Well-Posedness and Realization Theory for Delay Differential-Algebraic Equations

vorgelegt von
M. Sc.
Benjamin Unger
ORCID: 0000-0003-4272-1079

an der Fakultät II - Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Martin Henk
Gutachter:	Prof. Dr. Vu Hoang Linh
Gutachter:	Prof. Dr. Volker Mehrmann
Gutachter:	Prof. Dr. Wim Michiels

Tag der wissenschaftlichen Aussprache: 22.10.2020

Berlin 2020

Abstract

This thesis is dedicated to *delay differential-algebraic equations* (DDAEs), i.e., constraint dynamical systems where the rate of change depends on the current state and its past. Typical applications include

(i) feedback control, where the delay is a direct consequence of the time required to measure the current state, compute the feedback, and implement the control action,

(ii) hybrid numerical-experimental testing environments, used for instance in earthquake engineering,

(iii) transmission and propagation delays, encountered for example in chemical reactions connected in series and wide-area power-networks, and

(iv) as a mathematical tool to analyze hyperbolic equations and time-integration schemes.

The fact that algebraic equations can be included in the implicit system description fosters a rapid model development since complex models can be assembled from a library of existing models with well-defined interaction variables.

From a mathematical point of view, DDAEs do not only feature difficulties already known from the theory of *differential-algebraic equations* (DAEs) and *delay differential equations* (DDEs) but pose additional challenges. For instance, initial trajectory problems for DDAEs may not be causal. Thus, even in a distributional solution space, they may not have a solution for all initial trajectories. This fact, combined with the infinite-dimensional character of delay equations and the high sensitivity to perturbation known from the theory of DAEs, renders DDAEs a challenging mathematical object. Consequently, the analysis of DDAEs is far from complete. The aim of this thesis is to address some of the many open problems.

In the first part of the thesis, initial trajectory problems for *linear time-invariant* (LTI) and nonlinear DDAEs are discussed. We start our analysis with a distributional solution concept and establish the existence and uniqueness of solutions, whenever the DDAE is delay-regular. Jumps and Dirac-impulses in the solution can be avoided if the coefficient matrices of the LTI DDAE satisfy some algebraic conditions, which are obtained by tracking so-called primary discontinuities. We extend some of the results to the nonlinear setting resulting in existence and uniqueness results for a large class of nonlinear DDAEs.

The second part of the thesis is dedicated to constructing a DDAE solely from a prescribed set of data points. Having a time-delay in the realization allows us to build an infinite-dimensional system from finitely many points capable of reproducing the transcendental character of the transfer function of a distributed parameter subsystem that models convection or transport. We construct a realization so that it interpolates the data set in the frequency domain and demonstrate its applicability with several numerical examples.

Zusammenfassung

Diese Arbeit befasst sich mit *zeitverzögerten Differential-Algebraischen Gleichungen* (DDAEs), das heißt mit Differentialgleichungen mit Zwangsbedingungen, bei denen die Änderungsrate sowohl vom aktuellen Zustand als auch von der Vergangenheit abhängt. Typische Anwendungen umfassen

(i) Rückkopplungssteuerung, wobei sich die Verzögerung aus der benötigten Zeit zur Messung des aktuellen Zustands, Berechnung der Rückkopplung und Implementierung selbiger ergibt,

(ii) hybride numerisch-experimentelle Testverfahren, wie sie beispielsweise in der Erdbebenforschung eingesetzt werden,

(iii) Übertragungs- und Ausbreitungsverzögerungen, die beispielsweise bei in Reihe geschalteten chemischer Reaktionen sowie bei großflächigen Stromnetzen auftreten und

(iv) als mathematisches Werkzeug zur Analyse hyperbolischer Gleichungen und Zeitintegrationsverfahren.

Die Tatsache, dass algebraische Gleichungen in das implizite System aufgenommen werden können, begünstigt eine schnelle Modellentwicklung, da komplexe Modelle nach dem Baukastenprinzip aus einer Bibliothek von Modellen mit genau definierten Interaktionsvariablen zusammengesetzt werden können

Aus mathematischer Sicht weisen DDAEs nicht nur Schwierigkeiten auf, die bereits aus der Theorie der *Differential-Algebraischen Gleichungen* (DAEs) und *zeitverzögerten Differentialgleichungen* (DDEs) bekannt sind, sondern stellen zusätzliche Herausforderungen bereit. Beispielsweise können Anfangstrajektorienprobleme für DDAEs akausales Verhalten aufweisen. Dies führt dazu, dass selbst bei linearen DDAEs mit einem distributionellen Lösungskonzept nicht notwendigerweise für alle Anfangstrajektorien eine Lösung existiert. Dieses Phänomen in Kombination mit dem unendlich-dimensionalen Charakter von zeitverzögerten Differentialgleichungen sowie der aus der DAE-Theorie bekannten hohen Sensitivität gegenüber Störungen machen DDAEs zu einem herausfordernden mathematischen Objekt. Folglich gibt es zahlreiche nicht gelöste Forschungsfragen im

Zusammenhang mit DDAEs. Das Ziel dieser Arbeit ist es, einige dieser Fragen zu beantworten.

Im ersten Teil der Arbeit werden Anfangstrajektorienprobleme für lineare zeitinvariante (LTI) und nichtlineare DDAEs diskutiert. Wir beginnen unsere Analyse mit einem distributionellen Lösungskonzept und beweisen Existenz- und Eindeutigkeitsresultat für sogenannte *delay-reguläre* DDAEs. Um Sprünge, Dirac-Impulse und Ableitungen von Dirac-Impulsen in der Lösung ausschließen zu können, müssen die Koeffizientenmatrizen der LTI DDAE bestimmte algebraische Bedingungen erfüllen. Diese Bedingungen können durch eine Nachverfolgung sogenannter primärer Unstetigkeitsstellen hergeleitet werden. Teilweise können die erhaltenen Ergebnisse auf nichtlineare DDAEs verallgemeinert werden, was zu neuen Existenz- und Eindeutigkeitsresultaten führt.

Der zweite Teil der Arbeit befasst sich mit der Konstruktion einer DDAE aus einem vorgegebenen Datensatz, einer sogenannten Realisierung. Dabei liefert das Zeitverzögerungsglied in der Realisierung den Vorteil, dass aus endlich vielen Datenpunkten ein unendlich-dimensionales System konstruiert werden kann. Dieses System ist dann in der Lage, den transzendenten Charakter, der beispielsweise bei Transportgleichungen vorkommt, einer Übertragungsfunktion abzubilden. Unsere Konstruktion basiert darauf, dass die gegebenen Daten im Frequenzbereich interpoliert werden. Die Effektivität des Verfahrens wird anhand zahlreicher numerischer Beispiele demonstriert.

Acknowledgments

I would like to thank Prof. Volker Mehrmann for his guidance, supervision, and the provided freedom in the choice of my research directions. Further, I am thankful to Prof. Wim Michiels and Prof. Vu Hoang Linh for serving on my thesis committee. I also want to thank my co-authors Ines Ahres, Robert Altmann, Christopher Beattie, Felix Black, Elliot Fosong, Serkan Gugercin, Roland Maier, Philipp Schulze, and Stephan Trenn for many successful cooperations. Besides, I thank my colleagues Christian, Christoph, Daniel, Fredi, Marine, Matthias, Mathieu, Lena, Paul, Phi, Philipp, Riccardo, Ute, and Viola for valuable hints, lively discussions, and friendship. I acknowledge funding from the Deutsche Forschungsgemeinschaft through the Collaborative Research Center 910 *Control of self-organizing nonlinear systems: Theoretical methods and concepts of application* (project number 163436311) and the Berlin Mathematical School (BMS). Finally, I want to thank my wife and my parents for their constant support and unconditional love.

Contents

List of Abbreviations

BIBO	bounded-input/bounded-output
CSTR	continuous stirred-tank reactor
DAE	differential-algebraic equation
DDAE	delay differential-algebraic equation
DDE	delay differential equation
ETFE	empirical transfer function estimate
FFT	fast Fourier transform
ITP	initial trajectory problem
IVP	initial value problem
lsTFE	least-squares transfer function estimate
LTI	linear time-invariant
MBS	multibody system
MIMO	multiple-input/multiple-output
MOR	model order reduction
NDDE	neutral delay differential equation
ODE	ordinary differential equation
PDE	partial differential equation
RLC	resistor-inductor-capacitor
ROM	reduced order model
SISO	single-input/single-output
SVD	singular value decomposition

List of Figures

List of Tables

1 Introduction

Complex physical or chemical systems often comprise several subsystems that interact with each other and the environment. For instance, an electrical grid is a network of power generators, transmission lines, and consumers — each of which may again be a composition of subsystems. Let us emphasize that also the interaction with the environment, for instance, via external forces or dissipation of energy, might be represented as another (sub-)system that is interconnected with the physical system; see Figure 1.1 for a schematic representation where the interaction of the subsystems is represented with arrows. In a simulation-driven environment, it is standard to model the physical system under investigation in terms of (partial) differential equations that describe the evolution of the system. Instead of deriving the equations of motions for the complete system at once, a bottom-up approach models each subsystem separately and then connect the models for the subsystems via suitable interconnections. An easy way to model such an interconnection is given by an algebraic equation, thus making the complete model a (partial) *differential-algebraic equation* (DAE). Although, in principle, it may be possible to resolve the algebraic equations and hence rewrite the resulting system as a (partial) differential equation, it is a priori not clear, whether this would be reasonable from a computational perspective or a modeling perspective. An example

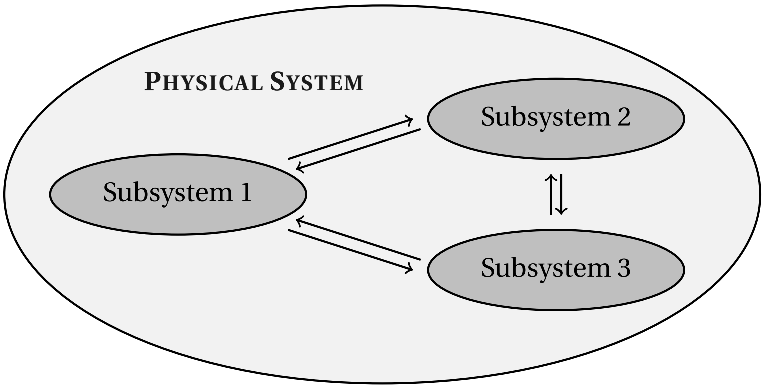


Figure 1.1 – Composition of a physical system by several subsystems

of the latter aspect is a chain of $n \geq 2$ mathematical pendulums; see, for instance, [12, Example 2.2]. As a consequence, we keep all algebraic constraints and work directly with the DAE.

If an extensive library of models with well-defined interaction ports of small components is available, then the modeling process of a network of some of these components can be automated, facilitating a quick modeling process. Examples of tools that use such an idea are SIMULINK and MODELICA. In some applications, for instance, in earthquake engineering, the simulation and model capacities are limited, such that there is still a need for actual experiments [218]. To remedy the high cost that comes with such experiments, it is common in the dynamical testing community to employ a hybrid numerical-experimental setup, see [45] and the references therein. In more detail, only a small physical subsystem that features the key region of interest is experimentally tested, while the remainder of the system is simulated numerically. The interconnection of the physical and numerical domain requires a transfer system, which is typically a set of actuators [44]. The transfer of information from the numerical system to the actuators is intrinsically non-instantaneous [129], which introduces a time delay in the overall system. A detailed example is presented in section 1.1.1. The complete system dynamics may thus be modeled as an implicit system of equations of the form

$$0 = F(t, x(t), \dot{x}(t), x(t-\tau)), \tag{1.1}$$

where $x(t)$ denotes the unknown state and $\tau > 0$ the time delay. We emphasize that in general, the partial derivative of F with respect to $\dot{x}$ is allowed to be singular, and hence the implicit function theorem may not be used directly to reformulate (1.1) as a *delay differential equation* (DDE). One possible source for the singularity of the partial derivative of F with respect to $\dot{x}$ is due to the interconnections, which are usually described via algebraic equations. Hence, we refer to (1.1) as a *delay differential-algebraic equation* (DDAE). Clearly, it is important to understand the effect of the delay that is introduced due to the hybrid numerical-experimental setting and thus a thorough analysis of the DDAE (1.1), which serves as the object under investigation in this thesis, is of paramount importance. We stress that further delays may arise in the modeling process of the system depicted in Figure 1.1:

(i) If the physical system itself is a controlled plant, then one may think of one of the subsystems as a controller. The controller interacts with the plant by measuring some quantity of interest, compute (in some sense) a control action, and implement this action for the system. If any of these steps requires some time, the controller induces an intrinsically necessary time delay. For instance, in a chemical process, one may take a sample, analyze it, and, based on the result, decide to modify the process. Another example from mechanical engineering is presented in section 1.1.2. We note that sometimes it may even be advantageous to implement a small time-delay to improve the control action [105, 149] or to uncover unstable periodic orbits in nonlinear dynamical systems [176].

(ii) If the subsystems are physically separated, then the interaction between subsystems in the form of exchange of energy, information, or data may require a non-neglectable amount of time, which introduces another source for a time-delay in the system. Such a communication or propagation delay appears for instance in modern electric power grids [2], satellite communication [195], or in a chemical process [64, 161]. The latter is illustrated in more detail in section 1.1.3.

(iii) One of the subsystems itself might be modeled with a delay equation. For instance, it is well-known that the linear advection equation can be rewritten as a delay equation. The reformulation of a hyperbolic equation as a delay equation not only offers a different approach to existence and uniqueness results [38], and the development of different mathematical models [37], but also is cheaper to solve numerically. This fact also serves as a motivation to realize a transport process with a DDAE [77,189,191]. The process of rewriting a hyperbolic equation as a delay equation is detailed in section 1.1.4.

Besides the relevance of DDAEs in modern modeling frameworks, we emphasize that DDAEs are also a powerful mathematical tool that can be used in the analysis and design of numerical algorithms, cf. section 1.1.5, or in assessing and classification of data, which is, for instance, used to relate finger tapping with the severity of Parkinson's disease [131,132].

Besides the already mentioned examples, futher applications include slow-fast systems, such as electro-optic oscillators [169,217] and optical networks [84], electric circuits [183], applications in biology [37,65], chemical kinetics [74], human balance control [160,198], and machine tool vibrations [119,163,164]. For additional phenomena that feature time delay we refer to the monographs [75,122] and the references therein.

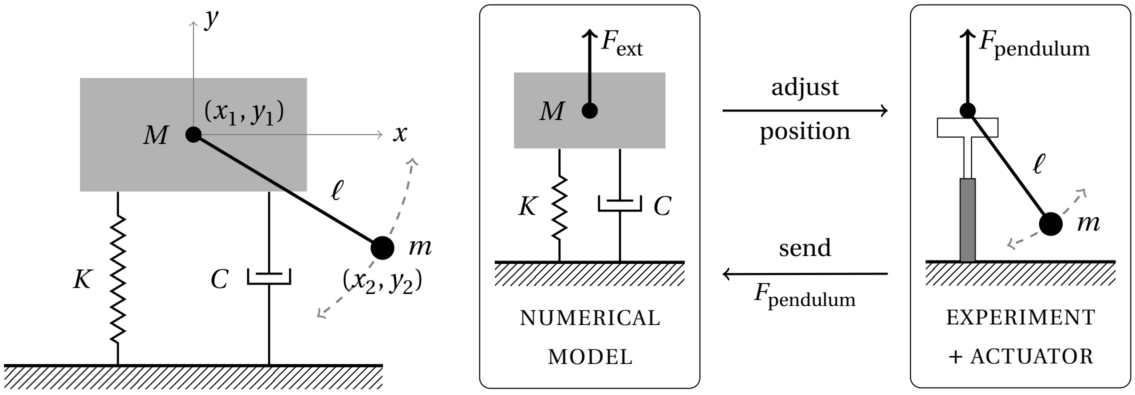
1.1 Motivating examples

We present some of the examples mentioned above in more detail in this section to stress the relevance of DDAEs in applications. The notation is simplified whenever it is deemed reasonable by omitting the explicit dependency on the time variable.

1.1.1 Real-time dynamic substructuring

In some applications, the description of a physical system with a mathematical model is difficult due to its complex nature or uncertainty [218]. Since testing of a complete prototype may be prohibitively expensive, it is desirable to incorporate the benefits of actual testing with the benefits of numerical simulation. This is accomplished by testing only a substructure (or subsystem in the sense of Figure 1.1) and connect the experiment via a transfer system with the remaining system, which is simulated numerically. Such a hybrid experimental-numerical approach is called *real-time dynamic substructuring* or *hardware-in-the-loop testing* [45]. The transfer system is typically realized with a set of hydraulic actuators. Since the dynamic behavior of any actuator includes a response delay [107,216], the resulting system is a DDAE. Let us emphasize that further delays might be present, which arise, for instance, from data acquisition, computation, or digital signal processing. In many applications, these delays are small compared to the actuator delay and may thus be neglected in the modeling process; for more details, we refer to [129] and the references therein.

We illustrate such a hybrid experimental-numerical setup with a coupled pendulum-mass-spring-damper system, as described in [129,212]. For our example, we consider the mass-spring-damper



(a) Fully coupled system

(b) Hybrid numerical-experimental setup

Figure 1.2 – Real-time dynamic substructuring for a coupled pendulum-mass-spring-damper system

system as the numerical simulation and the pendulum as the experiment; see Figure 1.2 for an illustration. For our numerical model, we assume that the mass M is mounted on a linear spring and a linear viscous damper. The resulting equation of motion for the mass-spring-damper system is given by

$$M\ddot{y}_1 + C\dot{y}_1 + Ky_1 = F_{\text{ext}}, \tag{1.2}$$

where C and K denote the damping and the stiffness coefficient, respectively, and y_1 denotes the vertical displacement of the center of mass with respect to the equilibrium position. We assume that there is no horizontal displacement, i.e., the mass M can only move upwards and downwards. We thus set $x_1 := 0$. The external force, which in this scenario will be provided by the pendulum, is given by F_{ext}. We assume that the pendulum is given by a point mass m that is attached to the spring-mass-damper system via a massless rod of length ℓ. Assuming no friction, the model for the pendulum is given by

$$\begin{aligned} m\ddot{x}_2 &= -2\lambda x_2, \\ m\ddot{y}_2 &= -2\lambda(y_2 - y_1) - m\mathrm{g}, \\ 0 &= x_2^2 + (y_2 - y_1)^2 - \ell^2, \end{aligned} \tag{1.3}$$

with gravitational constant g and Langrange multiplier λ. By Newton's second law, the force that the pendulum generates in y-direction is given by $F_{\text{pendulum}} = -2\lambda(y_2 - y_1) - m\mathrm{g}$. Consequently, the equations of motion for the fully coupled system (as depicted in Figure 1.2a) are given by

$$\begin{aligned} M\ddot{y}_1 + C\dot{y}_1 + Ky_1 &= -2\lambda(y_2 - y_1) - m\mathrm{g}, \\ m\ddot{x}_2 &= -2\lambda x_2, \\ m\ddot{y}_2 &= -2\lambda(y_2 - y_1) - m\mathrm{g}, \\ 0 &= x_2^2 + (y_2 - y_1)^2 - \ell^2, \end{aligned} \tag{1.4}$$

with unknown functions y_1, x_2, y_2, and λ. In the hybrid numerical-experimental setup (cf. Figure 1.2b), the actuator introduces a time-delay $\tau > 0$ into the system, which is assumed constant. The delay can be understood as an offset in time between the mass-spring-damper dynamics (1.2) and the pendulum dynamics (1.3). In particular, we have to replace t by $t - \tau$ in the pendulum dynamics (1.3) and in the force F_{pendulum}. Thus, the complete mathematical description for the hybrid numerical-experimental setup is given by

$$\begin{aligned}
M\ddot{y}_1 + C\dot{y}_1 + Ky_1 &= -2\lambda(\cdot-\tau)(y_2(\cdot-\tau) - y_1(\cdot-\tau)) - mg, && \text{(1.5a)}\\
m\ddot{x}_2(\cdot-\tau) &= -2\lambda(\cdot-\tau)x_2(\cdot-\tau), && \text{(1.5b)}\\
m\ddot{y}_2(\cdot-\tau) &= -2\lambda(\cdot-\tau)(y_2(\cdot-\tau) - y_1(\cdot-\tau)) - mg, && \text{(1.5c)}\\
0 &= x_2(\cdot-\tau)^2 + (y_2(\cdot-\tau) - y_1(\cdot-\tau))^2 - \ell^2. && \text{(1.5d)}
\end{aligned}$$

If we introduce new variables for $\dot{y}_1$, $\dot{x}_2$, and $\dot{y}_2$, then we can rewrite (1.5) in the form (1.1). Let us emphasize that in order to solve (1.5), for instance with a numerical method, we have to shift equations (1.5b) to (1.5d) in time.

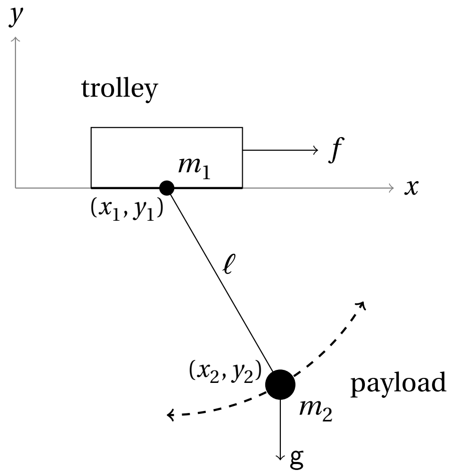
1.1.2 Time-delayed feedback control

In several (engineering) applications, it is a central goal to enforce a specific behavior within the system, for instance, to stabilize an unstable equilibrium, such as the upward position in an inverted pendulum. The goal is typically achieved by implementing a controller into the system. Given Figure 1.1, the controller can be interpreted as a subsystem. If the dynamics of the physical system are complex and thus the resulting model is subject to uncertainties or further external forces may act on the system, it is common to design the controller based on the current state of the system. There are two different sources for time-delays in such a *feedback* loop: (a) the controller requires some time to measure the quantity of interest, compute the feedback law, and implement the control action, or (b) a time-delay is implemented on purpose to facilitate some desired behavior. A popular control strategy that falls into the second category is called *Pyragas control* [175]. Prominent examples include stabilization of unstable periodic orbits in chaotic electrical networks [176], control of a Taylor-Couette Flow [140], microcantilevers [223], or semiconductor lasers [185], and sway reduction for container cranes [105, 149]. We present the latter application in more detail in the following.

A simplified model of a two-dimensional container crane (also called gantry crane) is given by a mathematical pendulum that is attached to a moving cart, the so-called trolley; see Figure 1.3. Hereby, both the payload and the cart are modeled as point masses. We may control the crane by applying a horizontal force to the cart and by changing the length of the rope. Assuming a frictionless movement of the cart, the simplified model in Figure 1.3 can be considered as a controlled *multibody system* (MBS). The equation of motions can be derived from a variational principle and Hamilton's principle of least action, resulting in the DAE

$$\begin{aligned}
m_1\ddot{x}_1 &= 2\lambda(x_2 - x_1) + f,\\
m_2\ddot{x}_2 &= -2\lambda(x_2 - x_1),\\
m_2\ddot{y}_2 &= -2\lambda(y_2 - y_1) - m_2 g,\\
0 &= (x_2 - x_1)^2 + (y_2 - y_1)^2 - \ell^2,\\
0 &= y_1,
\end{aligned} \tag{1.6}$$

see [12] and the references therein. A typical control task is to move the payload from an initial position $(x_2(0), y_2(0))$ to a given position $(\tilde{x}_2, \tilde{y}_2)$ as fast a possible. However, a rapid movement of the



m_1	mass of the trolley
m_2	mass of the payload
ℓ	length of the rope
x_1	horizontal position of the trolley
y_1	vertical position of the trolley
x_2	horizontal position of the payload
y_2	vertical position of the payload
g	gravity constant
f	horizontal applied force

Figure 1.3 – A simplified model for a two-dimensional gantry crane

crane may result in a swaying payload, maneuvering the gantry crane in a potentially dangerous state. Following [105, 148] the sway can be reduced by applying a delayed position feedback controller, i.e., the external force f is generated by a controller κ of the form

$$f(t) = \kappa(t, x_1(t-\tau), y_2(t-\tau), x_2(t-\tau)).$$

For details we refer to [75, Section 1.8].

1.1.3 Transmission and propagation delays

The exchange of energy, information, or data in the interaction between two or more subsystems in Figure 1.1 is often modeled as an instantaneous process to happen instantaneously. If, however, some of the subsystems are physically separated, then this approximation cannot reasonably be made, and the modeling process has to account for the resulting transmission, propagation, or communication delays. Examples are wide-area power networks [2], synchronization of distant brain regions [170], chemical processes [64, 161] or controlling a satellite in outer space [195]. In this subsection we illustrate a propagation delay via an irreversible reaction $A \rightarrow B$ that is coupled with the reversible reaction $B \rightleftharpoons C$ in a *continuous stirred-tank reactor* (CSTR) with reaction rates $r_{A\rightarrow B}$ and $r_{B\rightleftharpoons C}$, which depend on the reactant concentrations c_A, c_B, c_C, and the temperature T in the tank. The reversible reaction $B \rightleftharpoons C$ is assumed to happen much faster than the irreversible reaction $A \rightarrow B$, which implies that the fast reaction is essentially at equilibrium with equilibrium constant $K_{B\rightleftharpoons C}$. Notice that Le Chatelier's principle implies that the equilibrium constant depends on the temperature T in the tank, i.e., $K_{B\rightleftharpoons C} = K_{B\rightleftharpoons C}(T)$.

To facilitate the transformation from A to C, we process the reaction in two CSTRs. In the first reactor, we use a high temperature T_1 to promote the conversion from A to B. Since the thermodynamic equilibrium limits the production of C (see the discussion above), we prescribe a lower temperature T_2 in the second tank to promote the production of C. The two tanks are linked via a

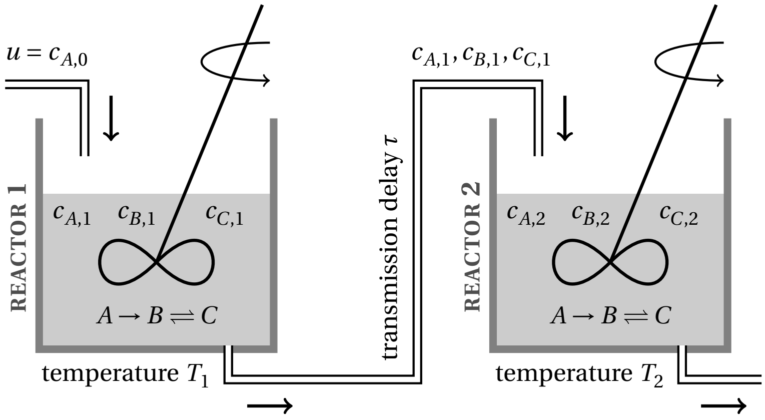

$u = c_{A,0}$
$c_{A,1}, c_{B,1}, c_{C,1}$
REACTOR 1
$c_{A,1}$ $c_{B,1}$ $c_{C,1}$
$A \rightarrow B \rightleftharpoons C$
temperature T_1
transmission delay τ
REACTOR 2
$c_{A,2}$ $c_{B,2}$ $c_{C,2}$
$A \rightarrow B \rightleftharpoons C$
temperature T_2

Figure 1.4 – A two-stage continuous stirred-tank reactor system

lossless transmission line (cf. Figure 1.4). As a consequence, the inflow of the second tank at time t equals the outflow of the first tank at time $t - \tau$. Hereby, $\tau > 0$ describes the time that is required to travel through the transmission line.

Assuming a reaction equilibrium for the fast reaction and constant volumes in both tanks, i.e., the inflow and outflow rates are identical, the set of DDAEs

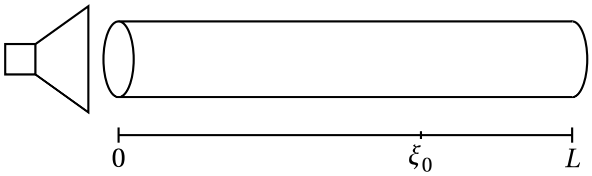
$$\begin{aligned}
\dot{c}_{A,1} &= \kappa_1(u - c_{A,1}) - r_{A\to B,1}, & \dot{c}_{A,2} &= \kappa_2(c_{A,1}(\cdot - \tau) - c_{A,2}) - r_{A\to B,2},\\
\dot{c}_{B,1} &= -\kappa_1 c_{B,1} + r_{A\to B,1} - r_{B\rightleftharpoons C,1}, & \dot{c}_{B,2} &= \kappa_2(c_{B,1}(\cdot - \tau) - c_{B,2}) + r_{A\to B,2} - r_{B\rightleftharpoons C,2},\\
\dot{c}_{C,1} &= -\kappa_1 c_{C,1} + r_{B\rightleftharpoons C,1}, & \dot{c}_{C,2} &= \kappa_2(c_{C,1}(\cdot - \tau) - c_{C,2}) + r_{B\rightleftharpoons C,2},\\
0 &= K_{B\rightleftharpoons C}(T_1)c_{B,1} - c_{C,1}, & 0 &= K_{B\rightleftharpoons C}(T_2)c_{B,2} - c_{C,2},\\
0 &= r_{A\to B,1} - k_{A\to B}(T_1)c_{A,1}, & 0 &= r_{A\to B,2} - k_{A\to B}(T_2)c_{A,2},
\end{aligned} \tag{1.7}$$

describes the complete model with unknowns $c_{A,i}, c_{B,i}, c_{C,i}, r_{A\to B,i}, r_{B\rightleftharpoons C,i}$ for $i = 1,2$. The constant κ_i is given by the flow rate divided by the volume. The reactant A is fed to the first CSTR with concentration $u = c_{A,0}$. Note that the reaction $r_{A\to B,i}$ is explicitly described by the algebraic equation and the prescribed function $k_{A\to B}$, which is for example given by the Arrhenius equation [130, p. 153]. In contrast, the reaction $r_{B\rightleftharpoons C,i}$ is only implicitly given by the equilibrium equation with prescribed equilibrium ratio $K_{B\rightleftharpoons C}(T)$.

Remark 1.1. The DDAE (1.7) can formally be obtained by accounting for the different time scales via a singular perturbation and a formal limit [123]. Such a process is typical for dynamics with fast and slow time scales, and it is essential to understand the limiting situation for the construction of numerical methods [127]. As a consequence, DDAEs also arise as the limiting situation of singularly perturbed DDEs. ♣

1.1.4 Reformulation of hyperbolic problems as delay equations

Not only the interaction between subsystems in Figure 1.1 may induce delays into the overall system dynamics, but also subsystems themselves might feature a description with delayed variables. For instance, it is well-known that many first-order hyperbolic *partial differential equations*



L	length of the duct
ρ_0	reference density
c	speed of sound
v	velocity
p	pressure

Figure 1.5 – Acoustic transmission in a fluid-filled duct

(PDEs) can be rewritten as delay difference equations by exploiting in some sense the method of characteristics [61, 120]. This technique is applied for instance to circuits that involve lossless transmission lines [39, 139], structured population models [36], mining ventilation [219], and blood flow systems [38]. We exemplify the transformation from a hyperbolic problem to a delay equation by considering acoustic transmission in a fluid-filled duct of length L that has an acoustic driver positioned at one end (cf. Figure 1.5). Following [63], the pressure $p(t,\xi)$ and the fluid velocity $v(t,\xi)$ at a point $(t,\xi) \in (0,T) \times (0,L)$ satisfy the coupled PDE

$$\frac{1}{c^2}\frac{p(t,\xi)}{\partial t} = -\rho_0 \frac{\partial v(t,\xi)}{\partial \xi}, \qquad \rho_0 \frac{\partial v(t,\xi)}{\partial t} = -\frac{\partial p(t,\xi)}{\partial \xi}, \qquad 0 < t < T, 0 < \xi < L, \tag{1.8a}$$

$$v(t,0) = u(t), \qquad p(t,L) = 0, \qquad 0 < t < T, \tag{1.8b}$$

where $c > 0$ denotes the speed of sound (which is assumed to be constant within the duct) and $\rho_0 > 0$ the reference density. The boundary condition imposed by the acoustic driver at $\xi = 0$ is given by the control input u. Since (1.8a) resembles a wave equation, the general solution of the PDE (1.8a) is given by

$$v(t,\xi) = \phi(\xi - ct) + \psi(\xi + ct), \qquad p(t,\xi) = \rho_0 c\big(\phi(\xi - ct) - \psi(\xi + ct)\big), \tag{1.9}$$

where ϕ is a wave traveling to the right and ψ a wave traveling to the left. Similar to [39, 102], we rewrite (1.9) to obtain

$$2\phi(\xi - ct) = v(t,\xi) + \frac{1}{\rho_0 c} p(t,\xi), \qquad 2\psi(\xi + ct) = v(t,\xi) - \frac{1}{\rho_0 c} p(t,\xi).$$

Using

$$\begin{aligned} 2\phi(-ct) = 2\phi\big(L - c\big(t + \tfrac{L}{c}\big)\big) &= v\big(t + \tfrac{L}{c}, L\big) + \frac{1}{\rho_0 c} p\big(t + \tfrac{L}{c}, L\big), \\ 2\psi(ct) = 2\psi\big(L + c\big(t - \tfrac{L}{c}\big)\big) &= v\big(t - \tfrac{L}{c}, L\big) - \frac{1}{\rho_0 c} p\big(t - \tfrac{L}{c}, L\big) \end{aligned}$$

and the boundary conditions (1.8b) at $t - L/C$ we obtain

$$\begin{aligned} u\big(t - \tfrac{L}{c}\big) = v(t - \tfrac{L}{c}, 0) &= \frac{1}{2}\Big(v(t,L) + v(t - \tfrac{2L}{c}, L) + \tfrac{1}{\rho_0 c}\big(p(t,L) - p(t - \tfrac{2L}{c}, L)\big)\Big) \\ &= \frac{1}{2}\big(v(t,L) + v(t - \tfrac{2L}{c}, L\big). \end{aligned}$$

Setting $x(t) := v(t,L)$, $\tilde{u}(t) := u(t - L/c)$, and $\tau := 2L/c$ we get the linear difference equation

$$x(t) = -x(t - \tau) + 2\tilde{u}(t),$$

which is a special case of a linear DDAE.

1.1.5 Delay differential-algebraic equations as mathematical tool

Besides the various application areas outlined above, DDAEs are also a powerful mathematical tool, which we highlight with several examples. For instance, DDAEs are used to design numerical time-integration methods for *neutral delay differential equations* (NDDEs), i.e., differential equations where the rate of change at time t depends on the rate of change at time $t-\tau$. If the NDDE features a particular structure [24], the DDAE formulation reveals that the problem consists of a differential equation coupled with an algebraic recursion for the delay. This reformulation benefits the theoretical and numerical investigation of the NDDE, see, for instance, [24, 25]. DDAEs are also used to establish existence and uniqueness of solutions. For example, in [38] the authors study a hyperbolic equation that is coupled via its boundary conditions to a switched DAE. By rewriting the hyperbolic equation as a delay equation (cf. section 1.1.4), the existence of solutions is a mere consequence of the existence of solutions of the corresponding switched DDAE. Another application of DDAEs is presented in [4, 5] in the analysis of a semi-explicit scheme for a linear poroelasticity problem that arises in the modeling of deformation resulting from tumor growth in the brain [182]. We detail this example to explain the line of reasoning. In more detail, the mathematical model of this problem is an elliptic equation for the displacement that is coupled to a parabolic equation for the pressure. Semi-discretization in space yields (under reasonable assumptions on the spatial discretization) to a DAE of the form

$$K_u u(t) - D^T p(t) = f(t), \tag{1.10a}$$

$$D\dot{u}(t) + M_p\dot{p}(t) + K_p p(t) = g(t). \tag{1.10b}$$

Hereby, u denotes the displacement, p the pressure, K_u and K_p the stiffness matrices corresponding to u and p, M_p the mass matrix for the pressure, and D the coupling matrix. For the numerical integration in time, we consider a step size $\tau > 0$ and the time grid $t_k := k\tau$. The equations (1.10a) and (1.10b) can be decoupled by employing a semi-explicit Euler discretization in time, given by

$$K_u u_{k+1} - D^T p_k = f_{k+1}, \tag{1.11a}$$

$$D(u_{k+1} - u_k) + M_p(p_{k+1} - p_k) + \tau K_p p_{k+1} = g_{k+1} \tag{1.11b}$$

with approximations $u_k \approx u(k\tau)$, $p_k \approx p(k\tau)$, and $f_k := f(k\tau)$, $g_k := g(k\tau)$. Note that p_k appears in (1.11a) (instead of p_{k+1}), which renders the scheme semi-explicit. Clearly, the decoupling has the advantage that two smaller subsystems with a nice structure have to be solved in each time-step. A key tool for the analysis of the semi-explicit scheme is the observation that (1.11) corresponds to an implicit Euler discretization of the DDAE

$$K_u u(t) - D^T p(t-\tau) = f(t), \tag{1.12a}$$

$$D\dot{u}(t) + M_p\dot{p}(t) + K_p p(t) = g(t). \tag{1.12b}$$

In particular, the semi-explicit scheme only converges if the DDAE (1.12) is asymptotically stable, which imposes a weak coupling condition, i.e., D is in some sense small compared to K_u and M_p. We refer to [4] for further details.

1.2 Scope and synopsis

Summarizing the motivating examples from the previous section, the main object under investigation in this thesis is the nonlinear DDAE

$$0 = F(t, x(t), \dot{x}(t), x(t-\tau), u(t)), \tag{1.13}$$

where $x(t) \in \mathbb{F}^{n_x}$ and $u(t) \in \mathbb{F}^{n_u}$ denote, respectively, the *state* and *control* of the system. Hereby, $\mathbb{F}$ denotes either the field of real or complex numbers, i.e., $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. The function F is defined on the time interval $\mathbb{I} := [t_0, t_\mathrm{f}]$ and open sets $\mathbb{D}_x, \mathbb{D}_{\dot{x}}, \mathbb{D}_{\sigma_\tau x} \subseteq \mathbb{F}^{n_x}$ and $\mathbb{D}_u \subseteq \mathbb{F}^{n_u}$ via

$$F\colon \mathbb{I} \times \mathbb{D}_x \times \mathbb{D}_{\dot{x}} \times \mathbb{D}_{\sigma_\tau x} \times \mathbb{D}_u \to \mathbb{F}^m$$

and is assumed to be sufficiently smooth. A special case of (1.13) is the so-called *linear time-varying DDAE*

$$E(t)\dot{x}(t) = A_1(t)x(t) + A_2(t)x(t-\tau) + B(t)u(t) + f(t), \tag{1.14}$$

with $E, A_1, A_2 \colon \mathbb{I} \to \mathbb{F}^{m \times n_x}$, $B\colon \mathbb{I} \to \mathbb{F}^{m \times n_u}$, and *inhomogeneity* $f\colon \mathbb{I} \to \mathbb{F}^m$. If the matrix functions $E, A, D,$ and B are constant on $\mathbb{I}$, then (1.14) is called *linear time-invariant* (LTI), and by abuse of notation written as

$$E\dot{x}(t) = A_1 x(t) + A_2 x(t-\tau) + Bu(t) + f(t), \tag{1.15}$$

with matrices $E, A_1, A_2 \in \mathbb{F}^{m \times n_x}$ and $B \in \mathbb{F}^{m \times n_u}$.

An important feature that distinguishes the DDAE (1.13) from a retarded DDE is that we allow $\frac{\partial}{\partial \dot{x}} F$ to be pointwise singular. While the potential singularity of $\frac{\partial}{\partial \dot{x}} F$ allows to include algebraic constraints and thus for a very flexible modeling approach, it comes with additional difficulties in the theoretical and numerical analysis. This is well-known for DAEs, see for instance [171], and accordingly, these difficulties are transferred directly to DDAEs [13, 53, 98].

Remark 1.2. The formulation of the DDAE (1.13) is not restricted to one single delay, since multiple commensurate delays [85] may be rewritten as a single delay by introducing new variables [94]. More precisely, a DDAE with multiple commensurate delays may be written as

$$0 = F(t, x(t), \dot{x}(t), x(t-\tau), x(t-2\tau), \ldots, x(t-k\tau), u(t)). \tag{1.16}$$

Introducing the new variables $z_i(t) = z_{i-1}(t-\tau)$ for $i = 1, \ldots, k$ with $z_0(t) = x(t)$ allows to write (1.16) as

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} F(t, z_0(t), \dot{z}_0(t), z_1(t), \ldots, z_k(t), u(t)) \\ z_1(t) - z_0(t-\tau) \\ \vdots \\ z_k(t) - z_{k-1}(t-\tau) \end{bmatrix},$$

which is again of the form (1.13). Note that if the DDAE depends also on derivatives of the past argument, i. e. on $x^{(\ell)}(t-\tau)$ for some $\ell \in \mathbb{N}$, one can use a similar procedure to recast this problem in the form (1.13). ♣

A standard question to ask is whether for a given control function u, the DDAE (1.13) (resp. (1.14)) possesses a unique solution. As it is already known from the theory of *ordinary differential equations* (ODEs), this is in general not the case. For ODEs, this can be fixed by imposing a constraint on the state in form of an initial or final condition (and some further technical assumptions). In this thesis we focus on prescribed initial conditions, which for delay equations take the form

$$x(t) = \phi(t) \qquad \text{for } t \in [-\tau, 0]. \tag{1.17}$$

The equations (1.13) and (1.17) together are referred to as *initial value problem* (IVP). After establishing conditions on F and ϕ for a solution to exist, the next question to ask is whether the solution is unique and whether its dependency on the data is continuous. More precisely, we are interested in the question, whether for a given control u the operator equation

$$\mathscr{K}(x) := \begin{bmatrix} F(\cdot, x, \dot{x}, \sigma_\tau x, u) \\ x_{[-\tau,0]} \end{bmatrix} = \begin{bmatrix} p \\ \phi \end{bmatrix}, \tag{1.18}$$

with shift operator $(\sigma_\tau x)(t) = x(t-\tau)$ is *well-posed* in the sense of Hadamard [88, 100, 147], i.e., if

(i) for each $(p, \phi) \in \mathscr{P} \times \Phi$, there is a solution $x \in \mathscr{X}$ of (1.18),

(ii) this solution is unique in $\mathscr{X}$ and

(iii) the dependency of x upon (p, ϕ) is continuous.

