \end{aligned} \tag{1.5}$$

where the algebraic constraint $\frac{Q_1}{C_1} - \frac{Q_2}{C_2} = 0$ represented in the last equation of (1.5) can be seen to give rise to a *constraint current* $[1\ -1\ 0]^T \lambda$, which is added to the first three ordinary differential equations in (1.5).

Next one may *eliminate* the Lagrange multiplier λ by pre-multiplying the differential equations by the matrix

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Together with the algebraic constraint this yields the *differential-algebraic system*

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{Q}_1 \\ \dot{Q}_2 \\ \dot{\varphi} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \frac{Q_1}{C_1} \\ \frac{Q_2}{C_2} \\ \frac{\varphi}{L} \end{bmatrix}. \tag{1.6}$$

Equations (1.5) and (1.6) are different representations of the same *port-Hamiltonian system* defined by the LC-circuit, which is geometrically (i.e., coordinate-free) described by a Dirac structure and constitutive relations corresponding to energy-storage. In this example the Dirac structure is given by the linear space

$$\begin{aligned} \mathcal{D} := \Bigg\{ (f, e) \in \mathbb{R}^3 \times \mathbb{R}^3 \mid f = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}, e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}, \\ \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} + \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = 0 \Bigg\} \end{aligned} \tag{1.7}$$

having the characteristic property that $e^T f = 0$ for all $(f, e) \in \mathcal{D}$ (total power is zero), and moreover having maximal dimension with regard to this property (in this case $\dim \mathcal{D} = 3$). The two representations (1.5) and (1.6) correspond to two different representations of this same Dirac structure.

Furthermore, the constitutive relations of energy-storage are given by $f = [f_1 \ f_2 \ f_3]^T = -[\dot{Q}_1 \ \dot{Q}_2 \ \dot{\varphi}]$, and $e = [e_1 \ e_2 \ e_3]^T = [\frac{Q_1}{C_1} \ \frac{Q_2}{C_2} \ \frac{\varphi}{L}]^T$, where the last vector is the gradient vector of the total stored energy, or *Hamiltonian*

$$H(Q_1, Q_2, \varphi) := \frac{Q_1^2}{2C_1} + \frac{Q_2^2}{2C_2} + \frac{\varphi^2}{2L}. \tag{1.8}$$

We may easily replace the linear constitutive relations of the capacitors and the inductor by more general nonlinear ones, corresponding to a general non-quadratic Hamiltonian

$$H(Q_1, Q_2, \varphi) = H_1(Q_1) + H_2(Q_2) + H_3(\varphi) \tag{1.9}$$

with $H_i(Q_i), i = 1, 2$, denoting the electric energies of the two capacitors, and $H_3(\varphi)$ the magnetic energy of the inductor. Then the resulting dynamics are given by the *nonlinear* differential-algebraic equations

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{Q}_1 \\ \dot{Q}_2 \\ \dot{\varphi} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \frac{dH_1}{dQ_1}(Q_1) \\ \frac{dH_2}{dQ_2}(Q_2) \\ \frac{dH_3}{d\varphi}(\varphi) \end{bmatrix}. \tag{1.10}$$

In the port-Hamiltonian description there is thus a clear *separation*[2] between the constitutive relations of the elementary subsystems (captured by the Hamiltonian H), and the interconnection structure (formalized by the Dirac structure $\mathcal{D}$). This has several advantages in terms of flexibility and standardization (e.g., one may replace linear subsystems by nonlinear ones, without changing the interconnection structure), and will give rise to a completely *compositional* theory of network models of physical systems: the interconnection of port-Hamiltonian systems defines another port-Hamiltonian system, where the Hamiltonians are simply added and the new Dirac structure results from the composition of the Dirac structures of the interconnected individual physical systems.

From a DAE perspective it may be noted that the algebraic constraint $\frac{Q_1}{C_1} = \frac{Q_2}{C_2}$ is of *index one*. In fact, under reasonable assumptions on the Hamiltonian this will turn out to be a general property of port-Hamiltonian differential-algebraic systems.

[2]Note that this separation is already present in the geometric description of Hamiltonian dynamics in classical mechanics; see e.g. [1]. There the dynamics is defined with the use of the Hamiltonian and the *symplectic structure* on the phase space of the system. Dirac structures form a generalization of symplectic structures, and allow the inclusion of *algebraic constraints*. Note furthermore that the symplectic structure in classical mechanics is commonly determined by the geometry of the *configuration space*, while the Dirac structure of a port-Hamiltonian system captures its *network topology*.

In the above example, the subsystems are all energy-storing elements (capacitors, inductors), and thus the total energy (Hamiltonian) is *preserved* along solutions of the differential-algebraic equations. The framework, however, extends to energy-dissipating elements (such as resistors), in which case the Hamiltonian will decrease along solutions. Furthermore, in the above example there are no external inputs to the system (such as voltage or current sources). In the port-Hamiltonian framework these are, however, immediately incorporated, and are in fact essential to describe the interconnection of port-Hamiltonian systems. By including external ports in the system description it will follow that along system trajectories of the port-Hamiltonian differential-algebraic system $\frac{dH}{dt}$ is always less than or equal than the power supplied to the system through these external ports, i.e., *passivity*.

Finally, the port-Hamiltonian formalism emphasizes the *analogy* between physical system models. The same system of equations as in (1.5) or (1.6) also results from the modeling of a system of two rigidly coupled masses connected to a single spring. In this case, the rigid coupling between the two masses with kinetic energies

$$H_i(p_i) = \frac{p_i^2}{2m_i}, \quad i = 1, 2$$

(where p_1, p_2 denote the momenta of the masses m_1, m_2) is given by the (index one) algebraic constraint

$$\frac{p_1}{m_1} = v_1 = v_2 = \frac{p_2}{m_2} \tag{1.11}$$

with v_1, v_2 denoting the velocities of the masses. Note that this is different from formulating the rigid coupling between two masses by the (index two) algebraic constraint

$$q_1 = q_2 \tag{1.12}$$

in terms of the *positions* q_i, $i = 1, 2$, of the two masses. In fact, the constraint (1.11) results from differentiation of (1.12). Indeed, in the port-Hamiltonian approach there is a preference for modeling constraints in mechanical systems as *kinematic constraints* (which can be holonomic, as in this simple example, or nonholonomic). This will be discussed in more detail later in this paper.

The contents of Sects. 2, 3, 5, 6, 7 of the present paper are a thoroughly reworked version of material that appeared before in [34], emphasizing and expanding the differential-algebraic nature of port-Hamiltonian systems.

2 Definition of Port-Hamiltonian Systems

In this section we will provide the general geometric (coordinate-free) definition of a finite-dimensional port-Hamiltonian system, and discuss different examples and subclasses.

A port-Hamiltonian system can be represented as in Fig. 1. Central in the definition of a port-Hamiltonian system is the notion of a *Dirac structure*, denoted in

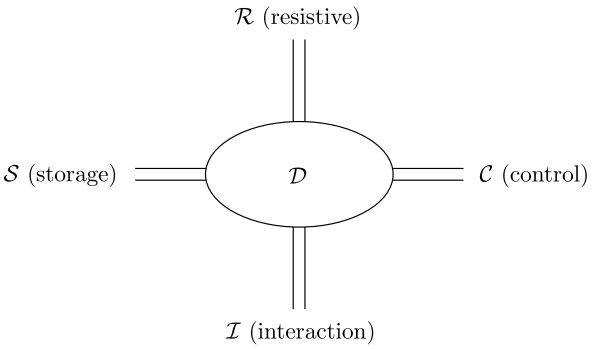


Fig. 1 Port-Hamiltonian system

Fig. 1 by $\mathcal{D}$. Basic property of a Dirac structure is *power-preservation*: the Dirac structure links the port variables in such a way that the total power associated with all the port-variables is zero.

The port variables entering the Dirac structure have been split in Fig. 1 in different parts. First, there are two *internal* ports. One, denoted by $\mathcal{S}$, corresponds to energy-storage and the other one, denoted by $\mathcal{R}$, corresponds to internal energy-dissipation (resistive elements). Second, two *external* ports are distinguished. The external port denoted by $\mathcal{C}$ is the port that is accessible for controller action. Also the presence of *sources* may be included in this port. Finally, the external port denoted by $\mathcal{I}$ is the interaction port, defining the interaction of the system with (the rest of) its environment.

2.1 Dirac Structures

We start with a finite-dimensional linear space of *flows* $\mathcal{F}$. The elements of $\mathcal{F}$ will be denoted by $f \in \mathcal{F}$, and are called *flow vectors*. The space of *efforts* is given by the *dual* linear space $\mathcal{E} := \mathcal{F}^*$, and its elements are denoted by $e \in \mathcal{E}$. In the case of $\mathcal{F} = \mathbb{R}^k$ the space of efforts is $\mathcal{E} = (\mathbb{R}^k)^*$, and as the elements $f \in \mathbb{R}^k$ are commonly written as *column* vectors the elements $e \in (\mathbb{R}^k)^*$ are appropriately represented as *row* vectors. Then the *total space* of flow and effort variables is $\mathcal{F} \times \mathcal{F}^*$, and will be called the space of *port variables*. On the total space of port variables, the *power* is defined by

$$P = \langle e \mid f \rangle, \quad (f, e) \in \mathcal{F} \times \mathcal{F}^*, \tag{2.1}$$

where $\langle e \mid f \rangle$ denotes the duality product, that is, the linear functional $e \in \mathcal{F}^*$ acting on $f \in \mathcal{F}$. Often we will write the flow f and effort e both as column vectors, in which case $\langle e \mid f \rangle = e^T f$.

Definition 2.1 A *Dirac structure* on $\mathcal{F} \times \mathcal{F}^*$ is a subspace $\mathcal{D} \subset \mathcal{F} \times \mathcal{F}^*$ such that

(i) $\langle e \mid f \rangle = 0$, for all $(f, e) \in \mathcal{D}$,
(ii) $\dim \mathcal{D} = \dim \mathcal{F}$.

Property (i) corresponds to *power-preservation*, and expresses the fact that the total power entering (or leaving) a Dirac structure is zero. It can be shown that the *maximal dimension* of any subspace $\mathcal{D} \subset \mathcal{F} \times \mathcal{F}^*$ satisfying property (i) is equal to $\dim \mathcal{F}$. Instead of proving this directly, we will give an equivalent definition of a Dirac structure from which this claim immediately follows. Furthermore, this equivalent definition of a Dirac structure has the advantage that it generalizes to the case of an *infinite-dimensional* linear space $\mathcal{F}$, leading to the definition of an infinite-dimensional Dirac structure. This will be instrumental in the definition of *distributed-parameter* port-Hamiltonian systems [39].

In order to give this equivalent characterization of a Dirac structure, let us look more closely at the geometric structure of the total space of flow and effort variables $\mathcal{F} \times \mathcal{F}^*$. Closely related to the definition of power, there exists a canonically defined *bilinear form* $\langle\langle\, , \rangle\rangle$ on the space $\mathcal{F} \times \mathcal{F}^*$, defined as

$$\big\langle\big\langle \big(f^a, e^a\big), \big(f^b, e^b\big)\big\rangle\big\rangle := \big\langle e^a \mid f^b\big\rangle + \big\langle e^b \mid f^a\big\rangle \tag{2.2}$$

with $(f^a, e^a), (f^b, e^b) \in \mathcal{F} \times \mathcal{F}^*$. Note that this bilinear form is *indefinite*, that is, $\langle\langle (f, e), (f, e)\rangle\rangle$ may be positive or negative. However, it is *non-degenerate*, that is, $\langle\langle (f^a, e^a), (f^b, e^b)\rangle\rangle = 0$ for all (f^b, e^b) implies that $(f^a, e^a) = 0$.

Proposition 2.1 ([8, 12]) *A (constant) Dirac structure on $\mathcal{F} \times \mathcal{F}^*$ is a subspace $\mathcal{D} \subset \mathcal{F} \times \mathcal{F}^*$ such that*

$$\mathcal{D} = \mathcal{D}^{\perp\!\!\!\perp}, \tag{2.3}$$

where $\perp\!\!\!\perp$ denotes the orthogonal complement with respect to the bilinear form $\langle\langle\, , \rangle\rangle$.

Proof Let $\mathcal{D}$ satisfy (2.3). Then for every $(f, e) \in \mathcal{D}$

$$0 = \big\langle\big\langle (f, e), (f, e)\big\rangle\big\rangle = \langle e \mid f\rangle + \langle e \mid f\rangle = 2\langle e \mid f\rangle.$$

By non-degeneracy of $\langle\langle\, , \rangle\rangle$

$$\dim \mathcal{D}^{\perp\!\!\!\perp} = \dim\big(\mathcal{F} \times \mathcal{F}^*\big) - \dim \mathcal{D} = 2 \dim \mathcal{F} - \dim \mathcal{D}$$

and hence property (2.3) implies $\dim \mathcal{D} = \dim \mathcal{F}$. Conversely, let $\mathcal{D}$ be a Dirac structure and thus satisfying properties (i) and (ii) of Definition 2.1. Let $(f^a, e^a), (f^b, e^b)$ be any vectors contained in $\mathcal{D}$. Then by linearity also $(f^a + f^b, e^a + e^b) \in \mathcal{D}$. Hence by property (i)

$$\begin{aligned} 0 &= \big\langle e^a + e^b \mid f^a + f^b\big\rangle \\ &= \big\langle e^a \mid f^b\big\rangle + \big\langle e^b \mid f^a\big\rangle + \big\langle e^a \mid f^a\big\rangle + \big\langle e^b \mid f^b\big\rangle \\ &= \big\langle e^a \mid f^b\big\rangle + \big\langle e^b \mid f^a\big\rangle = \big\langle\big\langle \big(f^a, e^a\big), \big(f^b, e^b\big)\big\rangle\big\rangle \end{aligned} \tag{2.4}$$

since by another application of property (i), $\langle e^a \mid f^a\rangle = \langle e^b \mid f^b\rangle = 0$. This implies that $\mathcal{D} \subset \mathcal{D}^{\perp\!\!\!\perp}$. Furthermore, by property (ii) and $\dim \mathcal{D}^{\perp\!\!\!\perp} = 2 \dim \mathcal{F} - \dim \mathcal{D}$ it follows that $\dim \mathcal{D} = \dim \mathcal{D}^{\perp\!\!\!\perp}$, thus yielding $\mathcal{D} = \mathcal{D}^{\perp\!\!\!\perp}$. □

Remark 2.1 Note that we have actually shown that property (i) implies $\mathcal{D} \subset \mathcal{D}^{\perp\!\!\perp}$. Together with the fact that $\dim \mathcal{D}^{\perp\!\!\perp} = 2\dim \mathcal{F} - \dim \mathcal{D}$ this implies that any subspace $\mathcal{D}$ satisfying property (i) has the property that $\dim \mathcal{D} \leq \dim \mathcal{F}$. Thus, as claimed before, a Dirac structure is a linear subspace *of maximal dimension* satisfying property (i).

Remark 2.2 The property $\mathcal{D} = \mathcal{D}^{\perp\!\!\perp}$ can be regarded as a generalization of Tellegen's theorem in circuit theory, since it describes a constraint between two *different* realizations of the port variables, in contrast to property (i).

From a mathematical point of view, there are a number of direct examples of Dirac structures $\mathcal{D} \subset \mathcal{F} \times \mathcal{F}^*$. We leave the proofs as an exercise to the reader.

(i) Let $J : \mathcal{F}^* \to \mathcal{F}$ be a skew-symmetric linear mapping, that is, $J = -J^*$, where $J^* : \mathcal{F}^* \to (\mathcal{F})^{**} = \mathcal{F}$ is the adjoint mapping. Then

$$\operatorname{graph} J := \big\{(f, e) \in \mathcal{F} \times \mathcal{F}^* \mid f = Je\big\}$$

is a Dirac structure.

(ii) Let $\omega : \mathcal{F} \to \mathcal{F}^*$ be a skew-symmetric linear mapping, then

$$\operatorname{graph} \omega := \big\{(f, e) \in \mathcal{F} \times \mathcal{F}^* \mid e = \omega f\big\}$$

is a Dirac structure.

(iii) Let $\mathcal{G} \subset \mathcal{F}$ be any subspace. Define

$$\mathcal{G}^{\perp} = \big\{e \in \mathcal{F}^* \mid \langle e \mid f\rangle = 0 \text{ for all } f \in \mathcal{G}\big\}.$$

Then $\mathcal{G} \times \mathcal{G}^{\perp} \subset \mathcal{F} \times \mathcal{F}^*$ is a Dirac structure. Special cases of such a Dirac structure are *ideal constraints*. Indeed, the ideal effort constraint

$$\mathcal{D} := \big\{(f, e) \in \mathcal{F} \times \mathcal{F}^* \mid e = 0\big\}$$

is defining a Dirac structure, and the same holds for the ideal flow constraint

$$\mathcal{D} := \big\{(f, e) \in \mathcal{F} \times \mathcal{F}^* \mid f = 0\big\}.$$

2.2 Energy Storage

The port variables associated with the internal storage port will be denoted by (f_S, e_S). They are interconnected to the energy storage of the system, which is defined by a finite-dimensional state space manifold $\mathcal{X}$ with coordinates x, together with a Hamiltonian function $H : \mathcal{X} \to \mathbb{R}$ denoting the energy. The flow variables of the energy storage are given by the *rate* $\dot{x}$ of the energy variables x. Furthermore, the

effort variables of the energy storage are given by the *co-energy* variables $\frac{\partial H}{\partial x}(x)$, resulting in the energy balance[3]

$$\frac{d}{dt}H = \left\langle \frac{\partial H}{\partial x}(x) \mid \dot{x} \right\rangle = \frac{\partial^T H}{\partial x}(x)\dot{x}. \tag{2.5}$$

The interconnection of the energy storing elements to the storage port of the Dirac structure is accomplished by setting

$$f_S = -\dot{x} \quad \text{and} \quad e_S = \frac{\partial H}{\partial x}(x). \tag{2.6}$$

Hence the energy balance (2.5) can be also written as

$$\frac{d}{dt}H = \frac{\partial^T H}{\partial x}(x)\dot{x} = -e_S^T f_S. \tag{2.7}$$

2.3 Energy Dissipation

The second internal port corresponds to internal energy dissipation (due to friction, resistance, etc.), and its port variables are denoted by (f_R, e_R). These port variables are terminated on a static resistive relation $\mathcal{R}$. In general, a static resistive relation will be of the form

$$R(f_R, e_R) = 0 \tag{2.8}$$

with the property that for all (f_R, e_R) satisfying (2.8)

$$\langle e_R \mid f_R \rangle \leq 0. \tag{2.9}$$

A typical example of such a nonlinear resistive relation will be given in Example 4.4. In many cases we may restrict ourselves to *linear* resistive relations in which case (f_R, e_R) satisfy relations of the form

$$R_f f_R + R_e e_R = 0. \tag{2.10}$$

The inequality (2.9) then corresponds to the square matrices R_f and R_e satisfying the properties

$$R_f R_e^T = R_e R_f^T \geq 0, \tag{2.11}$$

together with the dimensionality condition

$$\operatorname{rank}[R_f \mid R_e] = \dim f_R. \tag{2.12}$$

[3]Throughout we adopt the convention that $\frac{\partial H}{\partial x}(x)$ denotes the *column* vector of partial derivatives of H.

Indeed, by the dimensionality condition (2.12) and the symmetry (2.11) we can equivalently rewrite the kernel representation (2.10) of $\mathcal{R}$ into an image representation

$$f_R = R_e^T \lambda \quad \text{and} \quad e_R = -R_f^T \lambda. \tag{2.13}$$

That is, any pair (f_R, e_R) satisfying (2.10) can be written into the form (2.13) for a certain λ, and conversely any (f_R, e_R) for which there exists λ such that (2.13) holds is satisfying (2.10). Hence by (2.11) all f_R, e_R satisfying the resistive relation are such that

$$e_R^T f_R = -\left(R_f^T \lambda\right)^T R_e^T \lambda = -\lambda^T R_f R_e^T \lambda \leq 0. \tag{2.14}$$

Without the presence of additional external ports, the Dirac structure of the port-Hamiltonian system satisfies the power balance

$$e_S^T f_S + e_R^T f_R = 0 \tag{2.15}$$

which leads by substitution of equations (2.7) and (2.14) to

$$\frac{d}{dt} H = -e_S^T f_S = e_R^T f_R \leq 0. \tag{2.16}$$

An important special case of resistive relations between $f_R \in \mathbb{R}^{m_r}$ and $e_R \in \mathbb{R}^{m_r}$ occurs when the resistive relations can be expressed as an *input–output* mapping

$$f_R = -F(e_R), \tag{2.17}$$

where the resistive characteristic[4] $F : \mathbb{R}^{m_r} \to \mathbb{R}^{m_r}$ satisfies

$$e_R^T F(e_R) \geq 0, \quad e_R \in \mathbb{R}^{m_r}. \tag{2.18}$$

For *linear* resistive elements, (2.17) specializes to

$$f_R = -\tilde{R} e_R \tag{2.19}$$

for some positive semi-definite symmetric matrix $\tilde{R} = \tilde{R}^T \geq 0$.

2.4 External Ports

Now, let us consider in more detail the *external* ports to the system. We shall distinguish between two types of external port. One is the *control port* $\mathcal{C}$, with port variables (f_C, e_C), which are the port variables which are accessible for controller

[4] In many cases, F will be derivable from a so-called *Rayleigh dissipation function* $\mathfrak{R} : \mathbb{R}^{m_r} \to \mathbb{R}$, in the sense that $F(e_R) = \frac{\partial \mathfrak{R}}{\partial e_R}(e_R)$.

action. The other type of external port is the *interaction port* $\mathcal{I}$, which denotes the interaction of the port-Hamiltonian system with its environment. The port variables corresponding to the interaction port are denoted by (f_I, e_I). Taking both the external ports into account the power-balance (2.15) extends to

$$e_S^T f_S + e_R^T f_R + e_C^T f_C + e_I^T f_I = 0 \tag{2.20}$$

whereby (2.16) extends to

$$\frac{d}{dt} H = e_R^T f_R + e_C^T f_C + e_I^T f_I. \tag{2.21}$$

2.5 Resulting Port-Hamiltonian Dynamics

The port-Hamiltonian system with state space $\mathcal{X}$, Hamiltonian H corresponding to the energy storage port $\mathcal{S}$, resistive port $\mathcal{R}$ with relations (2.8), control port $\mathcal{C}$, interconnection port $\mathcal{I}$, and total Dirac structure $\mathcal{D}$ will be succinctly denoted by $\Sigma = (\mathcal{X}, H, \mathcal{R}, \mathcal{C}, \mathcal{I}, \mathcal{D})$. The dynamics of the port-Hamiltonian system is specified by considering the constraints on the various port variables imposed by the Dirac structure, that is,

$$(f_S, e_S, f_R, e_R, f_C, e_C, f_I, e_I) \in \mathcal{D}$$

and to substitute in these relations the equalities $f_S = -\dot{x}$ and $e_S = \frac{\partial H}{\partial x}(x)$. This leads to the implicitly defined dynamics

$$\left(-\dot{x}(t), \frac{\partial H}{\partial x}\big(x(t)\big), f_R(t), e_R(t), f_C(t), e_C(t), f_I(t), e_I(t)\right) \in \mathcal{D} \tag{2.22}$$

with $f_R(t), e_R(t)$ satisfying for all t the resistive relation

$$R\big(f_R(t), e_R(t)\big) = 0. \tag{2.23}$$

In many cases of interest, Eqs. (2.22) will *constrain* the state x. Thus in a coordinate representation (as will be treated in detail in the next section), port-Hamiltonian systems generally will consist of a mixed set of *differential* and *algebraic* equations (DAEs).

Example 2.1 (General RLC-circuits) We start by showing how Kirchhoff's laws define a Dirac structure on the space of currents and voltages of any electrical circuit. Consider a circuit-graph with m edges and n vertices, where the current through the ith edge is denoted by I_i and the voltage across the ith edge is V_i. Collect the currents in a single column vector I (of dimension m) and the voltages in an m-dimensional column vector V. Then Kirchhoff's *current* laws can be written as

$$\mathcal{B} I = 0, \tag{2.24}$$

where $\mathcal{B}$ is the $n \times m$ incidence matrix of the graph. Dually, Kirchhoff's *voltage* laws can be written as follows: all allowed vectors of voltages V in the circuit are given as

$$V = \mathcal{B}^T \lambda, \quad \lambda \in \mathbb{R}^n. \tag{2.25}$$

It is immediately seen that the total space of currents and voltages allowed by Kirchhoff's current and voltage laws

$$\mathcal{D} = \left\{ (I, V) \mid \mathcal{B} I = 0, \exists \lambda \text{ s.t. } V = \mathcal{B}^T \lambda \right\} \tag{2.26}$$

defines a Dirac structure. In particular $(V^a)^T I^b + (V^b)^T I^a = 0$ for all pairs $(I^a, V^a), (I^b, V^b) \in \mathcal{D}$. By taking V^a, I^b equal to zero, we obtain $(V^b)^T I^a = 0$ for all I^a satisfying (2.24) and all V^b satisfying (2.25), which amounts to *Tellegen's theorem*. Hence for an arbitrary RLC-circuit Kirchhoff's current and voltage laws take the form [35]

$$\begin{aligned} B_L I_L + B_C I_C + B_R I_R &= 0, \\ V_L &= B_L^T \lambda, \\ V_C &= B_C^T \lambda, \\ V_P &= B_P^T \lambda \end{aligned} \tag{2.27}$$

with $[B_L \ B_C \ B_R]$ denoting the incidence matrix of the circuit graph, where the edges have been ordered according to being associated to the inductors, capacitors, and resistors. Furthermore, I_L, I_C and I_R denote the currents through, respectively, the inductors, capacitors and resistors. Likewise, V_L, V_C and V_R denote the voltages across the inductors, capacitors and terminals. Kirchhoff's current and voltage laws define a Dirac structure $\mathcal{D}$ between the flows and efforts

$$\begin{aligned} f_S &= (I_C, V_L, I_R) = (-\dot{Q}, -\dot{\phi}, I_R), \\ e_S &= (V_C, I_L, V_R) = \left(\frac{\partial H}{\partial Q}, \frac{\partial H}{\partial \phi}, V_R \right) \end{aligned}$$

with Hamiltonian $H(Q, \phi)$ equal to the total energy. This leads to the port-Hamiltonian differential-algebraic system

$$\begin{aligned} -\dot{\phi} &= B_L^T \lambda, \\ \frac{\partial H}{\partial Q} &= B_C^T \lambda, \\ V_R &= B_R^T \lambda, \\ 0 &= B_L \frac{\partial H}{\partial \phi} - B_C \dot{Q} + B_R I_R, \\ V_R &= -R I_R \end{aligned}$$

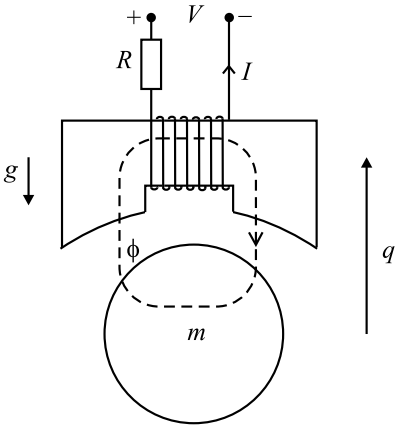


Fig. 2 Magnetically levitated ball

with state vector $x = (Q, \phi)$, where R is a positive diagonal matrix (Ohm's law describing the linear resistors). The equations can be easily extended to cover voltage or current sources, external ports or terminals [40].

Example 2.2 (Electro-mechanical system) Consider the dynamics of an iron ball in the magnetic field of a controlled inductor, as shown in Fig. 2. The port-Hamiltonian description of this system (with q the height of the ball, p the vertical momentum, and φ the magnetic flux of the inductor) is given as

$$\begin{bmatrix} \dot{q} \\ \dot{p} \\ \dot{\varphi} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -R \end{bmatrix} \begin{bmatrix} \frac{\partial H}{\partial q} \\ \frac{\partial H}{\partial p} \\ \frac{\partial H}{\partial \varphi} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} V, \qquad (2.28)$$
$$I = \frac{\partial H}{\partial \varphi}.$$

This is an example of a system where the *coupling* between two different physical domains (mechanical and magnetic) takes place via the Hamiltonian; in this case

$$H(q, p, \varphi) = mgq + \frac{p^2}{2m} + \frac{\varphi^2}{2k_1(1 - \frac{q}{k_2})},$$

where the last term depends both on the magnetic variable φ and the mechanical variable q.

2.6 Port-Hamiltonian Systems and Passivity

By the power-preserving property of the Dirac structure

$$e_S^T f_S + e_R^T f_R + e_C^T f_C + e_I^T f_I = 0.$$

Hence the port-Hamiltonian dynamics defined in (2.22) satisfies

$$\begin{aligned}\frac{dH}{dt} &= \frac{\partial^T H}{\partial x}(x)\dot{x} = -e_S^T f_S \\ &= e_R^T f_R + e_C^T f_C + e_I^T f_I \leq e_C^T f_C + e_I^T f_I,\end{aligned} \tag{2.29}$$

where the last inequality follows from the energy-dissipating property (2.9) of the resistive relation between f_R and e_R. Thus, whenever H is bounded from below (and thus can be changed into a non-negative function by adding a constant), the port-Hamiltonian system is *passive*. Furthermore, notice that in fact we may relax the requirement of H being bounded from below on the whole state space $\mathcal{X}$ by requiring that H is bounded from below on the part of $\mathcal{X}$ satisfying the algebraic constraints present in the system.

2.7 Modulated Dirac Structures and Port-Hamiltonian Systems on Manifolds

For many systems, especially those with 3-D mechanical components, the Dirac structure is actually *modulated* by the state variables. Furthermore, the state space $\mathcal{X}$ is a *manifold* and the flow vector $f_S = -\dot{x}$ corresponding to energy-storage are in the tangent space $T_x\mathcal{X}$ at the state $x \in \mathcal{X}$, while the effort vector e_S is in the co-tangent space $T_x^*\mathcal{X}$. The modulation of the Dirac structure is often intimately related to the underlying geometry of the system.

Example 2.3 (Spinning rigid body) Consider a rigid body spinning around its center of mass in the absence of gravity. The energy variables are the three components of the body angular momentum p along the three principal axes: $p = (p_x, p_y, p_z)$, and the energy is the kinetic energy

$$H(p) = \frac{1}{2}\left(\frac{p_x^2}{I_x} + \frac{p_y^2}{I_y} + \frac{p_z^2}{I_z}\right),$$

where I_x, I_y, I_z are the principal moments of inertia. Euler's equations describing the dynamics are

$$\begin{bmatrix}\dot{p}_x \\ \dot{p}_y \\ \dot{p}_z\end{bmatrix} = \underbrace{\begin{bmatrix}0 & -p_z & p_y \\ p_z & 0 & -p_x \\ -p_y & p_x & 0\end{bmatrix}}_{J(p)} \begin{bmatrix}\frac{\partial H}{\partial p_x} \\ \frac{\partial H}{\partial p_y} \\ \frac{\partial H}{\partial p_z}\end{bmatrix}. \tag{2.30}$$

The Dirac structure is given as the graph of the skew-symmetric matrix $J(p)$, and thus defines a subspace which is modulated by the state variables p.

This motivates to extend the definition of a *constant* Dirac structure $\mathcal{D} \subset \mathcal{F} \times \mathcal{F}^*$ (with $\mathcal{F}$ a linear space) as given before in Proposition 2.1 to *Dirac structures on manifolds*. Simply put, a Dirac structure on a manifold $\mathcal{X}$ is pointwise (that is, for every $x \in \mathcal{X}$) a constant Dirac structure $\mathcal{D}(x) \subset T_x\mathcal{X} \times T_x^*\mathcal{X}$.

Definition 2.2 Let $\mathcal{X}$ be a manifold. A Dirac structure $\mathcal{D}$ on $\mathcal{X}$ is a vector subbundle of the Whitney sum[5] $T\mathcal{X} \oplus T^*\mathcal{X}$ such that

$$\mathcal{D}(x) \subset T_x\mathcal{X} \times T_x^*\mathcal{X}$$

is for every $x \in \mathcal{X}$ a constant Dirac structure as before.

If, next to the energy storage port, there are additional ports (such as resistive, control or interaction ports) with port variables $f \in \mathcal{F}$ and $e \in \mathcal{F}^*$, then a modulated Dirac structure is pointwise specified by a constant Dirac structure

$$\mathcal{D}(x) \subset T_x\mathcal{X} \times T_x^*\mathcal{X} \times \mathcal{F} \times \mathcal{F}^*. \tag{2.31}$$

2.7.1 Kinematic Constraints in Mechanics

Modulated Dirac structures often arise as a result of *ideal constraints* imposed on the generalized velocities of the mechanical system by its environment, called *kinematic constraints*. In many cases, these constraints will be configuration dependent, leading to a Dirac structure modulated by the configuration variables.

Consider a mechanical system with n degrees of freedom, locally described by n configuration variables $q = (q_1, \ldots, q_n)$. Expressing the kinetic energy as $\frac{1}{2}\dot{q}^T M(q)\dot{q}$, with $M(q) > 0$ being the generalized mass matrix, we define in the usual way the Lagrangian function $L(q, \dot{q})$ as the *difference* of kinetic energy and potential energy $P(q)$, i.e.,

$$L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M(q)\dot{q} - P(q). \tag{2.32}$$

Suppose now that there are constraints on the generalized velocities $\dot{q}$, described as

$$A^T(q)\dot{q} = 0 \tag{2.33}$$

with $A(q)$ an $n \times k$ matrix of rank k everywhere (that is, there are k independent kinematic constraints). Classically, the constraints (2.33) are called *holonomic* if it is possible to find new configuration coordinates $\overline{q} = (\overline{q}_1, \ldots, \overline{q}_n)$ such that the constraints are equivalently expressed as

$$\dot{\overline{q}}_{n-k+1} = \dot{\overline{q}}_{n-k+2} = \cdots = \dot{\overline{q}}_n = 0 \tag{2.34}$$

[5]The Whitney sum of two vector bundles with the same base space is defined as the vector bundle whose fiber above each element of this common base space is the product of the fibers of each individual vector bundle.

in which case one may eliminate the configuration variables $\overline{q}_{n-k+1}, \ldots, \overline{q}_n$, since the kinematic constraints (2.34) are equivalent to the *geometric* constraints

$$\overline{q}_{n-k+1} = c_{n-k+1}, \quad \ldots, \quad \overline{q}_n = c_n \tag{2.35}$$

for certain constants $c_{n-k+1}, \ldots, c_n$ determined by the initial conditions. Then the system reduces to an *unconstrained* system in the $(n-k)$ remaining configuration coordinates $(\overline{q}_1, \ldots, \overline{q}_{n-k})$. If it is *not* possible to find coordinates $\overline{q}$ such that (2.34) holds (that is, if we are not able to *integrate* the kinematic constraints as above), then the constraints are called *nonholonomic*.