Hereby, the operator $\mathscr{K}$ maps the topological space $\mathscr{X}$ into the topological space $\mathscr{Z} := \mathscr{P} \times \Phi$. To answer the question whether (1.18) is well-posed, a number of other questions need to be answered first, ranging from the solution concept used for (1.13) (thus fixing the space $\mathscr{X}$), to the smoothness requirements for F. These questions are addressed in detail in the first part of the thesis. More precisely, we have the following results:

(i) The forthcoming Examples 1.4 and 1.5 reveal that in general we cannot expect a classical solution to exist. We thus start our analysis by seeking solutions in the space of piecewise-smooth distributions (see Definition 3.3). Our first main result — Theorem 3.5 — details that a regular matrix pair (E, A_1) in (1.15) is a sufficient condition for existence and uniqueness of solutions for any initial trajectory, input, and external forcing.

(ii) In order to obtain a necessary condition, we define the notion of *delay-regularity* (Definition 3.8) and establish in Theorem 3.20 existence and uniqueness of solutions for all external forcing signals if and only if $\det(sE - A_1 - \omega A_2) \not\equiv 0$. As a mere consequence, we show that a linear DAE can be regularized via feedback if and only if it can be regularized by a delayed feedback. The details are presented in section 3.2.

(iii) In Definition 3.28 we introduce a novel equivalence relation, called *delay-equivalence*, which allows us to establish that whenever the DDAE (1.15) is delay-regular, then it can be transformed to a delay-equivalent DDAE with regular matrix pencil, see Theorem 3.37 and Remark 3.39. The delay-equivalent DDAE can be obtained by a simple compress-and-shift algorithm, which was previously suggested in the literature [53]. Our analysis reveals that with a simple modification of the algorithm it can detect along the way if the DDAE is delay-regular, and terminate otherwise.

(iv) Using the previous results, we introduce a new classification of LTI DDAEs in section 4.1 that is based on the propagation of so-called primary discontinuities [26]. The main motivation for such a classification is to establish conditions that ensure existence of solutions that are continuously differentiable almost everywhere. We present a complete algebraic characterization of the different classes in Theorem 4.16, which in turn allows us to formulate a general existence result for DDAEs (cf. Theorem 4.18).

(v) If we impose some additional regularity assumptions on the history function ϕ, that is, we impose that the history function is linked smoothly to the solution of the DDAE, then we can further improve the result of Theorem 4.18. In more detail, the impact of so-called splicing conditions is analyzed in section 4.2, where we show the existence of a continuous solutions for a higher index DDAE in Theorem 4.27.

(vi) In chapter 5 we show that the results from the LTI case can in parts be translated to nonlinear DDAEs. The main observation is that the results in Chapter 4 can be formulated in terms of the underlying DDE, which can also be defined for nonlinear DDAEs. With this we establish existence and uniqueness results for a class of nonlinear DDAEs in Theorem 5.24.

(vii) We conclude our analysis with a detailed investigation of the hybrid numerical-experimental system presented in section 1.1.1. In particular, we show that the compress-and-shift algorithm (Algorithm 1) from chapter 3 can be applied to the hybrid system even in the nonlinear case. The algorithm terminates after a single shift with a regular DDAE, whenever the two subsystems are regular DAEs. The details are presented in Lemma 5.10, Theorem 5.15 and Theorem 5.17. Using the solution theory developed in chapter 5 we prove that the hybrid system is solvable whenever the experimental and numerical part are both strangeness-free, see Corollary 5.25.

In the second part of the thesis we invert the problem by asking whether we can determine an operator $\widetilde{\mathscr{K}}\colon \mathscr{X} \to \mathscr{Z}$ that minimizes

$$\|\widetilde{\mathscr{K}}(x) - p\| \tag{1.19}$$

for some given $x \in \mathscr{X}$, $p \in \mathscr{Z}$ and some suitable norm $\|\cdot\|$. In this case, $\widetilde{\mathscr{K}}$ is called a *realization* for the data pair (x, p). The simplest realization is of course to just map x onto p. However, this realization is only valid for the specific data pair (x, p) and any variation in the data pair results in a different realization. Instead, we may want to ask for a realization $\widetilde{\mathscr{K}}$ that minimizes (1.19) for all data pairs (x, p). Since the class of all possible realizations is hard to parameterize, we restrict ourselves to the case that $\widetilde{\mathscr{K}}$ is linear, i.e., we consider only linear time-invariant DDAEs of the form (1.15). Often in practical application, the state x itself may not be available (or of interest) and instead, only an observation of the state in form of an output equation

$$y(t) = Cx(t) \tag{1.20}$$

where the matrix $C \in \mathbb{F}^{n_y \times n_x}$ is available. Consequently, we assume that there exists a dynamical system, exemplified by an operator S that maps inputs u to outputs y. Hereby the standing assumption is that we can evaluate S for given inputs but do not have access to a state-space realization of S, i.e., S acts as a black-box. The DDAE realization problem can thus be stated as follows: For given trajectories $\widehat{u}$ and $\widehat{y}$, construct a DDAE

$$\widetilde{S}\colon \begin{cases} \widetilde{E}\dot{x}(t) = \widetilde{A}_1 x(t) + \widetilde{A}_2 x(t-\tau) + \widetilde{B}u(t), \\ \quad y(t) = \widetilde{C}x(t) \end{cases} \tag{1.21}$$

such that $\|S(u) - \widetilde{S}(u)\|$ is minimized for all admissible inputs u. If we choose the norm to be the $\mathscr{L}_\infty$ norm then it is well-known (cf. [20]) that the error is bounded by the $\mathscr{H}_2$ norm of the error system multiplied with the norm of the input, i.e.,

$$\left\| y - \widetilde{y} \right\|_{\mathscr{L}_\infty} \leq \|S - \widetilde{S}\|_{\mathscr{H}_2} \|u\|_{\mathscr{L}_2},$$

provided that all quantities are well-defined. If in addition we are seeking for a standard state-space system, i.e., $E = I_{n_x}$ and $A_2 = 0$, then the $\mathscr{H}_2$ error is minimized if the transfer function associated with $\widetilde{S}$ interpolates the transfer function associated with S at certain frequencies – see for instance [20]. Thus, we propose to tackle the DDAE realization problem by constructing a realization of the form (1.21), such that its transfer function

$$\widetilde{H}(s) = \widetilde{C}\left(s\widetilde{E} - \widetilde{A}_1 - \exp(-\tau s)\widetilde{A}_2\right)^{-1}\widetilde{B} \tag{1.22}$$

interpolates the transfer function associated with S at given frequencies. We refer to Problem 6.6 and Problem 6.9 for a precise problem description. We obtain the following results.

(i) We present necessary and sufficient conditions for interpolation in Theorem 7.1 by analyzing the more general class of realizations of the form $\widetilde{H}(s) = \widetilde{C}\left(\sum_{k=1}^{K} h_k(s)\widetilde{A}_k\right)^{-1}\widetilde{B}$ with linear independent family $\{h_1, \ldots, h_K\}$ of meromorphic functions mapping the complex plane into itself.

(ii) The interpolation conditions from Theorem 7.1 reveal that the situation is different for $K = 2$ and $K \geq 3$. For $K = 2$ we provide a direct solution of the problem (see Theorem 7.5) and show its close connection to the Loewner framework [150] in Corollary 7.7.

(iii) For the case $K \geq 3$, which is the case for the DDAE realization problem, we present two strategies to handle the remaining degrees of freedom: First by interpolation of additional data (cf. section 7.3.1), and second by interpolation of derivative information of the transfer function (cf. section 7.3.2). In both cases we do not increase the dimension of the involved matrices. The main results are presented in Theorem 7.12 and Theorem 7.16.

(iv) To obtain data in the frequency domain we consider in chapter 8 the estimation of frequency data from time series, i.e., from a sampling of u and y in the time-domain. We use the *least-squares transfer function estimate* (lsTFE) framework introduced in [168] and generalize the required results to our setting, which includes continuous time systems and general system structures. The resulting method is summarized in Algorithm 5.

(v) Based on additional data points, we present a least-squares approach (see section 8.3) to estimate possibly unknown parameters in the realization as, for instance, the delay parameter.

We demonstrate the results with a complete case study with a delay example in section 8.4.

1.3 Challenges and state of the art

Since DAEs and DDEs are special cases of the DDAE (1.13), it is clear that the problems specific to these sub classes are also inherent to DDAEs. For instance, DAEs require a so-called consistent initialization [127,167], since the class of possible initial conditions is restricted (see the forthcoming Chapter 2). For DDEs, one needs to put special emphasis on the tracking of so-called *breaking points* [90,91], which result from the fact that the history function ϕ may not be linked smoothly to the solution x at $t=0$, i.e. we have

$$\lim_{t\nearrow 0}\dot{\phi}(t) \neq \lim_{t\searrow 0}\dot{x}(t) \tag{1.23}$$

in general. The following examples suggest, that in some cases this so-called *primary discontinuity* [26] may be smoothed out over time, while in other cases, the discontinuity is propagated over time or even amplified.

Example 1.3. The IVP $\dot{x}(t) = -x(t-1)$ with history function $\phi \equiv 1$ can be solved by integration on successive time intervals (see chapter 2 for more details), which yields for $0 \le t \le 4$ the solution

$$x(t) = \begin{cases} 1-t, & \text{if } 0 \le t \le 1,\\ \frac{1}{2}t^2 - 2t + \frac{3}{2}, & \text{if } 1 \le t \le 2,\\ -\frac{1}{6}t^3 + \frac{3}{2}t^2 - 4t + \frac{17}{6}, & \text{if } 2 \le t \le 3,\\ \frac{1}{24}t^4 - \frac{2}{3}t^3 + \frac{15}{4}t^2 - \frac{17}{2}t + \frac{149}{24}, & \text{if } 3 \le t \le 4, \end{cases}$$

which is depicted as the blue line in Figure 1.6. Note that $\lim_{t\nearrow 0}\dot{\phi}(t) = 0 \neq -1 = \lim_{t\searrow 0}\dot{x}(0)$, i.e. the solution is not continuously differentiable at $t=0$. Straight forward computations show

$$\begin{aligned} \lim_{t\nearrow 1}\frac{\mathrm{d}}{\mathrm{d}t}(1-t) = -1 &= \lim_{t\searrow 1}\frac{\mathrm{d}}{\mathrm{d}t}\left(\tfrac{1}{2}t^2 - 2t + \tfrac{3}{2}\right),\\ \lim_{t\nearrow 2}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^2\left(\frac{1}{2}t^2 - 2t + \frac{3}{2}\right) = 1 &= \lim_{t\searrow 2}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^2\left(-\frac{1}{6}t^3 + \frac{3}{2}t^2 - 4t + \frac{17}{6}\right),\\ \lim_{t\nearrow 3}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^3\left(-\frac{1}{6}t^3 + \frac{3}{2}t^2 - 4t + \frac{17}{6}\right) = -1 &= \lim_{t\searrow 3}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^3\left(\frac{1}{24}t^4 - \frac{2}{3}t^3 + \frac{15}{4}t^2 - \frac{17}{2}t + \frac{149}{24}\right), \end{aligned}$$

and thus the solution becomes smoother over time. ♠

Example 1.4. The IVP $0 = x(t) + x(t-1) + 1$ with history function $\phi(t) = t$ has the solution

$$x(t) = \begin{cases} k-1-t, & \text{if } k-1 \le t \le k \text{ and } k \in \mathbb{N} \text{ odd},\\ t+k, & \text{if } k-1 \le t \le k \text{ and } k \in \mathbb{N} \text{ even}. \end{cases}$$

In particular, the solution x (plotted as dashed red line in Figure 1.6) is continuous but $\dot{x}$ is discontinuous at every $t = k$ and thus no smoothing occurs. ♠

Example 1.5. Consider the DDAE

$$\begin{bmatrix} 1 & 0\\ 0 & 0 \end{bmatrix}\begin{bmatrix} \dot{x}_1(t)\\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1(t)\\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 & 0\\ 0 & -1 \end{bmatrix}\begin{bmatrix} x_1(t-1)\\ x_2(t-1) \end{bmatrix} \tag{1.24}$$

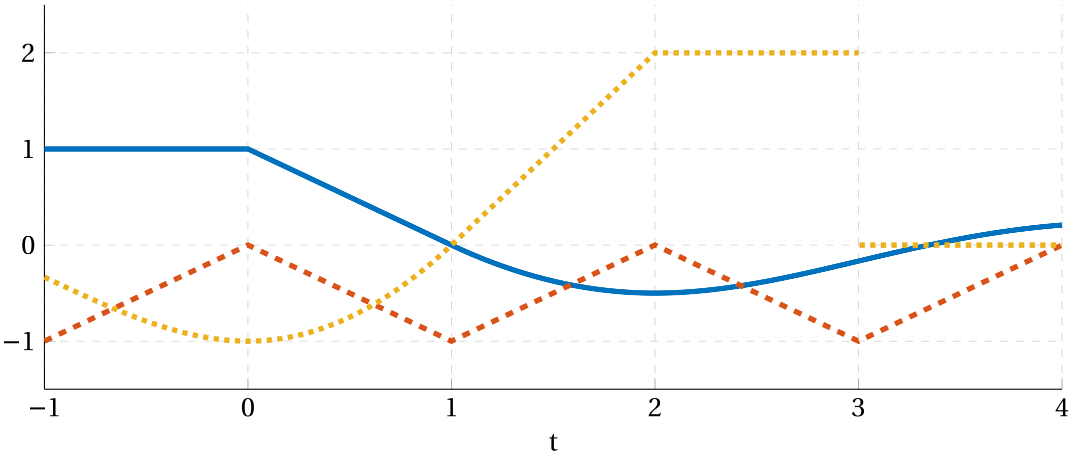


Figure 1.6 – Discontinuity propagation in DDAEs. The solid blue line represents the solution of the IVP in Example 1.3, the dashed red line the solution derived in Example 1.4, and the first component of the solution of Example 1.5 is presented with the dotted yellow line.

with initial condition

$$x(t) = \phi(t) = \begin{bmatrix} \frac{1}{3}t^3 + t^2 - 1, \\ \frac{1}{3}(t-1)^3 + (t-1)^2 - 1 \end{bmatrix} \qquad \text{for } -1 \le t \le 0.$$

Note that $x_1(t) = x_2(t-1)$ and thus it is sufficient to compute the solution x_1 (the dotted yellow line in Figure 1.6), which is given by

$$x_1(t) = \begin{cases} t^2 - 1, & t \in [0,1], \\ 2t - 2, & t \in [1,2], \\ 2, & t \in [2,3), \\ 0, & t \ge 3. \end{cases}$$

In particular, the solution becomes less smooth at multiples of the time delay and even discontinuous at $t = 3$. ♠

The study of primary discontinuities of the scalar DDE

$$a_0\dot{x}(t) + a_1\dot{x}(t-\tau) + b_0x(t) + b_1x(t-\tau) = f(t) \tag{1.25}$$

is based on the the classification proposed in [27]: The DDE (1.25) is said to be of *retarded type* if $a_0 \neq 0$ and $a_1 = 0$, of *neutral type* if $a_0 \neq 0$ and $a_1 \neq 0$, and of *advanced type* if $a_0 = 0$ and $a_1 \neq 0$. Following this classification, we observe that the DDAE in Example 1.3 is of retarded type, the DDAE Example 1.4 is of neutral type (if we differentiate the equation), while the first component in Example 1.5 satisfies a DDE of advanced type. In particular, the DDAE (1.13) may contain scalar delay differential equations of any of the three types. As a consequence, a history function ϕ for the IVP (1.13), (1.17) may be required to satisfy so-called *splicing conditions* [26]. We refer to section 4.2 for further details.

The classification of (1.25) was extended in the series of papers [97–99] to linear time-varying DDAEs of the form (1.14), using the so-called *underlying DDE* (see sections 4.1 and 4.3 for more details).

Loosely speaking, the underlying DDE is obtained by differentiating (and possibly shifting) parts of the DDAE (1.13) until one is able to solve for $\dot{x}$. Hereby, the number of differentiations that is required during this process is used as a classification for the difficulties associated with solving the DDAE (1.13) analytically or numerically. Different technical aspects and different ways of counting have led to several so-called *index* concepts for DAEs, for instance the *differentiation index* [54], the *perturbation index* [101], the *tractability index* [144–146], the *geometric index* [179, 181], the *structural index* [167, 174] and the *strangeness index* [124, 127]. For a comparison we refer to [153].

Besides the fact that neutral or even advanced equations may be hidden in the DDAE (1.13), the solution of the DDAE (1.13) may also depend on future evaluations of the DDAE. More precisely, the solution $x(t)$ at the time point t may depend on

$$0 = F(t + k\tau, x(t + k\tau), \dot{x}(t + k\tau), x(t + (k-1)\tau), u(t + k\tau))$$

with $k \in \mathbb{N}$. A simple example for such a situation is given if we choose $E = A_1 = 0$, and $A_2 = I_{n_x}$ in (1.15). Of course, in reality, a dependence on the future is not possible, and therefore, one may question the utility of a DDAE whose solution depends on future values. However, besides its mathematical relevance, the future evaluation of F may be interpreted as a *prediction* of that future value. In any case, the potential acausality causes some further difficulties in the analytical and numerical treatment of (1.13):

- The method of steps (see [26] and the forthcoming chapter 2), which is commonly used to solve the IVP (1.13), (1.17), cannot be used without pre-processing of the DDAE (1.13). The pre-processing requires to shift some of the equations of (1.13) to future time points [1, 53, 99]. The minimal number of shifts that is required to construct the solution is called the *shift index*. We refer to [98] for a precise definition.
- Due to the combination of differentiation and shifting the DDAE (1.13) may include higher-order differential equations [53, 97, 206].
- The shifting imposes restrictions on the set of history functions for which the IVP has a solution. In contrast to the DAE theory, where we expect a restriction only at the time points $t = 0$ and $t = -\tau$, the restriction applies to all $t \in [-\tau, 0]$.

So far, a general analysis of (1.13) is not available and most of the literature addresses only special cases. For instance, a classical solution theory for DDAEs that need not be shifted is developed in [13, 62] for nonlinear DDAEs with a special structure and for linear time-invariant DDAEs in [50]. Shifting and its consequences are studied in [1, 53, 94, 96–99, 173, 206]. Numerical time integration methods are developed and analyzed in [16, 33, 59, 90, 91, 98, 103, 108, 109, 136, 194, 201]. Most of the references for the numerical methods require that there is no need to shift equations and that the DAE that is obtained by substituting a smooth function parameter λ for the delayed variable has differentiation index one. Notable exceptions are provided for instance in [13, 16, 98, 103]. The stability and asymptotic stability of certain classes of DDAEs is studied in [41, 56, 69, 79, 95, 135, 138, 141, 151, 156, 222, 224, 225]. Surprisingly, it is not sufficient for asymptotic stability that all eigenvalues of the LTI system (1.15) have negative real part [68]. The main reason for this is that the solution fails to exists after some time. This is due to the fact that an advanced equation may be hidden in the DDAE (1.15), see the discussion above. Closely related to stability is the question wether a system can be stabilized via a suitable controller. This and further control theoretical topics are discussed in [3, 80, 81, 92, 157, 193, 194].

1.4 Previously published results and joint work

Some of the contents of this thesis are already published.

- The connection between the semi-explicit Euler discretization applied to linear poroelasticity and a suitable DDAE, as presented in section 1.1.5, is joint work with R. Altmann and R. Maier and published in [4]. An extension to nonlinear poroelasticity is considered in [5].
- The analysis of linear DDAEs within the space of piecewise-smooth distributions is a result of a collaboration with S. Trenn, which resulted in the conference proceedings [206] and the preprints [207, 208]. The results are presented in chapter 3.
- Most of the results from chapter 4 are published in [211], which additionally contains parts of section 2.1. The extension of these results to nonlinear DDAEs is covered in chapter 5 and in parts published in the preprint [212].
- The realization theory for DDAEs, discussed in chapter 7, is published in the journal articles [189] (together with P. Schulze) and [191] (with P. Schulze, C. Beattie, and S. Gugercin). The extension to time-domain data, covered in chapter 8, is based on the results of a collaboration with E. Fosong and P. Schulze, which are published in the preprint [77].

Results on structural analysis for DDAEs [1] (together with I. Ahrens), Kolmogorov n-widths for linear dynamical systems [213] (with S. Gugercin), model reduction for transport problems [35] (joint work with F. Black and P. Schulze), and model reduction for switched systems [190] (with P. Schulze) are only briefly mentioned in this thesis.

1.5 Notation

The symbols $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ denote the integers, the rational numbers, the real numbers, and the complex numbers, respectively. The natural numbers are the positive integers and are denoted with $\mathbb{N} := \{n \in \mathbb{Z} \mid n > 0\}$. For a field $\mathbb{F}$ and natural numbers $n, m \in \mathbb{N}$ we denote the set of all $n \times m$ matrices over $\mathbb{F}$ by $\mathbb{F}^{n \times m}$. The ring of polynomials with coefficients from a field $\mathbb{F}$ is denoted by $\mathbb{F}[s]$ with s being the indeterminate. The polynomial ring $\mathbb{F}[s]$ is naturally embedded in the field of rational functions, denoted by $\mathbb{F}(s)$. Therefore, we can also consider matrices with entries in the ring $\mathbb{F}[s]$. The set of n-dimensional nonsingular matrices over the field $\mathbb{F}$ is denoted with

$$\mathrm{GL}_n(\mathbb{F}) := \{A \in \mathbb{F}^{n \times n} \mid A \text{ nonsingular}\},$$

which together with the standard multiplication for matrices forms a group, the so-called *general linear group*. The neutral element of $\mathrm{GL}_n(\mathbb{F})$ is denoted with I_n and the inverse of $A \in \mathrm{GL}_n(\mathbb{F})$ is denoted with A^{-1}. In particular, we have $AA^{-1} = A^{-1}A = I_n$. The ith column of I_n, i.e., the ith unit vector, is denoted by $e_i \in \mathbb{F}^n$. If $A(s) \in \mathrm{GL}_n(\mathbb{F}[s])$ and $A(s)^{-1} \in \mathrm{GL}_n(\mathbb{F}[s])$, then $A(s)$ is called *unimodular* and it is easy to see that $A(s) \in \mathbb{F}[s]^{n\times n}$ is unimodular if and only if $\det(A(s))$ is a nonzero constant, i.e., $\det(A(s)) \in \mathbb{F}\setminus\{0\}$. The rank of a matrix $A \in \mathrm{GL}_n(\mathbb{F})$ is denoted with $\mathrm{rank}_{\mathbb{F}}(A)$ or simply with $\mathrm{rank}(A)$ if the field $\mathbb{F}$ is clear from the context. For polynomial matrices $A(s) \in \mathbb{F}[s]^{n\times m}$ we adopt the notation from the literature and write $\mathrm{rank}_{\mathbb{F}[s]}(A(s)) := \mathrm{rank}_{\mathbb{F}(s)}(A(s))$. The transpose and conjugate transpose of a matrix A are denoted with A^T and A^H.

Part I

Classification and well-posedness

2 Differential-algebraic equations and preliminary results

A standard approach to solve a differential equation with delayed argument, such as the *delay differential-algebraic equation* (DDAE)

$$0 = F(t, x(t), \dot{x}(t), x(t-\tau), u(t)) \qquad \text{for } t \in [t_0, t_\mathrm{f}) \tag{2.1a}$$

introduced in section 1.2 (cf. (1.13)) with initial condition

$$x(t) = \phi(t) \qquad \text{for } t \in [-\tau, 0] \tag{2.1b}$$

is via successive integration of (2.1) on the time intervals $[(i-1)\tau, i\tau)$, which is sometimes referred to as the *method of steps* [98], see also [26,50]. First we assume that M is the smallest integer such that $t_\mathrm{f} < M\tau$ and introduce for $i \in \mathscr{I} := \{1, \dots, M\}$ the functions

$$\begin{aligned}
x_{[i]} \colon [0,\tau] &\to \mathbb{F}^{n_x}, & t &\mapsto x(t + (i-1)\tau),\\
u_{[i]} \colon [0,\tau] &\to \mathbb{F}^{n_u}, & t &\mapsto u(t + (i-1)\tau),\\
F_{[i]} \colon [0,\tau] \times \mathbb{D}_x \times \mathbb{D}_{\dot{x}} \times \mathbb{D}_{\sigma_\tau x} \times \mathbb{D}_u &\to \mathbb{F}^m, & (t,x,y,z,u) &\mapsto F(t + (i-1)\tau, x, y, z, u),\\
x_{[0]} \colon [0,\tau] &\to \mathbb{F}^{n_x}, & t &\mapsto \phi(t - \tau).
\end{aligned} \tag{2.2}$$

Then for each $i \in \{1, \dots, M\}$ we have to solve the *differential-algebraic equation* (DAE)

$$0 = F_{[i]}(t, x_{[i]}(t), \dot{x}_{[i]}(t), x_{[i-1]}(t), u_{[i]}(t)), \qquad t \in [0,\tau), \tag{2.3a}$$

$$x_{[i]}(0) = x_{[i-1]}(\tau^-), \tag{2.3b}$$

with right continuation

$$x_{[i-1]}(\tau^-) := \lim_{t \nearrow \tau} x_{[i-1]}(t). \tag{2.4}$$

The analysis of DDAEs requires an in-depth understanding of the theory of DAE, which we thus recall in this chapter. For the analysis of (2.3) we employ the following solution concept from [127].

Definition 2.1. A function $x_{[i]} \in \mathscr{C}^1([0,\tau];\mathbb{F}^{n_x})$ is called a *(classical) solution* of (2.3a), if it satisfies (2.3a) pointwise. The function $x_{[i]} \in \mathscr{C}^1([0,\tau];\mathbb{F}^{n_x})$ is called a *(classical) solution of the initial value problem* (2.3) if it is a solution of (2.3a) and satisfies (2.3b). An initial condition $x_{[i-1]}(\tau)$ is called *consistent,* if the initial value problem (2.3) has at least one solution.

Note that the partial derivative of $F_{[i]}$ with respect to $x_{[i]}$ in (2.3a) is allowed to be singular, resulting in significant differences to the theory for *ordinary differential equations* (ODEs), see also [171]. The main differences are illustrated in the next example, taken from [205].

Example 2.2. Consider the *linear time-invariant* (LTI) DAE

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \end{bmatrix}.$$

The third equation implies $f_3 \equiv 0$, which means that even for arbitrarily smooth right-hand sides, we may not have existence of solutions. The second equation $x_2(t) = -f_2(t)$ shows that the set of initial conditions is restricted. Using the second equation we obtain the solution

$$x_1(t) = \dot{x}_2(t) - f_1(t) = -\dot{f}_2(t) - f_1(t)$$

for the first variable, which is thus less smooth than the inhomogeneity f. Note that x_3 is not specified at all and thus we do not have uniqueness of the solution. ♠

Since a solution of the DAE (2.3a) (and hence also a solution of the *initial value problem* (IVP) (2.1)) may depend on derivatives of $F_{[i]}$ and consequently on derivatives of the history function ϕ, we make the following assumption throughout the text.

Assumption 2.3. *The function $F_{[i]}$ in* (2.3a) *and the history function ϕ (and thus also the function F in* (2.1a)*) are sufficiently smooth.*

Remark 2.4. The theory of DAEs is already quite mature with a large body of literature, see for instance the monographs [42, 101, 127, 134, 197]. Fur further details we refer to the collection of survey articles [49, 111–114, 188] and the references therein. ♣

2.1 Classical solutions for linear time-invariant DAEs

As pointed out in Example 2.2, many aspects of the theory for DAEs are already present for LTI DAEs and hence we start our brief survey of DAE theory with linear systems of the form

$$E\dot{x} = A_1 x + \tilde{f}, \tag{2.5}$$

with $E, A_1 \in \mathbb{F}^{m \times n_x}$ and $\tilde{f} : [0, t_\mathrm{f}) \to \mathbb{F}^m$, where we omit the time dependency of x and f for the ease of presentation. As before, we frequently make the assumption that $\tilde{f}$ is smooth enough, i.e. that $\tilde{f}$ satisfies the following assumption.

Assumption 2.5. *The inhomogeneity $\tilde{f}$ is infinitely many times continuously differentiable.*

Remark 2.6. Applying the method of steps to the DDAE (1.15) results in the DAE (2.5) with $x = x_{[i]}$ and $\tilde{f} = A_2 x_{[i-1]} + f$. In particular, the inhomogeneity $\tilde{f}$ depends on the solution on the previous interval and hence Assumption 2.3 does not imply Assumption 2.5. ♣

The solvability of (2.5) is closely related to the properties of the matrix pair (E, A_1), see for instance [127] for further details. If $m = n_x$ and $\det(\lambda E - A_1) \in \mathbb{F}[\lambda] \setminus \{0\}$, then the matrix pair (E, A_1) is called *regular*. If a matrix pair is not regular, it is referred to as *singular*. The following result [127, Theorem 2.14] shows that the regularity of the matrix pair is a necessary condition to ensure existence and uniqueness of solutions for the DAE (2.5), see also [32].

Theorem 2.7. *Let $E, A \in \mathbb{C}^{m\times n_x}$ and suppose that (E, A_1) is a singular matrix pair.*

(i) If $\operatorname{rank}(\lambda E - A_1) < n_x$ for all $\lambda \in \mathbb{C}$, then the homogeneous initial value problem

$$E\dot{x} = A_1 x, \qquad x(0) = 0$$

has a nontrivial solution.

(ii) If $\operatorname{rank}(\lambda E - A_1) = n$ for some $\lambda \in \mathbb{C}$ and hence $m > n_x$, then there exist arbitrarily smooth inhomogeneities $\tilde{f}$ for which the corresponding DAE is not solvable.

Remark 2.8. If (E, A_1) is not regular, it is still possible that the IVP associated with the DDAE (1.15) has a unique solution (in the sense of [98]). In this case, the DDAE is called *noncausal* and under some technical assumptions [98] provides an algorithm to transform (3.1a) such that the transformed pencil $(\tilde{E}, \tilde{A})$ is regular. However, such a process adds additional restrictions on the history function. For more details, we refer to [52, 97] and the forthcoming section 3.1. ♣

A standard approach in studying linear differential equations is to introduce an equivalence relation on the system matrices that allows to characterize all solutions. In terms of matrix pencils, we say that (E, A_1) and $(\widetilde{E}, \widetilde{A}_1)$ are (strongly) *equivalent*, in symbols $(E, A_1) \sim (\widetilde{E}, \widetilde{A}_1)$, if and only if there exists nonsingular matrices $S \in \mathrm{GL}_m(\mathbb{F})$ and $T \in \mathrm{GL}_{n_x}(\mathbb{F})$ such that

$$\widetilde{E} = SET \qquad \text{and} \qquad \widetilde{A}_1 = SA_1T.$$

The canonical form for this equivalence relation is the *Kronecker canonical form* [82, Cha. XII, § 4] (assuming $\mathbb{F} = \mathbb{C}$). From the canonical form it is easy to determine whether the matrix pencil is regular, yielding a special form of the Kronecker canonical form that is known as the *Weierstraß canonical form* [82, Cha. XII, Thm. 3]. More precisely, we have the following characterization of regularity (which is sometimes referred to as the quasi-Weierstraß form [31]).

Theorem 2.9 (Quasi-Weierstraß form). *The matrix pencil $(E, A_1) \in \left(\mathbb{F}^{m\times n_x}\right)^2$ is regular if and only if $m = n_x$ and there exist matrices $S, T \in \mathrm{GL}_{n_x}(\mathbb{F})$ such that*

$$SET = \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & N \end{bmatrix} \qquad \text{and} \qquad SA_1T = \begin{bmatrix} J & 0 \\ 0 & I_{n_{x,\mathrm{a}}} \end{bmatrix}, \tag{2.6}$$

where $N \in \mathbb{F}^{n_{x,\mathrm{a}}\times n_{x,\mathrm{a}}}$ is a nilpotent matrix and $J \in \mathbb{F}^{n_{x,\mathrm{d}}\times n_{x,\mathrm{d}}}$.

Remark 2.10. If $\mathbb{F} = \mathbb{C}$, then the Weierstraß canonical form can be obtained by choosing S and T in Theorem 2.9 such that N and J are in Jordan canonical form. If $\mathbb{F} \neq \mathbb{C}$, then the Jordan canonical form of J may not exist and we cannot use the Weierstraß canonical form. For the analysis of the DAE (2.5) this is not important since the relevant feature of the (quasi-)Weierstraß form is the decoupling of the matrix pencil (see the forthcoming discussion). ♣

Note that the numbers $n_{x,\mathrm{d}}$ and $n_{x,\mathrm{a}}$ in Theorem 2.9 are independent of the specific choice of the matrices S and T (see for instance [31, 127]). In more detail, consider matrices $S_i, T_i \in \mathrm{GL}_{n_x}(\mathbb{F})$ for $i = 1,2$ that transform the regular pencil (E, A_1) into quasi-Weierstraß form, that is

$$S_i E T_i = \begin{bmatrix} I_{n_i} & 0 \\ 0 & N_i \end{bmatrix} \qquad \text{and} \qquad S_i A_1 T_i = \begin{bmatrix} J_i & 0 \\ 0 & I_{\tilde{n}_i} \end{bmatrix} \qquad \text{for } i = 1,2,$$

where for $i = 1,2$ the matrix $N_i \in \mathbb{F}^{\tilde{n}_i \times \tilde{n}_i}$ is nilpotent and $J_i \in \mathbb{F}^{n_i \times n_i}$. Then

$$\begin{aligned} \det(\lambda E - A_1) &= \det(S_i)^{-1} \det\left(\lambda \begin{bmatrix} I_{n_i} & 0 \\ 0 & N_i \end{bmatrix} - \begin{bmatrix} J_i & 0 \\ 0 & I_{\tilde{n}_i} \end{bmatrix}\right) \det(T_i)^{-1} \\ &= \det(S_i)^{-1} \det(T_i)^{-1} \det(\lambda I_{n_i} - J_i) \det(\lambda N_i - I_{\tilde{n}_i}) \qquad \text{for } i = 1,2. \end{aligned}$$

Since N_i is nilpotent, we have $\det(\lambda N_i - I_{\tilde{n}_i}) = (-1)^{\tilde{n}_i}$ for any $\lambda \in \mathbb{F}$ and $i = 1,2$. With the setting $c_i := (-1)^{\tilde{n}_i} \det(S_i)^{-1} \det(T_i)^{-1} \in \mathbb{F} \setminus \{0\}$ we obtain

$$\det(\lambda E - A_1) = c_i \det(\lambda I_{n_i} - J_i)$$

and thus $n_1 = n_2$ and $\tilde{n}_1 = n - n_1 = n - n_2 = \tilde{n}_2$.

The matrices S and T in Theorem 2.9 can be obtained from the so-called *Wong sequences* [220]

$$\mathcal{V}_0 := \mathbb{F}^{n_x}, \qquad \mathcal{V}_{i+1} := A_1^{-1}(E\mathcal{V}_i) := \{x \in \mathbb{F}^n \mid A_1 x \in E\mathcal{V}_i\}, \qquad \text{for } i \in \mathbb{N}, \tag{2.7a}$$

$$\mathcal{W}_0 := \{0\}, \qquad \mathcal{W}_{i+1} := E^{-1}(A_1\mathcal{W}_i) := \{x \in \mathbb{F}^n \mid Ex \in A_1\mathcal{W}_i\}, \qquad \text{for } i \in \mathbb{N}, \tag{2.7b}$$

where in this context E^{-1} and A_1^{-1} denote the preimage of E and A_1, respectively. Note that the sequences are nested, i.e., $\mathcal{V}_{i+1} \subseteq \mathcal{V}_i$ and $\mathcal{W}_i \subseteq \mathcal{W}_{i+1}$ and thus there exists a number $k \in \mathbb{N}$ such that

$$\mathcal{V} := \mathcal{V}_k = \mathcal{V}_{k+j} \qquad \text{and} \qquad \mathcal{W} := \mathcal{W}_k = \mathcal{W}_{k+j} \qquad \text{for all } j \in \mathbb{N}. \tag{2.8}$$

Following [31], the regularity of (E, A_1) implies $\dim(\mathcal{V}) = n_{x,\mathrm{d}}$ and $\dim(\mathcal{W}) = n_{x,\mathrm{a}}$ and for any matrices $V \in \mathbb{F}^{n_x \times n_{x,\mathrm{d}}}$ and $W \in \mathbb{F}^{n_x \times n_{x,\mathrm{a}}}$ that satisfy $\mathrm{im}(V) = \mathcal{V}$ and $\mathrm{im}(W) = \mathcal{W}$, the matrices

$$S := \begin{bmatrix} EV & A_1 W \end{bmatrix}^{-1} \qquad \text{and} \qquad T := \begin{bmatrix} V & W \end{bmatrix} \tag{2.9}$$

transform (E, A) into quasi-Weierstraß form (2.6). As a consequence (cf. [31, Remark 2.7]), we obtain

$$A_1 V = EVJ \qquad \text{and} \qquad EW = A_1 WN. \tag{2.10}$$

The next result shows that the converse direction is also true, i.e., that if $S, T \in \mathrm{GL}_{n_x}(\mathbb{F})$ transform the matrix pencil (E, A_1) into quasi-Weierstraß form, then S, T are of the form (2.9) with $\mathrm{im}(V) = \mathcal{V}$ and $\mathrm{im}(W) = \mathcal{W}$.

Proposition 2.11. *Consider a regular matrix pencil (E, A_1) and matrices $S_i, T_i \in \mathrm{GL}_{n_x}(\mathbb{F})$ for $i = 1, 2$ that transform (E, A_1) into quasi-Weierstraß form, that is*

$$S_i E T_i = \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & N_i \end{bmatrix} \quad \textit{and} \quad S_i A_1 T_i = \begin{bmatrix} J_i & 0 \\ 0 & I_{n_{x,\mathrm{a}}} \end{bmatrix} \quad \textit{for } i = 1, 2, \tag{2.11}$$

where $N_1, N_2 \in \mathbb{F}^{n_{x,\mathrm{a}} \times n_{x,\mathrm{a}}}$ are nilpotent and $J_1, J_2 \in \mathbb{F}^{n_{x,\mathrm{d}} \times n_{x,\mathrm{d}}}$. Then there exist matrices $P \in \mathrm{GL}_{n_{x,\mathrm{d}}}(\mathbb{F})$ and $Q \in \mathrm{GL}_{n_{x,\mathrm{a}}}(\mathbb{F})$ such that

$$S_2 = \begin{bmatrix} P^{-1} & 0 \\ 0 & Q^{-1} \end{bmatrix} S_1, \qquad T_2 = T_1 \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix}, \qquad J_2 = P^{-1} J_1 P, \qquad \textit{and} \qquad N_2 = Q^{-1} N_1 Q.$$

Proof. Partition $S_i = \begin{bmatrix} X_i & Y_i \end{bmatrix}^{-1}$ and $T_i = \begin{bmatrix} V_i & W_i \end{bmatrix}$ with $X_i, V_i \in \mathbb{F}^{n_x \times n_{x,\mathrm{d}}}$ and $Y_i, W_i \in \mathbb{F}^{n_x \times n_{x,\mathrm{a}}}$. We observe for $i = 1, 2$

$$\begin{bmatrix} EV_i & EW_i \end{bmatrix} = ET_i = S_i^{-1} \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & N_i \end{bmatrix} = \begin{bmatrix} X_i & Y_i \end{bmatrix} \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & N_i \end{bmatrix} = \begin{bmatrix} X_i & Y_i N_i \end{bmatrix} \qquad \text{and}$$

$$\begin{bmatrix} A_1 V_i & A_1 W_i \end{bmatrix} = A_1 T_i = S_i^{-1} \begin{bmatrix} J_i & 0 \\ 0 & I_{n_{x,\mathrm{a}}} \end{bmatrix} = \begin{bmatrix} X_i & Y_i \end{bmatrix} \begin{bmatrix} J_i & 0 \\ 0 & I_{n_{x,\mathrm{a}}} \end{bmatrix} = \begin{bmatrix} X_i J_i & Y_i \end{bmatrix}$$

and therefore $S_i^{-1} = \begin{bmatrix} EV_i & A_1 W_i \end{bmatrix}$ and $A_1 V_i = EV_i J_i$. We immediately observe $A_1 \operatorname{im}(V_i) \subseteq E \operatorname{im}(V_i)$. Thus [31, Proposition 2.13] together with $n_{x,\mathrm{d}} = \dim(\operatorname{im}(V_i))$ implies $\operatorname{im}(V_1) = \operatorname{im}(V_2)$. Hence there exists a matrix $P \in \mathrm{GL}_{n_{x,\mathrm{d}}}(\mathbb{F})$ with $V_2 = V_1 P$. Since T_1 is nonsingular, there exist matrices $\widetilde{P} \in \mathbb{F}^{n_x \times n_{x,\mathrm{d}}}$ and $Q \in \mathbb{F}^{n_x \times n_{x,\mathrm{a}}}$ with

$$T_2 = T_1 \begin{bmatrix} P & \widetilde{P} \\ 0 & Q \end{bmatrix} \qquad \text{and} \qquad S_2^{-1} = S_1^{-1} \begin{bmatrix} P & \widetilde{P} \\ 0 & Q \end{bmatrix}.$$

In particular, we have $Q \in \mathrm{GL}_{n_{x,\mathrm{a}}}(\mathbb{F})$. We obtain

$$\begin{aligned} S_1^{-1} \begin{bmatrix} P & \widetilde{P} N_2 \\ 0 & Q N_2 \end{bmatrix} &= S_1^{-1} \begin{bmatrix} P & \widetilde{P} \\ 0 & Q \end{bmatrix} \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & N_2 \end{bmatrix} = S_2^{-1} \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & N_2 \end{bmatrix} = ET_2 \\ &= ET_1 \begin{bmatrix} P & \widetilde{P} \\ 0 & Q \end{bmatrix} = S_1^{-1} \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & N_1 \end{bmatrix} \begin{bmatrix} P & \widetilde{P} \\ 0 & Q \end{bmatrix} = S_1^{-1} \begin{bmatrix} P & \widetilde{P} \\ () & N_1 Q \end{bmatrix}. \end{aligned}$$

Hence $\widetilde{P} = \widetilde{P} N_2$ and $N_2 = Q^{-1} N_1 Q$. As a consequence of the first equation and the fact that N_2 is nilpotent, we deduce $\widetilde{P} = 0$. It remains to show that $J_2 = P^{-1} J_1 P$ holds. This follows from

$$\begin{aligned} S_1^{-1} \begin{bmatrix} P J_2 & 0 \\ 0 & Q \end{bmatrix} &= S_2^{-2} \begin{bmatrix} J_2 & 0 \\ 0 & I_{n_{x,\mathrm{a}}} \end{bmatrix} = ET_2 = ET_1 \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \\ &= S_1^{-1} \begin{bmatrix} J_1 & 0 \\ 0 & I_{n_{x,\mathrm{a}}} \end{bmatrix} \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} = S_1^{-1} \begin{bmatrix} J_1 P & 0 \\ 0 & Q \end{bmatrix}. \end{aligned}$$

■

As a consequence of Proposition 2.11, the index of nilpotency of N in the quasi-Weierstraß form (4.5) is independent of the choice of the matrices S and T, which motivates the following definition (cf. [127]).