The equations of motion for a mechanical system with Lagrangian $L(q, \dot{q})$ and constraints (2.33) are given by the constrained Euler–Lagrange equations (derived from d'Alembert's principle) [22]

$$\begin{aligned} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} &= A(q)\lambda + B(q)u, \quad \lambda \in \mathbb{R}^k, u \in \mathbb{R}^m, \\ A^T(q)\dot{q} &= 0, \end{aligned} \tag{2.36}$$

where $B(q)u$ are the external forces (controls) applied to the system, for some $n \times m$ matrix $B(q)$, while $A(q)\lambda$ are the *constraint forces*. The Lagrange multipliers $\lambda(t)$ at any time t are uniquely determined by the requirement that the constraints $A^T(q(t))\dot{q}(t) = 0$ have to be satisfied for all t. Note that (2.36) defines a set of second-order differential-algebraic equations in the configuration variables q. Defining the generalized momenta

$$p = \frac{\partial L}{\partial \dot{q}} = M(q)\dot{q} \tag{2.37}$$

the constrained Euler–Lagrange equations (2.36) transform into the *constrained Hamiltonian equations*

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p}(q, p), \\ \dot{p} &= -\frac{\partial H}{\partial q}(q, p) + A(q)\lambda + B(q)u, \\ y &= B^T(q)\frac{\partial H}{\partial p}(q, p), \\ 0 &= A^T(q)\frac{\partial H}{\partial p}(q, p) \end{aligned} \tag{2.38}$$

with $H(q, p) = \frac{1}{2}p^T M^{-1}(q)p + P(q)$ the total energy. This defines a port-Hamiltonian differential-algebraic system with respect to the modulated Dirac structure

$$\begin{aligned} \mathcal{D} = \Big\{ (f_S, e_S, f_C, e_C) \mid 0 = A^T(q)e_S,\ e_C = B^T(q)e_S, \exists \lambda \in \mathbb{R}^k \text{ s.t.} \\ -f_S = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix} e_S + \begin{bmatrix} 0 \\ A(q) \end{bmatrix} \lambda + \begin{bmatrix} 0 \\ B(q) \end{bmatrix} f_c, \lambda \in \mathbb{R}^k \Big\}. \end{aligned} \tag{2.39}$$

Example 2.4 (Rolling euro) Let x, y be the Cartesian coordinates of the point of contact of the coin with the plane. Furthermore, φ denotes the heading angle, and θ the angle of Queen Beatrix' head.[6] With all constants set to unity, the constrained Euler–Lagrangian equations of motion are

$$\begin{aligned} \ddot{x} &= \lambda_1, \\ \ddot{y} &= \lambda_2, \\ \ddot{\theta} &= -\lambda_1 \cos\varphi - \lambda_2 \sin\varphi + u_1, \\ \ddot{\varphi} &= u_2 \end{aligned} \tag{2.40}$$

with u_1 the control torque about the rolling axis, and u_2 the control torque about the vertical axis. The total energy is $H = \frac{1}{2}p_x^2 + \frac{1}{2}p_y^2 + \frac{1}{2}p_\theta^2 + \frac{1}{2}p_\varphi^2$. The rolling constraints are the nonholonomic kinematic constraints $\dot{x} = \dot{\theta}\cos\varphi$ and $\dot{y} = \dot{\theta}\sin\varphi$, i.e., rolling without slipping, which can be written in the form (2.33) by defining

$$A^T(x, y, \theta, \phi) = \begin{bmatrix} 1 & 0 & -\cos\phi & 0 \\ 0 & 1 & -\sin\phi & 0 \end{bmatrix}.$$

2.8 Input–State–Output Port-Hamiltonian Systems

An important subclass of port-Hamiltonian systems is the class of *input–state–output port-Hamiltonian systems*, where there are no algebraic constraints on the state space variables, and the flow and effort variables of the resistive, control and interaction port can be split into conjugated input–output pairs.

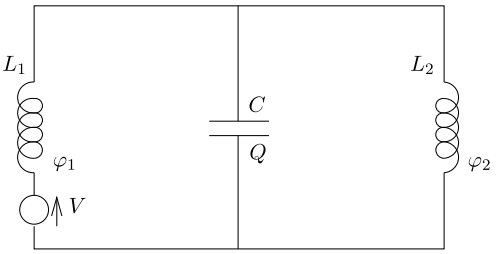
Input–state–output port-Hamiltonian systems are defined as dynamical systems of the following form:

$$\Sigma : \begin{aligned} \dot{x} &= \big[J(x) - R(x)\big]\frac{\partial H}{\partial x}(x) + g(x)u + k(x)d, \\ y &= g^T(x)\frac{\partial H}{\partial x}(x), \\ z &= k^T(x)\frac{\partial H}{\partial x}(x), \end{aligned} \qquad x \in \mathcal{X}, \tag{2.41}$$

where (u, y) are the input–output pairs corresponding to the control port $\mathcal{C}$, while (d, z) denote the input–output pairs of the interaction port $\mathcal{I}$. Note that $y^T u$ and $z^T d$ equal the power corresponding to the control, respectively, interaction port. Here the matrix $J(x)$ is skew-symmetric, that is, $J(x) = -J^T(x)$. The matrix $R(x) = R^T(x) \geq 0$ specifies the resistive structure. From a resistive port point of view, it is given as $R(x) = g_R^T(x)\tilde{R}(x)g_R(x)$ for some linear resistive relation $f_R = -\tilde{R}e_R$

[6] On the Dutch version of the euro.

Fig. 3 Controlled LC-circuit



with $\tilde{R}(x) = \tilde{R}^T(x) \geq 0$ and g_R representing the input matrix corresponding to the resistive port.

The underlying Dirac structure of the system is then given by the graph of the skew-symmetric linear map

$$\begin{bmatrix} -J(x) & -g_R(x) & -g(x) & -k(x) \\ g_R^T(x) & 0 & 0 & 0 \\ g^T(x) & 0 & 0 & 0 \\ k^T(x) & 0 & 0 & 0 \end{bmatrix}. \tag{2.42}$$

In general, the Dirac structure defined as the graph of the mapping (2.42) is a *modulated Dirac structure* since the matrices J, g_R, g, and k may all depend on the energy variables x.

Example 2.5 (LC-circuit with independent storage elements) Consider a controlled LC-circuit (see Fig. 3) consisting of two inductors with magnetic energies $H_1(\varphi_1)$ and $H_2(\varphi_2)$ (φ_1 and φ_2 being the magnetic flux linkages), and a capacitor with electric energy $H_3(Q)$ (Q being the charge). If the elements are linear, then

$$H_1(\varphi_1) = \frac{1}{2L_1}\varphi_1^2, \qquad H_2(\varphi_2) = \frac{1}{2L_2}\varphi_2^2, \qquad H_3(Q) = \frac{1}{2C}Q^2.$$

Furthermore, let $V = u$ denote a voltage source. Using Kirchhoff's laws, one immediately arrives at the input–state–output port-Hamiltonian system

$$\begin{bmatrix} \dot{Q} \\ \dot{\varphi}_1 \\ \dot{\varphi}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}}_{J} \begin{bmatrix} \frac{\partial H}{\partial Q} \\ \frac{\partial H}{\partial \varphi_1} \\ \frac{\partial H}{\partial \varphi_2} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u,$$

$$y = \frac{\partial H}{\partial \varphi_1} \quad (= \text{current through first inductor})$$

with $H(Q, \varphi_1, \varphi_2) := H_1(\varphi_1) + H_2(\varphi_2) + H_3(Q)$ the total energy. Clearly the matrix J is skew-symmetric. In this way, cf. [20], every LC-circuit with *independent* storage elements can be modeled as an input–state–output port-Hamiltonian system (with respect to a constant Dirac structure).

Input–state–output port-Hamiltonian systems with additional *feed-through terms* are given as (for simplicity we do not take the interaction port into account) [13, 31]

$$\begin{aligned} \dot{x} &= \big[J(x) - R(x)\big]\frac{\partial H}{\partial x}(x) + \big[g(x) - P(x)\big]u, \\ y &= \big[g(x) + P(x)\big]^\top \frac{\partial H}{\partial x}(x) + \big[M(x) + S(x)\big]u \end{aligned} \tag{2.43}$$

with the matrices P, R, S satisfying

$$Z = \begin{bmatrix} R(x) & P(x) \\ P^\top(x) & S(x) \end{bmatrix} \geq 0. \tag{2.44}$$

The relation between u, y and the storage port variables f_S, e_S is in this case given as

$$\begin{bmatrix} f_S \\ y \end{bmatrix} = \begin{bmatrix} -J(x) & -g(x) \\ g^T(x) & M \end{bmatrix} \begin{bmatrix} e_S \\ u \end{bmatrix} + \begin{bmatrix} R(x) & P(x) \\ P^T(x) & S(x) \end{bmatrix} \begin{bmatrix} e_S \\ u \end{bmatrix}. \tag{2.45}$$

It follows that

$$e_S^T f_S + u^T y = \begin{bmatrix} e_S^T & u^T \end{bmatrix} \begin{bmatrix} R(x) & P(x) \\ P^T(x) & S(x) \end{bmatrix} \begin{bmatrix} e_S \\ u \end{bmatrix} \geq 0$$

and thereby

$$\frac{d}{dt}H(x) = -e_S^T f_S = u^T y - \begin{bmatrix} e_S^T & u^T \end{bmatrix} \begin{bmatrix} R(x) & P(x) \\ P^T(x) & S(x) \end{bmatrix} \begin{bmatrix} e_S \\ u \end{bmatrix} \leq u^T y$$

thus recovering the basic energy balance for port-Hamiltonian systems. Port-Hamiltonian input–state–output systems with feed-through terms readily show up in the modeling of power converters [13], as well as in friction models (see e.g. [16] for a port-Hamiltonian description of the dynamic *LuGre* friction model).

Although the class of input–state–output port-Hamiltonian systems is a very important subclass, it is *not* closed under general power-preserving interconnections. Basically, only negative *feedback* interconnections of input–state–output port-Hamiltonian systems will result in another input–state–output port-Hamiltonian system, while otherwise algebraic constraints will arise, leading to port-Hamiltonian differential-algebraic systems. On the other hand, input–state–output port-Hamiltonian systems may arise from *solving the algebraic constraints* in a port-Hamiltonian differential-algebraic system. This will be discussed in Sect. 4.

3 Representations of Dirac Structures and Port-Hamiltonian Systems

In the preceding section, we have provided the geometric definition of a port-Hamiltonian system containing three main ingredients. First, the energy storage

which is represented by a state space manifold $\mathcal{X}$ specifying the space of state variables together with a Hamiltonian $H : \mathcal{X} \to \mathbb{R}$ defining the energy. Secondly, there are the static resistive elements, and thirdly there is the Dirac structure linking all the flows and efforts associated to the energy storage, resistive elements, and the external ports (e.g. control and interaction ports) in a power-conserving manner. This, together with the general formulation (2.2) of a Dirac structure, leads to a completely *coordinate-free* definition of a port-Hamiltonian system, because of three reasons: (a) we do not start with coordinates for the state space manifold $\mathcal{X}$, (b) we define the Dirac structure as a *subspace* instead of a set of equations, (c) the resistive relations are defined as a subspace constraining the port variables (f_R, e_R).

This geometric, coordinate-free, point of view has a number of advantages. It allows one to reason about port-Hamiltonian systems without the need to choose specific representations. For example, in Sect. 4 we will see that a number of properties of the port-Hamiltonian system, such as passivity, stability, existence of conserved quantities and algebraic constraints, can be analyzed without the need for choosing coordinates and equations. On the other hand, for many other purposes, including simulation, the need for a representation in coordinates of the port-Hamiltonian system is indispensable, in which case the emphasis shifts to finding the most convenient coordinate representation for the purpose at hand. The examples of the previous section have already been presented in this way. In this section, we will briefly discuss a number of possible representations of port-Hamiltonian systems. It will turn out that the key issue is the representation of the Dirac structure.

3.1 Representations of Dirac Structures

Dirac structures admit different *representations*. Here we just list a number of them. See e.g. [7–9, 15] for more information.

3.1.1 Kernel and Image Representation

Every Dirac structure $\mathcal{D} \subset \mathcal{F} \times \mathcal{F}^*$ can be represented in *kernel representation* as

$$\mathcal{D} = \{(f, e) \in \mathcal{F} \times \mathcal{F}^* \mid Ff + Ee = 0\} \tag{3.1}$$

for linear maps $F : \mathcal{F} \to \mathcal{V}$ and $E : \mathcal{F}^* \to \mathcal{V}$ satisfying

$$\begin{aligned} &\text{(i)} \quad EF^* + FE^* = 0, \\ &\text{(ii)} \quad \operatorname{rank}(F + E) = \dim \mathcal{F}, \end{aligned} \tag{3.2}$$

where $\mathcal{V}$ is a linear space with the same dimension as $\mathcal{F}$, and where $F^* : \mathcal{V}^* \to \mathcal{F}^*$ and $E^* : \mathcal{V}^* \to \mathcal{F}^{**} = \mathcal{F}$ are the adjoint maps of F and E, respectively.

It follows from (3.2) that $\mathcal{D}$ can be also written in *image representation* as

$$\mathcal{D} = \left\{(f, e) \in \mathcal{F} \times \mathcal{F}^* \mid f = E^*\lambda, e = F^*\lambda, \lambda \in \mathcal{V}^*\right\}. \tag{3.3}$$

Sometimes it will be useful to relax this choice of the linear mappings F and E by allowing $\mathcal{V}$ to be a linear space of dimension greater than the dimension of $\mathcal{F}$. In this case we shall speak of *relaxed* kernel and image representations.

Matrix kernel and image representations are obtained by choosing linear coordinates for $\mathcal{F}$, $\mathcal{F}^*$ and $\mathcal{V}$. Indeed, take any basis $f_1, \ldots, f_n$ for $\mathcal{F}$ and the *dual basis* $e_1 = f_1^*, \ldots, e_n = f_n^*$ for $\mathcal{F}^*$, where $\dim \mathcal{F} = n$. Furthermore, take any set of linear coordinates for $\mathcal{V}$. Then the linear maps F and E are represented by $n \times n$ matrices F and E satisfying

$$\begin{aligned} &\text{(i)} \quad EF^T + FE^T = 0, \\ &\text{(ii)} \quad \operatorname{rank}[F \mid E] = \dim \mathcal{F}. \end{aligned} \tag{3.4}$$

In the case of a relaxed kernel and image representation F and E will be $n' \times n$ matrices with $n' > n$.

A (constructive) proof for the existence of matrix kernel and image representations can be given as follows. Consider a Dirac structure $\mathcal{D} \subset \mathcal{F} \times \mathcal{F}^*$ where we have chosen linear coordinates for $\mathcal{F}$, $\mathcal{F}^*$ and $\mathcal{V}$. In particular, choose any basis $f_1, \ldots, f_n$ for $\mathcal{F}$ and the *dual basis* $e_1 = f_1^*, \ldots, e_n = f_n^*$ for $\mathcal{F}^*$, where $\dim \mathcal{F} = n$. Since $\mathcal{D}$ is a subspace of $\mathcal{F} \times \mathcal{F}^*$ it follows that there exist square $n \times n$ matrices F and E such that

$$\mathcal{D} = \operatorname{im} \begin{bmatrix} E^T \\ F^T \end{bmatrix},$$

where $\operatorname{rank}[F \mid E] = \dim \mathcal{F}$. Thus any element $(f, e) \in \mathcal{D}$ can be written as

$$f = E^T\lambda, \qquad e = F^T\lambda$$

for some $\lambda \in \mathbb{R}^n$. Since $e^T f = 0$ for every $(f, e) \in \mathcal{D}$ this implies that

$$\lambda^T F E^T \lambda = 0$$

for every λ, or equivalently, $EF^T + FE^T = 0$. Conversely, any subspace $\mathcal{D}$ given by (3.4) is a Dirac structure, since it satisfies $e^T f = 0$ for every $(f, e) \in \mathcal{D}$ and its dimension is equal to n.

Remark 3.1 A special type of kernel representation occurs if not only $EF^* + FE^* = 0$ but even more $FE^* = 0$. This implies that $\operatorname{im} E^* \subset \ker F$. Since it follows from the kernel/image representation of any Dirac structure that $\ker F \subset \operatorname{im} E^*$, we thus obtain $\operatorname{im} E^* = \ker F$. Hence the Dirac structure is the product of the subspace $\ker F \subset \mathcal{F}$ and the subspace $\ker F^\perp = \ker E \subset \mathcal{F}^*$. We have already encountered this special type of Dirac structure in the case of Kirchhoff's current and voltage laws.

3.1.2 Constrained Input–Output Representation

Every Dirac structure $\mathcal{D} \subset \mathcal{F} \times \mathcal{F}^*$ can be represented as

$$\mathcal{D} = \left\{(f, e) \in \mathcal{F} \times \mathcal{F}^* \mid f = Je + G\lambda, G^T e = 0\right\} \tag{3.5}$$

for a skew-symmetric mapping $J : \mathcal{F} \to \mathcal{F}^*$ and a linear mapping G such that $\operatorname{im} G = \{f \mid (f, 0) \in \mathcal{D}\}$. Furthermore, $\ker J = \{e \mid (0, e) \in \mathcal{D}\}$.

The proof that (3.5) defines a Dirac structure is straightforward. Indeed, for any (f, e) given as in (3.5) we have

$$e^T f = e^T (Je + G\lambda) = e^T Je + e^T G\lambda = 0$$

by skew-symmetry of J and $G^T e = 0$. Furthermore, let rank $G = r \leq n$. If $r = 0$ (or equivalently $G = 0$) then the dimension of $\mathcal{D}$ is clearly n since in that case it is the graph of the mapping J. For $r \neq 0$ the freedom in e will be reduced by dimension r, while at the other hand the freedom in f will be increased by dimension r (because of the term $G\lambda$). For showing that every Dirac structure $\mathcal{D} \subset \mathcal{F} \times \mathcal{F}^*$ can be represented in this way we refer to [8].

3.1.3 Hybrid Input–Output Representation

Let $\mathcal{D}$ be given in matrix kernel representation by square matrices E and F as in 1. Suppose rank $F = m(\leq n)$. Select m independent columns of F, and group them into a matrix F_1. Write (possibly after permutations) $F = [F_1 \mid F_2]$ and correspondingly $E = [E_1 \mid E_2]$,

$$f = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \quad \text{and} \quad e = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}.$$

Then it can be shown [4] that the matrix $[F_1 \mid E_2]$ is invertible, and

$$\mathcal{D} = \left\{ \left(\begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \right) \,\middle|\, \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = J \begin{bmatrix} e_1 \\ f_2 \end{bmatrix} \right\} \tag{3.6}$$

with $J := -[F_1 \mid E_2]^{-1}[F_2 \mid E_1]$ skew-symmetric.

It follows that any Dirac structure can be written as the graph of a skew-symmetric map. The vectors e_1, f_2 can be regarded as *input* vectors, while the complementary vectors f_1, e_2 can be seen as *output* vectors.[7]

[7] This is very much like the multi-port description of a passive linear circuit, where it is known that although it is *not* always possible to describe the port as an admittance or as an impedance, it *is* possible to describe it as a hybrid admittance/impedance transfer matrix, for a suitable selection of input voltages and currents and complementary output currents and voltages [3].

3.1.4 Canonical Coordinate Representation

For any constant Dirac structure there exist a basis for $\mathcal{F}$ and dual basis for $\mathcal{F}^*$, such that the vector (f, e), when partitioned as $(f, e) = (f_q, f_p, f_r, f_s, e_q, e_p, e_r, e_s)$, is in $\mathcal{D}$ if and only if

$$\begin{aligned} f_q &= -e_p, \\ f_p &= e_q, \\ f_r &= 0, \\ e_s &= 0. \end{aligned} \tag{3.7}$$

For a proof we refer to [8]. For a modulated Dirac structure the existence of canonical coordinates requires an additional *integrability condition*, for which we refer to Sect. 7. The representation of a Dirac structure by canonical coordinates is very close to the classical Hamiltonian equations of motion.

In [4, 9, 32] it is shown how one may convert any of the above representations into any other. An easy transformation that will be used frequently in the sequel is the transformation of the *constrained input–output* representation into the *kernel* representation. Consider the Dirac structure $\mathcal{D}$ given in constrained input–output representation by (3.5). Construct a linear mapping $G^\perp$ of maximal rank satisfying $G^\perp G = 0$. Then, pre-multiplying the first equation of (3.5) by $G^\perp$, one eliminates the Lagrange multipliers λ and obtains

$$\mathcal{D} = \left\{ (f, e) \in \mathcal{F} \times \mathcal{F}^* \mid G^\perp f = G^\perp J e, G^T e = 0 \right\} \tag{3.8}$$

which is easily seen to lead to a kernel representation. Indeed,

$$F = \begin{bmatrix} -G^\perp \\ 0 \end{bmatrix}, \qquad E = \begin{bmatrix} G^\perp J \\ G^T \end{bmatrix}$$

defines a kernel representation.

3.2 Representations of Port-Hamiltonian Systems

Coordinate representations of the port-Hamiltonian system (2.22) are obtained by choosing a specific coordinate representation of the Dirac structure $\mathcal{D}$. For example, if $\mathcal{D}$ is given in matrix kernel representation

$$\mathcal{D} = \left\{ (f_S, e_S, f, e) \in \mathcal{X} \times \mathcal{X}^* \times \mathcal{F} \times \mathcal{F}^* \mid F_S f_S + E_S e_S + F f + E e = 0 \right\} \tag{3.9}$$

with

$$\begin{aligned} &\text{(i)} \quad E_S F_S^T + F_S E_S^T + E F^T + F E^T = 0, \\ &\text{(ii)} \quad \operatorname{rank}[F_S \mid E_S \mid F \mid E] = \dim(\mathcal{X} \times \mathcal{F}) \end{aligned} \tag{3.10}$$

then the port-Hamiltonian system is given by the set of equations

$$F_S\dot{x}(t) = E_S\frac{\partial H}{\partial x}\big(x(t)\big) + Ff(t) + Ee(t). \tag{3.11}$$

Note that, in general, (3.11) consists of differential equations *and* algebraic equations in the state variables x (DAEs), together with equations relating the state variables and their time-derivatives to the external port variables (f, e).

Example 3.1 (1-D mechanical systems) Consider a *spring* with elongation q and energy function $H_s(q)$, which for a linear spring is given as $H_s(q) = \frac{1}{2}kq^2$. Let (v_s, F_s) represent the external port through which energy can be exchanged with the spring, where v_s is equal to the rate of elongation (velocity) and F_s is equal to the elastic force. This port-Hamiltonian system (without dissipation) can be written in kernel representation as

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} -\dot{q} \\ v_s \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} kq \\ F_s \end{bmatrix} = 0. \tag{3.12}$$

Similarly we can model a *moving mass m* with scalar momentum p and kinetic energy $H_m(p) = \frac{1}{2m}p^2$ as the port-Hamiltonian system

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} -\dot{p} \\ F_m \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} \frac{p}{m} \\ v_m \end{bmatrix} = 0, \tag{3.13}$$

where (F_m, v_m) are, respectively, the external force exerted on the mass and the velocity of the mass. The mass and the spring can be *interconnected* to each other using the symplectic gyrator

$$\begin{bmatrix} v_s \\ F_m \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} F_s \\ v_m \end{bmatrix}. \tag{3.14}$$

Collecting all equations we have obtained a port-Hamiltonian system with energy variables $x = (q, p)$, total energy $H(q, p) = H_s(q) + H_m(p)$ and with interconnected port variables (v_s, F_s, F_m, v_m). After elimination of the interconnection variables (v_s, F_s, F_m, v_m) one obtains the port-Hamiltonian system

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -\dot{q} \\ -\dot{p} \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} kq \\ \frac{p}{m} \end{bmatrix} = 0 \tag{3.15}$$

which is the ubiquitous mass–spring system. Note that the Dirac structure of this mass–spring system is derived from the Dirac structures of the spring system and the mass system together with their interconnection by means of the symplectic gyrator (which itself defines a Dirac structure). The systematic derivation of the resulting interconnected Dirac structure will be studied in Sect. 6.

In case of a Dirac structure modulated by the state variables x and the state space $\mathcal{X}$ being a manifold, the flows $f_S = -\dot{x}$ are elements of the tangent space $T_x\mathcal{X}$

at the state $x \in \mathcal{X}$, and the efforts e_S are elements of the co-tangent space $T_x^*\mathcal{X}$. Still, locally on $\mathcal{X}$, we obtain the kernel representation (3.11) for the resulting port-Hamiltonian system, but now the matrices F_S, E_S, F and E will depend on x.

The important special case of input–state–output port-Hamiltonian systems as treated before

$$\begin{aligned} \dot{x} &= J(x)\frac{\partial H}{\partial x}(x) + g(x)u, \\ y &= g^T(x)\frac{\partial H}{\partial x}(x), \end{aligned} \qquad x \in \mathcal{X},$$

can be interpreted as arising from a hybrid input–output representation of the Dirac structure (from e_S, u to f_S, y). If the matrices J, g depend on the energy variables x, then this is again a modulated Dirac structure.

In general, by a combination of the hybrid representation and the constrained input–output representation, it can be shown [9] that, locally, any port-Hamiltonian system can be represented in the following way:

$$\begin{aligned} \dot{x} &= J(x)\frac{\partial H}{\partial x}(x) + g(x)u + b(x)\lambda, \\ y &= g^T(x)\frac{\partial H}{\partial x}(x), \qquad\qquad x \in \mathcal{X}, \\ 0 &= b^T(x)\frac{\partial H}{\partial x}(x), \end{aligned} \tag{3.16}$$

where $y^T u$ denotes the power at the external port, and $0 = b^T(x)\frac{\partial H}{\partial x}(x)$ represents the algebraic constraints.[8] Note that the Hamiltonian formulation of mechanical systems with kinematic constraints, as discussed in Sect. 2.7.1, leads to this form; see in particular the constrained Hamiltonian equations (2.38) and its Lagrangian counterpart (2.36).

Example 3.2 (Coupled masses—Internal constraints) Consider two point masses m_1 and m_2 that are rigidly linked to each other, moving in one dimension. When decoupled, the masses are described by the input–state–output port-Hamiltonian systems

$$\begin{aligned} \dot{p}_i &= F_i, \\ v_i &= \frac{p_i}{m_i}, \end{aligned} \qquad i = 1, 2 \tag{3.17}$$

with F_i denoting the force exerted on mass m_i. Rigid coupling of the two masses is achieved by setting

$$v_1 = v_2, \qquad F_1 = -F_2. \tag{3.18}$$

[8] The equality $0 = b^T(x)\frac{\partial H}{\partial x}(x)$ also has the interpretation (well-known in a mechanical system context) that the constraint input $b(x)\lambda$ is 'workless'; i.e., the evolution of the value of the Hamiltonian H is not affected by this term.

This leads to the port-Hamiltonian differential-algebraic system

$$\begin{bmatrix} \dot{p}_1 \\ \dot{p}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \lambda,$$
$$0 = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{p_1}{m_1} \\ \frac{p_2}{m_2} \end{bmatrix} \tag{3.19}$$

where $\lambda = F_1 = -F_2$ now denotes the *internal* constraint force. The resulting interconnected system no longer has external ports. On the other hand, external ports for the interconnected system can be included by either extending (3.17) to

$$\begin{aligned} \dot{p}_i &= F_i + F_i^{\text{ext}}, \\ v_i &= \frac{p_i}{m_i}, \qquad i = 1, 2 \\ v_i^{\text{ext}} &= \frac{p_i}{m_i}, \end{aligned} \tag{3.20}$$

with F_i^{ext} and v_i^{ext} denoting the external forces and velocities, or by modifying the interconnection constraints (3.18) to e.g. $F_1 + F_2 + F^{\text{ext}} = 0$ and $v_1 = v_2 = v^{\text{ext}}$, with F^{ext} and v^{ext} denoting the external force exerted on the coupled masses, respectively the velocity of the coupled masses.

Remark 3.2 Note that in the above port-Hamiltonian description of the two coupled masses the *position* variables $q_i, i = 1, 2$, of the two masses do not come into play, while the interconnection is *not* described by the alternative formulation $q_1 = q_2, F_1 = -F_2$, but instead by $v_1 = v_2, F_1 = -F_2$. The positions $q_i, i = 1, 2$, *can* be included, albeit somewhat redundantly, by extending the port-Hamiltonian descriptions $\dot{p}_i = F_i, v_i = \frac{p_i}{m_i}$ of the two masses to the input–state–output port-Hamiltonian systems

$$\begin{aligned} \dot{q}_i &= \frac{p_i}{m_i}, \\ \dot{p}_i &= F_i, \\ v_i &= \frac{p_i}{m_i} \end{aligned}$$

with Hamiltonians $H_i(q_i, p_i) = \frac{2m_i}{p_i^2}$ (not depending on q_i!). Imposing again the 'port-Hamiltonian' interconnection constraints $v_1 = v_2, F_1 = -F_2$ this leads to a total system having as *conserved quantity* (Casimir) $q_1 - q_2$. Thus the fact that $v_1 = v_2$ implies $q_1 = q_2$ only up to a constant (since this constant disappears in differentiation) is reflected in the initial condition of the extended total system.

Note furthermore that specifying constraints as constraints on the velocities is in line with the use of kinematic constraints in mechanical systems. In general, such kinematic constraints can be *integrated* to geometric (position) constraints if

integrability conditions are satisfied (such as in the simple case $v_1 = v_2$); otherwise the kinematic constraints are called *nonholonomic*.

Since it is easy to eliminate the Lagrange multipliers in any constrained input–output representation of the Dirac structure, cf. (3.8), it is also relatively easy to eliminate the Lagrange multipliers in any port-Hamiltonian system. Indeed, consider the port-Hamiltonian system (3.16). The Lagrange multipliers λ are eliminated by constructing a matrix $b^{\perp}(x)$ of maximal rank such that

$$b^{\perp}(x)b(x) = 0.$$

Then, by pre-multiplication with this matrix $b^{\perp}(x)$, one obtains the equations

$$\begin{aligned} b^{\perp}(x)\dot{x} &= b^{\perp}(x)J(x)\frac{\partial H}{\partial x}(x) + b^{\perp}(x)g(x)u, \\ y &= g^T(x)\frac{\partial H}{\partial x}(x), \qquad\qquad x \in \mathcal{X} \\ 0 &= b^T(x)\frac{\partial H}{\partial x}(x), \end{aligned} \tag{3.21}$$

without Lagrange multipliers. This is readily seen to be a kernel representation of the port-Hamiltonian differential-algebraic system.

Example 3.3 (Example 3.2, continued) Consider the system of two coupled masses in Example 3.2. Pre-multiplication of the dynamic equations by the row vector [1 1] yields the equations

$$\dot{p}_1 + \dot{p}_2 = 0, \qquad \frac{p_1}{m_1} - \frac{p_2}{m_2} = 0 \tag{3.22}$$

which constitutes a kernel representation of the port-Hamiltonian DAE system, with matrices

$$F = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}. \tag{3.23}$$

Remark 3.3 Consider the representation (3.16) *without* the external port corresponding to the input and output variables u, y. Furthermore assume that $b(x)$ is given as

$$b(x) = -J(x)\frac{\partial \varphi}{\partial x}(x)$$

for a certain mapping $\varphi = (\varphi_1, \ldots, \varphi_m)^T$, where $m = \dim \lambda$, satisfying additionally

$$\frac{\partial \varphi_i}{\partial x}(x)J(x)\frac{\partial \varphi_j}{\partial x}(x) = 0, \quad i, j = 1, \ldots, m.$$

Then the constraints $b^T(x)\frac{\partial H}{\partial x}(x) = 0$ can be rewritten as

$$\frac{d\varphi}{dt} = 0.$$

Replacing the constraints $\frac{d\varphi}{dt} = 0$ by their *time-integrated* version $\varphi(x) = 0$, one obtains the constrained system

$$\begin{aligned} \dot{x} &= J(x)\frac{\partial H}{\partial x}(x) - J(x)\frac{\partial \varphi}{\partial x}(x)\lambda, \\ 0 &= \varphi(x). \end{aligned} \tag{3.24}$$

If additionally $J(x)$ is the standard symplectic form, then this is the starting point for the classical Dirac theory of Hamiltonian systems with constraints, leading to the concept of the *Dirac brackets* defined by $J(x)$, the Hamiltonian H, together with the constraint functions $\varphi_i(x), i = 1, \ldots, m$; see [11, 28].

4 Analysis of Port-Hamiltonian DAEs

In this section we will analyse a number of key aspects of port-Hamiltonian differential-algebraic systems, and discuss how the specific structure yields advantages as compared to general differential-algebraic (DAE) systems.

First of all, we will study the *index* of port-Hamiltonian differential-algebraic systems, and the possibilities to solve the algebraic constraints. Then we will study the algebraic constraints from a geometric point of view, directly based on the Dirac structure of the system. This coordinate-free approach shows how different coordinate representations can be chosen to express the algebraic constraints; each with their own advantages and disadvantages. Next it will be shown how the geometric theory of algebraic constraints can be dualized to the study of Casimir functions (conserved quantities independent of the Hamiltonian). In the last section we will show how the port-Hamiltonian structure naturally leads to stability analysis, using the Hamiltonian (or a combination of the Hamiltonian and a Casimir function) as a Lyapunov function for the differential-algebraic system. Finally, we will provide some observations concerning the (lack of) well-posedness of port-Hamiltonian differential-algebraic systems in case of nonlinear resistive characteristics.

4.1 Analysis and Elimination of Algebraic Constraints

An important problem concerns the possibility to solve for the *algebraic constraints* of a port-Hamiltonian differential-algebraic system. In case of the representation (3.21), the algebraic constraints are explicitly given by

$$0 = b^T(x)\frac{\partial H}{\partial x}(x). \tag{4.1}$$

In general, these equations will constrain the state variables x. However, the precise way this takes place depends on the properties of the Hamiltonian H as well as of the matrix $b(x)$. For example, if the Hamiltonian H is such that its gradient $\frac{\partial H}{\partial x}(x)$ happens to be contained in the kernel of the matrix $b^T(x)$ for all x, then the algebraic constraints (4.1) are automatically satisfied, and actually the state variables are not constrained.

In general, under constant rank assumptions, the set

$$\mathcal{X}_c := \left\{ x \in \mathcal{X} \mid b^T(x) \frac{\partial H}{\partial x}(x) = 0 \right\}$$

will define a submanifold of the total state space $\mathcal{X}$, called the *constrained state space*. In order that this constrained state space qualifies as the state space for a port-Hamiltonian system *without* further algebraic constraints, one needs to be able to restrict the dynamics of the port-Hamiltonian system to the constrained state space. This is always possible under the condition that the matrix

$$b^T(x) \frac{\partial^2 H}{\partial x^2}(x) b(x) \tag{4.2}$$

has full rank. Indeed, under this condition, the differentiated constraint equation

$$0 = \frac{d}{dt}\left(b^T(x) \frac{\partial H}{\partial x}(x) \right) = * + b^T(x) \frac{\partial^2 H}{\partial x^2}(x) b(x) \lambda \tag{4.3}$$

(with $*$ denoting unspecified terms) can always be uniquely solved for λ, leading to a uniquely defined dynamics on the constrained state space $\mathcal{X}_c$. Hence the set of *consistent states* for the port-Hamiltonian differential-algebraic system (the set of initial conditions for which the system has a unique ordinary solution) is equal to the constrained state space $\mathcal{X}_c$. Using terminology from the theory of DAEs, the condition that the matrix in (4.2) has full rank ensures that the *index* of the DAEs specified by the port-Hamiltonian system is equal to one. This can be summarized as

Proposition 4.1 *Consider the port-Hamiltonian differential-algebraic system represented as in* (3.21)*, with algebraic constraints* $b^T(x) \frac{\partial H}{\partial x}(x) = 0$*. Suppose that the matrix* $b^T(x) \frac{\partial^2 H}{\partial x^2}(x) b(x)$ *has full rank for all* $x \in \mathcal{X}_c$*. Then the system has index one, and the set of consistent states is equal to* $\mathcal{X}_c$*.*

Hence under the condition that $b^T(x) \frac{\partial^2 H}{\partial x^2}(x) b(x)$ has full rank, then the algebraic constraints $b^T(x) \frac{\partial H}{\partial x}(x) = 0$ can be eliminated, leading to a set of ordinary differential equations defined on the constrained state space $\mathcal{X}_c$. Of course, in the nonlinear case the *explicit* elimination of the algebraic constraints may be difficult, or even impossible.

If the matrix in (4.2) does not have full rank, then the index of the port-Hamiltonian differential-algebraic system will be larger than one, and it will be

necessary to further constrain the space $\mathcal{X}_c$ by considering apart from the 'primary' algebraic constraints (4.1), also their (repeated) time-derivatives (sometimes called *secondary constraints*). We refer to [23, 28] for a detailed treatment and conditions for reducing the port-Hamiltonian DAE system to a system without algebraic constraints in case $J(x)$ corresponds to a symplectic structure.

4.1.1 The Linear Index One Case

In the *linear* case the explicit elimination of the algebraic constraints under the assumption that the matrix in (4.2) has full rank proceeds as follows.

Consider a linear port-Hamiltonian system, without energy-dissipation and external ports, given in constrained input–output representation,

$$\begin{aligned} \dot{x} &= JQx + G\lambda, \quad & J &= -J^T,\ Q = Q^T, \\ 0 &= G^T Qx, & H(x) &= \frac{1}{2}x^T Qx. \end{aligned} \tag{4.4}$$

As before, the constraint forces $G\lambda$ are eliminated by pre-multiplying the first equation by the annihilating matrix $G^\perp$, leading to the DAE system

$$\begin{bmatrix} G^\perp \\ 0 \end{bmatrix} \dot{x} = \begin{bmatrix} G^\perp J \\ G^T \end{bmatrix} Qx.$$

The corresponding matrix pencil

$$s \begin{bmatrix} G^\perp \\ 0 \end{bmatrix} - \begin{bmatrix} G^\perp JQ \\ G^T Q \end{bmatrix} \tag{4.5}$$

is non-singular if $G^T QG$ has full rank, and in fact, the system has index one *if and only if* $G^T QG$ has full rank.