Definition 2.12 (Index of a regular matrix pencil)**.** Let (E, A_1) be a regular matrix pencil and let $N \in \mathbb{F}^{n_{x,\mathrm{a}} \times n_{x,\mathrm{a}}}$ denote the nilpotent matrix with index of nilpotency ν of the quasi-Weierstraß form from Theorem 2.9. Then the number

$$\operatorname{ind}(E, A_1) := \begin{cases} \nu, & \text{if } n_{x,\mathrm{a}} > 0, \\ 0, & \text{otherwise,} \end{cases}$$

is called the *index of the pencil* (E, A_1).

Theorem 2.9 allows us to decouple the DAE (2.5). In more detail, let $S, T \in \mathrm{GL}_{n_x}(\mathbb{F})$ be matrices that transform the pencil (E, A_1) in quasi-Weierstraß form (2.6). Since both matrices are nonsingular, we obtain a one-to-one correspondence between solutions of (2.3a) and solutions of

$$\dot{v} = Jv + \tilde{g}, \tag{2.12a}$$

$$N\dot{w} = w + \tilde{h}, \tag{2.12b}$$

with

$$\begin{bmatrix} v \\ w \end{bmatrix} := T^{-1}x \qquad \text{and} \qquad \begin{bmatrix} \tilde{g} \\ \tilde{h} \end{bmatrix} := S\tilde{f}.$$

While (2.12a) is a standard ordinary differential equation (ODE) in v that can be solved with the Duhamel integral, the so-called *fast subsystem* (2.12b) has the solution

$$w = -\sum_{k=0}^{\nu-1} N^k \tilde{h}^{(k)} \tag{2.13}$$

and hence the function $\tilde{h}$ must be ν times continuously differentiable for a classical solution to exist (cf. [127]). In addition, a consistent initial value $w(0)$ must satisfy equation (2.13). Similar to [200], we define the matrices

$$\begin{aligned} A^{\mathrm{diff}} &:= T \begin{bmatrix} J & 0 \\ 0 & 0 \end{bmatrix} T^{-1}, & A^{\mathrm{con}} &:= T \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & 0 \end{bmatrix} T^{-1}, \\ C_0 &:= T \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & 0 \end{bmatrix} S, & C_k &:= -T \begin{bmatrix} 0 & 0 \\ 0 & N^{k-1} \end{bmatrix} S \end{aligned} \tag{2.14}$$

for $k = 1, \ldots, \operatorname{ind}(E, A_1)$. As a consequence of Proposition 2.11 we notice that the matrices defined in (2.14) do not depend on the choice of the matrices S and T, see also [205].

Lemma 2.13. *Assume that the matrix pencil (E, A_1) is regular. Then the matrices A^{diff}, A^{con}, and C_k for $k = 0, 1, \ldots, \operatorname{ind}(E, A_1)$ defined in* (2.14) *do not depend on the matrices S, T that transform (E, A_1) into quasi-Weierstraß form* (2.6).

Proof. Consider matrices $S_i, T_i \in \mathrm{GL}_{n_x}(\mathbb{F})$ for $i = 1, 2$ that transform (E, A_1) into quasi-Weierstraß form, i.e., that satisfy (2.11). According to Proposition 2.11 there exist matrices $P \in \mathrm{GL}_{n_{x,\mathrm{d}}}(\mathbb{F})$ and $Q \in \mathrm{GL}_{n_{x,\mathrm{a}}}(\mathbb{F})$ such that

$$S_2 = \begin{bmatrix} P^{-1} & 0 \\ 0 & Q^{-1} \end{bmatrix} S_1, \qquad T_2 = T_1 \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix}, \qquad J_2 = P^{-1} J_1 P, \qquad \text{and} \qquad N_2 = Q^{-1} N_1 Q.$$

Thus

$$T_2 \begin{bmatrix} J_2 & 0 \\ 0 & 0 \end{bmatrix} T_2^{-1} = T_1 \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} J_2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P^{-1} & 0 \\ 0 & Q^{-1} \end{bmatrix} T_1^{-1} = T_1 \begin{bmatrix} PJ_2P^{-1} & 0 \\ 0 & 0 \end{bmatrix} T_1^{-1} = T_1 \begin{bmatrix} J_1 & 0 \\ 0 & 0 \end{bmatrix} T_1^{-1}.$$

The proof for the other matrices follows similarly. ■

Proposition 2.14. *Assume that the matrix pair* (E, A_1) *is regular and* $\tilde{f}$ *satisfies Assumption 2.5. Then any classical solution* x *of* (2.5) *fullfills the so called* underlying ODE

$$\dot{x} = A^{\text{diff}} x + \sum_{k=0}^{\operatorname{ind}(E,A_1)} C_k \tilde{f}^{(k)}. \tag{2.15}$$

Conversely, let x *be a classical solution of* (2.15). *Then* x *is a solution of* (2.5) *if and only if there exists* $s \in [0, t_{\mathrm{f}})$ *such that* $x(s)$ *satisfies*

$$x(s) = A^{\text{con}} x(s) + \sum_{k=1}^{\operatorname{ind}(E,A_1)} C_k \tilde{f}^{(k-1)}(s). \tag{2.16}$$

In this case (2.16) *is true for all* $s \in [0, t_{\mathrm{f}})$.

Proof. Let x be a classical solution of (2.5) and $S, T \in \mathrm{GL}_{n_x}(\mathbb{F})$ be matrices that satisfy (2.6) of the quasi-Weierstraß form and set $\nu := \operatorname{ind}(E, A_1)$. Differentiation of (2.13) yields

$$\begin{aligned}
\dot{x} &= T \begin{bmatrix} \dot{v} \\ \dot{w} \end{bmatrix} = T \begin{bmatrix} Jv + \tilde{g} \\ -\sum_{k=0}^{\nu-1} N^k \tilde{h}^{(k+1)} \end{bmatrix} \\
&= T \begin{bmatrix} J & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} + T \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{g} \\ \tilde{h} \end{bmatrix} - \sum_{k=1}^{\nu} T \begin{bmatrix} 0 & 0 \\ 0 & N^{k-1} \end{bmatrix} \begin{bmatrix} \tilde{g}^{(k)} \\ \tilde{h}^{(k)} \end{bmatrix} \\
&= A^{\text{diff}} x + \sum_{k=0}^{\nu} C_k \tilde{f}^{(k)}.
\end{aligned}$$

Conversely, let x be a classical solution of (2.15). Then for any $s \in [0, t_{\mathrm{f}})$ we have

$$x(t) = \mathrm{e}^{A^{\text{diff}}(t-s)} x(s) + \int_s^t \mathrm{e}^{A^{\text{diff}}(t-s-\tilde{t})} \sum_{k=0}^{\nu} C_k \tilde{f}^{(k)}\left(\tilde{t}\right) \mathrm{d}\tilde{t}. \tag{2.17}$$

Scaling (2.17) from the left by T^{-1} we obtain

$$w(t) = w(s) - \sum_{k=1}^{\nu} N^{k-1} \int_s^t \tilde{h}^{(k)}\left(\tilde{t}\right) \mathrm{d}\tilde{t} = w(s) - \sum_{k=0}^{\nu-1} N^k \tilde{h}^{(k)}(t) + \sum_{k=0}^{\nu-1} N^k \tilde{h}^{(k)}(s).$$

Suppose now that for a specific $s \in [0, t_{\mathrm{f}})$ the solution x satisfies (2.16), or equivalently (by scaling (2.16) from the left with T^{-1})

$$\begin{bmatrix} v(s) \\ w(s) \end{bmatrix} = \begin{bmatrix} v(s) \\ -\sum_{k=0}^{\nu-1} N^k \tilde{h}^{(k)}(s) \end{bmatrix}.$$

Together with (2.13) this implies that x is a solution of (2.5) and it is easy to see that (2.16) is satisfied for all $t \in [0, t_{\mathrm{f}})$. The remaining direction follows immediately from (2.13). ■

Setting $s = 0$ in the previous proposition yields the following requirement for an initial condition to be consistent.

Corollary 2.15. *Assume that the matrix pair (E, A_1) is regular and the inhomogeneity $\tilde{f}$ satisfies Assumption 2.5. Then the initial value $x(0)$ is consistent if and only if it satisfies the* consistency condition

$$x(0) = A^{\mathrm{con}} x(0) + \sum_{k=1}^{\operatorname{ind}(E,A_1)} C_k \tilde{f}^{(k-1)}(0). \tag{2.18}$$

In this case, the IVP (2.3) has a unique solution $x \in \mathscr{C}^\infty([0, t_\mathrm{f}); \mathbb{F}^{n_x})$.

2.2 Distributional solutions for inconsistent initial values

As we have clearly seen in the previous subsection, a classical solution does not exist for all initial conditions. Instead, a consistent initial condition has to satisfy the consistency condition (2.18) in Corollary 2.15. If the DAE under investigation is a result of applying the method of steps (2.3) to the DDAE (1.13), then it is a–priori not clear, if the initial value is consistent. Unfortunately, one cannot show that the initial value is always consistent. We have already seen a counterexample in Example 1.5. Thus, a general solution framework for DDAEs has to be able to deal with inconsistent initial values.

Inconsistent initial values appear also in other application areas, for instance when an electrical circuit is switched at a certain time [215]. As a consequence, a number of different approaches in the time and frequency domain have been proposed to deal with inconsistent initial values. For an overview we refer to [205]. All approaches have in common, that jumps or even Dirac impulses may occur in the solution, and hence a distributional solution space seems appropriate. Following [192], the space of test function

$$\mathscr{C}_0^\infty(\mathbb{R};\mathbb{R}) := \{ f \in \mathscr{C}^\infty(\mathbb{R};\mathbb{R}) \mid \operatorname{supp} f \text{ is bounded} \}$$

with $\operatorname{supp}(f) := \overline{\{x \in \mathbb{R} \mid f(x) \neq 0\}}$, can be equipped with a locally convex topology (see [117, § 12]), thus making it a topological space. The set of all linear and continuous maps from $\mathscr{C}_0^\infty(\mathbb{R};\mathbb{R})$ into the real numbers

$$\mathscr{D} := \{ f : \mathscr{C}_0^\infty(\mathbb{R};\mathbb{R}) \to \mathbb{R} \mid f \text{ is linear and continuous} \}, \tag{2.19}$$

i.e., the topological dual space of $\mathscr{C}_0^\infty(\mathbb{R};\mathbb{R})$, is called the space of *distributions.* Since the test functions are smooth and have compact support, we can define for any locally integrable function $f \in \mathscr{L}_{1,\mathrm{loc}}$ a distribution via

$$f_{\mathscr{D}} : \mathscr{C}_0^\infty(\mathbb{R};\mathbb{R}) \to \mathbb{R}, \qquad \varphi \mapsto \int_{-\infty}^{\infty} f(t)\varphi(t)\mathrm{d}t.$$

Consequently, the space $\mathscr{L}_{1,\mathrm{loc}}$ can be embedded into $\mathscr{D}$ via the injective homomorphism

$$\mathscr{L}_{1,\mathrm{loc}} \to \mathscr{D}, \qquad f \mapsto f_{\mathscr{D}}. \tag{2.20}$$

To describe the DAE (2.5) in a (yet to define) suitable distributional solution space, we need a derivative in $\mathscr{D}$. Following [117, § 19], we define the distributional derivative

$$\frac{\mathrm{d}}{\mathrm{d}t}\colon \mathscr{D}\to\mathscr{D}, \qquad f\mapsto\left(\frac{\mathrm{d}}{\mathrm{d}t}f\colon \mathscr{C}_0^\infty(\mathbb{R};\mathbb{R})\to\mathbb{R},\quad \varphi\mapsto -f\left(\frac{\mathrm{d}}{\mathrm{d}t}\varphi\right)\right). \tag{2.21}$$

Since $\frac{\mathrm{d}}{\mathrm{d}t}\varphi\in\mathscr{C}_0^\infty(\mathbb{R};\mathbb{R})$, this is indeed well-defined (see [117, Satz 19.1] for more details) and we immediately observe, that distributions are arbitrarily often differentiable. For notational convenience, we write

$$\dot{f}:=\frac{\mathrm{d}}{\mathrm{d}t}f \qquad \text{and} \qquad f^{(k)}:=\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^k f \qquad \text{for } f\in\mathscr{D}.$$

Note that we use the symbol $\frac{\mathrm{d}}{\mathrm{d}t}$ for the distributional derivative and the standard derivative. This is consistent, since for any differentiable (and thus locally integrable) function $f\colon\mathbb{R}\to\mathbb{R}$, we have

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}f\right)_{\mathscr{D}}=\frac{\mathrm{d}}{\mathrm{d}t}\left(f_{\mathscr{D}}\right).$$

Example 2.16. Consider for $\Omega\subseteq\mathbb{R}$ the indicator function

$$\mathbb{1}_\Omega\colon\mathbb{R}\to\mathbb{R},\qquad t\mapsto\begin{cases}1, & \text{if } t\in\Omega,\\ 0, & \text{otherwise.}\end{cases}$$

For any $s\in\mathbb{R}$ and $\varphi\in\mathscr{C}_0^\infty(\mathbb{R};\mathbb{R})$ we obtain

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\left(\left(\mathbb{1}_{[s,\infty)}\right)_{\mathscr{D}}\right)\right)(\varphi)=-\int_{-\infty}^{\infty}\mathbb{1}_{[s,\infty)}(t)\frac{\mathrm{d}}{\mathrm{d}t}\varphi(t)\mathrm{d}t=\varphi(s).$$

The distribution $\delta_s:=\frac{\mathrm{d}}{\mathrm{d}t}\left(\left(\mathbb{1}_{[s,\infty)}\right)_{\mathscr{D}}\right)$ is called the *Dirac impulse* at s. ♠

A generalization of $\mathscr{D}$ to vector-valued functions is straightforward, by defining

$$\mathscr{D}^k:=\left\{f=\begin{bmatrix}f_1 & \dots & f_k\end{bmatrix}^T \,\middle|\, f_i\in\mathscr{D} \text{ for } i=1,\dots,k\right\}$$

for any $k\in\mathbb{N}$. The multiplication of $f\in\mathscr{D}^k$ with a matrix $M\in\mathbb{R}^{p\times k}$ is then defined via

$$Mf\colon\mathscr{C}_0^\infty(\mathbb{R};\mathbb{R})\to\mathbb{R},\qquad \varphi\mapsto Mf(\varphi),$$

such that the DAE (2.5) can be interpreted as an equation in $\mathscr{D}^m$ with $x\in\mathscr{D}^{n_x}$ and $\tilde{f}\in\mathscr{D}^m$. However, embedding the DAE into a distributional framework does not resolve the issue of inconsistent initial conditions, since we cannot evaluate distributions at a point $t_0\in\mathbb{R}$. But even if we restrict the space of distributions such that the pointwise evalutation at certain points is well-defined, one can show, see [205], that solutions do not exist for arbitrary initial values. For instance, the trivial DAE $x=0$ posses the unique solution $x=0$ also in the distributional sense.

Instead, we assume that the DAE (2.5) only holds on $[0,\infty)$ (instead of the real axis) and the past, i.e., the behavior in the interval $(\infty,0)$, is prescribed as an initial trajectory. However, this requires us to define a distributional restriction to the interval $[0,\infty)$ and this is not possible for general distributions [203, Lemma 2.2.3]. This problem can be resolved by considering the space of impulsive-smooth distributions [83] or by the slightly bigger space of piecewise-smooth distributions [203]. The latter is also suitable for studying the DDAE (1.15), therefore we will use this space in the following as the underlying solution space for (2.5) and (1.15).

Definition 2.17 (Piecwise-smooth distributions)**.**

(i) A function $\alpha\colon \mathbb{R}\to\mathbb{R}$ is called *piecewise-smooth* if, and only if, there exists a family of real numbers $\{t_i\in\mathbb{R} \mid i\in\mathbb{Z}\}$ with $t_i < t_{i+1}$ for all $i\in\mathbb{Z}$ and $t_{\pm k}\to\pm\infty$ as $k\to\infty$ and smooth functions $\alpha_i\in\mathscr{C}^\infty(\mathbb{R};\mathbb{R})$ such that

$$\alpha = \sum_{i\in\mathbb{Z}} \mathbb{1}_{[t_i,t_{i+1})}\alpha_i.$$

The space of *piecewise-smooth functions* is defined as

$$\mathscr{C}^\infty_{\mathrm{pw}}(\mathbb{R};\mathbb{R}) := \{\alpha\colon\mathbb{R}\to\mathbb{R} \mid \alpha \text{ is piecewise-smooth}\}.$$

(ii) The space of *piecewise-smooth distributions* is defined as

$$\mathscr{D}_{\mathrm{pw}\mathscr{C}^\infty} := \left\{ \alpha_{\mathscr{D}} + \sum_{s\in S} D_s \;\middle|\; \begin{array}{l} \alpha\in\mathscr{C}^\infty_{\mathrm{pw}}(\mathbb{R};\mathbb{R}),\ S \text{ is a discrete set, and} \\ D_s\in\operatorname{span}\{\delta_s,\dot{\delta}_s,\ddot{\delta}_s,\ldots\} \text{ for } s\in S \end{array} \right\},$$

i.e., a piecewise-smooth distribution is the sum of a piecewise-smooth function and linear combinations of Dirac impulses and their derivatives at finitely many time instants in each compact interval.

For piecewise-smooth distributions the restriction

$$\mathscr{D}_{\mathrm{pw}\mathscr{C}^\infty}\times\mathscr{P}(\mathbb{R})\to\mathscr{D}_{\mathrm{pw}\mathscr{C}^\infty}, \qquad (f=\alpha_{\mathscr{D}}+\sum_{s\in S}D_s,\Omega)\mapsto f_\Omega := (\mathbb{1}_\Omega\alpha)_{\mathscr{D}} + \sum_{s\in S\cap\Omega} D_s$$

is well defined. Moreover, one can show (cf. [203, 204]) that each piecewise-smooth distribution $f\in\mathscr{D}_{\mathrm{pw}\mathscr{C}^\infty}$ posses a derivative and an anti-derivative in $\mathscr{D}_{\mathrm{pw}\mathscr{C}^\infty}$. It is important to note that the restriction operator and the distributional derivative operator do not commute. Instead, we have the following result.

Lemma 2.18 ([204, Proposition 12])**.** *For all* $-\infty\le t_1\le t_2\le\infty$ *and* $f=\alpha_{\mathscr{D}}+\sum_{s\in S}D_s\in\mathscr{D}_{\mathrm{pw}\mathscr{C}^\infty}$ *we have*

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(f_{[t_1,t_2)}\right) &= \left(\frac{\mathrm{d}}{\mathrm{d}t}f\right)_{[t_1,t_2)} + f(t_1^-)\delta_{t_1} - f(t_2^-)\delta_{t_2},\\
\frac{\mathrm{d}}{\mathrm{d}t}\left(f_{(t_1,t_2)}\right) &= \left(\frac{\mathrm{d}}{\mathrm{d}t}f\right)_{(t_1,t_2)} + f(t_1^+)\delta_{t_1} - f(t_2^-)\delta_{t_2},\\
\frac{\mathrm{d}}{\mathrm{d}t}\left(f_{(t_1,t_2]}\right) &= \left(\frac{\mathrm{d}}{\mathrm{d}t}f\right)_{[t_1,t_2)} + f(t_1^+)\delta_{t_1} - f(t_2^+)\delta_{t_2},\\
\frac{\mathrm{d}}{\mathrm{d}t}\left(f_{[t_1,t_2]}\right) &= \left(\frac{\mathrm{d}}{\mathrm{d}t}f\right)_{[t_1,t_2]} + f(t_1^-)\delta_{t_1} - f(t_2^+)\delta_{t_2},
\end{aligned}$$

where $\delta_{\pm\infty} = 0$ *and the left and right sided evaluation at a point* $t \in \mathbb{R}$ *are defined as*

$$f(t^-) := \lim_{h\searrow 0} \alpha(t-h) \qquad \textit{and} \qquad f(t^+) := \lim_{h\searrow 0} \alpha(t+h) = \alpha(t).$$

For further results on the space of piecewise-smooth distributions and its relation to other distributional solution concepts for DAEs we refer to [203, 204]. It is now possible to state an existence and uniqueness result for regular DAEs with possible inconsistent initial values. More precisely, we can interpret the DAE (2.5) in the space of piecewise-smooth distributions as the *initial trajectory problem* (ITP)

$$\begin{aligned} x_{(-\infty,t_0)} &= x^0_{(-\infty,t_0)}, \\ (E\dot{x})_{[t_0,\infty)} &= (A_1 x + \tilde{f})_{[t_0,\infty)}, \end{aligned} \tag{2.22}$$

with initial trajectory $x^0 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ and inhomogeneity $\tilde{f} \in \mathscr{D}^{m}_{\mathrm{pw}\mathscr{C}^\infty}$ and arbitrary $t_0 \in \mathbb{R}$.

Theorem 2.19 ([203, Theorem 3.5.2])**.** *Consider the ITP* (2.22) *with initial trajectory* $x^0 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ *and inhomogeneity* $\tilde{f} \in \mathscr{D}^{m}_{\mathrm{pw}\mathscr{C}^\infty}$. *If the matrix pair* (E, A_1) *is regular, then the ITP* (2.22) *has a unique solution* $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$.

In the study of DDAEs it turns out that even in the LTI setting, the DDAE (1.15) may contain higher-order differential equations and thus it is important to study also higher-order DAEs. To this end, consider a polynomial matrix $\mathscr{P}(s) \in \mathbb{R}[s]^{m\times n_x}$, i.e.,

$$\mathscr{P}(s) = \sum_{j=0}^{p} P_j s^j \qquad \text{with matrices } P_i \in \mathbb{F}^{m\times n_x} \text{ for } j = 0, 1, \ldots, p.$$

For a given polynomial matrix $\mathscr{P}(s) \in \mathbb{R}[s]^{m\times n_x}$ we consider the generalization of the DAE (2.5) given by the polynomial DAE

$$\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x = f. \tag{2.23}$$

Notice that the DAE (2.5) can be recast in the form (2.23) by introducing the polynomial matrix $\mathscr{P}(s) = Es - A_1$. Vice versa, by introducing new variables, we can easily recast the polynomial DAE (2.23) into the matrix form (2.5) and thus immediately obtain the following result (see e.g. the second part of the proof of Theorem 7 in [209] or the first part of the proof of Corollary 9 in [206]).

Corollary 2.20. *For given polynomial matrix* $\mathscr{P}(s) \in \mathbb{R}[s]^{n_x\times n_x}$, *initial trajectory* $x^0 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$, *inhomogeneity* $\tilde{f} \in \mathscr{D}^{m}_{\mathrm{pw}\mathscr{C}^\infty}$, *and initial time point* $t_0 \in \mathbb{R}$ *consider the ITP*

$$\begin{aligned} x_{(-\infty,t_0)} &= x^0_{(-\infty,t_0)}, \\ \left(\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x\right)_{[t_0,\infty)} &= (\tilde{f})_{[t_0,\infty)}. \end{aligned} \tag{2.24}$$

If $\det(\mathscr{P}(s)) \in \mathbb{R}[s] \setminus \{0\}$, *then the ITP* (2.24) *has a unique solution* $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$.

Proof. Let $\mathscr{P}(s) = \sum_{j=0}^{p} P_j s^j$. Then a standard companion form linearization of (2.23) yields the DAE

$$\mathscr{E}\dot{z} = \mathscr{A}z + \mathscr{F} \tag{2.25}$$

with $\mathscr{E}, \mathscr{A} \in \mathbb{R}^{pn_x \times pn_x}$, given by

$$\mathscr{E} = \begin{bmatrix} P_p & 0 & \cdots & 0 \\ 0 & I_n & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_n \end{bmatrix}, \quad \mathscr{A} = \begin{bmatrix} -P_{p-1} & -P_{p-2} & \cdots & -P_0 \\ I_n & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I_n & 0 \end{bmatrix}, \quad z = \begin{bmatrix} \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{p-1} x \\ \vdots \\ \frac{\mathrm{d}}{\mathrm{d}t}x \\ x \end{bmatrix}, \quad \mathscr{F} = \begin{bmatrix} f \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Note that there exists (cf. [142]) unimodular matrix polynomials $\mathscr{R}(s), \mathscr{S}(s) \in \mathbb{R}[s]^{pn_x \times pn_x}$ with

$$\mathscr{R}(s)(s\mathscr{E} - \mathscr{A})\mathscr{S}(s) = \begin{bmatrix} \mathscr{P}(s) & 0 \\ 0 & I_{(p-1)n} \end{bmatrix}.$$

Hereby, $\mathscr{R}$ and $\mathscr{S}$ are given as

$$\mathscr{R}(s) = \begin{bmatrix} I_\ell & sP_p + P_{p-1} & \cdots & \sum_{j=1}^{p} s^{j-1} P_j \\ & & & -I_n \\ & & \iddots & \\ & -I_n & & \end{bmatrix}, \qquad \mathscr{S}(s) = \begin{bmatrix} s^{p-1} I_n & \cdots & sI_n & I_n \\ \vdots & \iddots & \iddots & \\ sI_n & \iddots & & \\ I_n & & & \end{bmatrix}.$$

The proof now follows by the observation that there exists a constant $c \neq 0$ with

$$0 \not\equiv \det(\mathscr{P}(s)) = c \det(s\mathscr{E} - \mathscr{A})$$

and application of Theorem 2.19 to (2.25). ■

Remark 2.21. The initial trajectory $x^0_{(-\infty,t_0)}$ in (2.24) not only specifies the state $x_{(-\infty,t_0)}$ but also its (distributional) derivatives and thus providing the initial trajectories for the higher-order differential operator $\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)$ in (2.23). It should be noted that in general the standard companion form linearization in (2.25) may introduce additional smoothness requirements for the forcing term f (cf. [154, 187]) and instead a so-called *trimmed linearization* [47] should be used if we consider a classical solution concept. One of the main issues with higher-order differential equations is that there is no simple canonical form under strong equivalence if $\deg\mathscr{P} \geq 2$ [202]. ♣

Although in principle it is possible to rewrite the higher-order DAE (2.23) as a first-order DAE by introducing new variables, it is sometimes more efficient to work directly with the higher-order system. Consequently, we need a generalization of Lemma 2.18 to higher-order differential operators. For simplicity we consider only the restriction to the time intervals $(-\infty, 0)$ and $[0, \infty)$. Let $f \in \mathscr{D}^n_{\mathrm{pw}\mathscr{C}^\infty}$.

Repeated application of Lemma 2.18 yield

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^k (f_{(-\infty,0)}) = \left(f^{(k)}\right)_{(-\infty,0)} - \sum_{j=0}^{k-1} f^{(j)}(0^-)\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{k-1-j}\delta_0,$$
$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^k (f_{[0,\infty)}) = \left(f^{(k)}\right)_{[0,\infty)} + \sum_{j=0}^{k-1} f^{(j)}(0^-)\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{k-1-j}\delta_0$$

for $k \in \mathbb{N}$. For $\mathscr{P}(s) = \sum_{k=0}^{p} P_k s^k \in \mathbb{R}[s]^{m\times n_x}$ define $\mathscr{P}^{[0]}(s) := \mathscr{P}(s)$ and recursively

$$\mathscr{P}^{[i]}(s) := \frac{1}{s}\left(\mathscr{P}^{[i-1]}(s) - \mathscr{P}^{[i-1]}(0)\right) \in \mathbb{R}[s]^{m\times n_x},$$

i.e., $\mathscr{P}^{[i]} = \sum_{k=i}^{p} P_k s^{k-i}$ for $i = 0,1,\dots,k$. Then, we have proven the following generalization of Lemma 2.18.

Lemma 2.22. *Let* $f \in \mathscr{D}^n_{\mathrm{pw}\mathscr{C}^\infty}$ *and* $\mathscr{P}(s) \in \mathbb{R}[s]^{\ell\times n}$ *with* $\deg(\mathscr{P}) = d \geq 0$. *Then*

$$\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)(F_{(-\infty,0)}) = \left(\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)F\right)_{(-\infty,0)} - \sum_{j=0}^{d-1}\left(\mathscr{P}^{[j]}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)F\right)(0^-)\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{d-1-j}\delta_0,$$
$$\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)(F_{[0,\infty)}) = \left(\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)F\right)_{[0,\infty)} + \sum_{j=0}^{d-1}\left(\mathscr{P}^{[j]}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)F\right)(0^-)\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{d-1-j}\delta_0.$$

A crucial difference in the study of the polynomial DAE (2.23) in contrast to the DAE (2.5) is the fact that the entries of the polynomial matrices are elements of a ring and not a field. Although the ring of polynomials is embedded in the field of rational functions and thus it is straightforward to use concepts such as the rank, we have to ensure that whenever we substitute $\frac{\mathrm{d}}{\mathrm{d}t}$ for the indeterminate s, we only operate in the ring of polynomials. As a consequence, we cannot use nonsingular matrices (as for instance in the quasi-Weierstraß form), but have to restrict ourselves to unimodular matrices, i.e., polynomial matrices that are nonsingular and the inverse matrix is also a polynomial matrix (see also the proof of Corollary 2.20). The usage of unimodular matrices in the context of DAEs is not new: see for instance the work [116], where the authors construct a unimodular matrix to perform index-reduction. An important tool to study polynomial matrices is the *Smith canonical form.*

Theorem 2.23 (Smith canonical form, [118, Thm. 1.8.1])**.** *Let* $\mathscr{P}(s) \in \mathbb{F}[s]^{m\times n_x}$. *Then there exists unimodular matrices* $\mathscr{S}(s) \in \mathbb{F}[s]^{m\times m}$, $\mathscr{T}(s) \in \mathbb{F}[s]^{n_x\times n_x}$ *such that*

$$\mathscr{S}(s)\mathscr{P}(s)\mathscr{T}(s) = \begin{bmatrix} p_1(s) & & & & & \\ & \ddots & & & & \\ & & p_r(s) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}, \tag{2.26}$$

where $r := \operatorname{rank}_{\mathbb{F}[s]}(\mathscr{P}(s))$, $p_i(s) \in \mathbb{F}[s] \setminus \{0\}$, *and* $p_i(s)$ *divides* $p_{i+1}(s)$ *for* $i = 1, \ldots, r-1$.

A direct consequence of the Smith canonical form is that we can perform a rank revealing row-compression with a unimodular matrix, i.e., for $\mathscr{P}(s) \in \mathbb{F}[s]^{m \times n_x}$ with $r := \operatorname{rank}_{\mathbb{F}[s]}(\mathscr{P}(s))$, there exists a unimodular matrix $\mathscr{U}(s) \in \mathbb{F}[s]^{m \times m}$ and a polynomial matrix $\mathscr{P}_1(s) \in \mathbb{F}[s]^{r \times n_x}$ with $\operatorname{rank}_{\mathbb{F}[s]}(\mathscr{P}_1(s)) = r$ such that

$$\mathscr{U}(s)\mathscr{P}(s) = \begin{bmatrix} \mathscr{P}_1(s) \\ 0 \end{bmatrix}.$$

Similarly as for DAEs, we can study the properties of linear time-invariant DDAEs by analyzing pairs of matrix polynomials $(\mathscr{P}(s), \mathscr{Q}(s)) \in \left(\mathbb{R}[s]^{m \times n_x}\right)^2$. Although there is no equivalent to the Weierstraß canonical form for pairs of matrix polynomials, one can still use the condensed form approach from [46] to construct a condensed form for $(\mathscr{P}(s), \mathscr{Q}(s))$.

Theorem 2.24 ([96, Thm. 1])**.** *For any pair of polynomial matrices* $(\mathscr{P}(s), \mathscr{Q}(s)) \in (\mathbb{R}[s]^{m \times n_x})^2$ *there exists unimodular matrix polynomials* $\mathscr{U}(s) \in \mathbb{R}[s]^{m \times m}$ *and* $\mathscr{V}(s) \in \mathbb{R}[s]^{n_x \times n_x}$ *such that*

$$\mathscr{U}(s)\mathscr{P}(s)\mathscr{V}(s) = \begin{bmatrix} \widehat{\mathscr{P}}_{11}(s) & 0 & \widehat{\mathscr{P}}_{13}(s) \\ 0 & 0 & \widehat{\mathscr{P}}_{23}(s) \\ 0 & 0 & \widehat{\mathscr{P}}_{33}(s) \end{bmatrix}, \tag{2.27a}$$

$$\mathscr{U}(s)\mathscr{Q}(s)\mathscr{V}(s) = \begin{bmatrix} \widehat{\mathscr{Q}}_{11}(s) & \widehat{\mathscr{Q}}_{12}(s) & \widehat{\mathscr{Q}}_{13}(s) \\ 0 & 0 & \widehat{\mathscr{Q}}_{23}(s) \\ 0 & 0 & \widehat{\mathscr{Q}}_{33}(s) \end{bmatrix}, \tag{2.27b}$$

where $\widehat{\mathscr{P}}_{11}$ *is a nonsingular diagonal matrix,* $\widehat{\mathscr{P}}_{23}, \widehat{\mathscr{P}}_{33}, \widehat{\mathscr{Q}}_{33}$ *are block upper triangular matrices with zero diagonal blocks and* $\widehat{\mathscr{Q}}_{23}$ *is a nonsingular block upper triangular matrix.*

2.3 Strangeness-index for nonlinear DAEs

If F in (2.1a) is nonlinear, then the equation that we have to solve within the method of steps takes the form of a nonlinear DAE

$$F(t, x(t), \dot{x}(t), u(t)) = 0, \tag{2.28a}$$

where, as before, $x(t) \in \mathbb{R}^{n_x}$ and $u(t) \in \mathbb{R}^m$ denote, respectively, the *state* and *control* of the system, which is posed on the (compact) time interval $\mathbb{I} := [0, T]$. By abuse of notation, we use F in this section to denote the DAE (2.28a) and not the DDAE (1.13). The function

$$F \colon \mathbb{I} \times \mathbb{D}_x \times \mathbb{D}_{\dot{x}} \times \mathbb{D}_u \to \mathbb{R}^m$$

with open sets $\mathbb{D}_x, \mathbb{D}_{\dot{x}} \subseteq \mathbb{R}^{n_x}$, $\mathbb{D}_u \subset \mathbb{R}^{n_u}$ is assumed to be sufficiently smooth. The DAE (2.28a) is equipped with the initial condition

$$x(0) = x_0. \tag{2.28b}$$

for some $x_0 \in \mathbb{R}^{n_x}$. As for linear DAEs it is well-known that in general we cannot expect a unique solution if $m \neq n_x$ (cf. [127]). We therefore restrict ourselves to the case $m = n_x$, since one of the main goals within this thesis is to establish existence and uniqueness-results.

Definition 2.25. A function $x \in \mathscr{C}^1(\mathbb{I}, \mathbb{R}^{n_x})$ is called a *(classical) solution* of (2.28a) if x satisfies (2.28a) pointwise. An initial value $x_0 \in \mathbb{R}^{n_x}$ is called *consistent* if for a given control u, the associated IVP (2.28) has at least one solution. The DAE (2.28a) is called *regular*, if for every sufficiently smooth input u there exists a consistent initial value and for every consistent initial value, the solution of the ITP (2.28) is unique.

The control problem (2.28a) is often studied in the behavior framework [172], see for instance [55, 125]. Hereby, a new variable $\xi = [x, u]$ is introduced that includes the state and control variable such that the problem is reduced to the analysis of an underdetermined DAE [125], i. e., the meaning of the variables is not distinguished any more. One big advantage of this formalism is that the analysis determines the free variables in the system, which might not be the original control variables, and hence need to be reinterpreted. Since our main goal is to study the IVP (2.28) with a prescribed input function u this viewpoint is not possible. For given u we can study the restricted problem

$$\widetilde{F}(t, x(t), \dot{x}(t)) = 0, \qquad x(t_0) = x_0, \tag{2.29}$$

with $\widetilde{F}(t, x, \dot{x}) = F(t, x, \dot{x}, u)$.

If the partial derivative $\frac{\partial}{\partial \dot{x}} \widetilde{F}$ is singular, then the solution x of (2.29) may depend on derivatives of $\widetilde{F}$. The difficulties arising with these differentations are classified by so called *index* concepts (cf. [153] for a survey). In this paper, we make use of the *strangeness index* concept [127], which is – roughly speaking – a generalization of the *differentiation index* [42] to under- and overdetermined systems. The advantage of the strangeness index is that it preserves the algebraic constraints in the system, which in turn prevents numerical methods to drift away from the solution manifold [126]. The strangeness index is based on the *derivative array* [51] of level ℓ, defined as

$$\widetilde{\mathscr{D}}_\ell\left(t, x, \eta\right) := \begin{bmatrix} \widetilde{F}(t, x, \dot{x}) \\ \frac{\mathrm{d}}{\mathrm{d}t} \widetilde{F}(t, x, \dot{x}) \\ \vdots \\ \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^\ell \widetilde{F}(t, x, \dot{x}) \end{bmatrix} \in \mathbb{R}^{(\ell+1)n_x} \qquad \text{with } \eta := \begin{bmatrix} \dot{x} \\ \ddot{x} \\ \vdots \\ x^{(\ell+1)} \end{bmatrix}. \tag{2.30}$$

Since it is a-priori not clear, that a solution exists, we need to assume that the set

$$\widetilde{\mathscr{M}}_\ell := \left\{ \left(t, x, \eta\right) \in \mathbb{R}^{(\ell+2)n_x+1} \;\middle|\; \widetilde{\mathscr{D}}_\ell\left(t, x, \eta\right) = 0 \right\}$$

is nonempty. Similarly as in the theory for linear DAEs, we are interested in determining all algebraic constraints. In principal, the number of algebraic constraints may vary due to the nonlinearity of $\widetilde{F}$. To exclude this case we have to impose some constant rank assumptions. Following [124], we introduce the Jacobians

$$\begin{aligned} \widetilde{\mathscr{E}}_\ell(t, x, \dot{x}, \ldots, x^{(\ell+1)}) &:= \begin{bmatrix} \frac{\partial \widetilde{\mathscr{D}}_\ell}{\partial \dot{x}} & \ldots & \frac{\partial \widetilde{\mathscr{D}}_\ell}{\partial x^{(\ell+1)}} \end{bmatrix} \left(t, x, \dot{x}, \ldots, x^{(\ell+1)}\right) \in \mathbb{R}^{(\ell+1)n_x \times (\ell+1)n_x}, \\ \widetilde{\mathscr{A}}_\ell(t, x, \dot{x}, \ldots, x^{(\ell+1)}) &:= -\begin{bmatrix} \frac{\partial \widetilde{\mathscr{D}}_\ell}{\partial x} & 0 & \ldots & 0 \end{bmatrix} \left(t, x, \dot{x}, \ldots, x^{(\ell+1)}\right) \in \mathbb{R}^{(\ell+1)n_x \times (\ell+1)n_x}. \end{aligned}$$

In order to determine all algebraic equations that are encoded in (2.28a), we assume the following.

Assumption 2.26. *There exist integers μ and a such that the set*

$$\widetilde{\mathscr{M}_\mu} := \left\{ (t,x,\eta) \in \mathbb{R}^{(\mu+2)n_x+1} \;\middle|\; \widetilde{\mathscr{D}}_\mu(t,x,\eta) = 0 \right\}$$

associated with $\widetilde{F}$ is nonempty and such that for every $(t_0, x_0, \eta_0) \in \widetilde{\mathscr{M}_\mu}$, there exists a (sufficiently small) neighborhood $\widetilde{\mathscr{U}}$. Moreover we have $\operatorname{rank}(\widetilde{\mathscr{E}}_\mu) = (\mu+1)n_x - a$ *on* $\widetilde{\mathscr{M}_\mu} \cap \widetilde{\mathscr{U}}$*.*

The constant rank assumption allows us to define (via the smooth singular value decomposition [43], see also [127]) a matrix-valued function $\widetilde{Z}_\mathrm{A}$ of size $(\mu+1)n_x \times a$ and pointwise maximal rank that satisfies $\widetilde{Z}_\mathrm{A}^T \widetilde{\mathscr{E}}_\mu = 0$. The (linearized) algebraic equations are thus encoded in the matrix

$$\left(\widetilde{Z}_\mathrm{A}^T \frac{\partial \widetilde{\mathscr{D}}_\mu}{\partial x} \right)(t,x,\eta) \in \mathbb{R}^{a \times n_x}. \tag{2.31}$$

To ensure that the problem is regular, we need to be able to solve the algebraic equations for a unknowns, requiring that the matrix in (2.31) has rank a. We thus assume the following.

Assumption 2.27. *Let Assumption 2.26 hold, let $\widetilde{Z}_\mathrm{A}$ be constructed as above, and assume*

$$\operatorname{rank}\left(\left(\widetilde{Z}_\mathrm{A}^T \frac{\partial \widetilde{\mathscr{D}}_\mu}{\partial x} \right)(t,x,\eta) \right) = a \tag{2.32}$$

for all $(t,x,\eta) \in \widetilde{\mathscr{M}_\mu} \cap \widetilde{\mathscr{U}}$*.*

In view of the Weierstraß canonical form (cf. Theorem 2.9), we need to ensure that we have $d := n_x - a$ differential equations for the remaining d variables. Using (2.32) we deduce the existence of a smooth matrix function $\widetilde{T}_\mathrm{A}$ of size $n_x \times d$ with pointwise maximal rank satisfying

$$\left(\widetilde{Z}_\mathrm{A}^T \frac{\partial \widetilde{\mathscr{D}}_\mu}{\partial x} \widetilde{T}_\mathrm{A} \right)(t,x,\eta) = 0.$$

The remaining differential equations must be contained in the original DAE (in contrast to the algebraic equations, which are contained in the derivative array) and thus we assume the following to guarantee that we actually have d differential equations.