In this case the algebraic constraints $G^T Qx = 0$ are eliminated as follows. Assume throughout (without loss of generality) that G has full rank. Then define the linear coordinate transformation

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} G^\perp \\ G^T \end{bmatrix} x =: Vx \tag{4.6}$$

leading to the transformed system

$$\begin{aligned} \dot{z} &= \big(VJV^T\big)\big(V^{-T}QV^{-1}\big)(Vx) + VG\lambda = \tilde{J}\tilde{Q}z + \begin{bmatrix} 0 \\ G^T G \end{bmatrix}\lambda, \\ 0 &= G^T Qx = \begin{bmatrix} 0 & G^T G \end{bmatrix} \tilde{Q}z, \end{aligned} \tag{4.7}$$

where $\tilde{J} = VJV^T$, $\tilde{Q} = V^{-T}QV^{-1}$. Since $G^T G$ is assumed to have full rank, this means that the constraint $G^T Qx = 0$ amounts to $(\tilde{Q}z)_2 = 0$, where $\tilde{Q}z = \left[\begin{smallmatrix} (\tilde{Q}z)_1 \\ (\tilde{Q}z)_2 \end{smallmatrix}\right]$.

Hence the system reduces to the port-Hamiltonian system without algebraic constraints

$$\dot{z}_1 = J_{11}\big(\tilde{Q}_{11} - \tilde{Q}_{12}\tilde{Q}_{22}^{-1}\tilde{Q}_{21}\big)z_1, \tag{4.8}$$

where $z_1 = G^{\perp}x$ are coordinates for the constrained state space $\mathcal{X}_c = \{x \in \mathcal{X} \mid G^T Qx = 0\}$.

4.1.2 Elimination of Kinematic Constraints

An important example of differential-algebraic port-Hamiltonian systems are mechanical systems subject to kinematic constraints, as discussed in Sect. 2.7. The constrained Hamiltonian equations (2.38) define a port-Hamiltonian system with respect to the Dirac structure $\mathcal{D}$ (in constrained input–output representation)

$$\begin{aligned}\mathcal{D} = \Bigg\{ &(f_S, e_S, f_C, e_C) \mid 0 = \begin{bmatrix} 0 & A^T(q) \end{bmatrix} e_S, e_C = \begin{bmatrix} 0 & B^T(q) \end{bmatrix} e_S, \\ &-f_S = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix} e_S + \begin{bmatrix} 0 \\ A(q) \end{bmatrix} \lambda + \begin{bmatrix} 0 \\ B(q) \end{bmatrix} f_c, \lambda \in \mathbb{R}^k \Bigg\}. \end{aligned} \tag{4.9}$$

The algebraic constraints on the state variables (q, p) are thus given as

$$0 = A^T(q)\frac{\partial H}{\partial p}(q, p). \tag{4.10}$$

The *constrained state space* is therefore given as the following subset of the phase space (q, p):

$$\mathcal{X}_c = \left\{ (q, p) \mid A^T(q)\frac{\partial H}{\partial p}(q, p) = 0 \right\}. \tag{4.11}$$

We may solve for the algebraic constraints and eliminate the resulting constraint forces $A(q)\lambda$ in the following way [37]. Since rank $A(q) = k$, there exists locally an $n \times (n-k)$ matrix $S(q)$ of rank $n-k$ such that

$$A^T(q)S(q) = 0. \tag{4.12}$$

Now define $\tilde{p} = (\tilde{p}^1, \tilde{p}^2) = (\tilde{p}_1, \ldots, \tilde{p}_{n-k}, \tilde{p}_{n-k+1}, \ldots, \tilde{p}_n)$ as

$$\begin{aligned} \tilde{p}^1 &:= S^T(q)p, \\ \tilde{p}^2 &:= A^T(q)p, \end{aligned} \qquad \tilde{p}^1 \in \mathbb{R}^{n-k},\ \tilde{p}^2 \in \mathbb{R}^k. \tag{4.13}$$

It is readily checked that $(q, p) \mapsto (q, \tilde{p}^1, \tilde{p}^2)$ is a coordinate transformation. Indeed, by (4.12), the rows of $S^T(q)$ are orthogonal to the rows of $A^T(q)$. In the new

coordinates, the constrained Hamiltonian system (2.38) takes the form (see [37] for details), $*$ denoting unspecified elements,

$$\begin{aligned}\begin{bmatrix}\dot{q}\\ \dot{\tilde{p}}^1\\ \dot{\tilde{p}}^2\end{bmatrix} &= \begin{bmatrix}0_n & S(q) & *\\ -S^T(q) & (-p^T[S_i,S_j](q))_{i,j} & *\\ * & * & *\end{bmatrix}\begin{bmatrix}\frac{\partial \tilde{H}}{\partial q}\\ \frac{\partial \tilde{H}}{\partial \tilde{p}^1}\\ \frac{\partial \tilde{H}}{\partial \tilde{p}^2}\end{bmatrix}\\ &\quad + \begin{bmatrix}0\\ 0\\ A^T(q)A(q)\end{bmatrix}\lambda + \begin{bmatrix}0\\ B_c(q)\\ B(q)\end{bmatrix}u,\\ A^T(q)\frac{\partial H}{\partial p} &= A^T(q)A(q)\frac{\partial \tilde{H}}{\partial \tilde{p}^2} = 0\end{aligned} \tag{4.14}$$

with $\tilde{H}(q,\tilde{p})$ the Hamiltonian H expressed in the new coordinates $q,\tilde{p}$. Here, S_i denotes the ith column of $S(q)$, $i=1,\ldots,n-k$, and $[S_i,S_j]$ is the *Lie bracket* of S_i and S_j, in local coordinates q given as (see e.g. [1, 23])

$$[S_i,S_j](q) = \frac{\partial S_j}{\partial q}(q)S_i(q) - \frac{\partial S_i}{\partial q}S_j(q) \tag{4.15}$$

with $\frac{\partial S_j}{\partial q}$ and $\frac{\partial S_i}{\partial q}$ denoting the $n\times n$ Jacobian matrices. Since λ only influences the $\tilde{p}^2$-dynamics, and the constraints $A^T(q)\frac{\partial H}{\partial p}(q,p)=0$ are equivalently given by $\frac{\partial \tilde{H}}{\partial \tilde{p}^2}(q,\tilde{p})=0$, the constrained dynamics is determined by the dynamics of q and $\tilde{p}^1$, which serve as coordinates for the constrained state space $\mathcal{X}_c$:

$$\begin{bmatrix}\dot{q}\\ \dot{\tilde{p}}^1\end{bmatrix} = J_c\big(q,\tilde{p}^1\big)\begin{bmatrix}\frac{\partial H_c}{\partial q}(q,\tilde{p}^1)\\ \frac{\partial H_c}{\partial \tilde{p}^1}(q,\tilde{p}^1)\end{bmatrix} + \begin{bmatrix}0\\ B_c(q)\end{bmatrix}u, \tag{4.16}$$

where $H_c(q,\tilde{p}^1)$ equals $\tilde{H}(q,\tilde{p})$ with $\tilde{p}^2$ satisfying $\frac{\partial \tilde{H}}{\partial \tilde{p}^2}=0$, and where the skew-symmetric matrix $J_c(q,\tilde{p}^1)$ is given as the left-upper part of the structure matrix in (4.14), that is,

$$J_c\big(q,\tilde{p}^1\big) = \begin{bmatrix}O_n & S(q)\\ -S^T(q) & (-p^T[S_i,S_j](q))_{i,j}\end{bmatrix}, \tag{4.17}$$

where p is expressed as function of $q,\tilde{p}$, with $\tilde{p}^2$ eliminated from $\frac{\partial \tilde{H}}{\partial \tilde{p}^2}=0$. In fact, for the Hamiltonian $\tilde{H}$ given as

$$\tilde{H}(q,\tilde{p}) = \frac{1}{2}\tilde{p}^T\tilde{M}^{-1}(q)\tilde{p} + P(q)$$

with $\tilde{M}(q) =: N^{-1}(q)$ the transformed mass matrix for the pseudo-momenta, and $V(q)$ the potential energy, it follows that

$$\tilde{H}\big(q, \tilde{p}^1\big) = \frac{1}{2}\tilde{p}^{1T}\big(N_{11} - N_{12}(q)N_{22}^{-1}(q)N_{21}(q)\big)\tilde{p}^1 + P(q).$$

Furthermore, in the coordinates q, $\tilde{p}$, the output map is given in the form

$$y = \begin{bmatrix} B_c^T(q) & \overline{B}^T(q) \end{bmatrix} \begin{bmatrix} \frac{\partial \tilde{H}}{\partial \tilde{p}^1} \\ \frac{\partial \tilde{H}}{\partial \tilde{p}^2} \end{bmatrix} \tag{4.18}$$

which reduces on the constrained state space $\mathcal{X}_c$ to

$$y = B_c^T(q)\frac{\partial \tilde{H}}{\partial \tilde{p}^1}\big(q, \tilde{p}^1\big). \tag{4.19}$$

Summarizing, (4.16) and (4.19) define an *input–state–output* port-Hamiltonian system on $\mathcal{X}_c$, with Hamiltonian H_c given by the constrained total energy, and with structure matrix J_c given by (4.17).

Example 4.1 (Example 2.4, continued) Define according to (4.13) new p-coordinates

$$\begin{aligned} p_1 &= p_\varphi, \\ p_2 &= p_\theta + p_x\cos\varphi + p_y\sin\varphi, \\ p_3 &= p_x - p_\theta\cos\varphi, \\ p_4 &= p_y - p_\theta\sin\varphi. \end{aligned} \tag{4.20}$$

The constrained state space $\mathcal{X}_c$ is given by $p_3 = p_4 = 0$, and the dynamics on $\mathcal{X}_c$ is computed as

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \\ \dot{\varphi} \\ \dot{p}_1 \\ \dot{p}_2 \end{bmatrix} = \begin{bmatrix} & & & & 0 & \cos\varphi \\ & & & & 0 & \sin\varphi \\ & O_4 & & & 0 & 1 \\ & & & & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ -\cos\varphi & -\sin\varphi & -1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial H_c}{\partial x} \\ \frac{\partial H_c}{\partial y} \\ \frac{\partial H_c}{\partial \theta} \\ \frac{\partial H_c}{\partial \varphi} \\ \frac{\partial H_c}{\partial p_1} \\ \frac{\partial H_c}{\partial p_2} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}p_2 \\ p_1 \end{bmatrix}, \tag{4.21}$$

where $H_c(x, y, \theta, \varphi, p_1, p_2) = \frac{1}{2}p_1^2 + \frac{1}{4}p_2^2$.

4.2 The Geometric Description of Algebraic Constraints of Port-Hamiltonian DAEs

We will start by considering port-Hamiltonian differential-algebraic systems without external and resistive ports, described by a Dirac structure $\mathcal{D}$ and a Hamiltonian H. Define for every $x \in \mathcal{X}$ the subspace

$$P_{\mathcal{D}}(x) := \left\{\alpha \in T_x^*\mathcal{X} \mid \exists X \in T_x\mathcal{X} \text{ such that } (\alpha, X) \in \mathcal{D}(x)\right\}. \tag{4.22}$$

This defines a *co-distribution* on the manifold $\mathcal{X}$. Then it follows from the definition of a port-Hamiltonian system that the algebraic constraints are given in coordinate-free form as

$$\frac{\partial H}{\partial x}(x) \in P_{\mathcal{D}}(x). \tag{4.23}$$

Thus from a Dirac structure point of view algebraic constraints may only arise if the Dirac structure $\mathcal{D}$ is such that its associated co-distribution $P_{\mathcal{D}}$ is *not equal* to the whole cotangent bundle $T^*\mathcal{X}$, that is, if $P_{\mathcal{D}}(x)$ is a strict subspace of $T_x^*\mathcal{X}$.

The particular equational representation of the algebraic constraints depends on the chosen representation of the Dirac structure. For example, if the Dirac structure and the port-Hamiltonian system is given in constrained input–output representation (3.21), then the algebraic constraints are, as discussed above, given by $b^T(x)\frac{\partial H}{\partial x}(x) = 0$. On the other hand, if the Dirac structure is given in image representation as

$$\mathcal{D}(x) = \left\{(X, \alpha) \in T_x\mathcal{X} \times T_x^*\mathcal{X} \mid X = E^T(x)\lambda, \alpha = F^T(x)\lambda\right\} \tag{4.24}$$

then the algebraic constraints amount to the satisfaction of

$$\frac{\partial H}{\partial x}(x) \in \operatorname{im} F^T(x). \tag{4.25}$$

In the case of external ports, the algebraic constraints on the state variables x may also depend on the external port variables. A special case arises for resistive ports. Consider a Dirac structure

$$\left\{(X, \alpha, f_R, e_R) \in \mathcal{D}(x) \subset T_x\mathcal{X} \times T_x^*\mathcal{X} \times \mathcal{F}_R \times \mathcal{F}_R^*\right\} \tag{4.26}$$

with the resistive flow and effort variables satisfying a relation $R(f_R, e_R) = 0$. Then the gradient of the Hamiltonian has to satisfy the condition

$$\frac{\partial H}{\partial x}(x) \in \Big\{\alpha \in T_x^*\mathcal{X} \mid \exists X, f_R, e_R \in T_x\mathcal{X} \times \mathcal{F}_R \times \mathcal{F}_R^* \text{ such that } (X, \alpha, f_R, e_R) \in \mathcal{D}(x) \text{ and } R(f_R, e_R) = 0\Big\}.$$

Depending on the resistive relation $R(f_R, e_R) = 0$ this may again induce algebraic constraints on the state variables x.

4.2.1 Algebraic Constraints in the Canonical Coordinate Representation

A particular elegant representation of algebraic constraints arises from the *canonical coordinate representation*. We will only consider the case of a system without resistive and external ports. For a constant Dirac structure $\mathcal{D}$, there always exist (linear) canonical coordinates such that the Dirac structure is described by (3.7). If on the other hand $\mathcal{D}$ is a modulated Dirac structure on a manifold $\mathcal{X}$, then only if the Dirac structure $\mathcal{D}$ satisfies an additional *integrability condition*,[9] we can choose local coordinates $x = (q, p, r, s)$ for $\mathcal{X}$ (with $\dim q = \dim p$), such that, in the corresponding bases for (f_q, f_p, f_r, f_s) for $T_x\mathcal{X}$ and (e_q, e_p, e_r, e_s) for $T_x^*\mathcal{X}$, the Dirac structure on this coordinate neighborhood is still given by the relations (3.7).

Substituting in this case the flow and effort relations of the energy storage

$$\begin{aligned} f_q &= -\dot{q}, \qquad e_q = \frac{\partial H}{\partial q}, \\ f_p &= -\dot{p}, \qquad e_p = \frac{\partial H}{\partial p}, \\ f_r &= -\dot{r}, \qquad e_r = \frac{\partial H}{\partial r}, \\ f_s &= -\dot{s}, \qquad e_s = \frac{\partial H}{\partial s} \end{aligned} \tag{4.27}$$

into the canonical coordinate representation (3.7) of the Dirac structure yields the following dynamics:

$$\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p}(q, p, r, s), \\ \dot{p} &= -\frac{\partial H}{\partial q}(q, p, r, s), \\ \dot{r} &= 0, \\ 0 &= \frac{\partial H}{\partial s}(q, p, r, s). \end{aligned} \tag{4.28}$$

The variables q, p are the canonical coordinates known from classical Hamiltonian dynamics, while the variables r have the interpretation of Casimirs (conserved quantities independent of the Hamiltonian), see Sect. 4.3. The last equations $\frac{\partial H}{\partial s} = 0$ specify the algebraic constraints present in the system; in a form that is reminiscent of the first-order condition for optimality in the Maximum principle in optimal control theory.

The condition that the matrix in (4.2) has full rank (implying that the system has index one; cf. Proposition 4.1) is in the canonical coordinate representation equivalent to the partial Hessian matrix $\frac{\partial^2 H}{\partial s^2}$ being invertible. Solving, by the Implicit

[9] For more details regarding the precise form of the integrability conditions see Sect. 7.

Function theorem, the algebraic constraints $\frac{\partial H}{\partial s} = 0$ for s as a function $s(q, p, r)$ reduces the DAEs (4.28) to the ODEs

$$\begin{aligned} \dot{q} &= \frac{\partial \bar{H}}{\partial p}(q, p, r), \\ \dot{p} &= -\frac{\partial \bar{H}}{\partial q}(q, p, r), \\ \dot{r} &= 0, \end{aligned} \tag{4.29}$$

where $\bar{H}(q, p, r) := H(q, p, r, s(q, p, r))$.

4.3 Casimirs of Port-Hamiltonian DAEs

Consider a port-Hamiltonian differential-algebraic system without external and resistive ports, with Dirac structure $\mathcal{D}$ involving f_S, e_S. Similarly to (4.22) we may define the following, smaller, co-distribution:

$$\bar{P}_{\mathcal{D}}(x) := \left\{\alpha \in T_x^*\mathcal{X} \mid (\alpha, 0) \in \mathcal{D}(x)\right\} \tag{4.30}$$

This co-distribution will characterize the *conserved quantities* that are *independent* of the Hamiltonian H. In fact, it can be seen that $\bar{P}_{\mathcal{D}} = G_{\mathcal{D}}^{\perp}$, where $G_{\mathcal{D}}$ is the distribution on $\mathcal{X}$ defined as

$$G_{\mathcal{D}}(x) := \left\{X \in T_x\mathcal{X} \mid \exists \alpha \in T_x^*\mathcal{X} \text{ such that } (\alpha, X) \in \mathcal{D}(x)\right\}. \tag{4.31}$$

Now, let $C : \mathcal{X} \to \mathbb{R}$ be a function such that $\frac{\partial^T C}{\partial x}(x) \in \bar{P}_{\mathcal{D}}(x)$. Then

$$\frac{d}{dt}C\big(x(t)\big) = \frac{\partial^T C}{\partial x}\big(x(t)\big)\dot{x}(t) = 0, \tag{4.32}$$

for all possible vectors $\dot{x}(t)$ occurring in the system equations; independently of the Hamiltonian H. Such functions $C : \mathcal{X} \to \mathbb{R}$ are called the *Casimirs* of the system, and are very important for the analysis of the system. Note that the existence of finding functions C such that $\frac{\partial^T C}{\partial x}(x) \in \bar{P}_{\mathcal{D}}(x), x \in \mathcal{X}$ is related to the *integrability* of the co-distribution $\bar{P}_{\mathcal{D}}$, and thus to the integrability of the Dirac structure $\mathcal{D}$, cf. Sect. 7. Thus the Casimirs are completely characterized by the Dirac structure of the port-Hamiltonian system.

Similarly, we define the Casimirs of a port-Hamiltonian differential-algebraic system with a resistive relation to be all functions $C : \mathcal{X} \to \mathbb{R}$ satisfying $(0, e = \frac{\partial^T C}{\partial x}, 0, 0) \in \mathcal{D}$. Indeed, this will imply that

$$\frac{d}{dt}C = \frac{\partial^T C}{\partial x}\big(x(t)\big)\dot{x}(t) = 0 \tag{4.33}$$

for every possible derivative vector $\dot{x}$ occurring in the system equations, independently of the Hamiltonian and of the resistive relations.[10]

Example 4.2 In the case of a spinning rigid body (Example 2.3) the well-known Casimir is the total angular momentum $p_x^2 + p_y^2 + p_z^2$ (whose vector of partial derivatives is indeed in the kernel of the matrix $J(p)$ in (2.30)).

Similarly, in the LC-circuit of Example 2.5 the total flux $\phi_1 + \phi_2$ is a Casimir for $u = 0$.

Example 4.3 Consider a mechanical system with kinematic constraints (2.38) and $u = 0$. Then $(0, e) \in \mathcal{D}$ if and only if

$$0 = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ A(q) \end{bmatrix} \lambda, \qquad \begin{bmatrix} 0 & A^T(q) \end{bmatrix} e = 0.$$

Partioning $e = \begin{bmatrix} e_q \\ e_p \end{bmatrix}$ this means that $e_q = A(q)\lambda$, or equivalently, $e_q \in \operatorname{im} A(q)$. Since in general $A(q)$ is depending on q, finding Casimirs involves an additional *integrability condition*, see also Sect. 7. In fact, Casimirs correspond to vectors $e_q \in \operatorname{im} A(q)$ which additionally can be written as a vector of partial derivatives $\frac{\partial C}{\partial q}(q)$ for some function $C(q)$ (the Casimir). In the case of Example 4.1 it can be verified that this additional integrability condition is *not* satisfied, corresponding to the fact that the kinematic constraints in this example are completely *nonholonomic*.

In general it can be shown [37] that there exist as many independent Casimirs as the rank of the matrix $A(q)$ if and only if the kinematic constraints are holonomic, in which case the Casimirs are equal to the integrated kinematic constraints.

4.4 Stability Analysis of Port-Hamiltonian DAEs

As we have seen before, any port-Hamiltonian differential-algebraic system, without control and interaction ports, satisfies the energy balance (2.16), that is,

$$\frac{d}{dt} H = e_R^T f_R \leq 0. \tag{4.34}$$

This immediately follows from the power-conserving property of Dirac structures. As a consequence, the Hamiltonian H qualifies as a Lyapunov function if it is bounded from below.

Recently, the notion of Lyapunov functions for general nonlinear DAE systems was studied in depth in [18]; also providing a formal treatment of asymptotic stability. Let us show, again by exploiting the properties of Dirac structures, how the

[10] However it can be shown [26] that if (4.33) holds for *some* non-degenerate resistive relation then it has to hold for *all*.

Hamiltonian H of a port-Hamiltonian differential-algebraic system also defines a Lyapunov function in this set-up.[11] Consider a port-Hamiltonian system without control and interaction ports, whose Dirac structure $\mathcal{D}$ is given in *kernel representation* as, see (3.9),

$$\left\{(f_S, e_S, f_R, e_R) \mid F_S(x)f_S + E_S(x)e_S + F_R(x)f_R + E_R(x)e_R = 0\right\} \tag{4.35}$$

with

$$\begin{aligned} &F_S(x)E_S^T(x) + E_S(x)F_S^T(x) + F_R(x)E_R^T(x) + E_R(x)F_R^T(x) = 0, \\ &\operatorname{rank}\begin{bmatrix} F_S(x) & E_S(x) & F_R(x) & E_R(x) \end{bmatrix} = \dim f_S + \dim f_R. \end{aligned} \tag{4.36}$$

It follows that $\mathcal{D}$ is equivalently given in *image representation* as

$$\begin{aligned} f_S &= E_S^T(x)\lambda(x), \qquad & e_S &= F_S^T(x)\lambda(x), \\ f_R &= E_R^T(x)\lambda(x), \qquad & e_R &= F_R^T(x)\lambda(x). \end{aligned} \tag{4.37}$$

The resulting port-Hamiltonian system for a Hamiltonian H is thus given by the equations

$$\begin{aligned} \dot{x} &= -E_S^T(x)\lambda(x), \qquad & \frac{\partial H}{\partial x}(x) &= F_S^T(x)\lambda(x), \\ f_R &= E_R^T(x)\lambda(x), \qquad & e_R &= F_R^T(x)\lambda(x) \end{aligned} \tag{4.38}$$

(together with energy-dissipating constitutive relations). In particular

$$\frac{\partial^T H}{\partial x}(x)z = F_S^T(x)\lambda^T(x)F_S(x)z \tag{4.39}$$

for all vectors z; in agreement with one of the requirements for H being a Lyapunov function as stated in [18]. It also follows from here that (using the first line of (4.36))

$$\begin{aligned} \dot{H} &= -\lambda^T(x)F_S(x)\dot{x} = -\lambda^T(x)F_S(x)E_S^T(x)\lambda(x) \\ &= \lambda^T(x)F_R(x)E_R^T(x)\lambda(x) = e_R^T f_R \leq 0 \end{aligned} \tag{4.40}$$

being another condition in the formulation of [18].

Finally, if H does *not* have a minimum at a desired equilibrium x^*, then a well-known method in Hamiltonian dynamics, called the *Energy-Casimir method*, is to use in the Lyapunov analysis, next to the Hamiltonian function, additional *conserved quantities* of the system, in particular the Casimirs. Indeed, candidate Lyapunov functions can be sought within the class of *combinations* of the Hamiltonian H and the Casimirs. For more information we refer to e.g. [24, 25, 31]. Most of this literature is however on port-Hamiltonian systems *without* algebraic constraints, and the

[11] I thank Stephan Trenn for an enlightening discussion on this issue.

presence of algebraic constraints poses new questions. For example, the necessary conditions for a Lyapunov function only have to hold on the subset of the state space where the algebraic constraints are satisfied.

4.5 Ill-posedness Due to Nonlinear Resistive Characteristics

In the above we have largely confined ourselves to differential-algebraic port-Hamiltonian systems without resistive relations. The presence of such resistive relations, especially in the nonlinear case, may pose additional difficulties. In particular, well-posedness problems may arise for port-Hamiltonian systems where the flow variables of the resistive ports are input variables for the dynamics, while the resistive relation is *not* effort-controlled. We will not elaborate on this (difficult) topic, but confine ourselves to an example (taken from [10]) illustrating the problems which may arise.

Example 4.4 (Degenerate Van der Pol oscillator) Consider a degenerate form of the Van der Pol oscillator consisting of a unit capacitor

$$\dot{Q} = I, \qquad V = Q \tag{4.41}$$

in parallel with a nonlinear resistor given by the characteristic

$$\left\{ (f_R, e_R) = (I, V) \mid V = -\frac{1}{3} I^3 + I \right\}. \tag{4.42}$$

This resistive characteristic is *not* voltage-controlled, but instead is current-controlled. As a consequence, Eqs. (4.41) and (4.42) define an implicitly defined dynamics on the one-dimensional constraint submanifold R in (I, V) space given by

$$R = \left\{ (I, V) \mid V + \frac{1}{3} I^3 - I = 0 \right\}.$$

Difficulties in the dynamical interpretation arise at the points $(-1, -\frac{2}{3})$ and $(1, \frac{2}{3})$. At these points $\dot{V}$ is negative, respectively positive (while the corresponding time-derivative of I at these points tends to plus or minus infinity, depending on the direction along which these points are approached). Hence, because of the form of the constraint manifold R it is not possible to "integrate" the dynamics from these points (sometimes called *impasse points*) in a continuous manner along R.

For a careful analysis of the dynamics of this system we refer to [10]. In particular, it has been suggested in [10] that a suitable interpretation of the dynamics from the impasse points is given by the following *jump rules*:

$$\left(-1, -\frac{2}{3}\right) \to \left(2, -\frac{2}{3}\right), \qquad \left(1, \frac{2}{3}\right) \to \left(-2, \frac{2}{3}\right). \tag{4.43}$$

The resultant trajectory (switching from the region $I \leq -1$ to the region $I \geq 1$) is a 'limit cycle' that is known as a *relaxation oscillation*. For related examples in the context of constrained mechanical systems we refer to [5].

Existence and uniqueness of solutions *is* guaranteed if the resistive relation is well-behaved and the DAEs are of index one as discussed in the previous Sect. 4.1. Indeed, consider again the case of a port-Hamiltonian system given in the constrained input–output representation

$$\begin{aligned} \dot{x} &= \big[J(x) - R(x)\big]\frac{\partial H}{\partial x}(x) + g(x)u + b(x)\lambda, \\ y &= g^T(x)\frac{\partial H}{\partial x}(x), \qquad\qquad x \in \mathcal{X}. \\ 0 &= b^T(x)\frac{\partial H}{\partial x}(x), \end{aligned} \tag{4.44}$$

Imposing the same condition as before in Sect. 4.1, Proposition 4.1, namely that the matrix

$$b^T(x)\frac{\partial^2 H}{\partial x^2}(x)b(x) \tag{4.45}$$

has full rank, it can be seen that there is a unique solution starting from every feasible initial condition $x_0 \in \mathcal{X}_c$. Furthermore, this solution will remain in the constrained state space $\mathcal{X}_c$ for all time.

Example 4.5 A simple, but illustrative, example of a case where multiple solutions arise from feasible initial conditions can be deduced from the example of a linear LC-circuit with standard Hamiltonian $H(Q,\phi) = \frac{1}{2C}Q^2 + \frac{1}{2L}\phi^2$, where the voltage across the capacitor is constrained to be zero:

$$\begin{aligned} \dot{Q} &= \frac{1}{L}\phi + \lambda, \\ \dot{\phi} &= \frac{1}{C}Q, \\ 0 &= \frac{1}{C}Q. \end{aligned} \tag{4.46}$$

Here λ denotes the current through the external port whose voltage is set equal to zero. Since $b^T(x)\frac{\partial^2 H}{\partial x^2}(x)b(x)$ in this case reduces to $\frac{1}{C}$ it follows that there is a unique solution starting from every feasible initial condition. Indeed, the constrained state space $\mathcal{X}_c$ of the above port-Hamiltonian system is given by $\{(Q,\phi) \mid Q = 0\}$, while the Lagrange multiplier λ for any feasible initial condition $(0,\phi_0)$ is uniquely determined as $\lambda = -\frac{1}{L}\phi$. On the other hand, in the singular case where $C = \infty$ the Hamiltonian reduces to $H(Q,\phi) = \frac{1}{2L}\phi^2$ and the constraint equation $0 = \frac{1}{C}Q$ becomes vacuous, i.e., there are no constraints anymore. In this case the Lagrange

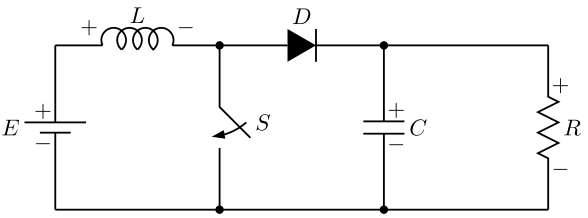


Fig. 4 Boost circuit with clamping diode

multiplier λ (the current through the external port) is not determined anymore, leading to multiple solutions $(Q(t), \phi(t))$ where $\phi(t)$ is constant (equal to the initial value ϕ_0) while $Q(t)$ is an *arbitrary* function of time.

5 Port-Hamiltonian Systems with Variable Topology

In many cases of interest it is useful to model fast transitions in physical systems as *instantaneous switches*. Examples include the description of elements like diodes and thyristors in electrical circuits, and impacts in mechanical systems. Within the port-Hamiltonian description, one obtains in all these cases an (idealized) model where the Dirac structure is not constant, but depends on the position of the switches. On the other hand, the Hamiltonian H and the resistive relations are usually *independent* of the position of the switches.

In both examples below, we thus obtain a *switching* port-Hamiltonian system, specified by a Dirac structure $\mathcal{D}_s$ depending on the switch position $s \in \{0, 1\}^n$ (here n denotes the number of independent switches), a Hamiltonian $H : \mathcal{X} \to \mathbb{R}$, and a resistive structure $\mathcal{R}$. Furthermore, every switching may be internally induced (like in the case of a diode in an electrical circuit or an impact in a mechanical system) or externally triggered (like an active switch in a circuit or mechanical system).

Example 5.1 (Boost converter) Consider the power converter in Fig. 4. The circuit consists of an inductor L with magnetic flux linkage ϕ_L, a capacitor C with electric charge q_C and a resistance load R, together with a diode D and an ideal switch S, with switch positions $s = 1$ (switch closed) and $s = 0$ (switch open). The diode is modeled as an ideal diode with voltage-current characteristic $v_D i_D = 0$, with $v_D \leq 0$ and $i_D \geq 0$.

The state variables are the electric charge Q_C and the magnetic flux linkage φ_L, and the stored energy (Hamiltonian) is the quadratic function $\frac{1}{2C} Q_C^2 + \frac{1}{2L} \varphi_L^2$. Note that there are four modes of operation of this system corresponding to the positions of the active switch (open or closed) and the diode (voltage- or current blocking). Two out of these four modes correspond to an algebraic constraint: namely $Q_C = 0$ if the switch is closed and the diode has $v_D = 0$, and $\varphi_L = 0$ if the switch is open and the diode has $i_D = 0$. (These two exceptional modes are sometimes called the *discontinuous modes* in the power converter literature.)

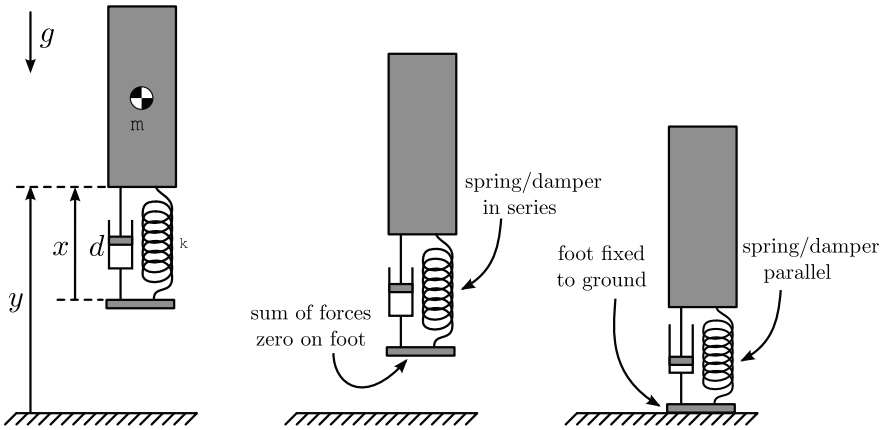


Fig. 5 Model of a bouncing pogo-stick: definition of the variables (*left*), situation without ground contact (*middle*), and situation with ground contact (*right*)

Example 5.2 (Bouncing pogo-stick) Consider the example of the vertically bouncing pogo-stick in Fig. 5: it consists of a mass m and a massless foot, interconnected by a linear spring (stiffness k and rest-length x_0) and a linear damper d. The mass can move vertically under the influence of gravity g until the foot touches the ground. The states of the system are taken as x (length of the spring), y (height of the bottom of the mass), and p (momentum of the mass, defined as $p := m\dot{y}$). Furthermore, the contact situation is described by a variable s with values $s = 0$ (no contact) and $s = 1$ (contact). The total energy (Hamiltonian) of the system equals

$$H(x, y, p) = \frac{1}{2}k(x - x_0)^2 + mg(y + y_0) + \frac{1}{2m}p^2, \tag{5.1}$$

where y_0 is the distance from the bottom of the mass to its center of mass.

When the foot is not in contact with the ground (middle figure), the total force on the foot is zero (since it is mass-less), which implies that the spring and damper forces must be equal but opposite. When the foot is in contact with the ground (right figure), the variables x and y remain equal, and hence also $\dot{x} = \dot{y}$. For $s = 0$ (no contact) the system can be described by the port-Hamiltonian system

$$\frac{d}{dt}\begin{bmatrix} x \\ y \\ p \end{bmatrix} = \begin{bmatrix} -\frac{1}{d} & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} k(x - x_0) \\ mg \\ \frac{p}{m} \end{bmatrix} \tag{5.2}$$

i.e. two independent systems (spring plus damper, and mass plus gravity), while for $s = 1$, the port-Hamiltonian description of the system is given as

$$\frac{d}{dt}\begin{bmatrix} x \\ y \\ p \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ -1 & -1 & -d \end{bmatrix} \begin{bmatrix} k(x - x_0) \\ mg \\ \frac{p}{m} \end{bmatrix}. \tag{5.3}$$

In this last case the resistive force $-d\dot{x}$ is added to the spring force and the gravitational force exerted on the mass, while for $s = 0$ the resistive force is equal to the spring force.

The two situations can be taken together into one port-Hamiltonian system with switching Dirac structure as follows

$$\frac{d}{dt}\begin{bmatrix} x \\ y \\ p \end{bmatrix} = \begin{bmatrix} \frac{s-1}{d} & 0 & s \\ 0 & 0 & 1 \\ -s & -1 & -sd \end{bmatrix} \begin{bmatrix} k(x - x_0) \\ mg \\ \frac{p}{m} \end{bmatrix}. \tag{5.4}$$

In addition, the conditions for switching of the contact are functions of the states, namely as follows: contact is switched from off to on when $y - x$ crosses zero in the negative direction, and contact is switched from on to off when the velocity $\dot{y} - \dot{x}$ of the foot is positive in the no-contact situation, i.e. when $\frac{p}{m} + \frac{k}{d}(x - x_0) > 0$.

In the present modeling of the system no algebraic constraints arise in any of the two modes. It should be noted, however, that this is critically on the assumption of a massless foot. Indeed, if the mass of the foot is taken into the account then another state variable (namely the momentum of the foot) needs to be taken into account, while the contact situation would correspond to this extra state variable being constrained to zero.