Assumption 2.28. *Let Assumptions 2.26 and 2.27 hold, set $d := n_x - a$, let $\widetilde{T}_\mathrm{A}$ be as above, and assume*

$$\operatorname{rank}\left(\left(\frac{\partial \widetilde{F}}{\partial \dot{x}} \widetilde{T}_\mathrm{A} \right)(t,x,\eta) \right) = d$$

for all $(t,x,\eta) \in \widetilde{\mathscr{M}_\mu} \cap \widetilde{\mathscr{U}}$*.*

To summarize the previous discussion, we make the following assumption, which for historical reasons (cf. [124]) is referred to as a hypothesis.

Hypothesis 2.29 ([124, Hypothesis 3.2])**.** *There exist integers μ and a such that the set*

$$\widetilde{\mathscr{M}}_\mu := \left\{(t,x,\eta) \in \mathbb{R}^{(\mu+2)n_x+1} \;\middle|\; \widetilde{\mathscr{D}}_\mu(t,x,\eta) = 0\right\}$$

associated with $\widetilde{\mathrm{F}}$ is nonempty and such that for every $(t_0, x_0, \eta_0) \in \widetilde{\mathscr{M}}_\mu$, there exists a (sufficiently small) neighborhood $\widetilde{\mathscr{U}}$ in which the following properties hold:

(i) We have $\operatorname{rank}(\frac{\partial}{\partial\eta}\mathscr{D}_\mu) = (\mu+1)n_x - a$ *on $\widetilde{\mathscr{M}}_\mu \cap \widetilde{\mathscr{U}}$ such that there exists a smooth matrix function $\widetilde{Z}_\mathrm{A}$ of size $(\mu+1)n_x \times a$ and pointwise maximal rank that satisfies $\widetilde{Z}_\mathrm{A}^T \frac{\partial}{\partial\eta}\widetilde{\mathscr{D}}_\mu = 0$.*

(ii) We have $\operatorname{rank}(\widetilde{Z}_\mathrm{A}^T \frac{\partial}{\partial x}\widetilde{\mathscr{D}}_\mu) = a$ *on $\widetilde{\mathscr{M}}_\mu \cap \widetilde{\mathscr{U}}$ such that there exists a smooth matrix function $\widetilde{T}_\mathrm{A}$ of size $n_x \times d$ with $d := n_x - a$ and pointwise maximal rank, satisfying $\widetilde{Z}_\mathrm{A}^T \left(\frac{\partial}{\partial x}\widetilde{\mathscr{D}}_\mu\right)\widetilde{T}_\mathrm{A} = 0$.*

(iii) We have $\operatorname{rank}(\frac{\partial \widetilde{F}}{\partial \dot{x}}\widetilde{T}_\mathrm{A}) = d$ *on $\widetilde{\mathscr{M}}_\mu \cap \widetilde{\mathscr{U}}$ such that there exists a smooth matrix function $\widetilde{Z}_\mathrm{D}$ of size $n_x \times d$ and pointwise maximal rank, satisfying* $\operatorname{rank}(\widetilde{Z}_\mathrm{D}^T \frac{\partial \widetilde{F}}{\partial \dot{x}}\widetilde{T}_\mathrm{A}) = d$.

Definition 2.30. The smallest possible μ for which Hypothesis 2.29 is satisfied is called *strangeness index* of the DAE (2.28a). If Hypothesis 2.29 is satisfied with $\mu = 0$, then the DAE (2.28a) is called *strangeness-free*.

The quantities a and d in Hypothesis 2.29 are, respectively, the numbers of algebraic and differential equations contained in the DAE (2.29). Using the matrix functions $\widetilde{Z}_\mathrm{D}$ and $\widetilde{Z}_\mathrm{A}$, the DAE (2.29) can (locally) be reformulated as

$$0 = \widetilde{D}(t,x,\dot{x}) := \left(\widetilde{Z}_\mathrm{D}^T \widetilde{F}\right)(t,x,\dot{x}), \tag{2.33a}$$

$$0 = \widetilde{A}(t,x) := \left(\widetilde{Z}_\mathrm{A}^T \widetilde{\mathscr{D}}_\mu\right)(t,x), \tag{2.33b}$$

which itself is strangeness-free and every solution of (2.29) also solves (2.33). Hereby we call (2.33a) the *differential part* of (2.29) and (2.33b) the *algebraic part*. Note that although $\widetilde{Z}_\mathrm{A}$ and $\widetilde{\mathscr{D}}_\mu$ may depend on derivatives of x it can be shown (cf. [124]) that their product only depends on t and x. Unfortunately, a solution of (2.33) is not necessarily a solution of (2.29). However, if we assume in addition, that Hypothesis 2.29 is satisfied with characteristic values μ, a, d and $\mu+1, a, d$, then for every initial value $x_{\mu+1,0} \in \mathscr{M}_{\mu+1}$ there exists a unique solution of (2.33) and this solution (locally) solves (2.29) (see [127, Theorem 4.13]). As a direct consequence, an initial value x_0 is consistent if and only if it is contained in the *consistency set*

$$(t_0, x_0) \in \widetilde{\mathbb{M}} := \left\{(t,x) \in \mathbb{R}^{n_x+1} \;\middle|\; \widetilde{A}(t,x) = 0\right\}. \tag{2.34}$$

If state transformations are allowed, then the implicit function theorem allows to (locally) rewrite the strangeness-free DAE (2.33) as

$$\dot{\xi} = \widetilde{\mathscr{L}}(t,\xi), \qquad \zeta = \widetilde{\mathscr{R}}(t,\xi) \tag{2.35}$$

with $\xi(t) \in \mathbb{R}^d$ and $\zeta(t) \in \mathbb{R}^a$. For the detailed derivation we refer to [127, Cha. 4.1]. Let $x = \mathscr{T}(t,\xi,\zeta)$ denote the transformation for the state. Then, the ordinary differential equation (ODE)

$$\dot{x} = \tilde{\mathfrak{f}}(t,x) := \mathscr{T}\left(t, \widetilde{\mathscr{L}}(t,\xi), \left(\tfrac{\partial}{\partial\xi}\widetilde{\mathscr{R}}\right)(t,\xi)\widetilde{\mathscr{L}}(t,\xi) + \left(\tfrac{\partial}{\partial t}\widetilde{\mathscr{R}}\right)(t,\xi)\right), \tag{2.36}$$

is called the *underlying ODE* for the DAE (2.29) and is the basis of the *differentiation index* [42], which is defined as $\mu + 1$ if $\frac{\partial}{\partial \dot{x}} \widetilde{F}$ is singular and 0 otherwise [127, Cor. 3.46].

Remark 2.31. In terms of the LTI DAE (2.5), the strangeness-index μ and the index of the matrix pencil ν (see Definition 2.12) satisfy

$$\nu = \begin{cases} 0, & \text{if } E \text{ is nonsingular,} \\ \mu + 1, & \text{otherwise.} \end{cases}$$

With the definition above, the differentiation index is thus a generalization of the index of the matrix pencil to nonlinear systems. ♣

If we want to solve the DAE (2.28a) numerically, we are not only interested in the existence of solutions but also that the solution of the initial value problem (2.29) is unique and depends continuously on the data. For DAEs, the so-called *well-posedness* can be be formulated as follows [127, Theorem 4.12].

Theorem 2.32. *Let* $\widetilde{F}$ *as in* (2.29) *be sufficiently smooth and satisfy Hypothesis 2.29. Let* $x^\star \in \mathscr{C}^1(\mathbb{I}, \mathbb{R}^{n_x})$ *be a sufficiently smooth solution of* (2.28). *Let the (nonlinear) operator* $\widetilde{\mathscr{F}} \colon \mathbb{D} \to \mathscr{Y}$, $\mathbb{D} \subseteq \mathscr{Z}$ *open, be defined by*

$$\widetilde{\mathscr{F}}(x)(t) = \begin{bmatrix} \dot{\xi} - \widetilde{\mathscr{L}}(t, \xi(t)) \\ \zeta - \widetilde{\mathscr{R}}(t, \xi(t)) \end{bmatrix}, \tag{2.37}$$

with the Banach spaces

$$\mathscr{Z} := \left\{ z \in \mathscr{C}\left(\mathbb{I}, \mathbb{R}^{n_x}\right) \;\middle|\; \xi \in \mathscr{C}^1(\mathbb{I}, \mathbb{R}^d),\ \xi(t_0) = 0 \right\}, \qquad \mathscr{Y} := \mathscr{C}\left(\mathbb{I}, \mathbb{R}^{n_x}\right)$$

according to (2.35). *Then* $x^\star$ *is a regular solution of the strangeness-free problem*

$$\widetilde{\mathscr{F}}(x) = 0$$

in the following sense. There exists a neighborhood $\mathscr{U}_x \subseteq \mathscr{Z}$ *of* $x^\star$ *and a neighborhood* $\mathscr{V} \subseteq \mathscr{Y}$ *of the origin such that for every* $f \in \mathscr{V}$ *the equation*

$$\widetilde{\mathscr{F}}(x) = f$$

has a unique solution $x \in \mathscr{U}_x$ *that depends continuously on* f. *In particular,* $x^\star$ *is the unique solution in* $\mathscr{U}_x$ *belonging to* $f = 0$.

In order to apply the theory to the original equation (2.28a) we have to ensure that the characteristic values μ, a, and d do not depend on the chosen input u. A simple way to guarantee this, is to ensure that the rank assumptions in Hypothesis 2.29 hold for all sufficiently smooth input functions. The derivative array (2.30) with explicit dependency on u takes the form

$$\mathscr{D}_\ell\left(t,x,\eta,u,\dot{u},\dots,u^{(\ell)}\right) := \begin{bmatrix} F(t,x,\dot{x},u) \\ \frac{\mathrm{d}}{\mathrm{d}t}F(t,x,\dot{x},u) \\ \vdots \\ \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^\ell F(t,x,\dot{x},u) \end{bmatrix} \in \mathbb{R}^{(\ell+1)n_x} \qquad \text{with } \eta := \begin{bmatrix} \dot{x} \\ \vdots \\ x^{(\ell+1)} \end{bmatrix}.$$

Hypothesis 2.33. *There exist integers μ and a, and matrix functions $Z_\mathrm{A}(\cdot) \in \mathbb{R}^{(\mu+1)n_x \times a}$, $T_\mathrm{A}(\cdot) \in \mathbb{R}^{n_x \times d}$, and $Z_\mathrm{D}(\cdot) \in \mathbb{R}^{n_x \times d}$ with pointwise maximal rank and $d := n - a$ such that for every sufficiently smooth u the set*

$$\mathscr{M}_\mu := \left\{\left(t,x,\eta,u,\dots,u^{(\mu)}\right) \in \mathbb{R}^{(\mu+2)n_x+(\mu+1)m+1} \;\middle|\; \mathscr{D}_\mu\left(t,x,\eta,u,\dots,u^{(\mu)}\right) = 0\right\}$$

associated with F is nonempty and such that for every $(t_0,x_0,\eta_0,u_0,\dots,u_0^{(\mu)}) \in \mathscr{M}_\mu$, there exists a (sufficiently small) neighborhood $\mathscr{U}$ in which the following properties hold:

(i) We have $\operatorname{rank}(\frac{\partial}{\partial\eta}\mathscr{D}_\mu) = (\mu+1)n_x - a$ *and* $Z_\mathrm{A}^T\frac{\partial}{\partial\eta}\mathscr{D}_\mu = 0$ *on* $\mathscr{M}_\mu \cap \mathscr{U}$.

(ii) We have $\operatorname{rank}(Z_\mathrm{A}^T\frac{\partial}{\partial x}\mathscr{D}_\mu) = a$ *and* $Z_\mathrm{A}^T\left(\frac{\partial}{\partial x}\mathscr{D}_\mu\right)T_\mathrm{A} = 0$ *on* $\mathscr{M}_\mu \cap \mathscr{U}$.

(iii) We have $\operatorname{rank}(\frac{\partial F}{\partial \dot{x}}T_\mathrm{A}) = d$ *and* $\operatorname{rank}(Z_\mathrm{D}^T\frac{\partial F}{\partial \dot{x}}T_\mathrm{A}) = d$ *on* $\mathscr{M}_\mu \cap \mathscr{U}$.

Remark 2.34. Note that similarly as in Hypothesis 2.29 the existence of the matrix functions Z_A, T_A, and Z_D in Hypothesis 2.33 follows from the constant rank assumptions and a smooth version of the singular value decomposition as in [127, Thm. 3.9 and Thm. 4.3]. ♣

Example 2.35. It is easy to see that the mass-spring-damper system (1.2) in section 1.1.1 with $M > 0$ satisfies Hypothesis 2.33 with $\mu = 0$. The equations for the pendulum (1.3) are in Hessenberg-form and therefore satisfy Hypothesis 2.33 with strangeness index $\mu = 2$ [127, Thm. 4.23]. ♠

Following the analysis in [124] that leads to the strangeness-free formulation (2.33) we observe that the functions D and A may depend on u and its derivatives. Due to the local character of Hypothesis 2.33 we can assume that D does not depend on derivatives of u. In any case, Hypothesis 2.33 yields the (local) reformulation

$$0 = D(t,x,\dot{x},u) := \left(Z_\mathrm{D}^T F\right)(t,x,\dot{x},u), \tag{2.38a}$$

$$0 = A\left(t,x,u,\dot{u},\dots,u^{(\mu)}\right) := \left(Z_\mathrm{A}^T\mathscr{D}_\mu\right)\left(t,x,u,\dot{u},\dots,u^{(\mu)}\right), \tag{2.38b}$$

which itself is strangeness-free. The corresponding explicit form (2.35) and the underlying ODE (2.36) therefore take the form

$$\dot{\xi} = \mathscr{L}(t,\xi,u), \qquad \zeta = \mathscr{R}(t,\xi,u,\dot{u},\dots,u^{(\mu)}) \tag{2.39}$$

and

$$\dot{x} = \mathfrak{f}\left(t,x,u,\dots,u^{(\mu+1)}\right). \tag{2.40}$$

Clearly, if a system satisfies Hypothesis 2.33, then it also satisfies Hypothesis 2.29 (with given u) and thus all previous results hold as well.

Remark 2.36. Let us emphasize that although derivatives of u up to order μ, respectively $\mu+1$ appear in the algebraic equation (2.38b), the explicit algebraic equation (2.39), and the underlying ODE (2.40), respectively, we may have

$$\frac{\partial}{\partial u^{(\ell)}} f\left(t, x, u, \ldots, u^{(\mu+1)}\right) \equiv 0$$

for some $\ell \in \{1, \ldots, \mu+1\}$, i.e., the underlying ODE (2.40) may not necessarily depend on all derivatives of u up to order $\mu+1$. ♣

3 Distributional solutions for linear time-invariant DDAEs

As outlined in the introduction (cf. section 1.2), one major aspect of this thesis is the development of general existence and uniqueness results for *delay differential-algebraic equations* (DDAEs), and a first starting point is to consider the *initial value problem* (IVP) for *linear time-invariant* (LTI) problems of the form (1.15). Recall that the IVP is given as

$$E\dot{x}(t) = A_1 x(t) + A_2 x(t-\tau) + f(t), \qquad \text{for } t \in \mathbb{I} := [0, t_{\mathrm{f}}), \tag{3.1a}$$

$$x(t) = \phi(t), \qquad \text{for } t \in [-\tau, 0], \tag{3.1b}$$

where $E, A_1, A_2 \in \mathbb{F}^{m \times n_x}$ are matrices over the field $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, $f\colon \mathbb{I} \to \mathbb{F}^m$ is the inhomogeneity, and $\phi\colon [-\tau, 0] \to \mathbb{F}^{n_x}$ is the history function or initial trajectory. For the ease of presentation we introduce the *shift operator* $(\sigma_\tau x)(t) := x(t-\tau)$ for $\tau > 0$ and thus, we can write

$$E\dot{x} = A_1 x + A_2 \sigma_\tau x + f \qquad \text{in } [0, t_{\mathrm{f}}) \tag{3.2}$$

instead of (3.1a). The Examples 1.4 and 1.5 demonstrate that we cannot expect the existence of a classical or even continuous solution for the IVP (3.1). Instead, we start our analysis with the distributional solution concept from section 2.2.

Most of the results in this section are obtained together with Stephan Trenn (University of Groningen) and published in [206, 207].

3.1 Distributional shift operator and delay-regularity

In order to interpret (3.1) within the space of piecewise-smooth distributions (cf. Definition 2.17), we need to define a distributional analogue of the time delay: For $\tau > 0$ we define the *distributional shift operator*

$$\sigma_\tau \colon \mathscr{D} \to \mathscr{D}, \qquad f \mapsto \left(\mathscr{C}_0^\infty(\mathbb{R};\mathbb{R}) \to \mathbb{R}, \quad \varphi \mapsto f(\varphi(\cdot + \tau))\right), \tag{3.3}$$

where $\mathscr{D}$ denotes the space of distributions as defined in (2.19). Note that for any continuous function $f \in \mathscr{C}(\mathbb{R};\mathbb{R})$ and any $\varphi \in \mathscr{C}_0^\infty(\mathbb{R};\mathbb{R})$ we have

$$\left(\sigma_\tau f\right)_{\mathscr{D}}(\varphi) = \int_{-\infty}^{\infty} f(t-\tau)\varphi(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} f(t)\varphi(t+\tau)\,\mathrm{d}t = \left(\sigma_\tau\left(f_{\mathscr{D}}\right)\right)(\varphi),$$

and thus $\left(\sigma_\tau f\right)_{\mathscr{D}} = \sigma_\tau\left(f_{\mathscr{D}}\right)$. Moreover, it is easy to see that σ_τ is a linear operator and $\sigma_\tau f \in \mathscr{D}_{\mathrm{pw}\mathscr{C}^\infty}$ for any $f \in \mathscr{D}_{\mathrm{pw}\mathscr{C}^\infty}$.

Lemma 3.1. *The distributional shift operator* σ_τ *defined in* (3.3) *and the distributional derivative* $\frac{\mathrm{d}}{\mathrm{d}t}$ *defined in* (2.21) *commute in* $\mathscr{D}$, *i.e.,* $\frac{\mathrm{d}}{\mathrm{d}t} \circ \sigma_\tau = \sigma_\tau \circ \frac{\mathrm{d}}{\mathrm{d}t}$.

Proof. Let $f \in \mathscr{D}$ and $\varphi \in \mathscr{C}_0^\infty(\mathbb{R};\mathbb{R})$. Then we obtain

$$\begin{aligned}\left(\frac{\mathrm{d}}{\mathrm{d}t}\circ\sigma_\tau\right)(f)(\varphi) &= \left(\frac{\mathrm{d}}{\mathrm{d}t}f\right)\left(\varphi(\cdot+\tau)\right) = -f\left(\frac{\mathrm{d}}{\mathrm{d}t}\varphi(\cdot+\tau)\right)\\ &= \left(\sigma_\tau\left(-f\right)\right)\left(\frac{\mathrm{d}}{\mathrm{d}t}\varphi\right) = \left(\sigma_\tau\circ\frac{\mathrm{d}}{\mathrm{d}t}\right)(f)(\varphi),\end{aligned}$$

where we have used that the derivative and the shift commute in $\mathscr{C}_0^\infty(\mathbb{R};\mathbb{R})$. ■

Lemma 3.2. *Let* $\Omega \subseteq \mathbb{R}$, $\beta \in \mathscr{C}^\infty(\mathbb{R};\mathbb{R})$, *and* $f \in \mathscr{D}_{\mathrm{pw}\mathscr{C}^\infty}$. *Then*

$$\sigma_\tau\left(\mathbb{1}_\Omega\right) = \mathbb{1}_{\Omega+\tau}, \qquad \sigma_\tau\left(\mathbb{1}_\Omega\beta\right) = \mathbb{1}_{\Omega+\tau}\sigma_\tau\beta, \qquad \textit{and} \qquad \sigma_\tau\left(f_\Omega\right) = \left(\sigma_\tau f\right)_{\Omega+\tau},$$

where $\Omega+\tau := \{\omega+\tau \mid \omega\in\Omega\}$.

Proof. Let $t \in \mathbb{R}$. Then

$$\begin{aligned}\left(\sigma_\tau\left(\mathbb{1}_\Omega\beta\right)\right)(t) = \left(\mathbb{1}_\Omega\beta\right)(t-\tau) &= \begin{cases}\beta(t-\tau), & \text{if } t-\tau\in\Omega,\\ 0, & \text{otherwise,}\end{cases}\\ &= \begin{cases}\beta(t-\tau), & \text{if } t\in\Omega+\tau,\\ 0, & \text{otherwise,}\end{cases} = \left(\mathbb{1}_{\Omega+\tau}\sigma_\tau\beta\right)(t)\end{aligned}$$

and thus $\sigma_\tau\left(\mathbb{1}_\Omega\beta\right) = \mathbb{1}_{\Omega+\tau}\sigma_\tau\beta$. The first assertion follows by choosing $\beta\equiv 1$. For the remaining assertion let $f = \alpha_{\mathscr{D}} + \sum_{s\in S} D_s$ with $\alpha \in \mathscr{C}^\infty_{\mathrm{pw}}(\mathbb{R};\mathbb{R})$, discrete set S and $D_s \in \operatorname{span}\{\delta_s, \dot{\delta}_s, \ddot{\delta}_s, \ldots\}$ for $s\in S$. Lemma 3.1 together with the already proven identities, and the definition of the Dirac impulse (cf. Example 2.16) imply $\sigma_\tau D_s = D_{s+\tau}$ for $s\in S$. We therefore conclude

$$\sigma_\tau\left(f_\Omega\right) = \sigma_\tau\left(\left(\mathbb{1}_\Omega\alpha\right)_{\mathscr{D}} + \sum_{s\in S\cap\Omega} D_s\right) = \left(\mathbb{1}_{\Omega+\tau}\sigma_\tau\alpha\right)_{\mathscr{D}} + \sum_{s\in S\cap(\Omega+\tau)} D_{s+\tau} = \left(\sigma_\tau f\right)_{\Omega+\tau}. \qquad ■$$

As in Section 2.2 for *differential-algebraic equations* (DAEs), we can now interpret the DDAE (3.1), respectively (3.2) in the space of piecewise-smooth distributions as the *initial trajectory problem* (ITP)

$$\begin{aligned}x_{(-\infty,0)} &= x^0_{(-\infty,0)},\\ (E\dot{x})_{[0,\infty)} &= (A_1x + A_2\sigma_\tau x + f)_{[0,\infty)},\end{aligned} \tag{3.4}$$

respectively the distributional DDAE

$$E\dot{x} = A_1x + A_2\sigma_\tau x + f, \tag{3.5}$$

with initial trajectory $x^0 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ and inhomogeneity $f \in \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty}$.

Definition 3.3. Consider the ITP (3.4) with $f \in \mathscr{D}^m_{\text{pw}\mathscr{C}^\infty}$. An initial trajectory $x^0 \in \mathscr{D}^{n_x}_{\text{pw}\mathscr{C}^\infty}$ is called *feasible* for the ITP (3.4), if there exists $x \in \mathscr{D}^{n_x}_{\text{pw}\mathscr{C}^\infty}$ that satisfies (3.4). In this case, x is called a *(distributional) solution* of (3.4). The ITP (3.4) is called *solvable* if there exists a feasible initial trajectory $x^0 \in \mathscr{D}^{n_x}_{\text{pw}\mathscr{C}^\infty}$ for the ITP (3.4).

Remark 3.4. Defining $f_{\text{ITP}} := f_{[0,\infty)} + \left(E\dot{x}^0 - A_1x^0 - A_2\sigma_\tau x^0\right)_{(-\infty,0)}$, then it is straightforward to see that every solution of the ITP (3.4) is also a solution of the distributional DDAE

$$E\dot{x} = A_1x + A_2\sigma_\tau x + f_{\text{ITP}}. \tag{3.6}$$

Conversely, let $x \in \mathscr{D}^{n_x}_{\text{pw}\mathscr{C}^\infty}$ satisfy (3.6) and define $x^0 := x$. Then x^0 is feasible and thus the ITP (3.4) is solvable. For a similar discussion for DAEs we refer to [178,205]. ♣

If the matrix pair (E, A_1) is regular, then we can use Theorem 2.19 to establish existence and uniqueness of solutions of the ITP (3.4) via integration on successive time intervals $[i\tau,(i+1)\tau)$, which is also referred to as method of steps (cf. chapter 2 and the forthcoming section 4.1).

Theorem 3.5. *Consider the ITP* (3.4) *with* $x^0 \in \mathscr{D}^{n_x}_{\text{pw}\mathscr{C}^\infty}$ *and* $f \in \mathscr{D}^m_{\text{pw}\mathscr{C}^\infty}$. *If the matrix pair* (E, A_1) *is regular, then the ITP* (3.4) *has a unique distributional solution* $x \in \mathscr{D}^{n_x}_{\text{pw}\mathscr{C}^\infty}$.

Proof. Applying the method of steps to (3.4) results in the sequence of DAE ITPs

$$\begin{aligned} x^i_{(-\infty,(i-1)\tau)} &= x^{i-1}_{(-\infty,(i-1)\tau)}, \\ \left(E\dot{x}^i\right)_{[(i-1)\tau,\infty)} &= \left(A_1x^i + \tilde{f}^i\right)_{[(i-1)\tau,\infty)}, \end{aligned} \tag{3.7}$$

with $\tilde{f}^i := A_2x^{i-1} + f$ and $i \in \mathbb{N}$. Theorem 2.19 implies recursively the existence of a unique solution $x^i \in \mathscr{D}^{n_x}_{\text{pw}\mathscr{C}^\infty}$ of (3.7) for each $i \in \mathbb{N}$. In particular, for each $i \in \mathbb{N}_0$ there exists $\alpha^i \in \mathscr{C}^\infty_{\text{pw}}(\mathbb{R};\mathbb{R})$, a discrete set $S^i \subseteq \mathbb{R}$ and distributions $D^i_s \in \operatorname{span}\{\delta_s, \dot{\delta}_s, \ddot{\delta}_s, \ldots\}$ for each $s \in S^i$ such that the jth component x^i_j of x^i is given by

$$x^i_j = {\alpha^i_j}_{\mathscr{D}} + \sum_{s \in S^i_j} D^i_{j;s}.$$

We show that $x := x^0_{(-\infty,0)} + \sum_{i=1}^\infty x^i_{[(i-1)\tau,i\tau)}$ is the solution of (3.4). First note that for the jth component the set

$$\left(S^0_j \cap (-\infty,0)\right) \cup \bigcup_{i\in\mathbb{N}} \left(S^i_j \cap [(i-1)\tau, i\tau)\right)$$

is discrete and $(\alpha^0_j)_{(-\infty,0)} + \sum_{i=1}^\infty (\alpha^i_j)_{[(i-1)\tau,i\tau)} \in \mathscr{C}^\infty_{\text{pw}}(\mathbb{R};\mathbb{R})$, which implies $x \in \mathscr{D}^{n_x}_{\text{pw}\mathscr{C}^\infty}$. Furthermore, by construction we have $x_{(-\infty,0)} = x^0_{(-\infty,0)}$. For $i \in \mathbb{N}$, Lemma 2.18 implies

$$\begin{aligned} (E\dot{x})_{[(i-1)\tau,i\tau)} &= \left(E\frac{\mathrm{d}}{\mathrm{d}t}\left(\left(x^0\right)_{(-\infty,0)} + \sum_{k=1}^\infty \left(x^k_{[(k-1)\tau,k\tau)}\right)\right)\right)_{[(i-1)\tau,i\tau)} \\ &= E\left(\dot{x}^i_{[(i-1)\tau,i\tau)} + \left(x^i((i-1)\tau^-) - x^{i-1}((i-1)\tau^-)\right)\delta_{(i-1)\tau}\right). \end{aligned}$$

Using $x^i_{(-\infty,(i-1)\tau)} = x^{i-1}_{(-\infty,(i-1)\tau)}$ and (3.7) we obtain

$$\begin{aligned}(E\dot{x})_{[(i-1)\tau,i\tau)} &= E\dot{x}^i_{[(i-1)\tau,i\tau)}\\ &= \left(A_1x^i + \tilde{f}^i\right)_{[(i-1)\tau,i\tau)}\\ &= A_1x^i_{[(i-1)\tau,i\tau)} + A_2x^{i-1}_{[(i-1)\tau,i\tau)} + f_{[(i-1)\tau,i\tau)}\\ &= A_1x^i_{[(i-1)\tau,i\tau)} + A_2\sigma_\tau x^i_{[(i-1)\tau,i\tau)} + f_{[(i-1)\tau,i\tau)}\\ &= \left(A_1x + A_2\sigma_\tau x + f\right)_{[(i-1)\tau,i\tau)}.\end{aligned}$$

Thus, x is a solution of the ITP (3.4). Since $x_{[(i-1)\tau,i\tau)} = x^i_{[(i-1)\tau,i\tau)}$ for each $i \in \mathbb{N}$ and x^i is the unique solution of (3.7), we conclude that x is the unique solution of (3.4). ■

Remark 3.6. The existence and uniqueness of distributional solutions for DDAEs was already hinted in [53] and [96]. Results for stronger solution concepts are presented for instance in [13,69,98] and the forthcoming chapter 4, although under much stronger assumptions on the history function and additional properties of the matrix pair (E,A_1). A generalization of Theorem 3.5 to switched DDAE is presented in [38]. ♣

Similarly to Theorem 2.7 we may expect that a singular matrix pencil (E,A_1) may result in an ITP that is either not uniquely solvable or not solvable at all. However, as the following example show, this is not the case; in addition see [94].

Example 3.7. Consider the scalar DDAE (3.1a) with $(E,A_1,A_2) = (0,0,1)$, i.e.

$$0 = \sigma_\tau x + f, \tag{3.8}$$

which clearly has the unique (acausal) solution $x = \sigma_{-\tau} f$ for any inhomogeneity f, although the matrix pair $(E,A_1) = (0,0)$ is not regular. Note however, that it is not possible to freely prescribe the initial trajectory for x on $[-\tau,0)$ because it is already fully specified by f given on $[0,\tau)$. ♠

The example shows that by introducing a time-delay term to a DAE with a singular matrix pair (E,A_1) we may arrive at a DDAE that is regular in a certain sense. Thus, we need to formalize the notion of regularity for DDAEs. Following [204], we give the following generalization of regularity.

Definition 3.8. The DDAE (3.2) is called *delay-regular*, if for all inhomogeneities $f \in \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty}$ with support in $[0,\infty)$ there exists a solution $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ and each solution for the same f is uniquely determined by the past, i.e., for two solutions $x_1, x_2 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ of (3.2), the implication

$$x_{1\,(-\infty,0)} = x_{2\,(-\infty,0)} \implies x_1 = x_2$$

holds. The matrix triple (E,A_1,A_2) is called *delay-regular* if and only if the corresponding DDAE is delay-regular.

In order to analyze the existence and uniqueness of solutions of the ITP (3.4), Example 3.7 reveals that in some sense the DDAE given by $(E,A_1,A_2) = (0,0,1)$ with singular matrix pair (E,A_1) is equivalent to the DDAE $(\widehat{E},\widehat{A},\widehat{D}) = (0,1,0)$ with a shifted inhomogeneity $\widehat{f} := \sigma_{-\tau} f$, where now $(\widehat{E},\widehat{A})$ is regular. We therefore want to define a notion of delay-equivalence. Unfortunately, it is not sufficient to consider matrix triplets only, since higher-order differential equations may be hidden in the DDAE (cf. [53] and [97]). This fact is illustrated with the following example.

Example 3.9. Consider the DDAE (3.1a) with

$$E = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \qquad A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \qquad f = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}.$$

Clearly, (E, A_1) is not regular, however differentiating the last equation twice and plugging in the first two equations yields

$$0 = \sigma_\tau \ddot{x}_1 + f_1 + \dot{f}_2 + \ddot{f}_3.$$

The same trick as applied in Example 3.7 can be used to shift the time-delay into the inhomogeneity, but the resulting equation cannot be written as a first-order DAE without increasing the dimension of the matrices (due to the presence of a second derivative). ♠

Example 3.9 motivates to study the more general DDAE

$$\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x = \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \sigma_\tau x + f, \tag{3.9}$$

respectively the ITP

$$x_{(-\infty,0)} = x^0_{(-\infty,0)}, \tag{3.10a}$$

$$\left(\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x\right)_{[0,\infty)} = \left(\mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) \sigma_\tau x + f\right)_{[0,\infty)}, \tag{3.10b}$$

with matrix polynomials $\mathscr{P}(s), \mathscr{Q}(s) \in \mathbb{R}[s]^{m \times n_x}$. Note that (3.10a) not only specifies the initial trajectory but also its (distributional) derivatives.

Definition 3.10. The DDAE (3.9) is called delay-regular, if for all inhomogeneities $f \in \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty}$ with support in $[0,\infty)$ there exists a solution $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ and each solution for the same f is uniquely determined by the past, i.e., for two solutions $x_1, x_2 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ of (3.9), the implication

$$x_{1\,(-\infty,0)} = x_{2\,(-\infty,0)} \implies x_1 = x_2$$

holds. The pair of matrix polynomials $(\mathscr{P}(s), \mathscr{Q}(s))$ is called *delay-regular* if and only if the corresponding DDAE is delay-regular.

Definition 3.11. An initial trajectory $x^0 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ is called *feasible* for the ITP (3.10), if there exists $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ that satisfies (3.10). In this case, x is called a *(distributional) solution* of (3.10). The ITP (3.10) is called *solvable* if there exists a consistent initial trajectory $x^0 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ for the ITP (3.10).

We first highlight the connection of delay-regularity with the solvability of the ITP (3.10).

Proposition 3.12. *If the DDAE* (3.9) *is delay-regular, then for each $f \in \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty}$ there exists an initial trajectory $x^0 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ such that the ITP* (3.10) *is uniquely solvable. Conversely, if the ITP* (3.10) *is uniquely solvable for $x^0 = 0$ and for any inhomogeneity $f \in \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty}$ then* (3.9) *is delay-regular.*

Proof. Let $f \in \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty}$ with support in $[0,\infty)$ and assume that (3.9) is delay-regular. Then there exists a solution $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ of (3.9). Setting $x^0 := x$ we immediately obtain a solution of the ITP (3.10). Suppose now that for $x^0 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ the ITP (3.10) has two solutions $x_1, x_2 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$. Then the difference $\widetilde{x} := x_1 - x_2$ satisfies the ITP (3.10) with initial trajectory $\widetilde{x}^0 := 0$ and zero inhomogeneity. Then $\widetilde{x}$ satisfies

$$\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\widetilde{x} = \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau\widetilde{x} + \left(\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\widetilde{x} - \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau\widetilde{x}\right)_{(-\infty,0)}.$$

Using

$$\left(\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\widetilde{x} - \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau\widetilde{x}\right)_{(-\infty,0)} = \left(\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\widetilde{x}_{(-\infty,0)} - \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau\left(\widetilde{x}_{(-\infty,0)}\right)\right)_{(-\infty,0)} = 0$$

and the delay-regularity of (3.9) we conclude $\widetilde{x} = 0$ and thus $x_1 = x_2$.

Now assume that the ITP (3.10) with $x_0 = 0$ has a unique solution x for all $f \in \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty}$. Then with the same argument as above it follows that x solves (3.9) with inhomogeneity $f_{[0,\infty)}$. To show uniqueness assume that (3.9) has two solutions, then the difference $\widetilde{x}$ satisfies $\widetilde{x}_{(-\infty,0)} = 0$ and therefore solves the ITP (3.10) with $x_0 = 0$ and $f = 0$. Hence $\widetilde{x}$ must coincide with the trivial solution of (3.10). ■

It is important to note the following for delay-regularity:

(i) Causality with respect to the inhomogeneity f is not assumed.

(ii) Existence of a solution for all initial trajectories is not assumed.

(iii) Unique solvability of the ITP with zero initial trajectory is only a *sufficient* condition for delay-regularity. In particular, delay-regularity does not imply in general that $x^0 = 0$ is a feasible initial trajectory for all inhomogeneities.

In fact, the second and third point is a consequence from the first point: because of the possible acausality the current inhomogeneity may determine the past (initial) state, see for instance Example 3.7.

Remark 3.13. In reality, a dependence on the future is not possible, and therefore one may question the utility of the notion of delay-regularity. However, besides its mathematical relevance, this notion may also be useful in practice if the future value of the inhomogeneity can be interpreted as a *prediction* of that future value. Additional applications are hybrid numerical-experimental systems [212], see also Chapter 5, and boundary value problems for DDAEs. ♣

Remark 3.14. Although the choice of $t_0 = 0$ in Definition 3.10, respectively Definition 3.8, seems arbitrary, it covers the situation that the support of f is in $[t_0,\infty)$ for some $t_0 \in \mathbb{R}$. To see this, let $(\mathscr{P}(s),\mathscr{Q}(s)) \in \left(\mathbb{R}[s]^{m\times n_x}\right)^2$ be a delay-regular pair of matrix polynomials and define $g := \sigma_{-t_0} f$. Using Lemma 3.2 we obtain

$$g_{(-\infty,0)} = \left(\sigma_{-t_0} f\right)_{(-\infty,0)} = \sigma_{-t_0}\left(f_{(-\infty,t_0)}\right) = 0$$

implying the existence of $z \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ that satisfies

$$\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) z = \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau z + g.$$

Setting $x := \sigma_{t_0} z$, Lemma 3.1 implies

$$\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x = \sigma_{t_0}\left(\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) z\right) = \sigma_{t_0}\left(\mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau z + g\right) = \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + f.$$

If $\widetilde{x} \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ is another solution of (3.9) satisfying $\widetilde{x}_{(-\infty,t_0)} = x_{(-\infty,t_0)}$, then similar arguments as before show $x = \widetilde{x}$. Using Proposition 3.12 we immediately conclude that the same arguments apply to the initial time point in the ITP (3.10). ♣

As a consequence of Theorem 3.5, we obtain the following sufficient condition for delay-regularity of the general DDAE (3.9) (cf. Corollary 2.20).

Theorem 3.15. *Consider the ITP* (3.10) *with* $m = n_x$ *and* $\det(\mathscr{P}(s)) \not\equiv 0$. *Then for any past trajectory* $x^0 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ *and any inhomogeneity* $f \in \mathscr{D}^{m}_{\mathrm{pw}\mathscr{C}^\infty}$, *there exists a unique solution* $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ *of* (3.10). *In particular, the DDAE* (3.9) *is delay-regular.*

Proof. The result follows as a consequence of Corollary 2.20 and Theorem 3.5. For the sake of completeness, we present the details here as well. Let $\mathscr{P}(s) = \sum_{j=0}^{p} P_j s^j$ and $\mathscr{Q}(s) = \sum_{j=0}^{q} Q_j s^j$. Since adding zero terms to $\mathscr{P}(s)$ does not alter the determinant of $\mathscr{P}(s)$, we may assume without loss of general $p = q+1$. Then a standard companion form linearization of (3.9) yields the DDAE

$$\mathscr{E}\dot{z} = \mathscr{A}z + \mathscr{D}\sigma_\tau z + \mathscr{F} \tag{3.11}$$

with $\mathscr{E},\mathscr{A},\mathscr{D} \in \mathbb{R}^{m+(p-1)n_x \times pn_x}$, given by

$$\mathscr{E} = \begin{bmatrix} P_p & 0 & \cdots & 0 \\ 0 & I_{n_x} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_{n_x} \end{bmatrix}, \qquad \mathscr{A} = \begin{bmatrix} -P_{p-1} & -P_{p-2} & \cdots & -P_0 \\ I_{n_x} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & I_{n_x} & 0 \end{bmatrix}, \qquad z = \begin{bmatrix} \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{p-1} x \\ \vdots \\ \frac{\mathrm{d}}{\mathrm{d}t} x \\ x \end{bmatrix},$$

$$\mathscr{D} = \begin{bmatrix} Q_{p-1} & \cdots & Q_0 \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}, \qquad \mathscr{F} = \begin{bmatrix} f \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Note that there exist (cf. [142]) unimodular matrix polynomials $\mathscr{R}(s) \in \mathbb{R}[s]^{m+(p-1)n_x \times m+(p-1)n_x}$ and $\mathscr{S}(s) \in \mathbb{R}[s]^{pn_x \times pn_x}$ with

$$\mathscr{R}(s)(s\mathscr{E} - \mathscr{A})\mathscr{S}(s) = \begin{bmatrix} \mathscr{P}(s) & 0 \\ 0 & I_{(p-1)n_x} \end{bmatrix}.$$

The proof now follows by the observation that there exists a constant $c \neq 0$ with

$$0 \not\equiv \det(\mathscr{P}(s)) = c\det(s\mathscr{E} - \mathscr{A})$$

and application of Theorem 3.5 to (3.11). ■

If $\det(\mathscr{P}(s)) \equiv 0$ then we cannot apply Theorem 3.15. Instead, we want to use the condensed form (2.27) from Theorem 2.24: there exist unimodular matrices $\mathscr{U}(s) \in \mathbb{R}[s]^{m\times m}$, $\mathscr{V}(s) \in \mathbb{R}[s]^{n_x \times n_x}$ such that

$$\mathscr{U}(s)\mathscr{P}(s)\mathscr{V}(s) = \begin{bmatrix} \widehat{\mathscr{P}}_{11}(s) & 0 & \widehat{\mathscr{P}}_{13}(s) \\ 0 & 0 & \widehat{\mathscr{P}}_{23}(s) \\ 0 & 0 & \widehat{\mathscr{P}}_{33}(s) \end{bmatrix} \quad \text{and} \quad \mathscr{U}(s)\mathscr{Q}(s)\mathscr{V}(s) = \begin{bmatrix} \hat{\mathscr{Q}}_{11}(s) & \hat{\mathscr{Q}}_{12}(s) & \hat{\mathscr{Q}}_{13}(s) \\ 0 & 0 & \hat{\mathscr{Q}}_{23}(s) \\ 0 & 0 & \hat{\mathscr{Q}}_{33}(s) \end{bmatrix},$$

where $\widehat{\mathscr{P}}_{11}$ is a nonsingular diagonal matrix, $\widehat{\mathscr{P}}_{23}, \widehat{\mathscr{P}}_{33}, \hat{\mathscr{Q}}_{33}$ are block upper triangular matrices with zero diagonal blocks and $\hat{\mathscr{Q}}_{23}$ is a nonsingular block upper triangular matrix. We therefore have to ensure that the transformation of $(\mathscr{P}(s), \mathscr{Q}(s))$ with unimodular matrices does not affect delay-regularity.