We note that because the Hamiltonian function is *common* to all the modes of the switching port-Hamiltonian system it still can be employed for the stability analysis, see e.g. [6, 14]. Clearly this presents enormous advantages as compared to the stability analysis of general switched differential-algebraic systems [18]. The presence of algebraic constraints in (some of) the modes poses another question: the specification of the instantaneous reset of the state at the moment of switching in order to satisfy the algebraic constraints of the mode which is active immediately after the switching time. This involves the determination of consistent state *reset rules*. For a rather complete analysis in the context of switching electrical circuits we refer to the treatment in [6, 35]. The study of reset rules and mode selection is a classical subject in mechanical systems; see [5] and the references therein. For the related theory of *complementarity* hybrid systems we refer to [41, 42].

6 Interconnection of Port-Hamiltonian Systems and Composition of Dirac Structures

Crucial feature of network modeling, analysis and control is 'interconnectivity' or 'compositionality', meaning that complex systems can be built from simpler parts, and that the complex system can be studied in terms of its constituent parts and the way they are interconnected. The class of port-Hamiltonian systems completely fits within this paradigm, in the sense that the power-conserving interconnection of port-Hamiltonian systems again defines a port-Hamiltonian system. Furthermore, it will turn out that the Hamiltonian of the interconnected system is simply the sum of the

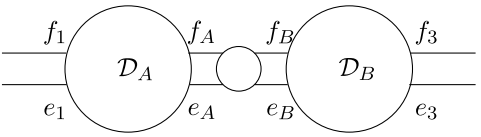


Fig. 6 The composition of $\mathcal{D}_A$ and $\mathcal{D}_B$

Hamiltonians of its parts, while the Dirac structure $\mathcal{D}$ of the interconnected system is solely determined by the Dirac structures of its components. This is clearly of immediate relevance for port-Hamiltonian differential-algebraic systems, since, as we have seen, the algebraic constraints are determined by the overall Dirac structure $\mathcal{D}$, in particular its co-distribution $P_{\mathcal{D}}$, and the overall Hamiltonian.

6.1 Composition of Dirac Structures

In this subsection, we investigate the *composition* or *interconnection* properties of Dirac structures. Physically it is clear that the composition of a number of power-conserving interconnections with partially shared variables should yield again a power-conserving interconnection. We show how this can be formalized within the framework of Dirac structures.

First, we consider the composition of *two* Dirac structures with partially shared variables. Once we have shown that the composition of two Dirac structures is again a Dirac structure, it is immediate that the power-conserving interconnection of any number of Dirac structures is again a Dirac structure.[12] Thus consider a Dirac structure $\mathcal{D}_A$ on a product space $\mathcal{F}_1 \times \mathcal{F}_2$ of two linear spaces $\mathcal{F}_1$ and $\mathcal{F}_2$, and another Dirac structure $\mathcal{D}_B$ on a product space $\mathcal{F}_2 \times \mathcal{F}_3$, with also $\mathcal{F}_3$ being a linear space. The linear space $\mathcal{F}_2$ is the space of shared flow variables, and $\mathcal{F}_2^*$ the space of shared effort variables; see 6.

In order to compose $\mathcal{D}_A$ and $\mathcal{D}_B$, a problem arises of *sign* convention for the power flow corresponding to the power variables $(f_2, e_2) \in \mathcal{F}_2 \times \mathcal{F}_2^*$. Indeed, if $\langle e \mid f \rangle$ denotes *incoming* power (see the previous section), then for

$$(f_1, e_1, f_A, e_A) \in \mathcal{D}_A \subset \mathcal{F}_1 \times \mathcal{F}_1^* \times \mathcal{F}_2 \times \mathcal{F}_2^*$$

the term $\langle e_A \mid f_A \rangle$ denotes the incoming power in $\mathcal{D}_A$ due to the power variables $(f_A, e_A) \in \mathcal{F}_2 \times \mathcal{F}_2^*$, while for

$$(f_B, e_B, f_3, e_3) \in \mathcal{D}_B \subset \mathcal{F}_2 \times \mathcal{F}_2^* \times \mathcal{F}_3 \times \mathcal{F}_3^*$$

the term $\langle e_B \mid f_B \rangle$ denotes the incoming power *in* $\mathcal{D}_B$. Clearly, the *incoming* power in $\mathcal{D}_A$ due to the power variables in $\mathcal{F}_2 \times \mathcal{F}_2^*$ should equal the *outgoing* power from

[12] See [2] for a direct approach to the composition of multiple Dirac structures.

$\mathcal{D}_B$. Thus we cannot simply equate the flows f_A and f_B and the efforts e_A and e_B, but instead we define the interconnection constraints as

$$f_A = -f_B \in \mathcal{F}_2, \qquad e_A = e_B \in \mathcal{F}_2^*. \tag{6.1}$$

Therefore, the *composition* of the Dirac structures $\mathcal{D}_A$ and $\mathcal{D}_B$, denoted $\mathcal{D}_A \parallel \mathcal{D}_B$, is defined as

$$\begin{aligned}\mathcal{D}_A \parallel \mathcal{D}_B := \big\{(f_1, e_1, f_3, e_3) \in \mathcal{F}_1 \times \mathcal{F}_1^* \times \mathcal{F}_3 \times \mathcal{F}_3^* \mid \exists (f_2, e_2) \in \mathcal{F}_2 \times \mathcal{F}_2^* \\ \text{s.t. } (f_1, e_1, f_2, e_2) \in \mathcal{D}_A \text{ and } (-f_2, e_2, f_3, e_3) \in \mathcal{D}_B\big\}.\end{aligned} \tag{6.2}$$

The fact that the composition of two Dirac structures is again a Dirac structure has been proved in [9, 30]. Here we follow the simpler alternative proof provided in [7] (inspired by a result in [21]), which, among other things, allows one to obtain explicit representations of the composed Dirac structure.

Theorem 6.1 *Let $\mathcal{D}_A$ and $\mathcal{D}_B$ be Dirac structures (defined with respect to $\mathcal{F}_1 \times \mathcal{F}_1^* \times \mathcal{F}_2 \times \mathcal{F}_2^*$, respectively $\mathcal{F}_2 \times \mathcal{F}_2^* \times \mathcal{F}_3 \times \mathcal{F}_3^*$, and their bilinear forms). Then $\mathcal{D}_A \parallel \mathcal{D}_B$ is a Dirac structure with respect to the bilinear form on $\mathcal{F}_1 \times \mathcal{F}_1^* \times \mathcal{F}_3 \times \mathcal{F}_3^*$.*

In the following theorem, an explicit expression is given for the composition of two Dirac structures in terms of a matrix kernel/image representation.

Theorem 6.2 *Let $\mathcal{F}_i, i = 1, 2, 3$ be finite-dimensional linear spaces with $\dim \mathcal{F}_i = n_i$. Consider Dirac structures $\mathcal{D}_A \subset \mathcal{F}_1 \times \mathcal{F}_1^* \times \mathcal{F}_2 \times \mathcal{F}_2^*$, $n_A = \dim \mathcal{F}_1 \times \mathcal{F}_2 = n_1 + n_2$, $\mathcal{D}_B \subset \mathcal{F}_2 \times \mathcal{F}_2^* \times \mathcal{F}_3 \times \mathcal{F}_3^*$, $n_B = \dim \mathcal{F}_2 \times \mathcal{F}_3 = n_2 + n_3$, given by relaxed matrix kernel/image representations $(F_A, E_A) = ([F_1 \mid F_{2A}], [E_1 \mid E_{2A}])$, with F_A and E_A $n'_A \times n_A$ matrices, $n'_A \geq n_A$, respectively $(F_B, E_B) = ([F_{2B} \mid F_3], [E_{2B} \mid E_3])$, with F_B and E_B $n'_B \times n_B$ matrices, $n'_B \geq n_B$. Define the $(n'_A + n'_B) \times 2n_2$ matrix*

$$M = \begin{bmatrix} F_{2A} & E_{2A} \\ -F_{2B} & E_{2B} \end{bmatrix} \tag{6.3}$$

and let L_A and L_B be $m \times n'_A$, respectively $m \times n'_B$, matrices ($m := \dim \ker M^T$), with

$$L = [L_A \mid L_B], \quad \ker L = \operatorname{im} M. \tag{6.4}$$

Then

$$F = [L_A F_1 \mid L_B F_3], \qquad E = [L_A E_1 \mid L_B E_3] \tag{6.5}$$

is a relaxed matrix kernel/image representation of $\mathcal{D}_A \parallel \mathcal{D}_B$.

Remark 6.1 The relaxed kernel/image representation (6.5) can be readily understood by pre-multiplying the equations characterizing the composition of $\mathcal{D}_A$

with $\mathcal{D}_B$

$$\begin{bmatrix} F_1 & E_1 & F_{2A} & E_{2A} & 0 & 0 \\ 0 & 0 & -F_{2B} & E_{2B} & F_3 & E_3 \end{bmatrix} \begin{bmatrix} f_1 \\ e_1 \\ f_2 \\ e_2 \\ f_3 \\ e_3 \end{bmatrix} = 0, \tag{6.6}$$

by the matrix $L := [L_A | L_B]$. Since $LM = 0$ this results in the relaxed kernel representation

$$L_A F_1 f_1 + L_A E_1 e_1 + L_B F_3 f_3 + L_B E_3 e_3 = 0 \tag{6.7}$$

corresponding to (6.5).

Instead of the canonical interconnection constraints $f_A = -f_B$, $e_A = e_B$ (cf. (6.1)), another standard power-conserving interconnection is the 'gyrative' interconnection

$$f_A = e_B, \qquad f_B = -e_A. \tag{6.8}$$

Composition of two Dirac structures $\mathcal{D}_A$ and $\mathcal{D}_B$ by this gyrative interconnection also results in a Dirac structure. In fact, the gyrative interconnection of $\mathcal{D}_A$ and $\mathcal{D}_B$ equals the interconnection $\mathcal{D}_A \parallel \mathcal{I} \parallel \mathcal{D}_B$, where $\mathcal{I}$ is the gyrative (or *symplectic*) Dirac structure

$$f_{IA} = -e_{IB}, \qquad f_{IB} = e_{IA} \tag{6.9}$$

interconnected to $\mathcal{D}_A$ and $\mathcal{D}_B$ via the canonical interconnections $f_{IA} = -f_A, e_{IA} = e_A$ and $f_{IB} = -f_B, e_{IB} = e_B$.

Example 6.1 (Feedback interconnection) The standard negative feedback interconnection of two input–state–output systems can be regarded as an example of a gyrative interconnection as above. Indeed, let us consider two input–state–output systems as in (2.41), for simplicity *without* external inputs d and external outputs z,

$$\Sigma_i : \begin{cases} \dot{x}_i = [J_i(x_i) - R_i(x_i)] \dfrac{\partial H_i}{\partial x_i}(x_i) + g_i(x_i) u_i, \\ y_i = g_i^T(x_i) \dfrac{\partial H_i}{\partial x_i}(x_i), \end{cases} \qquad x_i \in \mathcal{X}_i \tag{6.10}$$

for $i = 1, 2$. The standard feedback interconnection

$$u_1 = -y_2, \qquad u_2 = y_1 \tag{6.11}$$

is equal to the negative gyrative interconnection between the flows u_1, u_2 and the efforts y_1, y_2. The closed-loop system is the port-Hamiltonian system

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \end{bmatrix} = \begin{bmatrix} J_1(x) - R_1(x_1) & -g_1(x_1)g_2^T(x_2) \\ g_2(x_2)g_1^T(x_1) & J_2(x_2) - R_2(x_2) \end{bmatrix} \begin{bmatrix} \frac{\partial H_1}{\partial x_1}(x_1) \\ \frac{\partial H_2}{\partial x_2}(x_2) \end{bmatrix}$$

with state space $\mathcal{X}_1 \times \mathcal{X}_2$ and Hamiltonian $H_1(x_1) + H(x_2)$. This once more emphasizes the close connections of port-Hamiltonian systems theory with *passivity* theory.

6.2 Interconnection of Port-Hamiltonian Systems

The result derived in Sect. 6.1 concerning the compositionality of Dirac structures immediately leads to the result that any power-conserving interconnection of port-Hamiltonian systems again defines a port-Hamiltonian system. This can be regarded as a fundamental building block in the theory of port-Hamiltonian systems. The result not only means that the theory of port-Hamiltonian systems is a completely modular theory for modeling, but it also serves as a starting point for design and control.

Consider k port-Hamiltonian systems $(\mathcal{X}_i, \mathcal{F}_i, \mathcal{D}_i, H_i)$, $i = 1, \ldots, k$, interconnected by a Dirac structure $\mathcal{D}_I$ on $\mathcal{F}_1 \times \cdots \times \mathcal{F}_k \times \mathcal{F}$, with $\mathcal{F}$ a linear space of flow port variables. This can be seen to define a port-Hamiltonian system $(\mathcal{X}, \mathcal{F}, \mathcal{D}, H)$, where $\mathcal{X} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_k$, $H := H_1 + \cdots + H_k$, and where the Dirac structure $\mathcal{D}$ on $\mathcal{X} \times \mathcal{F}$ is determined by $\mathcal{D}_1, \ldots, \mathcal{D}_k$ and $\mathcal{D}_I$. Indeed, consider the *product* of the Dirac structures $\mathcal{D}_1, \ldots, \mathcal{D}_k$ on $(\mathcal{X}_1 \times \mathcal{F}_1) \times (\mathcal{X}_2 \times \mathcal{F}_2) \times \cdots \times (\mathcal{X}_k \times \mathcal{F}_k)$, and compose this with the Dirac structure $\mathcal{D}_I$ on $(\mathcal{F}_1 \times \cdots \times \mathcal{F}_k) \times \mathcal{F}$. This yields a total Dirac structure $\mathcal{D}$ modulated by $x = (x_1, \ldots, x_k) \in \mathcal{X} = \mathcal{X}_1 \times \cdots \mathcal{X}_k$ which is point-wise given as

$$\mathcal{D}(x_1, \ldots, x_k) \subset T_{x_1}\mathcal{X}_1 \times T^*_{x_1}\mathcal{X}_1 \times \cdots \times T_{x_k}\mathcal{X}_k \times T^*_{x_k}\mathcal{X}_k \times \mathcal{F} \times \mathcal{F}^*.$$

Finally we mention that the theory of composition of Dirac structures and the interconnection of port-Hamiltonian systems can be also extended to *infinite-dimensional* Dirac structures and port-Hamiltonian systems [17, 26].

7 Integrability of Modulated Dirac Structures

A key issue in the case of modulated Dirac structures is that of *integrability*. Loosely speaking, a Dirac structure is *integrable* if it is possible to find local coordinates for the state space manifold such that, in these coordinates, the Dirac structure becomes a *constant* Dirac structure, that is, it is *not* modulated anymore by the state variables. As we have seen before, in particular in the context of the canonical coordinate

representation (Sect. 3.1.4), this plays an important role in the representation of algebraic constraints (as well as in the existence of Casimirs).

First let us consider modulated Dirac structures which are given for every $x \in \mathcal{X}$ as the *graph* of a skew-symmetric mapping $J(x)$ from the co-tangent space $T_x^*\mathcal{X}$ to the tangent space $T_x\mathcal{X}$.

Integrability in this case means that the structure matrix J satisfies the conditions

$$\sum_{l=1}^{n}\left[J_{lj}(x)\frac{\partial J_{ik}}{\partial x_l}(x) + J_{li}(x)\frac{\partial J_{kj}}{\partial x_l}(x) + J_{lk}(x)\frac{\partial J_{ji}}{\partial x_l}(x)\right] = 0, \quad i, j, k = 1, \ldots, n. \tag{7.1}$$

In this case we may find, by Darboux's theorem (see e.g. [1]) around any point x_0 where the rank of the matrix $J(x)$ is constant, local coordinates $x = (q, p, r)$ in which the matrix $J(x)$ becomes the constant skew-symmetric matrix

$$\begin{bmatrix} 0 & -I_k & 0 \\ I_k & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{7.2}$$

Such coordinates are called *canonical*. A skew-symmetric matrix $J(x)$ satisfying (7.1) defines a *Poisson bracket* on $\mathcal{X}$, given for every $F, G : \mathcal{X} \to \mathbb{R}$ as

$$\{F, G\} = \frac{\partial^T F}{\partial x} J(x) \frac{\partial G}{\partial x}. \tag{7.3}$$

Indeed, by (7.1) the Poisson bracket satisfies the *Jacobi-identity*

$$\big\{F, \{G, K\}\big\} + \big\{G, \{K, F\}\big\} + \big\{K, \{F, G\}\big\} = 0 \tag{7.4}$$

for all functions F, G, K.

The choice of coordinates $x = (q, p, r)$ for the state space manifold also induces a basis for $T_x\mathcal{X}$ and a dual basis for $T_x^*\mathcal{X}$. Denoting the corresponding splitting for the flows by $f = (f_q, f_p, f_r)$ and for the efforts by $e = (e_q, e_p, e_r)$, the Dirac structure defined by J in canonical coordinates is seen to be given by

$$\mathcal{D} = \big\{(f_q, f_p, f_r, e_q, e_p, e_r) \mid f_q = -e_p,\ f_p = e_q,\ f_r = 0\big\}. \tag{7.5}$$

A similar story can be told for the case of a Dirac structure given as the graph of a skew-symmetric mapping $\omega(x)$ from the tangent space $T_x\mathcal{X}$ to the co-tangent space $T_x^*\mathcal{X}$. In this case the integrability conditions take the (slightly simpler) form

$$\frac{\partial \omega_{ij}}{\partial x_k}(x) + \frac{\partial \omega_{ki}}{\partial x_j}(x) + \frac{\partial \omega_{jk}}{\partial x_i}(x) = 0, \quad i, j, k = 1, \ldots, n. \tag{7.6}$$

The skew-symmetric matrix $\omega(x)$ can be regarded as the coordinate representation of a *differential two-form* ω on the manifold $\mathcal{X}$, that is, $\omega = \sum_{i=1,j=1}^{n} dx_i \wedge dx_j$, and the integrability condition (7.6) corresponds to the *closedness* of this two-form ($d\omega = 0$). The differential two-form ω is called a *pre-symplectic structure*, and a

symplectic structure if the rank of $\omega(x)$ is equal to the dimension of $\mathcal{X}$. If (7.6) holds, then again by a version of Darboux's theorem we may find, around any point x_0 where the rank of the matrix $\omega(x)$ is constant, local coordinates $x = (q, p, s)$ in which the matrix $\omega(x)$ becomes the constant skew-symmetric matrix

$$\begin{bmatrix} 0 & I_k & 0 \\ -I_k & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{7.7}$$

Such coordinates are again called *canonical*. The choice of coordinates $x = (q, p, s)$ as before induces a basis for $T_x\mathcal{X}$ and a dual basis for $T_x^*\mathcal{X}$. Denoting the corresponding splitting for the flows by $f = (f_q, f_p, f_s)$ and for the efforts by $e = (e_q, e_p, e_s)$, the Dirac structure corresponding to ω in canonical coordinates is seen to be given by

$$\mathcal{D} = \big\{(f_q, f_p, f_s, e_q, e_p, e_s) \mid f_q = -e_p,\ f_p = e_q,\ e_s = 0\big\}. \tag{7.8}$$

In case of a symplectic structure the variables s are absent and the Dirac structure reduces to

$$\mathcal{D} = \big\{(f_q, f_p, e_q, e_p) \mid f_q = -e_p,\ f_p = e_q\big\} \tag{7.9}$$

which is the standard *symplectic gyrator*.