Proposition 3.16. *Consider a pair of matrix polynomials $(\mathscr{P}(s), \mathscr{Q}(s)) \in (\mathbb{R}[s]^{m\times n_x})^2$ and unimodular matrices $\mathscr{U}(s) \in \mathbb{R}[s]^{m\times m}$ and $\mathscr{V}(s) \in \mathbb{R}[s]^{n_x\times n_x}$. Let*

$$\widehat{\mathscr{P}}(s) := \mathscr{U}(s)\mathscr{P}(s)\mathscr{V}(s) \qquad \textit{and} \qquad \hat{\mathscr{Q}}(s) := \mathscr{U}(s)\mathscr{Q}(s), \mathscr{V}(s).$$

Then $(\mathscr{P}(s), \mathscr{Q}(s))$ is delay-regular if, and only if $(\widehat{\mathscr{P}}(s), \hat{\mathscr{Q}}(s))$ is delay-regular.

Proof. First note that it is sufficient to show one direction. Let $\hat{f} \in \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty}$ with support in $[0,\infty)$ and consider

$$\widehat{\mathscr{P}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\hat{x} = \hat{\mathscr{Q}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau \hat{x} + \hat{f}. \tag{3.12}$$

Define $f := \mathscr{U}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{-1}\hat{f}$ and observe that

$$f_{(-\infty,0)} = \left(\mathscr{U}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{-1}\hat{f}\right)_{(-\infty,0)} = \left(\mathscr{U}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{-1}\hat{f}_{(-\infty,0)}\right)_{(-\infty,0)} = 0.$$

Delay-regularity of (3.9) thus implies the existence of a solution $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ of (3.9). The choice $\hat{x} = \mathscr{V}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{-1} x$ together with

$$\widehat{\mathscr{P}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\hat{x} = \mathscr{U}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\mathscr{V}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x = \mathscr{U}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\left(\mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + f\right) = \widetilde{\mathscr{Q}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau \hat{x} + \hat{f}.$$

shows existence of a solution for (3.12). Assume now that $\widehat{x}_1, \widehat{x}_2 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ are solutions of (3.12) for the same $\widehat{f} \in \mathscr{D}^{m}_{\mathrm{pw}\mathscr{C}^\infty}$ satisfying $(\widehat{x}_1)_{(-\infty,0)} = (\widehat{x}_2)_{(-\infty,0)}$. For $i = 1,2$ define $x_i := \mathscr{V}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\widehat{x}_i$ and observe that

$$\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x_i = \mathscr{U}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{-1}\widetilde{\mathscr{P}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\widehat{x}_i = \mathscr{U}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{-1}\left(\widetilde{\mathscr{Q}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau\widehat{x}_i + \widehat{f}\right) = \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x_i + \mathscr{U}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{-1}\widehat{f}.$$

Delay-regularity thus implies $x_1 = x_2$ and since $\mathscr{V}(s)$ is unimodular we conclude $\widehat{x}_1 = \widehat{x}_2$. ■

Theorem 3.17. *Consider $(\mathscr{P}(s), \mathscr{Q}(s)) \in \left(\mathbb{R}[s]^{m\times n_x}\right)^2$. Let $\mathscr{U}(s) \in \mathbb{R}[s]^{m\times m}$ and $\mathscr{V}(s) \in \mathbb{R}[s]^{n_x\times n_x}$ be the unimodular matrices from Theorem 2.24. Define $\widehat{\mathscr{P}}(s) := \mathscr{U}(s)\mathscr{P}(s)\mathscr{V}(s) \in \mathbb{R}[s]^{m\times n_x}$ and $\widehat{\mathscr{Q}}(s) := \mathscr{U}(s)\mathscr{Q}(s)\mathscr{V}(s) \in \mathbb{R}[s]^{m\times n_x}$, i.e.,*

$$\widehat{\mathscr{P}}(s) = \begin{bmatrix} \widehat{\mathscr{P}}_{11}(s) & 0 & \widehat{\mathscr{P}}_{13}(s) \\ 0 & 0 & \widehat{\mathscr{P}}_{23}(s) \\ 0 & 0 & \widehat{\mathscr{P}}_{33}(s) \end{bmatrix}, \qquad \widehat{\mathscr{Q}}(s) = \begin{bmatrix} \widehat{\mathscr{Q}}_{11}(s) & \widehat{\mathscr{Q}}_{12}(s) & \widehat{\mathscr{Q}}_{13}(s) \\ 0 & 0 & \widehat{\mathscr{Q}}_{23}(s) \\ 0 & 0 & \widehat{\mathscr{Q}}_{33}(s) \end{bmatrix}. \tag{3.13}$$

Then the following statements are true.

(i) The pair $(\widehat{\mathscr{P}}_{23}(s), \widehat{\mathscr{Q}}_{23}(s))$ is delay-regular.

(ii) The pair $(\widehat{\mathscr{P}}(s), \widehat{\mathscr{Q}}(s))$ is delay-regular if and only if the second column and the third block row are not present.

(iii) The pair $(\mathscr{P}(s), \mathscr{Q}(s))$ is delay-regular if and only if $(\widehat{\mathscr{P}}(s), \widehat{\mathscr{Q}}(s))$ is delay-regular.

Before we present the proof of Theorem 3.17 let us revisit Example 3.9.

Example 3.18. Consider the DDAE from Example 3.9, i.e.,

$$\mathscr{P}(s) = \begin{bmatrix} 0 & s & 0 \\ 0 & -1 & s \\ 0 & 0 & -1 \end{bmatrix} \qquad \text{and} \qquad \mathscr{Q}(s) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

We obtain

$$\begin{bmatrix} 0 & 1 & s \\ 0 & 0 & 1 \\ 1 & s & s^2 \end{bmatrix} \mathscr{P}(s) \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \left[\begin{array}{cc|c} -1 & 0 & 0 \\ 0 & -1 & 0 \\ \hline 0 & 0 & 0 \end{array}\right],$$

$$\begin{bmatrix} 0 & 1 & s \\ 0 & 0 & 1 \\ 1 & s & s^2 \end{bmatrix} \mathscr{Q}(s) \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \left[\begin{array}{cc|c} 0 & 0 & s \\ 0 & 0 & 1 \\ \hline 0 & 0 & s^2 \end{array}\right].$$

Clearly, the second block column and the third block-row in (3.13) are not present, such that Theorem 3.17 implies that $(\mathscr{P}(s), \mathscr{Q}(s))$ is delay-regular. ♠

For the proof of Theorem 3.17 we first need the following technical result, which generalizes the findings from Example 3.7.

Lemma 3.19. *Let $\mathscr{Q}(s) \in \mathbb{R}[s]^{n_x \times n_x}$ satisfy $\det(\mathscr{Q}(s)) \not\equiv 0$. Then the pair $(0, \mathscr{Q}(s))$ is delay-regular. In particular, for any $f \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ with support in $[0,\infty)$ there exists $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ with $x_{(-\infty,-\tau)} = 0$ that satisfies the DDAE*

$$0 = \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + f. \tag{3.14}$$

Proof. Let $f \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ with support in $[0,\infty)$. Theorem 3.15 implies that there exists a unique solution $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ of the ITP

$$\begin{aligned} x_{(-\infty,-\tau)} &= 0, \\ \left(\mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x\right)_{[-\tau,\infty)} &= -\left(\sigma_{-\tau}f\right)_{[-\tau,\infty)}, \end{aligned} \tag{3.15}$$

see also Remark 3.14. Lemmata 3.1 and 3.2 yield

$$\left(\mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x\right)_{(-\infty,0)} = \sigma_\tau\left(\mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x\right)_{(-\infty,-\tau)} = \sigma_\tau\left(\mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x_{(-\infty,-\tau)}\right)_{(-\infty,-\tau)} = 0,$$

and

$$\left(\mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + f\right)_{[0,\infty)} = \sigma_\tau\left(\mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x + \sigma_{-\tau}f\right)_{[-\tau,\infty)} = 0,$$

showing that x satisfies the DDAE (3.14). Suppose now that $x_1, x_2 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ solve (3.14) and satisfy $(x_1)_{(-\infty,0)} = (x_2)_{(-\infty,0)}$. Then the difference $\widetilde{x} := x_1 - x_2$ satisfies

$$0 = \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau \widetilde{x}$$

and thus also the ITP (3.15) with $f = 0$. Theorem 3.15 implies $\widetilde{x} = 0$, which completes the proof. ■

Proof of Theorem 3.17.

(i) Since $\widehat{\mathscr{P}}_{23}(s)$ and $\widehat{\mathscr{Q}}_{23}(s)$ are block upper triangular, we write (omitting s)

$$\widehat{\mathscr{P}}_{23} = \begin{bmatrix} 0 & \widetilde{\mathscr{P}}_{1,2} & \cdots & \widetilde{\mathscr{P}}_{1,k} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \widetilde{\mathscr{P}}_{k-1,k} \\ & & & 0 \end{bmatrix}, \quad \widehat{\mathscr{Q}}_{23} = \begin{bmatrix} \widetilde{\mathscr{Q}}_{1,1} & \cdots & \widetilde{\mathscr{Q}}_{1,k} \\ & \ddots & \vdots \\ & & \widetilde{\mathscr{Q}}_{k,k} \end{bmatrix}, \quad z = \begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix}, \quad g = \begin{bmatrix} g_1 \\ \vdots \\ g_k \end{bmatrix},$$

and study the DDAE

$$\widehat{\mathscr{P}}_{23}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)z = \widehat{\mathscr{Q}}_{23}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau z + g$$

with $g_{(-\infty,0)} = 0$. Since $\widehat{\mathscr{Q}}_{23}(s)$ is nonsingular, we conclude that $\widetilde{\mathscr{Q}}_{i,i}(s)$ is nonsingular for all $i = 1,\dots,k$. In particular, Lemma 3.19 implies that $(0, \widetilde{\mathscr{Q}}_{k,k}(s))$ is delay-regular, i.e., there exists z_k with $\left(z_k\right)_{(-\infty,-\tau)} = 0$ satisfying

$$0 = \widetilde{\mathscr{Q}}_{k,k}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau z_k + g_k.$$

Substituting z_k into the $(k-1)$th block equation yields the DDAE

$$0 = \widetilde{\mathscr{Q}}_{k-1,k-1}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau z_{k-1} + \widetilde{g}_{k-1}$$

with $\widetilde{g}_{k-1} := g_{k-1} - \widetilde{\mathscr{P}}_{k-1,k}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) z_k + \widetilde{\mathscr{Q}}_{k-1,k}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau z_k$. Having

$$\left(\widetilde{g}_{k-1}\right)_{(-\infty,-\tau)} = -\left(\widetilde{\mathscr{P}}_{k-1,k}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\left(z_k\right)_{(-\infty,-\tau)} + \widetilde{\mathscr{Q}}_{k-1,k}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau\left(\left(z_k\right)_{(-\infty,-2\tau)}\right)\right)_{(-\infty,-\tau)} = 0,$$

we conclude with the same line of arguing that

$$\left(\begin{bmatrix} 0 & \widetilde{\mathscr{P}}_{k-1,k}(s) \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} \widetilde{\mathscr{Q}}_{k-1,k-1}(s) & \widetilde{\mathscr{Q}}_{k-1,k}(s) \\ 0 & \widetilde{\mathscr{Q}}_{k,k}(s) \end{bmatrix}\right)$$

is delay-regular. Repeating this process shows that $(\widehat{\mathscr{P}}_{2,3}(s), \widehat{\mathscr{Q}}_{2,3}(s))$ is delay-regular.

(ii) For $\hat{f}^T = \begin{bmatrix} \hat{f}_1^T & \hat{f}_2^T & \hat{f}_3^T \end{bmatrix}$ consider the DDAE

$$\widehat{\mathscr{P}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\hat{x} = \widehat{\mathscr{Q}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau\hat{x} + \hat{f}. \tag{3.16}$$

Since $(\widehat{\mathscr{P}}_{23}(s), \widehat{\mathscr{Q}}_{23}(s))$ is delay-regular, there exists $\hat{x}_3$ satisfying

$$\widehat{\mathscr{P}}_{23}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\hat{x}_3 = \widehat{\mathscr{Q}}_{23}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau\hat{x}_3 + \hat{f}_2. \tag{3.17}$$

Assume first that $(\widehat{\mathscr{P}}(s), \widehat{\mathscr{Q}}(s))$ is delay-regular. Setting

$$\hat{f}_3 := \widehat{\mathscr{P}}_{33}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\hat{x}_3 - \widehat{\mathscr{Q}}_{33}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\hat{x}_3 + \mathbb{1}_{[0,\infty)}\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

yields $0 = \mathbb{1}_{[0,\infty)}$, which is true only if the third block row is not present. In addition, for any $\hat{x}_2$, and any $\hat{x}_1^0$ the ITP

$$\begin{aligned} (\hat{x}_1)_{(-\infty,0)} &= (\hat{x}_1^0)_{(-\infty,0)}, \\ \left(\widehat{\mathscr{P}}_{11}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\hat{x}_1\right)_{[0,\infty)} &= \left(\widehat{\mathscr{Q}}_{11}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau\hat{x}_1 + g\right)_{[0,\infty)} \end{aligned} \tag{3.18}$$

with $g := \hat{f}_1 - \widehat{\mathscr{P}}_{13}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\hat{x}_3 + \widehat{\mathscr{Q}}_{12}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau\hat{x}_2 + \widehat{\mathscr{Q}}_{13}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau\hat{x}_3$ has a unique solution (cf. Theorem 3.15). In particular we can choose $(\hat{x}_2)_{[0,\infty)}$ arbitrarily and thus the delay-regularity of $(\widehat{\mathscr{P}}(s), \widehat{\mathscr{Q}}(s))$ implies that the second block column is not present. Conversely, assume

$$\widehat{\mathscr{P}}(s) = \begin{bmatrix} \widehat{\mathscr{P}}_{11}(s) & \widehat{\mathscr{P}}_{13}(s) \\ 0 & \widehat{\mathscr{P}}_{23}(s) \end{bmatrix}, \quad \widehat{\mathscr{Q}}(s) = \begin{bmatrix} \widehat{\mathscr{Q}}_{11}(s) & \widehat{\mathscr{Q}}_{13}(s) \\ 0 & \widehat{\mathscr{Q}}_{23}(s) \end{bmatrix}.$$

We have already established that there exists $\hat{x}_3$ solving (3.17). Additionally, the ITP (3.18) has a unique solution for any initial trajectory $\hat{x}_1^0$ and thus $(\widehat{\mathscr{P}}(s), \widehat{\mathscr{Q}}(s))$ is delay-regular.

(iii) This is a consequence of Proposition 3.16. ∎

Theorem 3.20. *The pair* $(\mathscr{P}(s),\mathscr{Q}(s))$ *is delay-regular if and only if* $m = n_x$ *and there exists* $s,\omega \in \mathbb{C}$ *with*

$$\det(\mathscr{P}(s) - \omega\mathscr{Q}(s)) \neq 0. \tag{3.19}$$

In particular, if $m < n_x$ *then the ITP* (3.10) *possesses a nontrivial solution* $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ *with* $f = 0$ *and* $x_{(-\infty,0)} = 0$. *If instead* $m \geq n_x$ *and* $\operatorname{rank}(\mathscr{P}(s) - \omega\mathscr{Q}(s)) < m$ *for all* $s,\omega \in \mathbb{C}$ *then there exists an* $f \in \mathscr{D}^{m}_{\mathrm{pw}\mathscr{C}^\infty}$ *for which the ITP* (3.10) *has no solution.*

Proof. Theorem 3.17 implies that (3.9) is delay-regular if and only if the second block column and the third block row in (2.27) do not appear. We obtain

$$\begin{aligned}\det(\mathscr{P}(s) - \omega\mathscr{Q}(s)) &= c\det(\widehat{\mathscr{P}}(s) - \omega\widehat{\mathscr{Q}}(s))\\ &= c\det\left(\widehat{\mathscr{P}}_{11}(s) - \omega\widehat{\mathscr{Q}}_{11}(s)\right)\det\left(\widehat{\mathscr{P}}_{23}(s) - \omega\widehat{\mathscr{Q}}_{23}(s)\right),\end{aligned}$$

with $c := \det(\mathscr{U}(s))^{-1}\det(\mathscr{V}(s))^{-1} \neq 0$. Using the notation from the proof of Theorem 3.17 we obtain

$$\det\left(\widehat{\mathscr{P}}_{23}(s) - \omega\widehat{\mathscr{Q}}_{23}(s)\right) = \prod_{i=1}^{k}\det\left(-\omega\widetilde{\mathscr{Q}}_{i,i}(s)\right) = (-\omega)^{\rho}\prod_{i=1}^{k}\det(\widetilde{\mathscr{Q}}_{i,i}(s)) \not\equiv 0,$$

since $\widetilde{\mathscr{Q}}_{i,i}(s)$ is nonsingular for $i = 1,\dots,k$. Hereby, ρ denotes the dimension of the square matrix $\widehat{\mathscr{Q}}_{23}(s)$. The nonsingularity of $\widehat{\mathscr{P}}_{11}(s)$ thus shows that the delay-regularity of $(\mathscr{P}(s),\mathscr{Q}(s))$ implies (3.19). For the converse direction we observe that (3.19) immediately implies that the second block column cannot be present. From $n_x = m$ we infer that also the third block row cannot be present, such that Theorem 3.17 implies delay-regularity of $(\mathscr{P}(s),\mathscr{Q}(s))$. For the remaining assertions notice that $m < n_x$ implies that the second block column appears in (2.27). On the other hand, if $m \geq n_x$ and

$$\operatorname{rank}_{\mathbb{R}[s,\omega]}(\mathscr{P}(s) - \omega\mathscr{Q}(s)) < m,$$

then we conclude that the third block-row in (2.27) is present. The result follows from the proof of Theorem 3.17 (ii). ∎

Applying Theorem 3.20 to the DDAE (3.2) we obtain the following corollary.

Corollary 3.21. *The triplet* (E, A_1, A_2) *is delay-regular if and only if* $m = n_x$ *and there exists* $s,\omega \in \mathbb{C}$ *with*

$$\det(sE - A_1 - \omega A_2) \neq 0. \tag{3.20}$$

In particular, if $m < n_x$ *then the ITP* (3.4) *possesses a nontrivial solution* $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ *with* $f = 0$ *and* $x_{(-\infty,0)} = 0$. *If instead* $m \geq n_x$ *and* $\operatorname{rank}(\mathscr{P}(s) - \omega\mathscr{Q}(s)) < m$ *for all* $s,\omega \in \mathbb{C}$ *then there exists an* $f \in \mathscr{D}^{m}_{\mathrm{pw}\mathscr{C}^\infty}$ *for which the ITP* (3.4) *has no solution.*

Remark 3.22. The statements of Theorem 3.20 and Corollary 3.21 are a generalization of Theorem 2.7 to DDAEs. ♣

Let us emphasize that for delay-regularity we require existence of solutions for all $f \in \mathcal{D}^{m}_{\mathrm{pw}\mathcal{C}^\infty}$. Consequently, the ITP (3.10) may possess a unique solution even if $(\mathcal{P}(s), \mathcal{Q}(s))$ is not delay-regular. Consider for instance the DDAE

$$0 = \mathcal{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + \begin{bmatrix} f \\ \dot{f} \end{bmatrix} \tag{3.21}$$

with $\mathcal{Q}(s) := \left[\begin{smallmatrix} 1 \\ s \end{smallmatrix}\right]$ and $f \in \mathcal{D}_{\mathrm{pw}\mathcal{C}^\infty}$. Theorem 3.20 immediately implies that $(0, \mathcal{Q}(s))$ is not delay-regular. Still, the unique solution of (3.21) is given by $x = -\sigma_{-\tau} f$.

Remark 3.23. Equation (3.21) contradicts [96, Cor. 2 and Cor. 3], which claim that whenever a solution of the ITP (3.10) exists it is unique only if $m = n_x$ and (3.19) holds. Note that in view of the condensed form (2.27), uniqueness is related to the (non-)existence of the second block column, while existence is related to the (non-)existence of the third block row. ♣

3.2 Interlude: Feedback regularization of DAEs with delay

A standard control concept uses *feedback*, i.e., the control law depends on the current state or output of the system (see for instance Section 1.1.2). For linear time-invariant systems of the form

$$\begin{aligned} E\dot{x} &= A_1 x + Bu + f, \\ y &= Cx \end{aligned} \tag{3.22}$$

where $B \in \mathbb{R}^{m \times n_u}$, $C \in \mathbb{R}^{n_y \times n_x}$, u is the n_u-dimensional input, and y is the n_y-dimensional output, a simple feedback law takes the form

$$u = Fy = FCx \tag{3.23}$$

for some feedback matrix $F \in \mathbb{R}^{n_u \times n_y}$. The closed-loop system is thus given by the DAE

$$E\dot{x} = \left(A_1 + BFC\right)x + f, \tag{3.24}$$

showing that the feedback can be used to alter system properties. For instance, suppose that $(E, A_1) \in \left(\mathbb{R}^{n_x \times n_x}\right)^2$ is singular. We say that is is possible to *regularize* (3.22), if there exists some $F \in \mathbb{R}^{n_u \times n_y}$ such that the pencil $(E, A_1 + BFC)$ is regular. In fact the following result from [76] holds.

Lemma 3.24. *Consider the DAE* (3.22) *with $m = n_x$. There exists $F \in \mathbb{R}^{n_u \times n_y}$ such that the closed-loop system $(E, A_1 + BFC)$ is regular if and only if*

$$\operatorname{rank}\left(\begin{bmatrix} \lambda E - A_1 & B \end{bmatrix}\right) = \operatorname{rank}\left(\begin{bmatrix} \lambda E - A_1 \\ C^T \end{bmatrix}\right) = n_x$$

for some $\lambda \in \mathbb{C}$.

Although instantaneous feedback is a convenient theoretical approach, it is usually not implementable, in particular, when the signals have to be measured first, and some calculations have to be carried out, thus resulting in an intrinsically necessary time delay. We therefore are interested whether we can regularize (3.22) with a delayed feedback

$$u = F\sigma_\tau y = FC\sigma_\tau x, \tag{3.25}$$

i.e., if we can find $F \in \mathbb{R}^{n_u \times n_y}$ such that $(sE - A_1, BCF)$ is delay-regular. As important consequence of Theorem 3.20, respectively Corollary 3.21, we obtain the following result.

Theorem 3.25. *For $m = n_x$ consider the descriptor system* (3.22). *There exists $F \in \mathbb{R}^{n_u \times n_y}$ such that $(E, A_1 + BFC)$ is regular if and only if there exists $\widehat{F} \in \mathbb{R}^{n_u \times n_y}$ such that $(sE - A_1, B\widehat{F}C)$ is delay-regular.*

Proof. Assume first that there exists $F \in \mathbb{R}^{n_u \times n_y}$ such that $(E, A_1 + BFC)$ is regular. Then we have

$$n_x = \operatorname{rank}_{\mathbb{R}[s]}(sE - A_1 - BFC) \le \operatorname{rank}_{\mathbb{R}[s,\omega]}(sE - A_1 - \omega BFC) \le n_x,$$

i. e., $(sE - A_1, BFC)$ is delay-regular. On the other hand, assume the existence of $\widehat{F} \in \mathbb{R}^{n_u \times n_y}$ such that $(sE - A_1, B\widehat{F}C)$ is delay-regular. Then there exists $\widehat{\omega} \in \mathbb{R}$ such that

$$\det(sE - A_1 - \widehat{\omega}B\widehat{F}C) \not\equiv 0.$$

The choice $F = \widehat{\omega}\widehat{F}$ guarantees that $(E, A_1 + BFC)$ is regular. ■

The proof of Theorem 3.25 details that for any feedback matrix $F \in \mathbb{R}^{n_u \times n_y}$ that renders the pencil $(E, A_1 + BFC)$ regular, also the triplet (E, A_1, BFC) is delay-regular. The converse direction is however not true, as we can see from the following example:

Example 3.26. Consider the scalar DAE (3.22) with $E = 0$, $A_1 = 1$, $B = 1$, and $C = 1$. For $\widehat{F} = -1$ the pair of matrix polynomials $(sE - A_1, B\widehat{F}C) = (-1, -1)$ is delay-regular. However, the pencil $(E, A_1 + B\widehat{F}C) = (0, 0)$ is not regular. ♠

The reason for the behavior in Example 3.26 is due to the fact that the limit $\tau \to 0$ (implying $\omega \to 1$) may be singular, as for example pointed out in [214] in terms of stability of a neutral *delay differential equation* (DDE).

Corollary 3.27. *For $m = n_x$ consider the descriptor system* (3.22). *Then there exists a feedback matrix $F \in \mathbb{R}^{n_u \times n_y}$ such that $(sE - A_1, BFC)$ is delay-regular if and only if*

$$\operatorname{rank}\begin{bmatrix} \lambda E - A_1 & B \end{bmatrix} = \operatorname{rank}\begin{bmatrix} \lambda E - A_1 \\ C \end{bmatrix} = n_x$$

for some $\lambda \in \mathbb{C}$.

Proof. The proof follows from Theorem 3.25 and [76, Thm. 2]. ■

3.3 Delay-equivalence and the compress-and-shift algorithm

In the proof of the delay-regularity of $(\widehat{\mathscr{P}}_{23}(s), \widehat{\mathscr{Q}}_{23})$ in Theorem 3.17 (i), we used Lemma 3.19 to construct a solution of the associated ITP. The key idea in the proof of Lemma 3.19 is to shift the equations to obtain a transformed DDAE $(\widetilde{\mathscr{P}}(s), \widetilde{\mathscr{Q}}(s))$ with $\det(\widetilde{\mathscr{P}}(s)) \not\equiv 0$. To formalize this idea, we introduce the notion of delay-equivalence.

Definition 3.28. Two pairs of matrix polynomials $(\mathscr{P}(s), \mathscr{Q}(s)), (\widetilde{\mathscr{P}}(s), \widetilde{\mathscr{Q}}(s)) \in (\mathbb{R}[s]^{m\times n_x})^2$ are called *delay-equivalent* if and only if there exists a bijective map $\mathscr{T} : \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty} \to \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty}$ such that for all $(x, f) \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty} \times \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty}$ and $\widetilde{f} := \mathscr{T} f$ the equivalence

$$\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x = \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + f \qquad \Longleftrightarrow \qquad \widetilde{\mathscr{P}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x = \widetilde{\mathscr{Q}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + \widetilde{f}$$

holds. In this case we write $(\mathscr{P}(s), \mathscr{Q}(s)) \overset{\mathrm{d}}{\sim} (\widetilde{\mathscr{P}}(s), \widetilde{\mathscr{Q}}(s))$. We say that two DDAEs are delay-equivalent if the associated matrix polynomials are delay-equivalent.

It is easy to verify that delay-equivalence is indeed an equivalence relation. We note that delay-equivalence is a property of a distributional DDAE and delay-regularity is a property of an ITP. Hence we first have to establish a relation between delay-equivalence and delay-regularity.

Proposition 3.29. *Consider pairs of matrix polynomials* $(\mathscr{P}(s), \mathscr{Q}(s)), (\widetilde{\mathscr{P}}(s), \widetilde{\mathscr{Q}}(s)) \in (\mathbb{R}[s]^{n_x\times n_x})^2$ *and assume* $(\mathscr{P}(s), \mathscr{Q}(s)) \overset{\mathrm{d}}{\sim} (\widetilde{\mathscr{P}}(s), \widetilde{\mathscr{Q}}(s))$. *Then the pair* $(\mathscr{P}(s), \mathscr{Q}(s))$ *is delay-regular if and only if the pair* $(\widetilde{\mathscr{P}}(s), \widetilde{\mathscr{Q}}(s))$ *is delay-regular.*

Proof. First, let $(\widetilde{\mathscr{P}}(s), \widetilde{\mathscr{Q}}(s))$ be delay-regular. Assume that $(\mathscr{P}(s), \mathscr{Q}(s))$ is not delay-regular. Let $(\mathscr{U}(s), \mathscr{V}(s))$ denote the matrices from Theorem 2.24 that transform $(\mathscr{P}(s), \mathscr{Q}(s))$ to the condensed form (2.27), i.e.,

$$\mathscr{U}(s)\mathscr{P}(s)\mathscr{V}(s) = \begin{bmatrix} \widehat{\mathscr{P}}_{11}(s) & 0 & \widehat{\mathscr{P}}_{13}(s) \\ 0 & 0 & \widehat{\mathscr{P}}_{23}(s) \\ 0 & 0 & \widehat{\mathscr{P}}_{33}(s) \end{bmatrix} \quad \text{and} \quad \mathscr{U}(s)\mathscr{Q}(s)\mathscr{V}(s) = \begin{bmatrix} \widehat{\mathscr{Q}}_{11}(s) & \widehat{\mathscr{Q}}_{12}(s) & \widehat{\mathscr{Q}}_{13}(s) \\ 0 & 0 & \widehat{\mathscr{Q}}_{23}(s) \\ 0 & 0 & \widehat{\mathscr{Q}}_{33}(s) \end{bmatrix},$$

where $\widehat{\mathscr{P}}_{11}(s)$ is a nonsingular diagonal matrix, $\widehat{\mathscr{P}}_{23}(s), \widehat{\mathscr{P}}_{33}(s), \widehat{\mathscr{Q}}_{33}(s)$ are block upper triangular matrices with zero diagonal blocks and $\widehat{\mathscr{Q}}_{23}(s)$ is a nonsingular block upper triangular matrix. Note that $m = n_x$ implies that the second block column and the third block row are both present. Let $\tilde{f} \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ with support in $[0,\infty)$. Since $(\widetilde{\mathscr{P}}(s), \widetilde{\mathscr{Q}}(s))$ is delay-regular, there exists $\tilde{x} \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ solving the DDAE

$$\widetilde{\mathscr{P}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\tilde{x} = \widetilde{\mathscr{Q}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau \tilde{x} + \tilde{f}. \tag{3.26}$$

By assumption $\tilde{x}$ is a solution of

$$\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\tilde{x} = \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau \tilde{x} + \mathscr{T}^{-1}\tilde{f}.$$

Define

$$x := \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} := \mathscr{V}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^{-1} \tilde{x} \qquad \text{and} \qquad \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} := \mathscr{U}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\mathscr{T}^{-1}\tilde{f}.$$

Define $\hat{x}_2 := x_2 + \mathbb{1}_{[1,\infty)}$ and let $\hat{x}_1$ be the unique solution (cf. Theorem 3.15) of the ITP

$$\begin{aligned} \left(\hat{x}_1\right)_{(-\infty,0)} &= \left(x_1\right)_{(-\infty,0)}, \\ \left(\widehat{\mathscr{P}}_{11}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\hat{x}_1\right)_{[0,\infty)} &= \left(\widehat{\mathscr{Q}}_{11}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau \hat{x}_1 + \widehat{\mathscr{Q}}_{12}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau \hat{x}_2 + \widehat{\mathscr{Q}}_{13}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x_3 - \widehat{\mathscr{P}}_{13}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x_3 + f_1\right)_{[0,\infty)}. \end{aligned}$$

Define

$$\check{x} := \mathscr{V}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ x_3 \end{bmatrix}.$$

By construction (and Lemma 2.22) we obtain $\check{x}_{(-\infty,0)} = \hat{x}_{(-\infty,0)}$, $\check{x} \neq \hat{x}$, and

$$\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\check{x} = \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau \check{x} + \mathscr{T}^{-1}\tilde{f}.$$

Delay-equivalence implies that $\check{x}$ is another solution of the ITP (3.26) contradicting the delay-regularity of $(\widetilde{\mathscr{P}}(s), \widetilde{\mathscr{Q}}(s))$. Thus $(\mathscr{P}(s), \mathscr{Q}(s))$ is delay-regular. Interchanging the roles of $(\widetilde{\mathscr{P}}(s), \widetilde{\mathscr{Q}}(s))$ and $(\mathscr{P}(s), \mathscr{Q}(s))$ shows the converse direction. ■

Analyzing the situation in Lemma 3.19, respectively in Theorem 3.17 (ii) showcases that we shift equations if they depend solely on delayed variables. If the equations are not yet in such a form, one first has to transform them, for instance with a rank-revealing decomposition of $\mathscr{P}(s)$.

Lemma 3.30. *For* $(\mathscr{P}(s), \mathscr{Q}(s)) \in (\mathbb{R}[s]^{m\times n_x})^2$ *and inhomogeneity* $f \in \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty}$ *choose a unimodular matrix* $\mathscr{U}(s) = \left[\begin{smallmatrix} \mathscr{U}_1(s) \\ \mathscr{U}_2(s) \end{smallmatrix}\right] \in \mathbb{R}[s]^{m\times m}$ *such that*

$$\mathscr{U}(s)\mathscr{P}(s) =: \begin{bmatrix} \mathscr{P}_1(s) \\ 0 \end{bmatrix}, \quad \mathscr{U}(s)\mathscr{Q}(s) =: \begin{bmatrix} \mathscr{Q}_1(s) \\ \mathscr{Q}_2(s) \end{bmatrix},$$

where $\mathscr{P}_1(s), \mathscr{Q}_1(s) \in \mathbb{R}[s]^{k\times n_x}$ *with* $\operatorname{rank}_{\mathbb{R}[s]}\left(\mathscr{P}_1(s)\right) = k$*. Then the DDAE* (3.9) *is delay-equivalent to the* partially time-shifted *DDAE*

$$\begin{aligned} \mathscr{P}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x &= \mathscr{Q}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + f_1, \\ \mathscr{Q}_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x &= -\sigma_{-\tau} f_2 \end{aligned} \tag{3.27}$$

with $\left[\begin{smallmatrix} f_1 \\ f_2 \end{smallmatrix}\right] := \mathscr{U}(\frac{\mathrm{d}}{\mathrm{d}t})f$*. In particular,* $(\mathscr{P}(s), \mathscr{Q}(s))$ *is delay-regular if and only if* $\left(\left[\begin{smallmatrix} \mathscr{P}_1(s) \\ \mathscr{Q}_2(s) \end{smallmatrix}\right], \left[\begin{smallmatrix} \mathscr{Q}_1(s) \\ 0 \end{smallmatrix}\right]\right)$ *is delay-regular.*

Proof. Assume that $(x,f) \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty} \times \mathscr{D}^{m}_{\mathrm{pw}\mathscr{C}^\infty}$ satisfies the DDAE (3.9). Multiplication of (3.9) from the left with $\mathscr{U}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)$ shows that $\left(x, \mathscr{U}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f\right)$ solves

$$\begin{aligned} \mathscr{P}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x &= \mathscr{Q}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + f_1, \\ 0 &= \mathscr{Q}_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + f_2. \end{aligned} \tag{3.28}$$

Since $\mathscr{U}(s)$ is unimodular we can reverse the transformation such that (3.28) is delay-equivalent to (3.9). Applying a negative time-shift $\sigma_{-\tau}$ on the second equation and taking into account that differentiation and shifting commute (cf. Lemma 3.1) we obtain the DDAE

$$\begin{aligned} \mathscr{P}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x &= \mathscr{Q}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + f_1 \\ \mathscr{Q}_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x &= \bar{f}_2 \end{aligned}$$

where $\bar{f}_2 := -\sigma_{-\tau} f_2$. Clearly, the transformation of the inhomogeneity is reversible, hence this DDAE is delay-equivalent to (3.9). ■

The previous results suggest to perform a simple *compress-and-shift* algorithm (see for instance [53,206]) to determine whether a DDAE is delay-regular or not: If $\mathscr{P}(s)$ is rank deficient, perform a row compression of $\mathscr{P}(s)$. If $\mathscr{Q}_2(s)$ is rank deficient, then the DDAE is not delay-regular. Otherwise shift $\mathscr{Q}_2(s)$ and restart the algorithm with the transformed matrix polynomials. The details are outlined in Algorithm 1. Lemma 3.30 ensures that Algorithm 1 constructs a sequence of delay-equivalent polynomial matrix pairs $(\mathscr{P}_\nu(s), \mathscr{Q}_\nu(s)) \in (\mathbb{R}[s]^{m \times n_x})^2$.

Lemma 3.31. *Assume that Algorithm 1 terminates after* ν *iterations for the polynomial matrix pair* $(\mathscr{P}(s), \mathscr{Q}(s)) \in \mathbb{R}[s]^{n_x \times n_x}$. *Then the pair of polynomial matrices* $(\mathscr{P}(s), \mathscr{Q}(s))$ *is delay-regular if and only if* $k_\nu = n_x$.

Proof. If $k_\nu = n_x$, then $\det(\mathscr{P}_\nu(s)) \not\equiv 0$ and thus Theorem 3.15 implies that $(\mathscr{P}_\nu(s), \mathscr{Q}_\nu(s))$ is delay-regular. Using Proposition 3.29 we conclude that $(\mathscr{P}(s), \mathscr{Q}(s))$ is delay-regular. Conversely, assume $k_\nu < n_x$. Then

$$\operatorname{rank}_{\mathbb{R}[s,w]}\left(\mathscr{P}_\nu(s) - \omega \mathscr{Q}_\nu(s)\right) = \operatorname{rank}_{\mathbb{R}[s,w]}\left(\begin{bmatrix} \mathscr{P}_{\nu,1}(s) \\ 0 \end{bmatrix} - \omega \begin{bmatrix} \mathscr{Q}_{\nu,1}(s) \\ 0 \end{bmatrix}\right) = k_\nu < n_x.$$

Thus, Theorem 3.20 implies that $(\mathscr{P}_\nu(s), \mathscr{Q}_\nu(s))$ is not delay-regular, which completes the proof. ■

Example 3.32. Consider the matrix polynomials

$$\mathscr{P}(s) = \begin{bmatrix} s^2 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathscr{Q}(s) = \begin{bmatrix} 0 & s-1 \\ s & 0 \end{bmatrix}.$$

Algorithm 1 Compress-and-shift

```
   Input: $\mathscr{P}(s), \mathscr{Q}(s) \in \mathbb{R}[s]^{m\times n_x}$
1: Set $\nu = 1$ and $\mathscr{P}_\nu(s) := \mathscr{P}(s)$, $\mathscr{Q}_\nu(s) := \mathscr{Q}(s)$.
2: Choose unimodular
   $$\mathscr{U}_\nu(s) = \begin{bmatrix} \mathscr{U}_{\nu,1}(s) \\ \mathscr{U}_{\nu,2}(s) \\ \mathscr{U}_{\nu,3}(s) \end{bmatrix} \in \mathbb{R}[s]^{m\times m} \quad \text{with matrix polynomials} \begin{cases} \mathscr{U}_{\nu,1}(s) \in \mathbb{R}[s]^{k_\nu \times m}, \\ \mathscr{U}_{\nu,2}(s) \in \mathbb{R}[s]^{\rho_\nu \times m}, \\ \mathscr{U}_{\nu,3}(s) \in \mathbb{R}[s]^{m-k_\nu-\rho_\nu \times m}, \end{cases} \qquad (3.29)$$
   such that
   $$\mathscr{U}_\nu(s)\mathscr{P}_\nu(s) =: \begin{bmatrix} \mathscr{P}_{\nu,1}(s) \\ 0 \\ 0 \end{bmatrix} \qquad \text{and} \qquad \mathscr{U}_\nu(s)\mathscr{Q}_\nu(s) =: \begin{bmatrix} \mathscr{Q}_{\nu,1}(s) \\ \mathscr{Q}_{\nu,2}(s) \\ 0 \end{bmatrix},$$
   where $\mathscr{P}_{\nu,1}(s) := \mathscr{U}_{\nu,1}(s)\mathscr{P}_\nu(s) \in \mathbb{R}[s]^{k_\nu \times n_x}$ and $\mathscr{Q}_{\nu,2}(s) := \mathscr{U}_{\nu,2}(s)\mathscr{Q}_\nu(s) \in \mathbb{R}[s]^{\rho_\nu \times n_x}$ have full row rank.
3: if $\rho_\nu = 0$ then
4:     terminate
5: else
6:     Set $\mathscr{P}_{\nu+1}(s) = \begin{bmatrix} \mathscr{P}_{\nu,1}(s) \\ \mathscr{Q}_{\nu,2}(s) \end{bmatrix}$ and $\mathscr{Q}_{\nu+1}(s) = \begin{bmatrix} \mathscr{Q}_{\nu,1}(s) \\ 0 \end{bmatrix}$.
7:     Set $\nu \leftarrow \nu + 1$.
8:     Go to Line 2
9: end if
```

Applying Algorithm 1 to $(\mathscr{P}(s), \mathscr{Q}(s))$ yields $\mathscr{U}_1(s) = I$ with $\mathscr{Q}_{1,2}(s) = \begin{bmatrix} s & 0 \end{bmatrix}$ in Line 2. We obtain

$$\mathscr{P}_2(s) = \begin{bmatrix} s^2 & 0 \\ s & 0 \end{bmatrix}, \qquad \mathscr{Q}_2(s) = \begin{bmatrix} 0 & s-1 \\ 0 & 0 \end{bmatrix}, \qquad \mathscr{U}_2(s) = \begin{bmatrix} 0 & 1 \\ 1 & -s \end{bmatrix},$$

and $\rho_2 = 1$. The new matrix polynomials

$$\mathscr{P}_3(s) = \begin{bmatrix} s & 0 \\ 0 & s-1 \end{bmatrix} \quad \text{and} \quad \mathscr{Q}_3(s) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

satisfy $\operatorname{rank}_{\mathbb{R}[s]}\left(\mathscr{P}_3(s)\right) = 2 = n_x$ and thus Lemma 3.31 implies that $(\mathscr{P}(s), \mathscr{Q}(s))$ is delay-regular. ♠

Example 3.33. Applying Algorithm 1 to the matrix polynomials

$$\mathscr{P}(s) = \begin{bmatrix} s & 1 & 0 \\ s & 0 & s^2 \\ s^2 & s & 0 \end{bmatrix} \quad \text{and} \quad \mathscr{Q}(s) = \begin{bmatrix} 0 & 0 & s \\ 0 & 0 & 0 \\ -s & 0 & 0 \end{bmatrix}$$

results in the sequence

$$\mathscr{U}_1(s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -s & 0 & 1 \end{bmatrix}, \qquad \mathscr{U}_1(s)\mathscr{P}_1(s) = \begin{bmatrix} s & 1 & 0 \\ s & 0 & s^2 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \mathscr{U}_1(s)\mathscr{Q}_1(s) = \begin{bmatrix} 0 & 0 & s \\ 0 & 0 & 0 \\ -s & 0 & -s^2 \end{bmatrix}$$

$$\mathscr{U}_2(s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \qquad \mathscr{U}_2(s)\mathscr{P}_2(s) = \begin{bmatrix} s & 1 & 0 \\ s & 0 & s^2 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \mathscr{U}_2(s)\mathscr{Q}_2(s) = \begin{bmatrix} 0 & 0 & s \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

such that Algorithm 1 terminates with $v = 2$. From Lemma 3.31 we deduce that $(\mathscr{P}(s), \mathscr{Q}(s))$ is not delay-regular. ♠

The important assumption in Lemma 3.31 is that Algorithm 1 terminates after a finite number of steps. We immediately observe that by construction

$$\operatorname{rank}_{\mathbb{R}[s]}(\mathscr{P}_{v+1}(s)) \geq \operatorname{rank}_{\mathbb{R}[s]}(\mathscr{P}_v(s)), \tag{3.30a}$$

$$\operatorname{rank}_{\mathbb{R}[s]}(\mathscr{Q}_{v+1}(s)) \leq \operatorname{rank}_{\mathbb{R}[s]}(\mathscr{Q}_v(s)). \tag{3.30b}$$

If in each iteration of Algorithm 1 one of these inequalities is strict, then Algorithm 1 terminates after a finite number of iterations. Or equivalently, Algorithm 1, does not terminate if and only if after finitely many iterations the ranks in (3.30) remain constant in all further iterations of the algorithm. The following example shows that this indeed can happen.

Example 3.34. Consider the input data

$$\mathscr{P}(s) = \begin{bmatrix} s & 1 \\ s^2 & s \end{bmatrix}, \quad \mathscr{Q}(s) = \begin{bmatrix} -s & -1 \\ 0 & 0 \end{bmatrix},$$

for Algorithm 1. We obtain

$$\mathscr{U}_1(s) = \begin{bmatrix} 1 & 0 \\ -s & 1 \end{bmatrix}, \qquad \mathscr{P}_2(s) = \begin{bmatrix} s & 1 \\ s^2 & s \end{bmatrix} = \mathscr{P}(s), \qquad \mathscr{Q}_2(s) = \begin{bmatrix} -s & -1 \\ 0 & 0 \end{bmatrix} = \mathscr{Q}(s).$$

Hence each iteration of Algorithm 1 works with the same pair of polynomial matrices and consequently Algorithm 1 does not terminate. ♠

From this example one may conjecture that once the ranks in (3.30) remain constant, they also will remain constant in future iterations so that at least the algorithm can be terminated with a warning. Unfortunately, this is not true as the following example shows.

Example 3.35. Applying Algorithm 1 to the matrices

$$\mathscr{P}(s) = \begin{bmatrix} s & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathscr{Q}(s) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

yields in the first iteration

$$\mathscr{U}_1(s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \mathscr{P}_2(s) = \begin{bmatrix} s & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathscr{Q}_2(s) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Note that neither of the rank inequalities in (3.30) is strict. However, we continue with

$$\mathscr{U}_2(s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}, \qquad \mathscr{P}_3(s) = \begin{bmatrix} s & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \qquad \mathscr{Q}_3(s) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and conclude that $(\mathscr{P}(s), \mathscr{Q}(s))$ is delay-regular. ♠

Another issue with Algorithm 1 is that the rank-revealing decomposition (3.29) is not unique and that the non-uniqueness of $\mathscr{U}_\nu(s)$ may influence the termination of Algorithm 1 as the following examples illustrates.

Example 3.36. Consider the matrix polynomials

$$\mathscr{P}(s) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad \mathscr{Q}(s) = \begin{bmatrix} -1 & -1 \\ 0 & 0 \end{bmatrix}.$$

Picking $\mathscr{U}_\nu(s) = \left[\begin{smallmatrix} 1 & 0 \\ -1 & 1 \end{smallmatrix}\right]$ we obtain for all $\nu \in N$ the equality $(\mathscr{P}_\nu(s), \mathscr{Q}_\nu(s)) = (\mathscr{P}(s), \mathscr{Q}(s))$ and consequently $\rho_\nu = 1$ for all $\nu \in N$. If we use $\mathscr{U}_1(s) = \left[\begin{smallmatrix} 0 & 1 \\ 1 & -1 \end{smallmatrix}\right]$ we obtain

$$\mathscr{P}_2(s) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad \mathscr{Q}_2(s) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

and thus $\rho_3 = 0$ and thus Algorithm 1 terminates with $k_2 = 1 < 2 = n_x$. ♠

To summarize the previous discussion there are two major issues with Algorithm 1:

(i) It may happen that the algorithm does not terminate and

(ii) the choice of $\mathscr{U}_\nu(s)$ may influence the termination of the algorithm.

From (3.30) we immediately observe that the algorithm fails to terminate if and only if there exists an index $\tilde{\nu} \in \mathbb{N}$ such that

$$\operatorname{rank}_{\mathbb{R}[s]}(\mathscr{P}_\nu(s)) = \operatorname{rank}_{\mathbb{R}[s]}(\mathscr{P}_{\tilde{\nu}}(s)), \tag{3.31a}$$

$$\operatorname{rank}_{\mathbb{R}[s]}(\mathscr{Q}_\nu(s)) = \operatorname{rank}_{\mathbb{R}[s]}(\mathscr{Q}_{\tilde{\nu}}(s)) \tag{3.31b}$$

for all $\nu \geq \tilde{\nu}$ and $\operatorname{rank}_{\mathbb{R}[s]}(\mathscr{P}_\nu(s)) < m$. The following theorem details that this cannot happen if the DDAE (3.9) is delay-regular.

Theorem 3.37. *Algorithm 1 terminates for any delay-regular DDAE* (3.9). *In particular, the DDAE* (3.9) *is delay-equivalent to a DDAE*

$$\widetilde{\mathscr{P}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x = \widetilde{\mathscr{Q}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + \widetilde{f}$$

with $\det(\widetilde{\mathscr{P}}(s)) \not\equiv 0$.

In order to prove Theorem 3.37 we observe that it suffices to apply Algorithm 1 directly to the condensed polynomial matrices (2.27). Since we assume that the DDAE (3.9) is delay-regular, we can simplify (2.27) as follows.

Lemma 3.38. *Consider a delay-regular pair of matrix polynomials* $(\mathscr{P}(s), \mathscr{Q}(s)) \in (\mathbb{R}[s]^{n_x \times n_x})^2$. *There exist unimodular matrices* $\mathscr{U}(s), \mathscr{V}(s) \in \mathbb{R}[s]^{n_x \times n_x}$ *such that*

$$\mathscr{U}(s)\mathscr{P}(s)\mathscr{V}(s) = \begin{bmatrix} \mathscr{P}_{1,1}(s) & \mathscr{P}_{1,2}(s) & \cdots & \cdots & \mathscr{P}_{1,k}(s) \\ 0 & 0 & \mathscr{P}_{2,3}(s) & \cdots & \mathscr{P}_{2,k}(s) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \mathscr{P}_{k-1,k}(s) \\ 0 & 0 & \cdots & \cdots & 0 \end{bmatrix}, \tag{3.32a}$$

$$\mathscr{U}(s)\mathscr{Q}(s)\mathscr{V}(s) = \begin{bmatrix} \mathscr{Q}_{1,1}(s) & \mathscr{Q}_{1,2}(s) & \cdots & \mathscr{Q}_{1,k}(s) \\ 0 & \mathscr{Q}_{2,2}(s) & \cdots & \mathscr{Q}_{2,k}(s) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathscr{Q}_{k,k}(s) \end{bmatrix} \tag{3.32b}$$

where $\mathscr{P}_{1,1}(s)$, $\mathscr{Q}_{i,i}(s)$ *for* $i = 2, \ldots, k$ *are nonsingular and the matrices* $\mathscr{P}_{i,i+1}(s)$ *have full row rank for* $i = 2, \ldots, k-1$.

Proof. The form (3.32) follows directly from Theorem 2.24 and Theorem 3.17 (ii) with nonsingular blocks $\mathscr{P}_{1,1}(s)$, $\mathscr{Q}_{i,i}(s)$ for $i = 2, \ldots, k$. Let $j \in \{2, \ldots, k-1\}$ denote the largest number such that the polynomial matrix $\mathscr{P}_{j,j+1}(s)$ does not have full row rank. Then there exists a unimodular matrix $\mathscr{U}_j(s)$ such that

$$\mathscr{U}_j(s)\mathscr{P}_{j,j+1}(s) = \begin{bmatrix} \widehat{\mathscr{P}}_{j,j+1}(s) \\ 0 \end{bmatrix}$$

where $\widehat{\mathscr{P}}_{j,j+1}(s)$ has full row rank. Since $\mathscr{Q}_{j,j}(s)$ is nonsingular, there exists a unimodular matrix $\mathscr{V}_j(s)$ such that

$$\mathscr{U}_j(s)\mathscr{Q}_{j,j}(s)\mathscr{V}_j(s) = \begin{bmatrix} \check{\mathscr{Q}}_{j,j}(s) & \check{\mathscr{Q}}_{j,j}(s) \\ 0 & \tilde{\mathscr{Q}}_{j,j}(s) \end{bmatrix}$$

with nonsingular matrices $\check{\mathscr{Q}}_{j,j}(s)$ and $\tilde{\mathscr{Q}}_{j,j}(s)$. For the corresponding sub block matrices in (3.32) we thus obtain

$$\begin{bmatrix} \mathscr{U}_j(s) & \\ & \mathrm{Id} \end{bmatrix} \left[\begin{array}{c|c} 0 & \mathscr{P}_{j,j+1}(s) \\ \hline & 0 \end{array}\right] \begin{bmatrix} \mathscr{V}_j(s) & \\ & \mathrm{Id} \end{bmatrix} = \left[\begin{array}{c|cc} 0 & 0 & \widehat{\mathscr{P}}_{j,j+1}(s) \\ \hline & 0 & 0 \\ & 0 & 0 \end{array}\right]$$

and

$$\begin{bmatrix} \mathscr{U}_j(s) & \\ & \mathrm{Id} \end{bmatrix} \left[\begin{array}{c|c} \mathscr{Q}_{j,j}(s) & \mathscr{Q}_{j,j+1}(s) \\ \hline & \mathscr{Q}_{j+1,j+1}(s) \end{array}\right] \begin{bmatrix} \mathscr{V}_j(s) & \\ & \mathrm{Id} \end{bmatrix} = \left[\begin{array}{c|cc} \check{\mathscr{Q}}_{j,j}(s) & \check{\mathscr{Q}}_{j,j}(s) & \check{\mathscr{Q}}_{j,j+1}(s) \\ \hline 0 & \tilde{\mathscr{Q}}_{j,j}(s) & \tilde{\mathscr{Q}}_{j,j+1}(s) \\ & 0 & \mathscr{Q}_{j+1,j+1}(s) \end{array}\right],$$

where clearly the blocks have the desired properties. Repeating this procedure for the remaining rank defective blocks yields the desired result. ■

Proof of Theorem 3.37. It suffices to show that we can ensure that the situation in (3.31) cannot happen. Lemma 3.38 implies that applying Algorithm 1 to (3.32) yield a shift of the last block row. After shifting the last block row in (3.32), we observe that the compression step affects only the last two rows. Since $\mathscr{P}_{k-1,k}(s)$ has full row rank, this implies that there exists a unimodular matrix

$$\widehat{\mathscr{U}}(s) = \begin{bmatrix} \hat{\mathscr{U}}_1(s) & \hat{\mathscr{U}}_2(s) \\ \hat{\mathscr{U}}_3(s) & \hat{\mathscr{U}}_4(s) \end{bmatrix} \qquad \text{with} \qquad \widehat{\mathscr{U}}(s) \begin{bmatrix} \mathscr{P}_{k-1,k}(s) \\ \mathscr{Q}_{k,k}(s) \end{bmatrix} = \begin{bmatrix} 0 \\ \widehat{\mathscr{Q}}_{k,k}(s) \end{bmatrix}.$$

Since $\mathscr{P}_{k-1,k}(s)$ has full row rank and $\mathscr{Q}_{k,k}(s)$ is nonsingular, we conclude that $\hat{\mathscr{U}}_1(s)$ has full row rank implying that $\hat{\mathscr{U}}_1(s)\mathscr{Q}_{k-1,k-1}(s)$ is nonsingular. We can thus repeat the above procedure with the submatrices that are obtained by removing the last block column and last block row. Proceeding iteratively, we conclude that Algorithm 1 terminates after $k-1$ shifts. ■

Remark 3.39. Suppose that the DDAE (3.1a) is delay-regular. Then Theorem 3.37 implies that Algorithm 1 constructs a delay-equivalent DDAE

$$\widetilde{\mathscr{P}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x = \widetilde{\mathscr{Q}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + \widetilde{f}$$

with $\det(\widetilde{\mathscr{P}}(s)) \not\equiv 0$. Performing a first-order reformulation as for instance in the proof of Theorem 3.15 yields a DDAE

$$\mathscr{E}\dot{z} = \mathscr{A}_1 z + \mathscr{A}_2 \sigma_\tau z + \mathscr{F}$$

with regular matrix pencil $(\mathscr{E}, \mathscr{A}_1)$. ♣

Let us emphasize that shifting enlarges the set of feasible initial trajectories for the ITP, and thus solutions of the ITP for the partially time-shifted DDAE (3.27) may not be solutions of the original ITP, see for instance [212, Ex. 4.7].

Lemma 3.40. *Let the notation be as in Lemma 3.30 and set*

$$\widetilde{\mathscr{P}}(s) := \begin{bmatrix} \mathscr{P}_1(s) \\ \mathscr{Q}_2(s) \end{bmatrix}, \quad \widetilde{\mathscr{Q}}(s) := \begin{bmatrix} \mathscr{Q}_1(s) \\ 0 \end{bmatrix}.$$

For $f \in \mathscr{D}^m_{\mathrm{pw}\mathscr{C}^\infty}$ and $x^0 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ define

$$f_{\mathrm{ITP}} := f_{[0,\infty)} + \left(\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x^0 - \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x^0\right)_{(-\infty,0)}, \quad \tilde{f}_{\mathrm{ITP}} := \begin{bmatrix} \mathscr{U}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) f_{\mathrm{ITP}} \\ -\sigma_{-\tau}\mathscr{U}_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) f_{\mathrm{ITP}} \end{bmatrix}, \quad \tilde{f} := (\tilde{f}_{\mathrm{ITP}})_{[0,\infty)}.$$

Assume that $x \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$ is a solution of the ITP

$$\begin{aligned} x_{(-\infty,0)} &= x^0_{(-\infty,0)}, \\ \left(\widetilde{\mathscr{P}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x\right)_{[0,\infty)} &= \left(\widetilde{\mathscr{Q}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + \tilde{f}\right)_{[0,\infty)}. \end{aligned}$$

If x_0 satisfies

$$\left(\mathscr{U}_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x^0\right)_{[0,\tau)} = -\left(\mathscr{U}_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f_{\mathrm{ITP}}\right)_{[0,\tau)}, \tag{3.33}$$

then x is a solution of the ITP (3.10).

Proof. From the definition of $\tilde{f}$ we immediately obtain

$$(\tilde{f}_{\mathrm{ITP}})_{[0,\infty)} = \tilde{f} = \tilde{f}_{[0,\infty)}.$$

Further, Lemma 2.22 implies

$$\begin{aligned} \left(\mathscr{U}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f_{\mathrm{ITP}}\right)_{(-\infty,0)} &= \left(\mathscr{P}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)x^0 - \mathscr{Q}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x^0\right)_{(-\infty,0)}, \\ \left(\mathscr{U}_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f_{\mathrm{ITP}}\right)_{(-\infty,0)} &= -\left(\mathscr{Q}_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x^0\right)_{(-\infty,0)}, \end{aligned}$$

such that (3.33) yields

$$\left(\mathscr{U}_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)f_{\mathrm{ITP}}\right)_{(-\infty,\tau)} = -\left(\mathscr{Q}_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x^0\right)_{(-\infty,\tau)}.$$

Thus Lemma 3.2 implies

$$\begin{aligned}(\tilde{f}_{\mathrm{ITP}})_{(-\infty,0)} &= \begin{bmatrix} \mathscr{U}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) f_{\mathrm{ITP}} \\ -\sigma_{-\tau}\mathscr{U}_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) f_{\mathrm{ITP}} \end{bmatrix}_{(-\infty,0)} = \begin{bmatrix} \mathscr{P}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x^0 - \mathscr{Q}_1\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x^0 \\ \mathscr{Q}_2\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x^0 \end{bmatrix}_{(-\infty,0)} \\ &= \left(\widetilde{\mathscr{P}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x - \widetilde{\mathscr{Q}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x^0\right)_{(-\infty,0)}.\end{aligned}$$

We conclude that x satisfies

$$\widetilde{\mathscr{P}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x = \widetilde{\mathscr{Q}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + \tilde{f}_{\mathrm{ITP}}.$$

Delay-equivalence thus implies that x solves

$$\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x = \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x + f_{\mathrm{ITP}}$$

and thus is a solution of the ITP (3.10). ■

Applying Lemma 3.40 several times yields, together with Theorems 3.15 and 3.37, the following sufficient condition for an initial trajectory to be consistent.

Theorem 3.41. *Consider the ITP* (3.10) *with delay-regular pair* $(\mathscr{P}(s),\mathscr{Q}(s)) \in (\mathbb{R}[s]^{n_x\times n_x})^2$*, external forcing* $f \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$*, and* $x^0 \in \mathscr{D}^{n_x}_{\mathrm{pw}\mathscr{C}^\infty}$*. Let Algorithm 1 applied to* $(\mathscr{P}(s),\mathscr{Q}(s))$ *terminate after* $\nu \in \mathbb{N}$ *iterations. Define*

$$\begin{aligned} f_{\mathrm{ITP},1} &:= f_{[0,\infty)} + \left(\mathscr{P}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) x^0 - \mathscr{Q}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\sigma_\tau x^0\right)_{(-\infty,0)}, \\ f_{\mathrm{ITP},k} &:= \begin{bmatrix} \mathscr{U}_{k,1}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) f_{\mathrm{ITP},k-1} \\ -\sigma_\tau \mathscr{U}_{k,2}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) f_{\mathrm{ITP},k-1} \end{bmatrix} \end{aligned}$$

for $k = 2,\dots,\nu-1$*. If* x^0 *satisfies*

$$\left(\mathscr{U}_{k,2}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) f_{\mathrm{ITP,k}}\right)_{[0,\tau)} = -\left(\mathscr{U}_{k,2}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)\mathscr{Q}_k(\frac{\mathrm{d}}{\mathrm{d}t})\sigma_\tau x^0\right)_{[0,\tau)},$$

for $k = 1,\dots,\nu-1$*, then* x *is a solution of the ITP* (3.10).

It remains to analyze the situation what happens with Algorithm 1 when the DDAE is not delay-regular. Checking the proof of Theorem 3.37 reveals that Algorithm 1 terminates also in the case that the second block column in (2.27) is present. Thus, the only reason for non-termination is hidden in the third block row in (2.27). A modification of Algorithm 1 to prevent non-termination must thus be able to recognize this case. We have already seen in Example 3.35 that it is not sufficient to terminate Algorithm 1 whenever the ranks in (3.30) do not change from one iteration to another. This can also be seen from the proof of Theorem 3.37. However, we observe that the image of $\mathscr{Q}_{\nu,2}(s)$ is different for every ν in the delay-regular case whenever the ranks remain constant. It thus suffices to check that the matrix

$$\begin{bmatrix} \mathscr{Q}_{1,2}(s) \\ \mathscr{Q}_{2,2}(s) \\ \vdots \\ \mathscr{Q}_{\nu,2}(s) \end{bmatrix}$$

has full row rank. A modified version of Algorithm 1 is presented in Algorithm 2.

Algorithm 2 Compress-and-shift (modified)

```
    Input: $\mathscr{P}(s), \mathscr{Q}(s) \in \mathbb{R}[s]^{m\times n_x}$

 1: Set $\nu = 1$ and $\mathscr{P}_\nu(s) := \mathscr{P}(s)$, $\mathscr{Q}_\nu(s) := \mathscr{Q}(s)$.
 2: Set $\mathscr{K}(s) := [] \in \mathbb{R}[s]^{0\times n_x}$
 3: Choose unimodular matrix
    $$\mathscr{U}_\nu(s) = \begin{bmatrix} \mathscr{U}_{\nu,1}(s) \\ \mathscr{U}_{\nu,2}(s) \\ \mathscr{U}_{\nu,3}(s) \end{bmatrix} \in \mathbb{R}[s]^{m\times m} \quad \text{with matrix polynomials} \begin{cases} \mathscr{U}_{\nu,1}(s) \in \mathbb{R}[s]^{k_\nu\times m}, \\ \mathscr{U}_{\nu,2}(s) \in \mathbb{R}[s]^{\rho_\nu\times m}, \\ \mathscr{U}_{\nu,3}(s) \in \mathbb{R}[s]^{m-k_\nu-\rho_\nu\times m}, \end{cases}$$
    such that
    $$\mathscr{U}_\nu(s)\mathscr{P}_\nu(s) =: \begin{bmatrix} \mathscr{P}_{\nu,1}(s) \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathscr{U}_\nu(s)\mathscr{Q}_\nu(s) =: \begin{bmatrix} \mathscr{Q}_{\nu,1}(s) \\ \mathscr{Q}_{\nu,2}(s) \\ 0 \end{bmatrix},$$
    where $\mathscr{P}_{\nu,1}(s) := \mathscr{U}_{\nu,1}(s)\mathscr{P}_\nu(s) \in \mathbb{R}[s]^{k_\nu\times n_x}$ and $\mathscr{Q}_{\nu,2}(s) := \mathscr{U}_{\nu,2}(s)\mathscr{Q}_\nu(s) \in \mathbb{R}[s]^{\rho_\nu\times n_x}$ have full row rank.
 4: if $\rho_\nu = 0$ then
 5:     terminate
 6: else
 7:     Set $\mathscr{P}_{\nu+1}(s) = \begin{bmatrix} \mathscr{P}_{\nu,1}(s) \\ \mathscr{Q}_{\nu,2}(s) \end{bmatrix}$ and $\mathscr{Q}_{\nu+1}(s) = \begin{bmatrix} \mathscr{Q}_{\nu,1}(s) \\ 0 \end{bmatrix}$.
 8:     Set $\mathscr{K}(s) = \begin{bmatrix} \mathscr{K}(s) \\ \mathscr{Q}_{\nu,2}(s) \end{bmatrix}$.
 9:     if $\mathscr{K}(s)$ does not have full row rank then
10:         terminate (not delay-regular)
11:     end if
12:     Set $\nu \leftarrow \nu + 1$.
13:     Go to Line 3
14: end if
```

The discussion above yields our final result of this chapter.

Theorem 3.42. *Algorithm 2 terminates for any DDAE* (3.9).

We conclude this chapter by revisiting Example 3.34, where Algorithm 1 failed to terminate. The matrix pair in Example 3.34 is given by

$$\mathscr{P}(s) = \begin{bmatrix} s & 1 \\ s^2 & s \end{bmatrix}, \quad \mathscr{Q}(s) = \begin{bmatrix} -s & -1 \\ 0 & 0 \end{bmatrix}.$$

Applying Algorithm 2 to $(\mathscr{P}(s), \mathscr{Q}(s))$ yields in the first iteration

$$\mathscr{U}_1(s) = \begin{bmatrix} 1 & 0 \\ -s & 1 \end{bmatrix}, \qquad \mathscr{P}_2(s) = \begin{bmatrix} s & 1 \\ s^2 & s \end{bmatrix} = \mathscr{P}(s), \qquad \mathscr{Q}_2(s) = \begin{bmatrix} -s & -1 \\ 0 & 0 \end{bmatrix} = \mathscr{Q}(s),$$

and $\mathscr{K}(s) = \begin{bmatrix} s^2 & s \end{bmatrix}$. The previous computations show that in the next iteration, we obtain

$$\mathscr{K}(s) = \begin{bmatrix} s^2 & s \\ s^2 & s \end{bmatrix},$$

such that Algorithm 2 terminates with the information that $(\mathscr{P}(s), \mathscr{Q}(s))$ is not delay-regular. This is the correct result, which can be easily verified with Theorem 3.20.

4 Classical solutions and discontinuity propagation

Having established the existence and uniqueness of solutions for the *linear time-invariant* (LTI) *delay differential-algebraic equation* (DDAE) (1.15) in a distributional solution space (cf. Theorem 3.20), we now turn our attention to a more classical solution concept, namely solutions that are continuously differentiable almost everywhere. Following our analysis in chapter 3 it is sufficient to focus on delay-regular DDAEs. Invoking Theorem 3.37 and Algorithm 1, we can thus restrict our analysis to linear DDAEs

$$E\dot{x}(t) = A_1 x(t) + A_2 x(t-\tau) + f(t) \tag{4.1a}$$

with regular matrix pencil $(E, A_1) \in \left(\mathbb{F}^{n_x \times n_x}\right)^2$. As before, the DDAE (4.1a) is equipped with the initial trajectory

$$x(t) = \phi(t) \qquad \text{for } t \in [-\tau, 0]. \tag{4.1b}$$

Already in the case of *differential-algebraic equations* (DAEs), i.e., in the case $A_2 = 0$, a necessary condition for the existence of a classical solution is that f is sufficiently smooth (cf. Assumption 2.5). For DDAEs this in additional implies that also the history function ϕ needs to be sufficiently smooth, which we assume for the remainder of this chapter. In summary, we invoke the following assumption for the upcoming analysis.

> **Assumption 4.1.** *The matrix pair* $(E, A_1) \in \left(\mathbb{F}^{n_x \times n_x}\right)^2$ *in* (4.1a) *is regular, i.e., there exists* $\lambda \in \mathbb{F}$ *such that* $\det(\lambda E - A_1) \neq 0$. *Moreover, we assume that the history function* $\phi\colon [-\tau, 0] \to \mathbb{F}^{n_x}$ *and the inhomogeneity* $f : \mathbb{I} \to \mathbb{F}^{n_x}$ *are infinitely many times continuously differentiable.*

4.1 Continuous solutions and classification

Similarly as in the proof of Theorem 3.5, we can apply the DAE theory to the sequence of DAEs (2.3) that arises from applying the method of steps to (1.13). The corresponding sequence of DAEs for the initial trajectory problem (4.1) is given by

$$E\dot{x}_{[i]}(t) = A_1 x_{[i]}(t) + \tilde{f}_{[i]}(t), \qquad t \in [0,\tau), \tag{4.2a}$$

$$x_{[i]}(0) = x_{[i-1]}(\tau^-) \tag{4.2b}$$

with $x_{[0]}(t) = \phi(t-\tau)$ and $\tilde{f}_{[i]}(t) = A_2 x_{[i-1]}(t) + f(t+(i-1)\tau)$ for $t \in [0,\tau]$ and $i \in \mathscr{I} = \{1,\ldots,M\}$. We can use the quasi-Weierstraß form (cf. Theorem 2.9) for our analysis: There exist matrices $S, T \in \mathrm{GL}_{n_x}(\mathbb{F})$ such that

$$SET = \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & N \end{bmatrix} \qquad \text{and} \qquad SA_1T = \begin{bmatrix} J & 0 \\ 0 & I_{n_{x,\mathrm{a}}} \end{bmatrix} \tag{4.3}$$

with matrix $J \in \mathbb{F}^{n_{x,\mathrm{d}} \times n_{x,\mathrm{d}}}$ and nilpotent matrix $N \in \mathbb{F}^{n_{x,\mathrm{a}} \times n_{x,\mathrm{a}}}$. For the upcoming analysis we introduce

$$\begin{bmatrix} B_\mathrm{d} \\ B_\mathrm{a} \end{bmatrix} := SA_2, \quad \begin{bmatrix} B_{\mathrm{d},1} & B_{\mathrm{d},2} \\ B_{\mathrm{a},1} & B_{\mathrm{a},2} \end{bmatrix} := SA_2T, \quad \begin{bmatrix} v \\ w \end{bmatrix} := T^{-1}x, \quad \begin{bmatrix} g \\ h \end{bmatrix} := Sf, \quad \text{and} \quad \begin{bmatrix} \psi \\ \eta \end{bmatrix} := T^{-1}\phi, \tag{4.4}$$

where we use the same block dimensions as in (4.3). Applying the matrices S, T to (4.1a) yields

$$\dot{v} = Jv + B_{\mathrm{d},1}\sigma_\tau v + B_{\mathrm{d},2}\sigma_\tau w + g, \tag{4.5a}$$

$$N\dot{w} = w + B_{\mathrm{a},1}\sigma_\tau v + B_{\mathrm{a},2}\sigma_\tau w + h. \tag{4.5b}$$

Example 4.2. For the DDAE (1.24) in Example 1.5 we directly observe that the matrices $S = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ and $T = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ transform the associated matrix pair to quasi-Weierstraß form with $n_{x,\mathrm{d}} = 0$ and $n_{x,\mathrm{a}} = 2$. The according form (4.5) is given with the matrices

$$N = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad \text{and} \qquad B_{\mathrm{a},2} = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}.$$

♠

Even with Assumption 4.1 we cannot expect a continuously differentiable solution, as is illustrated in the Examples 1.4 and 1.5. This is mainly due to the fact that the identity

$$\lim_{t \searrow 0} \dot{x}(t) = \lim_{t \nearrow 0} \dot{\phi}(t),$$

which can be written in the form $\dot{x}(0) = \dot{\phi}(0^-)$, is not satisfied in general and this discontinuity in the first derivative at $t = 0$ may propagate over time (cf. [26] and Examples 1.4 and 1.5), which is the reason for analyzing solutions in the space of piecewise smooth distributions (see section 3.1). If we are interested in solutions that are at least continuous, then we can search for a solution in the space of absolutely continuous functions, i.e., functions that are continuous and differentiable almost everywhere. Assuming that the history function ϕ and the inhomogeneity f are sufficiently smooth, we expect discontinuities only at integer multiples of the time delay τ and thus consider the space of piecewise continuously differentiable functions as solution space. More precisely, we employ the following solution concept for the remainder of this chapter.

Definition 4.3 (Solution concept)**.** Assume that the matrix pair (E, A_1) in the DDAE (2.1a) is regular and the history function ϕ and the inhomogeneity f are infinitely many times continuously differentiable. We call $x \in \mathscr{C}(\mathbb{I}, \mathbb{F}^{n_x})$ a *solution* of (4.1) if for all $i \in \mathscr{I}$ the restriction $x_{[i]}$ of x as in (2.2) is a solution of (2.3). We call the history function $\phi \colon [-\tau, 0] \to \mathbb{F}^{n_x}$ *consistent* if the initial value problem (4.1) has at least one solution.

Since our solution concept is inherently related to the method of steps we immediately obtain the following relation between the DDAE (4.1a) and the sequence of DAEs (4.2).

Proposition 4.4. *Let the DDAE* (4.1a) *satisfy Assumption 4.1. If x is a solution of the* initial trajectory problem *(ITP)* (4.1)*, then the restriction $x_{[i]}(t) = x(t + (i-1)\tau)$ for $i \in \mathscr{I}$ is a solution of* (4.2)*. Conversely, if the sequence $(x_{[i]})_{i \in \mathscr{I}}$ is a solution of* (4.2)*, then*

$$x(t) = \begin{cases} x_{[i]}(t - (i-1)\tau), & \textit{if } (i-1)\tau \leq t < i\tau \textit{ for some } i \in \mathbb{N}, \\ \phi(t), & \textit{otherwise,} \end{cases}$$

is a solution of (4.1)*.*

The three introductory examples, namely Examples 1.3 to 1.5, show that solutions of linear DDAEs may have very different smoothness properties. Since the standard classification for delay equations is only valid for scalar equations, we pursue the following strategy: We first introduce a new classification, which is based on the worst possible smoothing behavior, and then give an algebraic characterization of the different types in terms of the matrices of the DDAE. To this end we recall Proposition 2.14, which establishes a connection of the DAE (4.2a) and the so-called underlying *ordinary differential equation* (ODE) (2.15)

$$\dot{x} = A^{\mathrm{diff}} x + \sum_{k=0}^{\mathrm{ind}(E,A_1)} C_k \tilde{f}^{(k)} \tag{4.6}$$

via the consistency condition (2.16). Hereby, the matrices are defined as (see (2.14))

$$A^{\mathrm{diff}} = T \begin{bmatrix} J & 0 \\ 0 & 0 \end{bmatrix} T^{-1}, \quad A^{\mathrm{con}} = T \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & 0 \end{bmatrix} T^{-1}, \quad C_0 = T \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & 0 \end{bmatrix} S, \quad C_k = -T \begin{bmatrix} 0 & 0 \\ 0 & N^{k-1} \end{bmatrix} S,$$

where $S, T \in \mathrm{GL}_{n_x}(\mathbb{F})$ denote matrices that transform (E, A_1) into quasi-Weierstraß form (2.6). Introducing the matrices

$$B_k := C_k A_2 \qquad \text{for } k = 0, \ldots, \mathrm{ind}(E, A_1)$$

allows us to re-substitute $\tilde{f}_{[i]}(t) = A_2 x_{[i-1]}(t) + f(t + (i-1)\tau)$. This yields the *delay differential equation* (DDE)

$$\dot{x} = A^{\mathrm{diff}} x + \sum_{k=0}^{\mathrm{ind}(E,A_1)} \left(B_k \sigma_\tau x^{(k)} + C_k f^{(k)} \right), \tag{4.7}$$

which we call the the *underlying DDE* for the DDAE (4.1a). From Corollary 2.15 and the discussion thereafter we immediately observe the following.

Lemma 4.5. *Let* (4.1) *satisfy Assumption 4.1. A necessary condition for the history function ϕ to be consistent is that ϕ satisfies the equation*

$$\phi(0) = A^{\mathrm{con}}\phi(0) + \sum_{k=1}^{\mathrm{ind}(E,A_1)} \Big(B_k\phi^{(k-1)}(-\tau) + C_k f^{(k-1)}(0)\Big). \tag{4.8}$$

Unfortunately, as Example 1.5 suggests, this condition is not sufficient for consistency, and even worse, the consistency of a history function depends on the time interval for which we want to solve the DDAE. Since our main goal is to analyze the propagation of primary discontinuities, it is sufficient to ensure that a solution exists for some time, which gives rise to the following definition.

Definition 4.6. Let the ITP (4.1) with history function $\phi\colon [-\tau,0] \to \mathbb{F}^{n_x}$ satisfy Assumption 4.1. Then ϕ is called *admissible* for the ITP (4.1) if $x_{[1]}(0) = \phi(0)$ is consistent for the DAE

$$E\dot{x}_{[1]}(t) = A_1 x_{[1]}(t) + A_2\phi(t-\tau) + f_{[1]}(t) \qquad \text{for } t \in [0,\tau),$$

i.e. ϕ satisfies (4.8). Similarly, $x_{[0]}\colon [0,\tau] \to \mathbb{F}^{n}_{x}$ is called *admissible* for the sequence of DAEs (4.2) if the DAE $E\dot{x}_{[1]}(t) = A_1 x_{[1]}(t) + A_2 x_{[0]}(t) + f(t)$ with $x_{[1]}(0) = x_{[0]}(\tau)$ has a solution on $[0,\tau)$.

Let $\phi\colon [-\tau,0] \to \mathbb{F}^{n_x}$ be admissible. As a consequence of Assumption 4.1 there exists a number $M \in \mathbb{N}$ and a unique sequence $(x_{[i]})_{i\in\{0,\ldots,M\}}$ that satisfies (4.2) (cf. Corollary 2.15). Hence for any $i \in \{1,\ldots,M\}$ we can define the level ℓ_i of the primary discontinuity as

$$\ell_i := \min_{\substack{x_{[0]}\in\mathscr{C}^\infty([0,\tau],\mathbb{F}^{n_x})\\ x_{[0]}\text{ admissible}}} \; \min_{f\in\mathscr{C}^\infty(\mathbb{I},\mathbb{F}^{n_x})} \max\left\{\ell\in\mathbb{N}_0 \;\middle|\; \begin{array}{c} x_{[j]} \text{ solves (4.2) for } j = 1,\ldots,i \text{ and} \\ x_{[i]}^{(\ell)}(0) = x_{[i-1]}^{(\ell)}(\tau^-) \end{array}\right\}. \tag{4.9}$$

If for some $j \in \mathbb{N}$ the initial condition $x_{[j]}(0) = x_{[j-1]}(\tau)$ is not consistent and thus no solution of (4.2) exists, we formally set $\ell_i := -\infty$ for all $i \geq j$. Note that this definition is independent of the specific choice of the inhomogeneity f and the history ϕ and hence serves as the worst-case scenario. To simplify the computation of the numbers ℓ_i we observe the following, which is a generalization of [102, Theorem 7.1]

Proposition 4.7. *Let the ITP* (4.1) *satisfy Assumption 4.1. Then the solution x of* (4.1) *is continuously differentiable on* $[-\tau,\tau)$ *if and only if ϕ satisfies*

$$\dot{\phi}(0^-) = A^{\mathrm{diff}}\phi(0) + \sum_{k=0}^{\mathrm{ind}(E,A_1)} \Big(B_k\phi^{(k)}(-\tau) + C_k f^{(k)}(0)\Big). \tag{4.10}$$

The solution x of (4.1) *is κ times continuously differentiable on* $[-\tau,\tau)$ *if and only if ϕ satisfies*

$$\phi^{(p+1)}(0^-) = A^{\mathrm{diff}}\phi^{(p)}(0^-) + \sum_{k=0}^{\mathrm{ind}(E,A_1)} \Big(B_k\phi^{(k+p)}(-\tau) + C_k f^{(k+p)}(0)\Big) \tag{4.11}$$

for $p = 0,1,\ldots,\kappa-1$.

Proof. Since ϕ is admissible, the initial condition $x_{[1]}(0) = \phi(0)$ is consistent and following Corollary 2.15 the solution x exists on $[-\tau, \tau)$. Thus, it is sufficient to check the point $t = 0$. Using Proposition 2.14 we can consider (2.15) and thus obtain

$$\begin{aligned} \dot{x}_{[1]}(0) &= A^{\text{diff}} x_{[1]}(0) + \sum_{k=0}^{\text{ind}(E,A_1)} \left(B_k x_{[0]}^{(k)}(0) + C_k f_{[1]}^{(k)}(0) \right) \\ &= A^{\text{diff}} \phi(0) + \sum_{k=0}^{\text{ind}(E,A_1)} \left(B_k \phi^{(k)}(-\tau) + C_k f_{[1]}^{(k)}(0) \right) \end{aligned}$$

and hence x is continuously differentiable on $[-\tau, \tau)$ if and only if ϕ satisfies (4.10). For arbitrary $\kappa \in \mathbb{N}$ we invoke Proposition 2.14, which guarantees that the solution x exists on the interval $[0, \tau)$ and allows us to consider the underlying DDE (4.7) instead of the DDAE. Since the assumption guarantees that x is sufficiently smooth on $[0, \tau)$ we can differentiate (4.7) $p \in \mathbb{N}$ times to obtain

$$\begin{aligned} x_{[1]}^{(p+1)}(0) &= A^{\text{diff}} x_{[1]}^{(p)}(0) + \sum_{k=0}^{\text{ind}(E,A_1)} \left(B_k x_{[0]}^{(k+p)}(0) + C_k f_{[1]}^{(k+p)}(0) \right) \\ &= A^{\text{diff}} \phi^{(p)}(0^-) + \sum_{k=0}^{\text{ind}(E,A_1)} \left(B_k \phi^{(k+p)}(-\tau) + C_k f_{[1]}^{(k+p)}(0) \right), \end{aligned}$$

which implies the assertion. ■

Since we require $\phi \in \mathscr{C}^\infty([-\tau, 0], \mathbb{F}^{n_x})$ to be admissible we immediately obtain $\ell_1 \geq 0$. On the other hand assume that we have given the values $\phi(0)$ and $\phi^{(k)}(-\tau)$ for $k = 0, \ldots, \nu$ such that ϕ is admissible. Then we can always construct (via Hermite interpolation) ϕ in such a way that (4.10) is not satisfied and hence $\ell_1 \leq 0$, which yields $\ell_1 = 0$. Thus, the questions about propagation of discontinuities can be rephrased as whether

- there exists $k \in \mathbb{N}$ with $\ell_k > 0$ (i.e. the solution becomes smoother), or
- there exists $k \in \mathbb{N}$ with $\ell_k = -\infty$ (i.e. the solution becomes less smooth),
- or if $\ell_i = \ell_1$ for all $i \in \mathbb{N}$.

We notice that the smoothing may not start immediately (i.e. we cannot ask for $\ell_1 = 1$), as the following example suggests.

Example 4.8. Consider the DDAE given by $\mathbb{F} = \mathbb{R}$, $n_x = 2$, $f \equiv 0$, $\tau = 1$, and

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad \phi(t) = \begin{bmatrix} t, \\ -1 \end{bmatrix}.$$

Since (E, A) is already in Weierstraß form, it is easy to see that the DDAE corresponds to the DDE

$$\dot{v}(t) = v(t - 2\tau) \tag{4.12}$$

with coupled equation $w(t) = v(t - \tau)$. Straightforward calculations show that $\ell_1 \leq 0$ (using the specified history function ϕ) and $\ell_1 \geq 0$ implying $\ell_1 = 0$. On the other hand, (4.12) is a scalar delay equation and it is well-known that the solution is continuously differentiable at $t = 2\tau$, thus we have $\ell_2 \geq 1$. ♠

Definition 4.9. Consider the DDAE (4.1a) on the interval $\mathbb{I} = [0, M\tau]$, set $\mathcal{I} := \{1,\ldots,M\}$, and suppose that (4.1) satisfies Assumption 4.1. We say that (1.13) is of

- smoothing type if there exists $j \in \mathcal{I}$, $j > 1$ such that $\ell_j = 1$ and $\ell_i = 0$ for $i < j$,
- discontinuity invariant type if $\ell_i = 0$ for all $i \in \mathcal{I}$, and
- de-smoothing type if there exists $j \in \mathcal{I}$, $j > 1$ such that $\ell_j = -\infty$ and $\ell_i = 0$ for $i < j$.

In the following, we analyze the DDAE (4.1a) in detail and derive conditions for the matrices E, A_1, and A_2, from which the type can be determined. Before we analyze the general DDAE case we focus on the case of $\operatorname{ind}(E, A_1) \le 1$, i.e., the system is a pure DDE or $N = 0$ in (2.12b).