For general Dirac structures, integrability is defined in the following way.

Definition 7.1 ([12]) A Dirac structure $\mathcal{D}$ on $\mathcal{X}$ is *integrable* if for arbitrary pairs of smooth vector fields and differential one-forms $(X_1, \alpha_1), (X_2, \alpha_2), (X_3, \alpha_3) \in \mathcal{D}$ we have

$$\langle L_{X_1}\alpha_2 \mid X_3\rangle + \langle L_{X_2}\alpha_3 \mid X_1\rangle + \langle L_{X_3}\alpha_1 \mid X_2\rangle = 0 \tag{7.10}$$

with L_{X_i} denoting the Lie-derivative.

Remark 7.1 (Pseudo-Dirac structures) In the usual definition of Dirac structures on manifolds (see [8, 12]), this *integrability* condition is *included* in the definition. Dirac structures that do *not* satisfy this integrability condition are therefore sometimes called *pseudo*-Dirac structures.

The above integrability condition for Dirac structures generalizes properly the closedness of symplectic forms and the Jacobi identity for Poisson brackets as discussed before. In particular, for Dirac structures given as the graph of a symplectic or Poisson structure, the notion of integrability is equivalent to the Jacobi-identity or closedness condition as discussed above (see e.g. [8, 9, 12] for details).

Note that a *constant* Dirac structure trivially satisfies the integrability condition. Conversely, a Dirac structure satisfying the integrability condition together with an additional constant rank condition can be represented *locally* as a *constant* Dirac

structure. The precise form of the constant rank condition can be stated as follows. Recall that for any Dirac structure $\mathcal{D}$, we may define the distribution

$$G_{\mathcal{D}}(x) = \left\{ X \in T_x\mathcal{X} \mid \exists \alpha \in T_x^*\mathcal{X} \text{ s.t. } (X, \alpha) \in D(x) \right\}$$

and the co-distribution

$$P_{\mathcal{D}}(x) = \left\{ \alpha \in T_x^*\mathcal{X} \mid \exists X \in T_x\mathcal{X} \text{ s.t. } (X, \alpha) \in D(x) \right\}.$$

We call x_0 a *regular* point for the Dirac structure if both the distribution $G_{\mathcal{D}}$ and the co-distribution $P_{\mathcal{D}}$ have constant dimension around x_0.

If the Dirac structure is integrable and x_0 is a regular point, then, again by a version of Darboux's theorem, we can choose local coordinates $x = (q, p, r, s)$ for $\mathcal{X}$ (with $\dim q = \dim p$), such that, in the resulting bases for (f_q, f_p, f_r, f_s) for $T_x\mathcal{X}$ and (e_q, e_p, e_r, e_s) for $T_x^*\mathcal{X}$, the Dirac structure on this coordinate neighborhood is given as (see (3.7))

$$\begin{cases} f_q = -e_p, \\ f_p = e_q, \\ f_r = 0, \\ e_s = 0. \end{cases} \tag{7.11}$$

Coordinates $x = (q, p, r, s)$ as above are again called *canonical*. Note that the choice of canonical coordinates for a Dirac structure satisfying the integrability condition encompasses the choice of canonical coordinates for a Poisson structure and for a (pre-)symplectic structure as above.

Explicit conditions for integrability of a Dirac structure can be readily stated in terms of a kernel/image representation. Indeed, let

$$\begin{aligned} \mathcal{D} &= \left\{ (f, e) \mid F(x) f + E(x) e = 0 \right\} \\ &= \left\{ (f, e) \mid f = E^T(x)\lambda,\ e = F^T(x)\lambda,\ \lambda \in \mathbb{R}^n \right\}. \end{aligned}$$

Denote the transpose of ith row of $E(x)$ by $Y_i(x)$ and the transpose of the ith row of $F(x)$ by $\beta_i(x)$. The vectors $Y_i(x)$ are naturally seen as coordinate representations of *vector fields* while the vectors $\beta_i(x)$ are coordinate representations of *differential forms*. Then integrability of the Dirac structure is equivalent to the condition

$$\langle L_{Y_i}\beta_j \mid Y_k \rangle + \langle L_{Y_j}\beta_k \mid Y_i \rangle + \langle L_{Y_k}\beta_i \mid Y_j \rangle = 0 \tag{7.12}$$

for all indices $i, j, k = 1, \ldots, n$.

Another form of the integrability conditions can be obtained as follows. In [8, 9, 12] it has been shown that a Dirac structure on a manifold $\mathcal{X}$ is integrable if and only if, for all pairs of smooth vector fields and differential one-forms $(X_1, \alpha_1), (X_2, \alpha_2) \in \mathcal{D}$, we have

$$\left([X_1, X_2], i_{X_1} d\alpha_2 - i_{X_2} d\alpha_1 + d\langle \alpha_2 \mid X_1 \rangle \right) \in \mathcal{D}. \tag{7.13}$$

Using the definition of the vector fields Y_i and differential forms β_i, $i = 1, \ldots, n$, as above, it follows that the Dirac structure is integrable if and only if

$$\big([Y_i, Y_j], i_{Y_i} d\beta_j - i_{Y_j} d\beta_i + d\langle \beta_j \mid Y_i \rangle\big) \in \mathcal{D} \tag{7.14}$$

for all $i, j = 1, \ldots, n$. This can be more explicitly stated by requiring that

$$F(x)[Y_i, Y_j](x) + E(x)\big(i_{Y_i} d\beta_j(x) - i_{Y_j} d\beta_i(x) + d\langle \beta_j \mid Y_i \rangle(x)\big) = 0 \tag{7.15}$$

for all $i, j = 1, \ldots, n$ and for all $x \in \mathcal{X}$. See for more details [9].

Example 7.1 (Kinematic constraints) Recall from the discussion in Sect. 2.7.1 that the modulated Dirac structure corresponding to an actuated mechanical system subject to kinematic constraints $A^T(q)\dot{q} = 0$ is given by

$$\mathcal{D} = \Bigg\{ (f_S, e_S, f_C, e_C) \mid 0 = \begin{bmatrix} 0 & A^T(q) \end{bmatrix} e_S,\ e_C = \begin{bmatrix} 0 & B^T(q) \end{bmatrix} e_S,$$
$$-f_S = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix} e_S + \begin{bmatrix} 0 \\ A(q) \end{bmatrix} \lambda + \begin{bmatrix} 0 \\ B(q) \end{bmatrix} f_c,\ \lambda \in \mathbb{R}^k \Bigg\}.$$

Complete necessary and sufficient conditions for integrability of this Dirac structure have been derived in [9]. Here we only state a slightly simplified version of this result, also detailed in [9]. We assume that the actuation matrix $B(q)$ has the special form (often encountered in examples) where every jth column ($j = 1, \ldots, m$) is given as

$$\begin{bmatrix} 0 \\ \frac{\partial C_j}{\partial q}(q) \end{bmatrix}$$

for some function $C_j(q)$ only depending on the configuration variables q. In this case, the Dirac structure $\mathcal{D}$ is integrable *if and only if the kinematic constraints are holonomic*.

It has been shown in Sect. 2.7.1 that, after elimination of the Lagrange multipliers and the algebraic constraints, the constrained mechanical system reduces to a port-Hamiltonian system on the constrained submanifold defined with respect to a Poisson structure matrix J_c. As has been shown in [37], J_c satisfies the integrability condition (7.1) again if and only if the constraints (2.33) are *holonomic*. In fact, if the constraints are holonomic, then the coordinates s as in (3.7), (4.28) can be taken to be equal to the 'integrated constraint functions' $\overline{q}_{n-k+1}, \ldots, \overline{q}_n$ of (2.35).

It can be verified that the structure matrix J_c obtained in 2.4, see (4.21), does not satisfy the integrability conditions, in accordance with the fact that the rolling constraints in this example are *nonholonomic*.

8 Conclusions

In this paper we have surveyed how the port-Hamiltonian formalism offers a systematic framework for modeling and control of large-scale multi-physics systems, emphasizing at the same time the network structure of the system (captured by its Dirac structure) and the energy-storage and energy-dissipation (formalized with the help of Hamiltonian functions and resistive relations). In many cases of interest this will lead to the description of the system dynamics by a mixed set of differential and algebraic equations (DAEs); however, endowed with a (generalized) Hamiltonian structure. We have shown how the identification of the underlying Hamiltonian structure offers additional insights and tools for analysis and control, as compared to general differential-algebraic systems.

In this paper we have confined ourselves to *lumped-parameter*, i.e., finite-dimensional, models. The port-Hamiltonian framework, however, has been successfully extended to distributed-parameter models (see e.g. [39]), corresponding to infinite-dimensional Dirac structures. Therefore an important venue for further research concerns the analysis within the port-Hamiltonian framework of mixed systems of differential, algebraic, as well as of *partial differential* equations.

Acknowledgements This survey article is based on joint work with many colleagues, whom I thank for a very stimulating collaboration. In particular I thank Bernhard Maschke for continuing joint efforts over the years.

References

1. Abraham, R., Marsden, J.E.: Foundations of Mechanics, 2nd edn. Benjamin/Cummings, Reading (1978)
2. Battle, C., Massana, I., Simo, E.: Representation of a general composition of Dirac structures. In: Proc. 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), Orlando, FL, USA, December 12–15, pp. 5199–5204 (2011)
3. Belevitch, V.: Classical Network Theory. Holden-Day, San Francisco (1968)
4. Bloch, A.M., Crouch, P.E.: Representations of Dirac structures on vector spaces and nonlinear LC circuits. In: Proceedings Symposia in Pure Mathematics, Differential Geometry and Control Theory, pp. 103–117. American Mathematical Society, Providence (1999)
5. Brogliato, B.: Nonsmooth Mechanics. Communications and Control Engineering, 2nd edn. Springer, London (1999)
6. Camlibel, M.K., Heemels, W.P.M.H., van der Schaft, A.J., Schumacher, J.M.: Switched networks and complementarity. IEEE Trans. Circuits Syst. I, Fundam. Theory Appl. **50**, 1036–1046 (2003)
7. Cervera, J., van der Schaft, A.J., Banos, A.: Interconnection of port-Hamiltonian systems and composition of Dirac structures. Automatica **43**, 212–225 (2007)
8. Courant, T.J.: Dirac manifolds. Trans. Am. Math. Soc. **319**, 631–661 (1990)
9. Dalsmo, M., van der Schaft, A.J.: On representations and integrability of mathematical structures in energy-conserving physical systems. SIAM J. Control Optim. **37**(1), 54–91 (1999)
10. Desoer, C.A., Sastry, S.S.: Jump behavior of circuits and systems. IEEE Trans. Circuits Syst. **28**, 1109–1124 (1981)
11. Dirac, P.A.M.: Generalized Hamiltonian dynamics. Can. J. Math. **2**, 129–148 (1950)

12. Dorfman, I.: Dirac Structures and Integrability of Nonlinear Evolution Equations. Wiley, Chichester (1993)
13. Escobar, G., van der Schaft, A.J., Ortega, R.: A Hamiltonian viewpoint in the modelling of switching power converters. Automatica **35**, 445–452 (1999). Special Issue on Hybrid Systems
14. Gerritsen, K., van der Schaft, A.J., Heemels, W.P.M.H.: On switched Hamiltonian systems. In: Gilliam, D.S., Rosenthal, J. (eds.) Proceedings 15th International Symposium on Mathematical Theory of Networks and Systems (MTNS2002), South Bend, August 12–16 (2002)
15. Golo, G., van der Schaft, A.J., Breedveld, P.C., Maschke, B.M.: Hamiltonian formulation of bond graphs. In: Nonlinear and Hybrid Systems in Automotive Control, pp. 351–372 (2003)
16. Koopman, J., Jeltsema, D., Verhaegen, M.: Port-Hamiltonian formulation and analysis of the LuGre friction model. In: 47th IEEE Conference on Decision and Control, Cancun, Mexico, pp. 3181–3186. IEEE, Control Systems Society, New York (2008)
17. Kurula, M., Zwart, H., van der Schaft, A.J., Behrndt, J.: Dirac structures and their composition on Hilbert spaces. J. Math. Anal. Appl. **372**, 402–422 (2010)
18. Liberzon, D., Trenn, S.: Switched nonlinear differential algebraic equations: solution theory, Lyapunov functions, and stability. Automatica **48**(5), 954–963 (2012)
19. Maschke, B.M., van der Schaft, A.J.: Port-controlled Hamiltonian systems: modelling origins and system theoretic properties. In: 2nd IFAC Symposium on Nonlinear Control Systems Design (NOLCOS), Bordeaux, June, pp. 359–365 (1992)
20. Maschke, B.M., van der Schaft, A.J., Breedveld, P.C.: An intrinsic Hamiltonian formulation of network dynamics: non-standard Poisson structures and gyrators. J. Franklin Inst. **329**(5), 923–966 (1992)
21. Narayanan, H.: Some applications of an implicit duality theorem to connections of structures of special types including Dirac and reciprocal structures. Syst. Control Lett. **45**, 87–96 (2002)
22. Neimark, J.I., Fufaev, N.A.: Dynamics of Nonholonomic Systems. Translations of Mathematical Monographs, vol. 33. American Mathematical Society, Providence (1972)
23. Nijmeijer, H., van der Schaft, A.J.: Nonlinear Dynamical Control Systems. Springer, New York (1990)
24. Ortega, R., van der Schaft, A.J., Mareels, Y., Maschke, B.M.: Putting energy back in control. Control Syst. Mag. **21**, 18–33 (2001)
25. Ortega, R., van der Schaft, A.J., Maschke, B.M., Escobar, G.: Interconnection and damping assignment passivity-based control of port-controlled Hamiltonian systems. Automatica **38**(4), 585–596 (2002)
26. Pasumarthy, R., van der Schaft, A.J.: Achievable Casimirs and its implications on control of port-Hamiltonian systems. Int. J. Control **80**, 1421–1438 (2007)
27. Paynter, H.M.: Analysis and Design of Engineering Systems. MIT Press, Cambridge (1961)
28. van der Schaft, A.J.: Equations of motion for Hamiltonian systems with constraints. J. Phys. A, Math. Gen. **20**, 3271–3277 (1987)
29. van der Schaft, A.J.: Implicit Hamiltonian systems with symmetry. Rep. Math. Phys. **41**, 203–221 (1998)
30. van der Schaft, A.J.: Interconnection and geometry. In: Polderman, J.W., Trentelman, H.L. (eds.) The Mathematics of Systems and Control: From Intelligent Control to Behavioral Systems, Groningen, The Netherlands, pp. 203–218 (1999)
31. van der Schaft, A.J.: L_2-Gain and Passivity Techniques in Nonlinear Control. Communications and Control Engineering. Springer, Berlin (2000)
32. van der Schaft, A.J.: Port-Hamiltonian systems: network modeling and control of nonlinear physical systems. In: Advanced Dynamics and Control of Structures and Machines. CISM Courses and Lectures, vol. 444. Springer, New York (2004)
33. van der Schaft, A.J.: Port-Hamiltonian systems: an introductory survey. In: Sanz-Sole, M., Soria, J., Verona, J.L., Verdura, J. (eds.) Proc. of the International Congress of Mathematicians, vol. III, Invited Lectures, Madrid, Spain, pp. 1339–1365 (2006)
34. van der Schaft, A.J.: Port-Hamiltonian systems. In: Duindam, V., Macchelli, A., Stramigioli, S., Bruyninckx, H. (eds.) Modeling and Control of Complex Physical Systems; the

Port-Hamiltonian Approach, pp. 53–130. Springer, Berlin (2009). ISBN 978-3-642-03195-3. Chap. 2
35. van der Schaft, A.J., Camlibel, M.K.: A state transfer principle for switching port-Hamiltonian systems. In: Proc. 48th IEEE Conf. on Decision and Control, Shanghai, China, December 16–18, pp. 45–50 (2009)
36. van der Schaft, A.J., Cervera, J.: Composition of Dirac structures and control of port-Hamiltonian systems. In: Gilliam, D.S., Rosenthal, J. (eds.) Proceedings 15th International Symposium on Mathematical Theory of Networks and Systems, August (2002)
37. van der Schaft, A.J., Maschke, B.M.: On the Hamiltonian formulation of nonholonomic mechanical systems. Rep. Math. Phys. **34**, 225–233 (1994)
38. van der Schaft, A.J., Maschke, B.M.: The Hamiltonian formulation of energy conserving physical systems with external ports. Arch. Elektron. Übertragungstech. **45**, 362–371 (1995)
39. van der Schaft, A.J., Maschke, B.M.: Hamiltonian formulation of distributed-parameter systems with boundary energy flow. J. Geom. Phys. **42**, 166–194 (2002)
40. van der Schaft, A.J., Maschke, B.: Port-Hamiltonian systems on graphs. SIAM J. Control Optim. (2013, to appear)
41. van der Schaft, A.J., Schumacher, J.M.: The complementary-slackness class of hybrid systems. Math. Control Signals Syst. **9**, 266–301 (1996)
42. van der Schaft, A.J., Schumacher, J.M.: Complementarity modelling of hybrid systems. IEEE Trans. Autom. Control **AC-43**, 483–490 (1998)

Index

A. Ilchmann, T. Reis (eds.), *Surveys in Differential-Algebraic Equations I*,
Differential-Algebraic Equations Forum,
DOI 10.1007/978-3-642-34928-7, © Springer-Verlag Berlin Heidelberg 2013