Remark 4.10. The case $\operatorname{ind}(E, A_1) \le 1$ includes DDEs of the form

$$\dot{\hat{x}}(t) = \widehat{A}_1\hat{x}(t) + \widehat{A}_2\hat{x}(t-\tau) + \widehat{D}\dot{\hat{x}}(t-\tau) + \hat{f}(t), \tag{4.13}$$

with arbitrary matrices $\widehat{A}_1, \widehat{A}_2, \widehat{D} \in \mathbb{F}^{n_x\times n_x}$, since (4.13) can be recast in the form (4.1a) by introducing the new state variable $x(t) = \left[\begin{smallmatrix}\hat{x}(t)\\ \hat{x}(t-\tau)\end{smallmatrix}\right]$ and

$$E := \begin{bmatrix} -I_{n_x} & -\widehat{D} \\ 0 & 0 \end{bmatrix} \qquad A_1 := \begin{bmatrix} \widehat{A}_1 & 0 \\ 0 & I_{n_x} \end{bmatrix}, \qquad A_2 := \begin{bmatrix} \widehat{A}_2 & 0 \\ -I_{n_x} & 0 \end{bmatrix}, \qquad f := \begin{bmatrix} \hat{f} \\ 0 \end{bmatrix}.$$

♣

If $\operatorname{ind}(E, A_1) = 0$, then the matrix E is nonsingular and the DDAE is of the form

$$\dot{x}(t) = E^{-1}A_1x(t) + E^{-1}A_2x(t-\tau) + E^{-1}f(t) \tag{4.14}$$

and the ODE solution formula together with Proposition 4.7 directly implies $\ell_1 = 1$, i.e., (4.14) is of smoothing type.

Theorem 4.11. *Consider the DDAE* (4.1a) *on the interval* $\mathbb{I} = [0, M\tau]$ *and suppose that Assumption 4.1 holds. If* $\operatorname{ind}(E, A_1) = 1$*, then* (4.1a) *is of smoothing type if and only if* $B_{\mathrm{a},2}$ *in* (4.4) *is nilpotent with index of nilpotency* ν_B *and furthermore we have* $\nu_B \le M - 1$.

Proof. Let $S, T \in \mathrm{GL}_n(\mathbb{F})$ be matrices that transform (4.1a) into quasi-Weierstraß form (4.5). Applying the method of steps yields

$$\dot{v}_{[i+1]} = Jv_{[i+1]} + B_{\mathrm{d},1}v_{[i]} + B_{\mathrm{d},2}v_{[i]} + g_{[i+1]} \qquad \text{and} \qquad w_{[i+1]} = -B_{\mathrm{a},1}v_{[i]} - B_{\mathrm{a},2}w_{[i]} - h_{[i+1]}.$$

Since $\ell_1 = 0$, we have

$$\begin{aligned} w_{[1]}(\tau) &= -B_{\mathrm{a},1}v_{[0]}(\tau) - B_{\mathrm{a},2}w_{[0]}(\tau) - h_{[1]}(\tau) \\ &= -B_{\mathrm{a},1}v_{[1]}(0) - B_{\mathrm{a},2}w_{[1]}(0) - h_{[2]}(0) = w_{[2]}(0) \end{aligned}$$

and thus $\ell_2 \ge 0$. By induction we conclude $\ell_i \ge 0$ for $i \in \mathcal{I}$. Moreover, we have

$$\begin{aligned} \dot{w}_{[i+1]} &= -B_{\mathrm{a},1}\dot{v}_{[i]} - B_{\mathrm{a},2}\dot{w}_{[i]} - \dot{h}_{[i+1]} \\ &= -B_{\mathrm{a},1}\left(Jv_{[i]} + B_{\mathrm{d},1}v_{[i-1]} + B_{\mathrm{d},2}w_{[i-1]} + g_{[i]}\right) - B_{\mathrm{a},2}\dot{w}_{[i]} - \dot{h}_{[i+1]} \end{aligned}$$

which implies $\dot{w}_{[i+1]}(0^+) - \dot{w}_{[i]}(\tau^-) = B_{\mathrm{a},2}\left(\dot{w}_{[i-1]}(\tau^-) - \dot{w}_{[i]}(0^+)\right)$ holds. By induction we have

$$\dot{w}_{[i+1]}(0^+) - \dot{w}_{[i]}(\tau^-) = (-1)^i B_{\mathrm{a},2}^i\left(\dot{w}_{[1]}(0^+) - \dot{\eta}(0^-)\right) \qquad \text{for } i = 1,\dots,M-1.$$

Thus $\ell_{i+1} \geq 1$ holds if and only if $B_{\mathrm{a},2}^i = 0$. ∎

Applying Theorem 4.11 to the DDAE in Example 4.8 shows that this DDAE is of smoothing type, since it is already in quasi-Weierstraß form with $B_{\mathrm{a},2} = 0$. Conversely, if the DDAE (4.1a) with $\mathrm{ind}(E, A_1) = 1$ is of smoothing type, then the index of nilpotency indicates the number of delays present in the system. More precisely, we have the following result.

Corollary 4.12. *Suppose that the DDAE* (4.1a) *satisfies Assumption 4.1 and is of smoothing type with* $\mathrm{ind}(E, A_1) \leq 1$. *Furthermore let ν_B denote the index of nilpotency of $B_{\mathrm{a},2}$ if $n_{x,\mathrm{a}} > 0$ and $\nu_B = 0$ otherwise. Then there exists matrices $D_k \in \mathbb{F}^{n_{x,\mathrm{d}} \times n_{x,\mathrm{d}}}$ ($k = 0,\dots,\nu_B$) and an inhomogeneity ϑ such that the solution v of* (4.5a) *is a solution of the ITP*

$$\begin{aligned}
\dot{z}(t) &= Jz(t) + \sum_{k=0}^{\nu_B} D_k z(t-(k+1)\tau) + \vartheta(t) && \text{for } t \in [\nu_B\tau, t_\mathrm{f}), && (4.15\mathrm{a})\\
z(t) &= v(t), && \text{for } t \in [-\tau, \nu_B\tau]. && (4.15\mathrm{b})
\end{aligned}$$

Proof. The result is trivial for $\mathrm{ind}(E, A_1) = 0$, i.e., assume $\mathrm{ind}(E, A_1) = 1$, which implies that $N = 0$ in (4.5). Let $\Delta_{[t_0,t_1)}$ denote the characteristic function for the interval $[t_0, t_1)$, i.e.

$$\Delta_{(t_0,t_1]}(t) = \begin{cases} 1, & \text{if } t \in [t_0, t_1),\\ 0, & \text{otherwise.}\end{cases}$$

Combination of the fast subsystem (4.5b) and the initial condition yields

$$(I_{n_{x,\mathrm{a}}} + B_{\mathrm{a},2}\Delta_{[\tau,t_\mathrm{f})}\sigma_\tau)w = -B_\mathrm{a}\Delta_{[0,\tau)}\sigma_\tau\phi - B_{\mathrm{a},1}\Delta_{[\tau,t_\mathrm{f})}\sigma_\tau v - h. \tag{4.16}$$

By induction we obtain $(\Delta_{[\tau,t_\mathrm{f})}(t)\sigma_\tau)^k = \Delta_{[k\tau,t_\mathrm{f})}(t)\sigma_{k\tau}$ and from $B_{\mathrm{a},2}^{\nu_B} = 0$ we deduce

$$\left(\sum_{k=0}^{\nu_B-1}(-1)^k\left(B_{\mathrm{a},2}\Delta_{[\tau,t_\mathrm{f})}\sigma_\tau\right)^k\right)\left(I_{n_{x,\mathrm{a}}} + B_{\mathrm{a},2}\Delta_{[\tau,t_\mathrm{f})}\sigma_\tau\right) = I_{n_{x,\mathrm{a}}}$$

such that w in (4.16) is given by

$$\begin{aligned}
w &= \sum_{k=0}^{\nu_B-1}(-1)^{k+1}\left(B_{\mathrm{a},2}\Delta_{[\tau,t_\mathrm{f})}\sigma_\tau\right)^k\left(B_\mathrm{a}\Delta_{[0,\tau)}\sigma_\tau\phi + B_{\mathrm{a},1}\Delta_{[\tau,t_\mathrm{f})}\sigma_\tau v + h\right)\\
&= \sum_{k=0}^{\nu_B-1}(-1)^{k+1}B_{\mathrm{a},2}^k\left(B_\mathrm{a}\Delta_{[k\tau,(k+1)\tau)}\sigma_{(k+1)\tau}\phi + B_{\mathrm{a},1}\Delta_{[k\tau,t_\mathrm{f})}\sigma_{(k+1)\tau}v + \Delta_{[k\tau,t_\mathrm{f})}\sigma_{k\tau}h\right).
\end{aligned}$$

Inserting this identity in (4.5a) and introducing for $k = 1,\dots,\nu_B$ the matrices

$$D_0 := B_{\mathrm{d},1}, \qquad D_k := (-1)^k B_{\mathrm{d},2}B_{\mathrm{a},2}^{k-1}B_{\mathrm{a},1}$$

implies that the solution v of (4.5a) is a solution of the ITP (4.15), where ϑ is given by

$$\vartheta(t) := g(t) + \sum_{k=0}^{\nu_B-1} (-1)^{k+1} B_{\mathrm{d},2} B_{\mathrm{a},2}^{k} h(t-(k+1)\tau). \qquad \blacksquare$$

Example 4.13. Consider the DDE (4.13) in the DDAE formulation given in Remark 4.10. The matrices $S := \begin{bmatrix} I_{n_x} & -\widehat{A_1}\widehat{D} \\ 0 & I_{n_x} \end{bmatrix}$ and $T := \begin{bmatrix} I_{n_x} & \widehat{D} \\ 0 & I_{n_x} \end{bmatrix}$ transform the DDAE to quasi-Weierstraß form given by

$$\begin{bmatrix} I_{n_x} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{v}(t) \\ \dot{w}(t) \end{bmatrix} = \begin{bmatrix} \widehat{A_1} & 0 \\ 0 & I_{n_x} \end{bmatrix} \begin{bmatrix} v(t) \\ w(t) \end{bmatrix} + \begin{bmatrix} \widehat{A_2} + \widehat{A_1}\widehat{D} & (\widehat{A_2} + \widehat{A_1}\widehat{D})\widehat{D} \\ -I_{n_x} & -\widehat{D} \end{bmatrix} \begin{bmatrix} v(t-\tau) \\ w(t-\tau) \end{bmatrix} + \begin{bmatrix} f(t) \\ 0 \end{bmatrix}.$$

Hence, the DDE (4.13) is of smoothing type if and only if $\widehat{D}$ is nilpotent. In this case, the corresponding retarded equation (4.15a) is given by

$$\dot{z}(t) = \widehat{A_1} z(t) + (\widehat{A_2} + \widehat{A_1}\widehat{D}) z(t-\tau) + \sum_{k=1}^{\nu_{\widehat{D}}-1} (-1)^k (\widehat{A_2} + \widehat{A_1}\widehat{D}) \widehat{D}^k z(t-(k+1)\tau) + g(t),$$

where $\nu_{\widehat{D}}$ is the index of nilpotency of $\widehat{D}$. ♠

Remark 4.14. The delay equation (4.15) in Corollary 4.12 may be used to determine whether the DDAE (4.13) is stable (which can be done for example via DDE-biftool [73, 196]). This provides an alternative way to the theory outlined in [68, 69]. ♣

For the analysis of the general DDAE case with arbitrary index, we use the following preliminary result.

Proposition 4.15. *Suppose that the ITP* (4.1) *satisfies Assumption 4.1 and let* $S, T \in \mathrm{GL}_{n_x}(\mathbb{F})$ *be matrices that transform* (E, A_1) *to quasi-Weierstraß form* (2.6)*, such that* (4.1a) *is transformed to* (4.5) *with* $x = T\left[\begin{smallmatrix} v \\ w \end{smallmatrix}\right]$. *Then for any* $m \in \mathbb{N}$ *and any* $\tilde{v} \in \mathbb{F}^{n_{x,\mathrm{d}}}$, $\tilde{w} \in \mathbb{F}^{n_{x,\mathrm{a}}}$ *there exists an admissible history function* $\phi = T^{-1}\left[\begin{smallmatrix} \psi \\ \eta \end{smallmatrix}\right]$ *that is analytic and satisfies*

$$\psi^{(p)}(0^-) = v^{(p)}(0) \qquad \textit{for } p = 0, 1, \ldots, m-1, \tag{4.17a}$$

$$\eta^{(p)}(0^-) = w^{(p)}(0) \qquad \textit{for } p = 0, 1, \ldots, m-1, \tag{4.17b}$$

$$\tilde{v} = \psi^{(m)}(0^-) - v^{(m)}(0), \qquad \textit{and} \tag{4.17c}$$

$$\tilde{w} = \eta^{(m)}(0^-) - w^{(m)}(0). \tag{4.17d}$$

Proof. Let $m \in \mathbb{N}$. Proposition 4.7 implies that the solution x of the ITP (4.1) is m times continuously differentiable on $[-\tau, \tau)$ if and only if ϕ satisfies (4.11) for $p = 0, 1, \ldots, m-1$. Multiply (4.11) from the left with T^{-1} to obtain

$$\psi^{(p+1)}(0^-) = J\psi^{(p)}(0^-) + B_{\mathrm{d},1}\psi^{(p)}(-\tau) + B_{\mathrm{d},2}\eta^{(p)}(-\tau) + g^{(p)}(0), \tag{4.18a}$$

$$\eta^{(p+1)}(0^-) = -\sum_{k=0}^{\mathrm{ind}(E,A_1)-1} N^k \left(B_{\mathrm{a},1}\psi^{(k+p+1)}(-\tau) + B_{\mathrm{a},2}\eta^{(k+p+1)}(-\tau) + h^{(k+p+1)}(0) \right) \tag{4.18b}$$

for $p = 0,\ldots,m-1$. We then can proceed as follows to construct ψ and η that satisfy the conditions (4.17). Choose any value for $\psi^{(p)}(-\tau)$ and $\eta^{(p)}(-\tau)$ for $p = 0,\ldots,\mathrm{ind}(E,A)+m$, and compute $\eta^{(p+1)}(0^-)$ for $p = 0,\ldots,m-2$ according to (4.18b). For an arbitrary $\psi(0)$, set $\psi^{(p+1)}(0^-)$ according to (4.18a) for $p = 0,\ldots,m-2$. Finally, set

$$\psi^{(m)}(0^-) = \tilde{v} + \left(J\psi^{(m-1)}(0^-) + B_{\mathrm{d},1}\eta^{(m-1)}(-\tau) + B_{\mathrm{d},2}\eta^{(m-1)}(-\tau) + g^{(m)}(0)\right) \quad \text{and}$$

$$\eta^{(m)}(0^-) = \tilde{w} - \sum_{k=0}^{\mathrm{ind}(E,A_1)-1} N^k \left(B_{\mathrm{a},1}\psi^{(k+p+1)}(-\tau) + B_{\mathrm{a},2}\eta^{(k+p+1)}(-\tau) + h^{(k+p+1)}(0)\right).$$

The desired history functions are then given via Hermite interpolation. ■

Applying the method of steps and the solution formula (2.13) for the fast subsystem yields

$$w_{[i+1]} = -\sum_{k=0}^{\mathrm{ind}(E,A_1)-1} N^k \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^k \left(B_{\mathrm{a},1}v_{[i]} + B_{\mathrm{a},2}w_{[i]} + h_{[i+1]}\right). \tag{4.19}$$

Since Assumption 4.1 implies that all functions are sufficiently smooth we obtain

$$\begin{aligned} w_{[2]}(0) - w_{[1]}(\tau^-) &= \sum_{k=0}^{\mathrm{ind}(E,A_1)-1} N^k \left(B_{\mathrm{a},1}\left(\psi^{(k)}(0^-) - v_{[1]}^{(k)}(0)\right) + B_{\mathrm{a},2}\left(\eta^{(k)}(0^-) - w_{[1]}^{(k)}(0)\right)\right) \\ &= \sum_{k=0}^{\mathrm{ind}(E,A_1)-1} N^k B_{\mathrm{a}} T \begin{bmatrix} \psi^{(k)}(0^-) - v_{[1]}^{(k)}(0) \\ \eta^{(k)}(0^-) - w_{[1]}^{(k)}(0) \end{bmatrix} \\ &= \sum_{k=1}^{\mathrm{ind}(E,A_1)-1} N^k B_{\mathrm{a}} T \begin{bmatrix} \psi^{(k)}(0^-) - v_{[1]}^{(k)}(0) \\ \eta^{(k)}(0^-) - w_{[1]}^{(k)}(0) \end{bmatrix}, \end{aligned}$$

where the last identity follows from the fact the ϕ is assumed to be admissible. Proposition 4.15 implies that (4.1a) is of de-smoothing type if there exists $k \in \{1,\ldots,\mathrm{ind}(E,A_1)-1\}$ such that $N^k B_{\mathrm{a}} \neq 0$. If we conversely assume $NB_{\mathrm{a}} = 0$ then (4.19) is given by

$$w_{[i+1]} = -B_{\mathrm{a},1}v_{[i]} - B_{\mathrm{a},2}w_{[i]} - \sum_{k=0}^{\mathrm{ind}(E,A_1)-1} N^k h_{[i+1]}^{(k)},$$

which implies $\ell_i \geq 0$. Together with Theorem 4.11, this proves the following theorem.

Theorem 4.16. *Consider the DDAE* (4.1a) *on the interval* $\mathbb{I} = [0, M\tau]$ *and suppose that the associated ITP* (4.1) *satisfies Assumption 4.1. Let* N, B_{a} *and* $B_{\mathrm{a},2}$ *be the matrices that are associated with the quasi-Weierstraß form* (4.5). *Then* (4.1a) *is of*

- *smoothing type if* $NB_{\mathrm{a}} = 0$ *and* $B_{\mathrm{a},2}$ *is nilpotent with nilpotency index* $\nu_B < M$,
- *de-smoothing type if there exists* $k \in \mathbb{N}$ *such that* $N^k B_{\mathrm{a}} \neq 0$, *and*
- *discontinuity invariant type, otherwise.*

Example 4.17. Introducing the shifted variable $z(t) = x(t-\tau)$ shows that the DDAE associated with

$$x(t) = A_2 x(t-\tau) + D\dot{x}(t-\tau) + f(t) \tag{4.20}$$

is of de-smoothing type if and only if $D \neq 0$. ♠

As a direct consequence of Definition 4.9 and Theorem 4.16 we can formulate our main result about the existence and uniqueness of continuous solutions.

Theorem 4.18. *Let the ITP* (4.1) *satisfy Assumption 4.1. Let N and B_{a} be the matrices associated with the quasi-Weierstraß form as defined in* (4.3) *and* (4.4).

(i) If the history function ϕ is admissible, i.e., ϕ satisfies (4.8)*, then $NB_{\mathrm{a}} = 0$ is a sufficient condition for the existence of a solution (in the sense of Definition 4.3). In this case, the solution is unique.*

(ii) If $NB_{\mathrm{a}} \neq 0$, then there exists an admissible history function ϕ, such that a solution exists only on the time interval $\mathbb{I} = [0,\tau)$.

Remark 4.19. Checking the proof of Corollary 4.12, we immediately infer from Theorem 4.16 that Corollary 4.12 is also true for an arbitrary index $\operatorname{ind}(E, A_1)$. As a consequence, if the DDAE (4.1a) is of smoothing type, then there exists a sequence $j_k \in \mathbb{N}$ such that $\ell_{j_k} = k$ and hence the solution becomes arbitrarily smooth over time justifying the terminology *smoothing type.* ♣

Note that $N^k B_{\mathrm{a}} \neq 0$ for some $k \in \mathbb{N}$ implies

$$B_{k+1} = C_{k+1} A_2 = -T \begin{bmatrix} 0 & 0 \\ 0 & N^k \end{bmatrix} S A_2 = -T \begin{bmatrix} 0 \\ N^k B_{\mathrm{a}} \end{bmatrix} \neq 0,$$

i.e. the DDAE (4.1a) is of de-smoothing type if $B_k \neq 0$ for some $k \geq 2$. Using

$$\begin{aligned} B_k(I_{n_x} - A^{\mathrm{con}}) &= -T \begin{bmatrix} 0 & 0 \\ N^{k-1} B_{\mathrm{a},1} & N^{k-1} B_{\mathrm{a},2} \end{bmatrix} T^{-1} T \left(\begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & I_{n_{x,\mathrm{a}}} \end{bmatrix} - \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & 0 \end{bmatrix} \right) T^{-1} \\ &= -T \begin{bmatrix} 0 & 0 \\ 0 & N^{k-1} B_{\mathrm{a},2} \end{bmatrix} T^{-1} \end{aligned} \tag{4.21}$$

we immediately see that $B_{\mathrm{a},2}$ is nilpotent if and only if $B_1(I_{n_x} - A^{\mathrm{con}})$ is nilpotent, which shows that the results of Theorem 4.16 can be formulated in terms of the underlying DDE (4.7). As a consequence of Lemma 2.13 this shows that the previous results are independent of the particular choice of the matrices S, T used to transform (E, A_1) to quasi-Weierstraß form. In more detail, we have the following two results.

Corollary 4.20. *Consider the ITP* (4.1) *with associated underlying DDE* (4.7) *on the interval $\mathbb{I} = [0, M\tau)$ and suppose that Assumption 4.1 applies. Then* (4.1a) *is of*

- *smoothing type if $B_2 = 0$ and $B_1(I_{n_x} - A^{\mathrm{con}})$ is nilpotent with nilpotency index $\nu_{B_1} \leq M$,*
- *de-smoothing type if $B_k \neq 0$ for some $k \geq 2$, and*
- *discontinuity invariant type otherwise.*

Theorem 4.21. *Consider the ITP* (4.1) *with associated underlying DDE* (4.7) *on the interval* $\mathbb{I} = [0, M\tau)$ *and suppose that Assumption 4.1 applies.*

(i) If the history function ϕ is admissible, i.e., ϕ satisfies (4.8)*, then $B_2 = 0$ is a sufficient condition for the existence of a solution (in the sense of Definition 4.3). In this case, the solution is unique.*

(ii) If $B_2 \neq 0$, then there exists an admissible history function ϕ, such that a solution exists only on the time interval $\mathbb{I} = [0, \tau)$.

A common approach to analyze the (exponential) stability of the DDAE (4.1a) is to compute the *spectral abscissa*, which is defined as

$$\alpha(E, A_1, A_2) = \sup\{\mathrm{Re}(\lambda) \mid \det(\lambda E - A_1 - \exp(-\lambda\tau)A_2) = 0\}.$$

Surprisingly, the condition $\alpha(E, A, D) < 0$ is not sufficient for a DDAE to be exponentially stable [69]. However, based on the new classification we have the following result.

Corollary 4.22. *Suppose that the DDAE* (4.1a) *is not of de-smoothing type. Then the DDAE* (4.1a) *is exponentially stable if and only if $\alpha(E, A_1, A_2) < 0$.*

Proof. Since the DDAE (4.1a) is not of de-smoothing type, we have $NB_{\mathrm{a}} = 0$. The result follows directly from [69, Proposition 3.4 and Theorem 3.4]. ■

Note that we refrain from using the terminology *retarded*, *neutral*, and *advanced* in Definition 4.9, although these terms are widely used in the delay literature [26, 27, 98, 102], see section 1.3 for a definition. The reason is that in the classical definition, a retarded DDE becomes advanced if it is solved backwards in time, an advanced equation becomes retarded and a neutral equation stays neutral. This is no longer true for the classification introduced in Definition 4.9. To see this, we introduce the new variable $\xi(t - \tau) = x(-t)$ such that (4.1a) transforms to

$$E\dot{\xi}(t - \tau) = -A_2\xi(t) - A_1\xi(t - \tau) - f(-t).$$

Definition 4.23. Consider the DDAE (4.1a) and define

$$\mathscr{E} := \begin{bmatrix} 0 & E \\ 0 & 0 \end{bmatrix} \in \mathbb{F}^{2n_x, 2n_x}, \qquad \mathscr{A}_1 := \begin{bmatrix} -A_2 & 0 \\ 0 & I_{n_x} \end{bmatrix} \in \mathbb{F}^{2n_x, 2n_x}, \qquad \mathscr{A}_2 := \begin{bmatrix} -A_1 & 0 \\ -I_{n_x} & 0 \end{bmatrix} \in \mathbb{F}^{2n_x, 2n_x}.$$

Then we call the DDAE

$$\mathscr{E}\dot{\zeta}(t) = \mathscr{A}_1\zeta(t) + \mathscr{A}_2\zeta(t - \tau) + \mathscr{F}(t) \tag{4.22}$$

with $\mathscr{F} : \mathbb{I} \to \mathbb{F}^{2n_x}$ the *backward system* for the DDAE (4.1a).

The matrix pair $(\mathscr{E}, \mathscr{A}_1)$ is regular if and only if $\det(A_2) \neq 0$. In this case, we can transform the backward system (4.22) to quasi-Weierstraß form via the matrices

$$S = \begin{bmatrix} -A_2^{-1} & 0 \\ 0 & I_{n_x} \end{bmatrix} \qquad \text{and} \qquad T = I_{2n_x}.$$

In particular, we have

$$(S\mathscr{E}T)(S\mathscr{A}_2T) = \begin{bmatrix} 0 & -A_2^{-1}E \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A_2^{-1}A_1 & 0 \\ -I_{n_x} & 0 \end{bmatrix} = \begin{bmatrix} -A_2^{-1}E & 0 \\ 0 & 0 \end{bmatrix}.$$

Thus Theorem 4.16 implies that $E = 0$ is a necessary condition for the backward system (4.22) to be of smoothing type or discontinuity invariant type, which implies that the DDAE (4.1a) cannot be of de-smoothing type.

Example 4.24. Consider the DDAE given by $\mathbb{F} = \mathbb{R}$, $n_x = 2$, $f \equiv 0$, $\tau = 1$, and

$$E = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Since (E, A_1) is already in Weierstraß form and $EA_2 \neq 0$, Theorem 4.16 implies that the DDAE is of de-smoothing type. Since $E \neq 0$ also the backward system is of de-smoothing type. ♠

Let us mention that if $\det(A_2) = 0$, then the method of steps (2.3) cannot be used to determine the solution of the backward system. Instead, one may use the shift-index concept defined in [98, 99] to make the pencil $(\mathscr{E}, \mathscr{A}_1)$ regular. This can be achieved for instance with Algorithm 2, since

$$\begin{aligned} \operatorname{rank}_{\mathbb{R}(s,\omega)}(s\mathscr{E} + \mathscr{A}_1 + \omega\mathscr{A}_2) &= \operatorname{rank}_{\mathbb{R}(s,\omega)}\left(\begin{bmatrix} -A_2 - \omega A_1 & sE \\ -\omega I_{n_x} & I_{n_x} \end{bmatrix}\right) \\ &= \operatorname{rank}_{\mathbb{R}(s,\omega)}\left(\begin{bmatrix} s\omega E - A_2 - \omega A_1 & sE \\ 0 & I_{n_x} \end{bmatrix}\right) \\ &= \operatorname{rank}_{\mathbb{R}(s,\omega)}\left(\begin{bmatrix} s\omega E - A_2 - \omega A_1 & 0 \\ 0 & I_{n_x} \end{bmatrix}\right) \\ &= \operatorname{rank}_{\mathbb{R}(s,\omega)}\left(sE - A_1 - \omega A_2\right) + n_x \\ &= 2n_x \end{aligned}$$

implies that $(\mathscr{E}, \mathscr{A}_1, \mathscr{A}_2)$ is delay-regular according to Theorem 3.20.

4.2 Impact of splicing conditions

In the previous section algebraic criteria were established to check whether a discontinuity in the derivative of $\dot{x}$ at $t = 0$ is smoothed out, is propagated to $t = \tau$ or is amplified in the sense that x becomes discontinuous at $t = \tau$. While the definition of the discontinuity invariant type is valid for all integer multiples of the delay time, the definitions of smoothing type and de-smoothing type are based on single time points and hence the question whether the (de-)smoothing continues is imminent. For DDAEs of smoothing type, this can be answered positively (see Remark 4.19). For DDAEs of de-smoothing type, the question can be rephrased as follows: If we restrict the set of admissible history functions such that the *splicing condition* (cf. [26])

$$\phi^{(k)}(0^-) = x^{(k)}(0) \qquad \text{for } k = 0, \ldots, \kappa \tag{4.23}$$

is satisfied for some $\kappa \in \mathbb{N}$, is there an integer $j \in \mathbb{N}$ such that the inital condition

$$x_{[j]}(0) = x_{[j-1]}(\tau^-)$$

is not consistent for the DAE (4.2)? Similarly, we can ask if for DDAEs of discontinuity invariant type the smoothness at integer multiples of the delay time stays invariant. Before we answer these questions, we note that in order to check if the splicing condition (4.23) is satisfied, it seems that one has to solve the DDAE (4.1a) first. That is however not necessary, since the splicing condition (4.23) can be checked by solely investigating the history function ϕ with Proposition 4.7.

Lemma 4.25. *Suppose that the DDAE* (4.1a) *is of discontinuity invariant type and the admissible history function* $\phi \in \mathscr{C}^\infty([-\tau,0],\mathbb{F}^{n_x})$ *satisfies the splicing condition* (4.23)*. Then*

$$x_{[i]}^{(k)}(0) = x_{[i-1]}^{(k)}(\tau^-) \qquad \textit{for all } i \in \mathbb{N},\ k = 0,\ldots,\kappa.$$

Proof. Since (4.1a) is of discontinuity invariant type, we have $NB_\mathrm{a} = 0$ in (4.5) according to Theorem 4.16. It suffices to show that

$$x_{[2]}^{(j)}(0) = x_{[1]}^{(j)}(\tau^-) \qquad \text{for all } j = 0,\ldots,\kappa.$$

Since ϕ is admissible and the DDAE is of discontinuity invariant type, equation (4.5a) implies that

$$\dot{v}_{[2]}(0) - \dot{v}_{[1]}(\tau^-) = J\left(v_{[2]}(0) - v_{[1]}(\tau)\right) + B_\mathrm{d}\left(x_{[1]}(0) - \phi(0)\right) = 0.$$

Iteratively, we obtain

$$v_{[2]}^{(k+1)}(0) - v_{[1]}^{(k+1)}(\tau^-) = J\left(v_{[2]}^{(k)}(0) - v_{[1]}^{(k)}(\tau^-)\right) + B_\mathrm{d}\left(x_{[1]}^{(k)}(0) - \phi^{(k)}(0)\right) = 0$$

for $k = 2,\ldots,\kappa$. For the fast system (4.5b) we directly infer

$$w_{[2]}^{(k)}(0) - w_{[1]}^{(k)}(\tau^-) = B_\mathrm{a}\left(\phi^{(k)}(0^-) - x_{[1]}^{(k)}(0)\right) = 0$$

for $k = 0,1,\ldots,\kappa$, which completes the proof. ■

Lemma 4.25 guarantees that the solution of the DDAE is at least as smooth as the initial transition from the history function to the solution. Conversely, assume that the Jordan canonical form of $B_{\mathrm{a},2}$ exists and let $\tilde{w} \in \mathbb{F}^{n_{x,\mathrm{a}}} \setminus \{0\}$ be an eigenvector of $B_{\mathrm{a},2}$ for the eigenvalue $\lambda \neq 0$. Then Proposition 4.15 implies (with $m = \kappa + 1$) the existence of a history function ϕ such that the solution of the ITP (4.1) satisfies

$$w_{[2]}^{(\kappa+1)}(0) - w_{[1]}^{(\kappa+1)}(\tau^-) = B_{\mathrm{a},2}\left(\eta^{(\kappa+1)}(0^-) - w_{[1]}^{(\kappa+1)}(0)\right) = \lambda\tilde{w} \neq 0.$$

Thus, in general, we cannot expect the solution of a DDAE of discontinuity invariant type to get any smoother, which again justifies the terminology. For DDAEs of de-smoothing type, Example 1.5 might suggest that the solution becomes less and less smooth until it becomes discontinuous. This is however not necessarily the case as the following example demonstrates.

Example 4.26. Suppose that the ITP (4.1) satisfies Assumption 4.1 and additionally satisfies $NB_{\mathrm{a},2} = 0$, $NB_{\mathrm{a}} \neq 0$, and $N^2 B_{\mathrm{a}} = 0$, i.e., the DDAE is of de-smoothing type according to Theorem 4.16. Suppose that the history function ϕ satisfies (4.10). Then

$$\begin{aligned} w_{[2]}(0) - w_{[1]}(0) &= \sum_{k=0}^{\operatorname{ind}(E,A_1)-1} N^k B_{\mathrm{a}} \left(\phi^{(k)}(0^-) - x_{[1]}^{(k)}(0)\right) \\ &= \sum_{k=0}^{1} N^k B_{\mathrm{a}} \left(\phi^{(k)}(0^-) - x_{[1]}^{(k)}(0)\right) = 0. \end{aligned}$$

However, we have $\dot{v}_{[2]}(0) - \dot{v}_{[1]}(\tau^-) = 0$ by the definition of the slow system (4.5a) and by induction we infer

$$w_{[i+1]}(0) - w_{[i]}(\tau^-) = NB_{\mathrm{a},1}\left(\dot{v}_{[i-1]}(\tau^-) - \dot{v}_{[i]}(0)\right) = 0.$$

Thus the initial condition $x_{[i]}(0) = x_{[i-1]}(\tau^-)$ is consistent for (4.2) and hence the solution exists for all $t_{\mathrm{f}} > 0$. ♠

For a general analysis we assume that the ITP (4.1) satisfies Assumption 4.1, is of de-smoothing type, and the history function ϕ satisfies the splicing condition (4.23) for some $\kappa \in \mathbb{N}$. From (4.5a) we inductively infer

$$v_{[2]}^{(k)}(0) = Jv_{[2]}^{(k-1)}(0) + B_{\mathrm{d}} x_{[1]}^{(k-1)}(0) + g_{[2]}^{(k)}(0) = v_{[1]}^{(k)}(\tau^-)$$

for $k = 1, \ldots, \kappa + 1$. For the fast subsystem (4.5b), the splicing condition (4.23) implies

$$w_{[2]}(0) - w_{[1]}(\tau^-) = \sum_{k=\kappa+1}^{\operatorname{ind}(E,A_1)-1} N^k B_{\mathrm{a}} \left(\phi^{(k)}(0^-) - x_{[1]}^{(k)}(0)\right)$$

and hence a sufficient condition for the initial condition $w_{[2]}(0) = w_{[1]}(\tau^-)$ to be consistent is to assume $N^k B_{\mathrm{a}} = 0$ for $k \geq \kappa + 1$. This is immediately satisfied for $\operatorname{ind}(E, A_1) \leq \kappa + 1$. To analyze the next interval, we compute

$$\begin{aligned} w_{[3]}(0) - w_{[2]}(\tau^-) &= \sum_{k=1}^{\kappa} N^k B_{\mathrm{a}} T \begin{bmatrix} v_{[1]}^{(k)}(\tau^-) - v_{[2]}^{(k)}(0) \\ w_{[1]}^{(k)}(\tau^-) - w_{[2]}^{(k)}(0) \end{bmatrix} \\ &= \sum_{k=1}^{\kappa} N^k B_{\mathrm{a},2} \left(w_{[1]}^{(k)}(\tau^-) - w_{[2]}^{(k)}(0)\right). \end{aligned}$$

The assumption $NB_{\mathrm{a},2} = 0$ implies $w_{[3]}(0) - w_{[2]}(\tau^-) = 0$. Unfortunately, we have

$$v_{[3]}^{(2)}(0) - v_{[2]}^{(2)}(\tau^-) = B_{\mathrm{d},2}\left(\dot{w}_{[2]}(0) - \dot{w}_{[1]}(\tau^-)\right),$$

and thus cannot show that the initial condition $w_{[4]}(0) = w_{[3]}(\tau)$ is consistent without posing further assumptions on the matrices E, A_1, and A_2. Since this becomes quite technical, we summarize our findings only for the case $\operatorname{ind}(E, A_1) \leq 3$.

Theorem 4.27. *Suppose that the ITP* (4.1) *satisfies Assumption 4.1 and* $\operatorname{ind}(E, A_1) \leq 3$. *Moreover, assume* $NB_{\mathrm{a},2} = 0$ *and* $N^2 B_{\mathrm{a},1} B_{\mathrm{d},2} = 0$. *Then for every admissible history function* ϕ *that satisfies* (4.11) *for* $\kappa = 2$, *the ITP* (4.1) *has a unique solution.*

Proof. The assumptions on ϕ imply that the splicing condition (4.23) is satisfied for $\kappa = 2$ (see Proposition 4.7). Since $\operatorname{ind}(E, A_1) \leq 3$, we have $N^3 = 0$. Together with $NB_{\mathrm{a},2} = 0$ the previous discussion guarantees that a solution exists on the interval $[-\tau, 3\tau]$. Using $NB_{\mathrm{a},2} = 0$, we observe (inductively)

$$\begin{aligned} w_{[i+1]}(0) - w_{[i]}(\tau^-) &= \sum_{k=0}^{2} N^k B_{\mathrm{a},1} \left(v_{[i-1]}^{(k)}(\tau^-) - v_{[i]}^{(k)}(0) \right) \\ &= N^2 B_{\mathrm{a},1} B_{\mathrm{d},2} \left(\dot{w}_{[i-2]}(\tau^-) - \dot{w}_{[i-1]}(0) \right) = 0 \end{aligned}$$

and thus the initial condition $x_{[i+1]}(0) = x_{[i]}(\tau^-)$ is consistent for all $i \in \mathbb{N}$. The result follows from Corollary 2.15. ■

The assumptions in Theorem 4.27 can also be formulated in terms of the underlying DDE (4.7) and the matrices defined in (2.14). More precisely, (4.21) and

$$B_0(I_{n_x} - A^{\mathrm{con}}) = T \begin{bmatrix} B_{\mathrm{d},1} & B_{\mathrm{d},2} \\ 0 & 0 \end{bmatrix} T^{-1} T \left(\begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & I_{n_{x,\mathrm{a}}} \end{bmatrix} - \begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & 0 \end{bmatrix} \right) T^{-1} = T \begin{bmatrix} 0 & B_{\mathrm{d},2} \\ 0 & 0 \end{bmatrix} T^{-1}$$

imply that $NB_{\mathrm{a},2} = 0$ and $N^2 B_{\mathrm{a},1} B_{\mathrm{d},2} = 0$ if and only if

$$B_2(I_{n_x} - A^{\mathrm{con}}) = 0 \qquad \text{and} \qquad B_3 A^{\mathrm{con}} B_0 (I_{n_x} - A^{\mathrm{con}}) = 0,$$

respectively.

Remark 4.28. The proof of Theorem 4.27 shows that the result can be further improved by requiring different splicing conditions for the history function ψ for the slow state v and for the history function η of the fast state w. ♣

4.3 Comparison to the existing classification

In [98] the authors replace the delayed argument in the DDAE (4.1a) with a function parameter $\lambda \colon \mathbb{I} \to \mathbb{F}^{n_x}$ and obtain the *initial value problem* (IVP)

$$\begin{aligned} E\dot{x}(t) &= A_1 x(t) + A_2 \lambda(t) + f(t), \\ x(t) &= \phi(0), \end{aligned} \tag{4.24}$$

on the time interval $\mathbb{I}$. They call the function parameter λ *consistent* if there exists a consistent initial condition $\phi(0)$ for the IVP (4.24). Based on the function parameter λ the following classification for DDAEs [98] is introduced.

Definition 4.29. The DDAE (4.1a) is called *retarded*, *neutral*, or *advanced*, if the minimum smoothness requirement for a consistent function parameter λ is that $\lambda \in \mathscr{C}(\mathbb{I},\mathbb{F}^{n_x})$, $\lambda \in \mathscr{C}^1(\mathbb{I},\mathbb{F}^{n_x})$, or $\lambda \in \mathscr{C}^k(\mathbb{I},\mathbb{F}^{n_x})$ for some $k \geq 2$.

To compare the classification based on propagation of primary discontinuities (cf. Definition 4.9) with the classification given in [98], we need to understand Definition 4.29 in terms of the quasi-Weierstraß form.

Proposition 4.30. *Suppose that the matrix pair* (E, A_1) *in the DDAE* (4.1a) *is regular and the inhomogeneity* f *is sufficiently smooth. Then the DDAE* (4.1a) *is*

- *retarded if and only if* $B_{\mathrm{a}} = 0$*,*
- *neutral if and only if* $B_{\mathrm{a}} \neq 0$ *and* $NB_{\mathrm{a}} = 0$*, and*
- *advanced otherwise,*

where B_{a} *and* N *are the matrices from the quasi-Weierstraß form (Theorem 2.9) and* (4.5).

Proof. The smoothness requirements for λ can be directly seen from the underlying DDE (4.7). We have

$$B_0 = T\begin{bmatrix} I_{n_{x,\mathrm{d}}} & 0 \\ 0 & 0 \end{bmatrix} SA_2 = T\begin{bmatrix} B_{\mathrm{d}} \\ 0 \end{bmatrix} \qquad \text{and}$$

$$B_k = -T\begin{bmatrix} 0 & 0 \\ 0 & N^{k-1} \end{bmatrix} SA_2 = -T\begin{bmatrix} 0 \\ N^{k-1}B_{\mathrm{a}} \end{bmatrix}$$

for $k = 1,\dots,\mathrm{ind}(E, A_1)$. Hence (4.1a) is retarded if and only if $N^{k-1}B_{\mathrm{a}} = 0$ for all $k = 1,\dots,\mathrm{ind}(E, A_1)$, which is equivalent to $B_{\mathrm{a}} = 0$. The DDAE is neutral, if $N^{k-1}B_{\mathrm{a}} = 0$ for all $k = 2,\dots,\mathrm{ind}(E, A_1)$, which is equivalent to $NB_{\mathrm{a}} = 0$ and otherwise advanced. ■

With the characterization, we immediately see that the classification introduced [98] provides an upper bound for the new definition in the following sense.

Corollary 4.31. *Suppose that the ITP* (4.1) *satisfies Assumption 4.1.*

- *If the DDAE* (4.1a) *is not advanced, then it is not of de-smoothing type.*
- *If the DDAE* (4.1a) *is advanced, then it is of de-smoothing type.*

Since the classification introduced in this paper is based on the worst-case scenario, the numerical method described in [98], which is formulated for DDAEs that are not advanced, is safe to use.

Remark 4.32. The numerical method introduced in [98] is tailored to DDAEs that are not advanced and cannot be used for advanced DDAEs. However, if it is known that the history function satisfies the splicing condition (4.23) for some $\kappa > 0$, then also advanced DDAEs may be solved (cf. Theorem 4.27). Thus, there is a need for numerical integration schemes that can handle such situations. This is subject to further research. ♣

5 Nonlinear DDAEs

Having established the existence and uniqueness theory for linear *delay differential-algebraic equations* (DDAEs) in a distributional (chapter 3) and classical solution framework (chapter 4), we are now ready to turn our attention to nonlinear *initial value problems* (IVPs) of the form

$$0 = F(t, x(t), \dot{x}(t), x(t-\tau)), \qquad t \geq 0, \tag{5.1a}$$

$$x(t) = \phi(t), \qquad t \in [-\tau, 0] \tag{5.1b}$$

with (nonlinear) function $F\colon \mathbb{I} \times \mathbb{D}_x \times \mathbb{D}_{\dot{x}} \times \mathbb{D}_{\sigma_\tau x} \to \mathbb{F}^m$ defined on open sets $\mathbb{D}_x, \mathbb{D}_{\dot{x}}, \mathbb{D}_{\sigma_\tau x} \subseteq \mathbb{R}^{n_x}$ and time interval $\mathbb{I} = [t_0, t_\mathrm{f})$. The function $\phi\colon [-\tau, 0] \to \mathbb{R}^{n_x}$ is called initial trajectory or history function. Examples are for instance the hybrid pendulum-mass-spring-damper system discussed in section 1.1.1 and the delayed feedback control for the container crane in section 1.1.2. In both cases, the complete physical systems can be separated in two components, namely the numerical and experimental part for the hybrid testing approach, and the plant and the controller for the feedback control system. The decomposition is illustrated in Figure 5.1, which is a special case of Figure 1.1.

As in the previous chapters, we frequently use the shift operator σ_τ defined via $(\sigma_\tau x)(t) = x(t-\tau)$. In particular, the DDAE (5.1a) takes the form

$$0 = F(t, x, \dot{x}, \sigma_\tau x).$$

Let us mention that direct extensions to time-variable or state-dependent delays may be possible via the transformation described in [165, 166]. This is, however, beyond the scope of this thesis and requires further investigation. In general the solution of (5.1) depends on derivatives of F and ϕ and thus we make the following assumption for the remainder of this chapter.

Assumption 5.1. *The functions F and ϕ in* (5.1) *are sufficiently smooth.*

If F is linear and time-independent, then the analysis in chapter 3 reveals that the DDAE is delay-regular if and only if it can be transformed to a DDAE, where the *differential-algebraic equation* (DAE) that is obtained from the method of steps (cf. (2.3a)) is regular. We therefore restrict ourselves

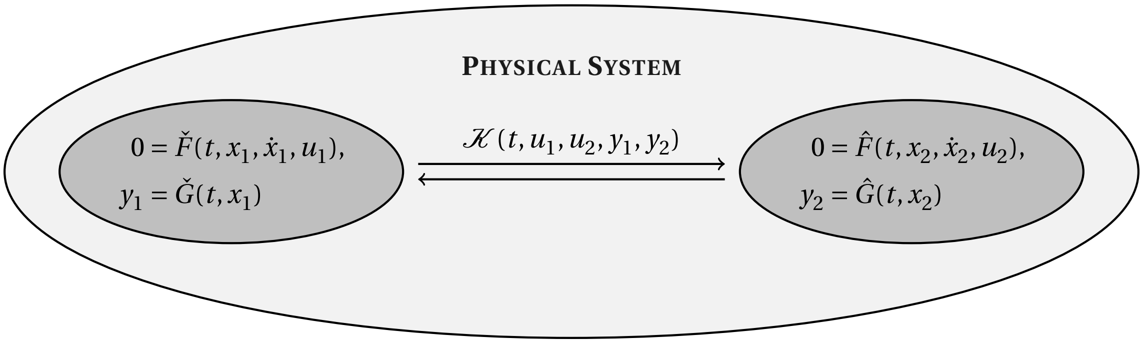

PHYSICAL SYSTEM
$0 = \check{F}(t, x_1, \dot{x}_1, u_1),$
$y_1 = \check{G}(t, x_1)$
$\mathscr{K}(t, u_1, u_2, y_1, y_2)$
$0 = \hat{F}(t, x_2, \dot{x}_2, u_2),$
$y_2 = \hat{G}(t, x_2)$

Figure 5.1 – Decomposition of a physical system into substructures

to this situation — in particular we assume $m = n_x$ — and analyze the well-posedness of (5.1) in this chapter with a classical solution concept. We illustrate our theoretical findings for the hybrid numerical-experimental system introduced in section 1.1.1, and hence first discuss this system in more detail in section 5.1.

The main tool to establish existence and uniqueness results for linear time-invariant DDAEs in Chapter 4 is the Weierstraß canonical form (cf. Theorem 2.9), which allows to decouple a linear time-invariant DAE into a differential equation and an algebraic equation. For nonlinear DAEs, the separation into differential and algebraic equations requires the implicit function theorem (cf. [127, Theorem 4.12]) and hence the analysis of the propagation of discontinuities in terms of the original DDAE becomes difficult. Instead, we make use of the fact that the results of Chapter 4 can also be stated in terms of the underlying *delay differential equation* (DDE) — see Theorem 4.21.

The two main contributions in this chapter are the following:

(i) We show that the compress-and-shift algorithm (Algorithm 1) from Chapter 3 can be applied to the nonlinear hybrid numerical-experimental system and terminates with a regular DDAE whenever the two subsystems are represented by regular DAEs. The details are presented in Lemma 5.10, Theorem 5.15 and Theorem 5.17.

(ii) We establish existence and uniqueness results for a class of nonlinear DDAEs in Theorem 5.24 and conclude that the hybrid system is solvable whenever the subsystems are strangeness-free (cf. Corollary 5.25).

5.1 Hybrid numerical-experimental system

For the general description of the model equations, we assume that we have already subdivided the complete model into two sub-models, which later on represent the numerical part and the experimental part. For an illustration we refer to Figure 5.1. The first subsystem is described by the descriptor system

$$0 = \check{F}(t, x_1, \dot{x}_1, u_1), \tag{5.2a}$$

$$y_1 = \check{G}(t, x_1) \tag{5.2b}$$

with state $x_1(t) \in \mathbb{R}^{n_{x,1}}$, input $u_1(t) \in \mathbb{R}^{m_1}$, and output $y_1(t) \in \mathbb{R}^{p_1}$. The second subsystem is given by

$$\begin{aligned} 0 &= \hat{F}(t, x_2, \dot{x}_2, u_2), && (5.3a) \\ y_2 &= \hat{G}(t, x_2) && (5.3b) \end{aligned}$$

with $x_2(t) \in \mathbb{R}^{n_{x,2}}$, $u_2(t) \in \mathbb{R}^{m_2}$, and $y_2(t) \in \mathbb{R}^{p_2}$. The complete model is given by imposing the interconnection

$$0 = \mathscr{K}(t, u_1, y_1, u_2, y_2). \tag{5.4}$$

In order to solve this relation for u_1 and u_2 it is common to require that

$$\begin{bmatrix} \frac{\partial \mathscr{K}}{\partial u_1} & \frac{\partial \mathscr{K}}{\partial u_2} \end{bmatrix} (t, u_1, y_1, u_2, y_2)$$

is nonsingular for all (t, u_1, y_1, u_2, y_2). For simplicity, we restrict ourselves to the case that the interconnection is given by

$$u_1(t) = y_2(t) \qquad \text{and} \qquad u_2(t) = y_1(t). \tag{5.5}$$

In particular, we assume $m_1 = p_2$ and $m_2 = p_1$. The complete model as depicted in Figure 5.1 is thus given by the implicit equation

$$0 = \begin{bmatrix} \check{F}(t, x_1, \dot{x}_1, \hat{G}(t, x_2)) \\ \hat{F}(t, x_2, \dot{x}_2, \check{G}(t, x_1)) \end{bmatrix} \tag{5.6}$$

with initial conditions

$$x_1(0) = \zeta_1 \qquad \text{and} \qquad x_2(0) = \zeta_2. \tag{5.7}$$

Example 5.2. To recast the coupled pendulum-mass-spring-damper system from section 1.1.1 in this form, we first have to transform the systems to first order. By introducing new variables for the velocities and after renaming we obtain

$$\begin{aligned} \check{F}(t, x_1, \dot{x}_1, u_1) &= \begin{bmatrix} \dot{x}_{1,1} - x_{1,2} \\ M\dot{x}_{1,2} + Cx_{1,2} + Kx_{1,1} - u_1 \end{bmatrix}, & \check{G}(t, x_1) &= x_{1,1}, \\ \hat{F}(t, x_2, \dot{x}_2, u_2) &= \begin{bmatrix} \dot{x}_{2,1} - x_{2,4} \\ \dot{x}_{2,2} - x_{2,5} \\ x_{2,6} - u_2 \\ m\dot{x}_{2,4} + 2x_{2,3}x_{2,1} \\ m\dot{x}_{2,5} + 2x_{2,3}(x_{2,2} - u_2) + mg \\ x_{2,1}^2 + (x_{2,2} - u_2)^2 - L^2 \end{bmatrix}, & \hat{G}(t, x_2) &= -2x_{2,3}(x_{2,2} - x_{2,6}) - mg. \end{aligned}$$

Note that we have introduced the artificial variable $x_{2,6}$ to account for the feedthrough, i.e., the fact that the force F_{pendulum} depends on the vertical position of the mass-spring-damper, which itself is used as input for the mathematical pendulum. ♠

If both subsystems are linear time-invariant, then we write

$$\begin{aligned} \check{F}(t, x_1, \dot{x}_1, u_1) &= \check{E}\dot{x}_1 - \check{A}x_1 - \check{B}u_1 + \check{f}(t), & \check{G}(t, x_1) &= \check{C}x_1, \\ \hat{F}(t, x_2, \dot{x}_2, u_2) &= \hat{E}\dot{x}_2 - \hat{A}x_2 - \hat{B}u_2 + \hat{f}(t), & \hat{G}(t, x_2) &= \hat{C}x_2, \end{aligned} \tag{5.8}$$

with external forcing functions $\check{f}$ and $\hat{f}$, such that the complete model (5.6) is given by

$$\begin{bmatrix} \check{E} & 0 \\ 0 & \hat{E} \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \check{A} & \check{B}\hat{C} \\ \hat{B}\check{C} & \hat{A} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \check{f} \\ \hat{f} \end{bmatrix}. \tag{5.9}$$

Before we continue our discussion, let us emphasize that, in general, there is no relation between the regularity of the subsystems (5.2) and (5.3) and the regularity of the coupled system (5.6). Also, the index from the subsystems might differ from the index of the coupled system. As an immediate consequence, the splitting of the system into smaller subsystems is a delicate task that must be performed carefully.

Example 5.3. Consider the linear DAE

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & c & 0 \\ c & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} \tag{5.10}$$

with external forcing function $f = [f_1, f_2, f_3]$ and parameter $c \in \mathbb{R}$. It is easy to see that for any $c \in \mathbb{R}$ the system has differentiation index $\nu = 1$. Splitting the system into $z_1 = [x_1, x_2]$ and $z_2 = x_3$ we obtain the two subsystems

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & c \\ c & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_1 + \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \tag{5.11a}$$

$$0 = -x_3 + u_2 + f_3. \tag{5.11b}$$

The second subsystem (5.11b) has differentiation index $\nu = 1$. For the first subsystem (5.11a) we observe that for $c = 0$ the pencil of the DAE is singular. For $c \neq 0$ the pencil is regular with index $\nu = 2$, which is higher than the index of the coupled system. ♠

Example 5.4. For $i = 1, 2$ we consider the subsystems

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \dot{x}_i = \begin{bmatrix} a_i & 0 \\ 0 & 1 \end{bmatrix} x_i + \begin{bmatrix} b_{i,1} & b_{i,2} \\ c_{i,1} & c_{i,2} \end{bmatrix} u_i,$$

which are already in Weierstraß canonical form (2.6) with index $\nu = 1$. The coupled system with coupling relations $u_1 = x_2$ and $u_2 = x_1$ is given by the linear DAE $E\dot{x} = Ax$ with

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} a_1 & 0 & b_{1,1} & b_{1,2} \\ 0 & 1 & c_{1,1} & c_{1,2} \\ b_{2,1} & b_{2,2} & a_2 & 0 \\ c_{2,1} & c_{2,2} & 0 & 1 \end{bmatrix}, \quad \text{and} \quad x = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix}.$$

Using strong equivalence, see section 2.1, we obtain

$$(E, A) \sim \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} a_1 & b_{1,1} & 0 & b_{1,2} \\ b_{2,1} & a_2 & b_{2,2} & 0 \\ 0 & c_{1,1} & 1 & c_{1,2} \\ c_{2,1} & 0 & c_{2,2} & 1 \end{bmatrix} \right),$$

and immediately observe that (E, A) has differentiation index $\nu = 1$ if and only if $c_{1,2}c_{2,2} \neq 1$. Otherwise, we obtain

$$(E, A) \sim \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} a_1 & b_{1,1} & b_{1,2} & 0 \\ b_{2,1} & a_2 - c_{1,1}b_{2,1} & -b_{21}c_{12} & 0 \\ c_{2,1} & -c_{1,1}c_{2,2} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right)$$

showing that also $\nu = 2$, $\nu = 3$, and (E, A) singular are possible. ♠

Remark 5.5. If both subsystems are port-Hamiltonian systems [21], then, under reasonable conditions, the coupled system itself is again a port-Hamiltonian system. In this case, [152, Thm. 4.3] implies that the differentiation index of the coupled system is at most $\nu = 2$. ♣

Our standing assumption is that the first model is simulated numerically, while the second model is tested experimentally. Following the discussion in section 1.1.1, the transfer system that realizes the numerical results in real-time within the experiment is delayed, such that the second model technically acts at a different time point. The hybrid numerical-experimental model, which we study in this paper, is thus given by

$$0 = \begin{bmatrix} \check{F}(t, x_1(t), \dot{x}_1(t), \hat{G}(t-\tau, x_2(t-\tau))) \\ \hat{F}(t-\tau, x_2(t-\tau), \dot{x}_2(t-\tau), \check{G}(t-\tau, x_1(t-\tau))) \end{bmatrix}, \tag{5.12}$$

which in the linear case simplifies to

$$\begin{aligned} \begin{bmatrix} \check{E} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \hat{E} \end{bmatrix} \begin{bmatrix} \dot{x}_1(t-\tau) \\ \dot{x}_2(t-\tau) \end{bmatrix} &= \\ \begin{bmatrix} \check{A} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 & \check{B}\hat{C} \\ \hat{B}\check{C} & \hat{A} \end{bmatrix} \begin{bmatrix} x_1(t-\tau) \\ x_2(t-\tau) \end{bmatrix} &+ \begin{bmatrix} \check{f}(t) \\ \hat{f}(t-\tau) \end{bmatrix}. \end{aligned} \tag{5.13}$$

Note that if the hybrid model is initialized at time t_0, then the numerical simulation starts at t_0, while the experimental part starts at $t_0 + \tau$. In particular, it is sufficient to prescribe an initial trajectory solely for the experimental part, i.e., only for x_2.

5.2 The method of steps

The standard procedure to solve initial trajectory problems for delay equations is via successive integration on the time intervals $[(i-1)\tau, i\tau)$ with $i = 1, \ldots, M$, where $M \in \mathbb{N}$ is the smallest integer such that $T \leq M\tau$. This approach is already discussed in chapter 2 and used in chapters 3 and 4 to establish existence and uniqueness results for linear time-invariant DDAEs. For the sake of presentation, the method is recalled in detail. For the DDAE (5.1a) we introduce for $i \in \mathscr{I} := \{1, \ldots, M\}$

$$\begin{aligned} x_{[i]}\colon [0,\tau] &\to \mathbb{R}^{n_x}, & t &\mapsto x(t + (i-1)\tau), \\ F_{[i]}\colon [0,\tau] \times \mathbb{D}_x \times \mathbb{D}_{\dot{x}} \times \mathbb{D}_{\sigma_\tau x} &\to \mathbb{R}^{n_x}, & (t, x, y, z) &\mapsto F(t + (i-1)\tau, x, y, z), \\ x_{[0]}\colon [0,\tau] &\to \mathbb{R}^{n_x}, & t &\mapsto \phi(t-\tau). \end{aligned} \tag{5.14}$$

Then we have to solve for each $i \in \{1,\dots,M\}$ the DAE

$$0 = F_{[i]}(t, x_{[i]}, \dot{x}_{[i]}, x_{[i-1]}), \qquad t \in [0,\tau), \tag{5.15a}$$

$$x_{[i]}(0) = x_{[i-1]}(\tau^-), \tag{5.15b}$$

with right continuation

$$x_{[i-1]}(\tau^-) := \lim_{t \nearrow \tau} x_{[i-1]}(t).$$

If (5.15) is uniquely solvable (provided that the initial value $x_{[i-1]}(\tau^-)$ is consistent), then we can construct the solution of (5.1) on the successive time intervals $[(i-1)\tau, i\tau)$. As outlined in the introduction we cannot expect a smooth transition of the solution between these intervals, see for instance Examples 1.4 and 1.5. We therefore extend the solution concept from Definition 4.3 to the nonlinear case.

Definition 5.6 (Solution concept)**.** Assume that F in the DDAE (5.1) and the initial trajectory ϕ are sufficiently smooth. We call $x \in \mathscr{C}(\mathbb{I}, \mathbb{R}^{n_x})$ a *solution* of (5.1) if for all $i \in \mathscr{I}$ the restriction $x_{[i]}$ of x as in (5.14) is a solution of (5.15). We call the initial trajectory $\phi\colon [-\tau, 0] \to \mathbb{R}^{n_x}$ *consistent* if the initial value problem (5.1) has at least one solution.

We emphasize that in order to check if an initial trajectory is consistent, we actually have to compute a solution of the initial value problem (5.1), see also section 4.1. This is in contrast to the DAE theory, where it suffices to compute the consistency set (2.34). To account for this issue, we adopt Definition 4.6, which ensures that we can at least ensure a solution in the interval $[0,\tau)$.

Definition 5.7 (Admissible initial trajectory)**.** The initial trajectory ϕ is called *admissible* for the DDAE (5.1a) if the initial condition

$$x_{[1]}(0) = \phi(0)$$

is consistent for the DAE (5.15) with $i = 1$.

Following the discussion in section 2.3, consistent initial values are characterized by the consistency set (2.34). We therefore have to assume that the DAE

$$0 = F_{[1]}(t, x_{[1]}, \dot{x}_{[1]}, \phi(t-\tau)) \tag{5.16}$$

satisfies Hypothesis 2.29. In order to simplify the discussion, we make the following definition, which is motivated from the discussion in [98].

Definition 5.8. The DAE that is obtained from the DDAE (5.1a) by substituting $x(t-\tau)$ with a control function $u(t)$ is called the *associated DAE* for the DDAE (5.1a). We say that the DDAE (5.1a) satisfies Hypothesis 2.33 if its associated DAE satisfies Hypothesis 2.33.

Suppose now that the DDAE (5.1a) satisfies Hypothesis 2.33 with strangeness index μ. Then the strangeness-free reformulation for the associated DAE as discussed in (2.38) is given by

$$0 = D(t, x, \dot{x}, u), \tag{5.17a}$$

$$0 = A\left(t, x, u, \dot{u}, \dots, u^{(\mu)}\right). \tag{5.17b}$$

Although formally, the algebraic equation depends on derivatives of u up to order μ, it may happen that

$$\frac{\partial A}{\partial u^{(\ell)}}\left(t, x, u, \dot{u}, \ldots, u^{(\mu)}\right) \equiv 0$$

for some $\ell \leq \mu$. Following the classification in Definition 4.9, respectively, Theorem 4.16 and Corollary 4.20, it is essential to know the largest number s such that

$$\frac{\partial A}{\partial u^{(s)}}\left(t, x, u, \dot{u}, \ldots, u^{(\mu)}\right) \not\equiv 0.$$

Consequently, from this point forward, we work with

$$0 = A\left(t, x, u, \dot{u}, \ldots, u^{(s)}\right) \tag{5.18}$$

instead of (5.17b), with the understanding that the algebraic equation does not depend on u, i.e.,

$$0 = A(t, x)$$

if $s = -\infty$. Replacing the control input u with the delayed argument results in the difference equation

$$0 = A\left(t, x, \sigma_\tau x, \sigma_\tau \dot{x}, \ldots, \sigma_\tau x^{(s)}\right). \tag{5.19}$$

Since the set of consistent initial values is described by (5.19), we immediately obtain the following result.

Lemma 5.9. *Assume that the history function ϕ is sufficiently smooth and the DDAE* (5.1a) *satisfies Hypothesis 2.33 with strangeness index μ. Let* (5.19) *denote the difference equation that results from the strangeness-free reformulation. Then ϕ is admissible for the DDAE* (5.1a) *if and only if*

$$0 = A\left(t, \phi(0), \phi(-\tau), \dot{\phi}(-\tau), \ldots, \phi^{(s)}(-\tau)\right). \tag{5.20}$$

Lemma 5.9 requires that the DDAE satisfies Hypothesis 2.33, which in turn implies that the associated DAE is regular. Unfortunately, this is only a sufficient condition for the existence of a unique solution for the IVP (5.1), as discussed in detail in section 3.1, see for instance Theorem 3.17. It is easy to see that the associated DAE for the hybrid numerical-experimental model (5.12) is not regular and therefore does not satisfy Hypothesis 2.33 and hence Lemma 5.9 does not apply to (5.12).

One strategy to resolve this issue is to find a reformulation of the DDAE (1.13) by shifting certain equations. This is achieved either by a combined shift-and-derivative array and the so-called *shift index* [94, 98], or by the compress-and-shift algorithm (Algorithm 2) presented in section 3.3. The latter algorithm's idea is to identify (after a potential transformation of the equations – the compression step), which equations do not depend on the current state but solely on the past state. These equations are then shifted in time, and the procedure is iterated. Let us emphasize that neither the shift-and-derivative array approach nor the compress-and-shift algorithm is readily available for general nonlinear DDAEs. Still, the special structure of the hybrid numerical-experimental model immediately suggests to shift the second block row of equations, yielding

$$0 = F(t,x,\dot{x},\sigma_\tau x) := \begin{bmatrix} \check{F}(t,x_1,\dot{x}_1,\sigma_\tau \hat{G}(t,x_2)) \\ \hat{F}(t,x_2,\dot{x}_2,\check{G}(t,x_1)) \end{bmatrix}, \tag{5.21}$$

with $x(t) := \left[\begin{smallmatrix} x_1(t) \\ x_2(t) \end{smallmatrix}\right] \in \mathbb{R}^{n_x}$, $n_x := n_{x,1} + n_{x,2}$. In the linear case (5.21) simplifies to

$$\begin{bmatrix} \check{E} & 0 \\ 0 & \hat{E} \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \check{A} & 0 \\ \hat{B}\check{C} & \hat{A} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 & \check{B}\hat{C} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_\tau x_1 \\ \sigma_\tau x_2 \end{bmatrix} + \begin{bmatrix} \check{f} \\ \hat{f} \end{bmatrix}. \tag{5.22}$$

We immediately obtain

$$\det\left(\begin{bmatrix} s\check{E} - \check{A} & 0 \\ -\hat{B}\check{C} & s\hat{E} - \hat{A} \end{bmatrix}\right) = \det(s\check{E} - \check{A})\det(s\hat{E} - \hat{A})$$

and thus have proven the next result.

Lemma 5.10. *The matrix pencil of the associated DAE for the linear shifted hybrid numerical-experimental system* (5.22) *is regular if and only if the associated DAEs of the linear subsystems* (5.8) *are regular.*

Remark 5.11. In the terminology of [98], the hybrid numerical-experimental system (5.13) has *shift index* $\kappa = 1$. In the literature, shifting of equations, i.e., systems with shift index $\kappa > 0$, are often referred to as noncausal and hence not physical. The hybrid numerical-experimental setup details that the shifting of equations can also occur if the dynamics of the subsystems affect the overall dynamic at different time instants. In particular, the requirement to shift parts of the equations may be a result of how a system is modeled. ♣

Before we proceed let us emphasize that shifting of equations potentially enlarges the solution space of the IVP for the differential equation.

Example 5.12. Consider the DDAE

$$\dot{x}_1(t) = x_2(t-\tau) + f(t), \tag{5.23a}$$

$$0 = x_2(t-\tau) - g(t). \tag{5.23b}$$

Notice that the second equation constitutes a restriction for the initial trajectory. Indeed, if we prescribe the initial trajectory

$$x_1(t) = \phi_1(t), \qquad x_2(t) = \phi_2(t), \qquad \text{for } t \leq 0, \tag{5.24}$$

then a solution cannot exist if $\phi_2(t) \neq g(t+\tau)$ for $t \in [-\tau, 0]$. If $\phi_2(t) = g(t+\tau)$ for $t \in [-\tau, 0]$, then the solution of the initial trajectory problem (5.23),(5.24) is given by

$$x_1(t) = \phi_1(0) + \int_0^t g(s) + f(s)\,\mathrm{d}s, \qquad x_2(t) = g(t+\tau) \qquad \text{for } t \geq 0.$$

In particular, the solution space for x_1 is parameterized by $\phi_1(0)$ and thus a one-dimensional vector space. If we, however, replace (5.23b) with the shifted equation

$$x_2(t) = g(t+\tau) \tag{5.25}$$

and consider the initial trajectory problem (5.23a),(5.25),(5.24), then for any initial trajectory ϕ that satisfies $\phi_2(0) = g(\tau)$ the solution of (5.23a),(5.25),(5.24) for $t \in [0,\tau]$ is given by

$$x_1(t) = \phi_1(0) + \int_0^t \phi_2(s-\tau) + f(s)\,\mathrm{d}s, \qquad x_2(t) = g(t+\tau),$$

such that the solution space for x_1 is infinite-dimensional. ♠

Remark 5.13. The shifted hybrid system (5.21) showcases, that only an initial trajectory for the experimental system $\hat{F}$ is required, which is in agreement with the discussion after (5.13). This is no contradiction to Example 5.12, since the numerical and experimental part are initialized at different time points. ♣

If the linear subsystems are regular, then Lemma 5.10 together with Theorem 2.19 immediately implies existence and uniqueness of solutions of the *initial trajectory problem* (ITP) for the DDAE (5.22) in the space of piecewise-smooth distributions, see [203] and section 2.2. Let us emphasize that $\tau > 0$ is a crucial assumption in Lemma 5.10, since Example 5.4 showcases that a similar result cannot be obtained if $\tau = 0$. Unfortunately, it is not immediately clear, what the index of the matrix pencil of the associated DAE is.

Example 5.14. Consider the matrix pencil

$$\left(\left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right], \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right]\right) \sim \left(\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ c & 0 & 0 & 1 \end{bmatrix}\right)$$

of the associated DAE for the hybrid numerical-experimental system (5.22), where both subsystems have differentiation index $\nu = 2$. If $c = 0$, then the pencil also has index $\nu = 2$, otherwise the index is $\nu = 3$. ♠

The index of the shifted hybrid numerical-experimental model depends on the coupling functions $\check{G}$ and $\hat{G}$. As a direct consequence, Hypothesis 2.33 has to be checked for each example separately, since it is not clear a-priori, what the corresponding strangeness index μ is. A notable exception is provided in the case that both subsystems are strangeness-free.

Theorem 5.15. *Suppose that the subsystems* (5.2) *and* (5.3) *are strangeness-free, i.e., satisfy Hypothesis 2.33 with characteristic values $\check{\mu} = \hat{\mu} = 0$, $\check{a}, \hat{a}, \check{d}$, and $\hat{d}$, respectively. If $\tau > 0$, then the shifted hybrid numerical-experimental model* (5.21) *satisfies Hypothesis 2.33 with characteristic values $\mu = 0$, $a = \check{a} + \hat{a}$, and $d = \check{d} + \hat{d}$.*

Proof. Let $\check{Z}_{\mathrm{A}}$, $\check{T}_{\mathrm{A}}$, $\check{Z}_{\mathrm{D}}$, and $\hat{Z}_{\mathrm{A}}$, $\hat{T}_{\mathrm{A}}$, $\hat{Z}_{\mathrm{D}}$ denote the matrix functions from Hypothesis 2.33 for the subsystems (5.2) and (5.3), respectively. Define $a := \check{a} + \hat{a}$ and accordingly

$$d = n - a = n_1 - \check{a} + n_2 - \hat{a} = \check{d} + \hat{d}.$$

Choose $\hat{T}_{\mathrm{A}}^{\star}$ such that $\begin{bmatrix} \hat{T}_{\mathrm{A}} & \hat{T}_{\mathrm{A}}^{\star} \end{bmatrix}$ is nonsingular. From Hypothesis 2.33 we deduce that

$$\left(\hat{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial x_2} \hat{T}_{\mathrm{A}}^{\star}\right)(t, x_2, \dot{x}_2, \hat{G}(t, x_1))$$

is nonsingular. Define (omitting arguments) the matrix functions

$$Z_{\mathrm{A}} := \begin{bmatrix} \check{Z}_{\mathrm{A}} & 0 \\ 0 & \hat{Z}_{\mathrm{A}} \end{bmatrix}, \qquad T_{\mathrm{A}} := \begin{bmatrix} \check{T}_{\mathrm{A}} & 0 \\ -\hat{T}_{\mathrm{A}}^{\star}\left(\hat{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial x_2} \hat{T}_{\mathrm{A}}^{\star}\right)^{-1} \check{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial u_2} \frac{\partial \check{G}}{\partial x_1} \hat{T}_{\mathrm{A}} & \hat{T}_{\mathrm{A}} \end{bmatrix}, \qquad Z_{\mathrm{D}} := \begin{bmatrix} \check{Z}_{\mathrm{D}} & 0 \\ 0 & \hat{Z}_{\mathrm{D}} \end{bmatrix}.$$

We have to check the different items from Hypothesis 2.33 for the shifted hybrid numerical-experimental model (5.21). We notice that $\check{\mu} = 0 = \hat{\mu}$ implies $\check{\mathscr{D}}_{\mu} = \check{F}$ and $\hat{\mathscr{D}}_{\mu} = \hat{F}$ and observe

$$\operatorname{rank}\left(\frac{\partial F}{\partial \dot{x}}\right) = \operatorname{rank}\left(\begin{bmatrix} \frac{\partial \check{F}}{\partial x_1} & 0 \\ 0 & \frac{\partial \hat{F}}{\partial x_2} \end{bmatrix}\right) = \operatorname{rank}\left(\frac{\partial \check{F}}{\partial x_1}\right) + \operatorname{rank}\left(\frac{\partial \hat{F}}{\partial x_2}\right) = \check{a} + \hat{a} = a.$$

We immediately conclude

$$\left(Z_{\mathrm{A}}^T \frac{\partial F}{\partial \dot{x}}\right)(t, x, \dot{x}, \sigma_\tau x) = \begin{bmatrix} \left(\check{Z}_{\mathrm{A}}^T \frac{\partial \check{F}}{\partial \dot{x}_1}\right)(t, x_1, \dot{x}_1, \sigma_\tau \hat{G}(t, x_2)) & 0 \\ 0 & \left(\hat{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial \dot{x}_2}\right)(t, x_2, \dot{x}_2, \check{G}(t, x_1)) \end{bmatrix} = 0$$

such that the first item from Hypothesis 2.33 is satisfied. For the second item we obtain (omitting arguments)

$$\hat{a} = \operatorname{rank}\left(\hat{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial x_2}\right) \leq \operatorname{rank}\left(\begin{bmatrix} \check{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial u_2} \frac{\partial \check{G}}{\partial x_1} & \hat{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial x_2} \end{bmatrix}\right) \leq \hat{a}$$

and thus

$$\begin{aligned} \operatorname{rank}\left(Z_{\mathrm{A}}^T \frac{\partial F}{\partial x}\right) &= \operatorname{rank}\left(\begin{bmatrix} \check{Z}_{\mathrm{A}}^T \frac{\partial \check{F}}{\partial x_1} & 0 \\ \check{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial u_2} \frac{\partial \check{G}}{\partial x_1} & \hat{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial x_2} \end{bmatrix}\right) \\ &= \operatorname{rank}\left(\check{Z}_{\mathrm{A}}^T \frac{\partial \check{F}}{\partial x_1}\right) + \operatorname{rank}\left(\hat{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial x_2}\right) = \check{a} + \hat{a} = a. \end{aligned}$$

We conclude

$$Z_{\mathrm{A}}^T \frac{\partial F}{\partial z} T_{\mathrm{A}} = \begin{bmatrix} \check{Z}_{\mathrm{A}}^T \frac{\partial \check{F}}{\partial z_1} \check{T}_{\mathrm{A}} & 0 \\ \check{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial u_2} \frac{\partial \check{G}}{\partial x_1} \hat{T}_{\mathrm{A}} - \hat{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial z_2} \hat{T}_{\mathrm{A}}^{\star}\left(\hat{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial x_2} \hat{T}_{\mathrm{A}}^{\star}\right)^{-1} \check{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial u_2} \frac{\partial \check{G}}{\partial x_1} \hat{T}_{\mathrm{A}} & \hat{Z}_{\mathrm{A}}^T \frac{\partial \hat{F}}{\partial z_2} \hat{T}_{\mathrm{A}} \end{bmatrix} = 0.$$

Similarly as before we have

$$\operatorname{rank}\left(\frac{\partial F}{\partial \dot{x}} T_{\mathrm{A}}\right) = \operatorname{rank}\left(\frac{\partial \check{F}}{\partial \dot{x}_1} \check{T}_{\mathrm{A}}\right) + \operatorname{rank}\left(\frac{\partial \hat{F}}{\partial \dot{x}_2} \hat{T}_{\mathrm{A}}\right) = \check{d} + \hat{d} = d.$$

The proof follows from

$$\operatorname{rank}\left(Z_{\mathrm{D}}^T \frac{\partial F}{\partial \dot{x}} T_{\mathrm{A}}\right) = \operatorname{rank}\left(\check{Z}_{\mathrm{D}}^T \frac{\partial \check{F}}{\partial \dot{x}_1} \check{T}_{\mathrm{A}}\right) + \operatorname{rank}\left(\hat{Z}_{\mathrm{D}}^T \frac{\partial \hat{F}}{\partial \dot{x}_2} \hat{T}_{\mathrm{A}}\right) = d. \qquad \blacksquare$$

Remark 5.16. The assumption $\tau > 0$ is crucial in Theorem 5.15. In the case $\tau = 0$, we have already seen in Example 5.4 that even if both subsystems are strangeness-free, the coupled system might have strangeness-index $\mu > 0$. ♣

In the case that either of the subsystems is not strangeness-free we can proceed as follows. Let

$$\begin{aligned} 0 &= \check{D}(t, x_1, \dot{x}_1, u_1), & 0 &= \hat{D}(t, x_2, \dot{x}_2, u_2),\\ 0 &= \check{A}\left(t, x_1, u_1, \dot{u}_1, \ldots, u_1^{(\check{\mu})}\right), & 0 &= \hat{A}\left(t, x_2, u_2, \dot{u}_2, \ldots, u_2^{(\hat{\mu})}\right),\\ \dot{x}_1 &= \check{\mathfrak{f}}\left(t, x_1, u_1, \dot{u}_1, \ldots, u^{(\check{\mu}+1)}\right), & \dot{x}_2 &= \hat{\mathfrak{f}}\left(t, x_2, u_2, \dot{u}_2, \ldots, u^{(\hat{\mu}+1)}\right) \end{aligned}$$

denote the strangeness-free reformulations and the underlying ODEs for (5.2) and (5.3), respectively. Recall the coupling conditions

$$u_1 = \sigma_\tau\big(\hat{G}(t, x_2)\big) \qquad \text{and} \qquad u_2 = \check{G}(t, x_1),$$

which we have to differentiate $\hat{\mu} + 1$, respectively $\check{\mu} + 1$ times. We observe that in the interval $[0, \tau)$ the coupling condition for u_1 does not depend on x_2 but on the history ϕ_2. In particular, we obtain (assuming that $\hat{G}$ is sufficiently smooth)

$$\begin{aligned} \dot{u}_1 &= \sigma_\tau\left(\frac{\partial \hat{G}}{\partial t}(t, \phi_2) + \frac{\partial \hat{G}}{\partial x_2}(t, \phi_2)\dot{\phi}_2\right),\\ \ddot{u}_1 &= \sigma_\tau\left(\frac{\partial^2 \hat{G}}{\partial t^2}(t, \phi) + 2\frac{\partial^2 \hat{G}}{\partial t \partial x_2}(t, \phi_2)\dot{\phi}_2 + \frac{\partial^2 \hat{G}}{\partial x_2^2}(t, \phi_2)\dot{\phi}_2 + \frac{\partial \hat{G}}{\partial x_2}(t, \phi_2)\ddot{\phi}_2\right), \end{aligned}$$

and similarly for higher derivatives. In particular, there exist functions $\check{\check{D}}$, $\check{\check{A}}$, and $\check{\check{\mathfrak{f}}}$

$$\begin{aligned} 0 &= \check{\check{D}}(t, x_1, \dot{x}_1, \sigma_\tau \phi_2),\\ 0 &= \check{\check{A}}\left(t, x_1, \sigma_\tau \phi_2, \sigma_\tau \dot{\phi}_2, \ldots, \sigma_\tau \phi_2^{(\check{\mu})}\right),\\ \dot{x}_1 &= \check{\check{\mathfrak{f}}}\left(t, x_1, \sigma_\tau \phi_2, \sigma_\tau \dot{\phi}_2, \ldots, \sigma_\tau \phi_2^{(\check{\mu}+1)}\right) \end{aligned}$$

for $t \in [0, \tau)$. Consequently, we can (locally) solve for x_1, provided that the initial trajectory ϕ_2 is sufficiently smooth and $x_1(0)$ satisfies the consistency condition

$$0 = \check{\check{A}}\left(0, x_1(0), \phi_2(-\tau), \dot{\phi}_2(-\tau), \ldots, \phi_2^{(\check{\mu})}(-\tau)\right).$$

On the other hand, the input relation for u_2 implies

$$\begin{aligned} \dot{u}_2 &= \frac{\partial \hat{G}}{\partial t}(t, x_1) + \frac{\partial \hat{G}}{\partial x_1}(t, x_1)\dot{x}_1\\ &= \frac{\partial \hat{G}}{\partial t}(t, x_1) + \frac{\partial \hat{G}}{\partial x_1}(t, x_1)\check{\check{\mathfrak{f}}}\left(t, x_1, \sigma_\tau \phi_2, \sigma_\tau \dot{\phi}_2, \ldots, \sigma_\tau \phi_2^{(\check{\mu}+1)}\right). \end{aligned}$$

Note that although derivatives of ϕ_2 up to order $\check{\mu} + 1$ appear, $\dot{u}_2$ does not necessarily depend on all of them (see for instance Example 5.14 and the discussion after Definition 5.8). In any case, there exists functions $\hat{\hat{D}}$, $\hat{\hat{A}}$, and $\hat{\hat{\hat{\mathfrak{f}}}}$ such that

$$
\begin{aligned}
0 &= \hat{\hat{D}}(t, x_2, \dot{x}_2, x_1),\\
0 &= \hat{\hat{A}}\left(t, x_2, x_1, \sigma_\tau \phi_2, \sigma_\tau \dot{\phi}_2, \dots, \sigma_\tau \phi_2^{(\check{\mu}+\hat{\mu})}\right),\\
\dot{x}_2 &= \hat{\hat{\hat{\mathfrak{f}}}}\left(t, x_1, x_2, \sigma_\tau \phi_2, \sigma_\tau \dot{\phi}_2, \dots, \sigma_\tau \phi_2^{(\check{\mu}+\hat{\mu}+1)}\right).
\end{aligned}
$$

Thus, the underlying delay differential equation for the shifted hybrid numerical-experimental system (5.21) in $[0, \tau)$ is given by

$$
\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} \check{\check{\mathfrak{f}}}\left(t, x_1, \sigma_\tau \phi_2, \sigma_\tau \dot{\phi}_2, \dots, \sigma_\tau \phi_2^{(\check{\mu}+1)}\right) \\ \hat{\hat{\hat{\mathfrak{f}}}}\left(t, x_1, x_2, \sigma_\tau \phi_2, \sigma_\tau \dot{\phi}_2, \dots, \sigma_\tau \phi_2^{(\check{\mu}+\hat{\mu}+1)}\right) \end{bmatrix} \tag{5.26}
$$

and the differentiation index is at most $\check{\mu} + \hat{\mu} + 1$ and we have shown the following result.

Theorem 5.17. *Assume that the subsystems* (5.2) *and* (5.3) *satisfy Hypothesis 2.33 with strangeness index $\check{\mu}$, $\hat{\mu}$, respectively. Then the shifted hybrid numerical-experimental system* (5.21) *has a well-defined differentiation index, which is at most $\check{\mu} + \hat{\mu} + 1$.*

Example 5.18. Shifting the equations for the pendulum in the hybrid version of the coupled pendulum-spring-mass-damper system in (1.5) and introducing new variables $v_1 := \dot{y}_1$, $v_2 := \dot{x}_2$, and $v_3 := \dot{y}_2$ for the velocities, yields the system

$$
\begin{aligned}
\dot{y}_1 &= v_1, && (5.27a)\\
\dot{x}_2 &= v_2, && (5.27b)\\
\dot{y}_2 &= v_3, && (5.27c)\\
M\dot{v}_1 + Cv_1 + Ky_1 &= f(\sigma_\tau y_1, \sigma_\tau y_2, \sigma_\tau \lambda), && (5.27d)\\
m\dot{v}_2 &= -2\lambda x_2, && (5.27e)\\
m\dot{v}_3 &= -2\lambda(y_2 - y_1) - m\mathrm{g}, && (5.27f)\\
0 &= x_2^2 + (y_2 - y_1)^2 - L^2, && (5.27g)
\end{aligned}
$$

which is a multibody system with forcing term $f(y_1, y_2, \lambda) = -2\lambda(y_2 - y_1) - m\mathrm{g}$ that solely depends on delayed variables. Since multibody systems are special instances of Hessenberg systems, we conclude from [127, Sec. 4.2] that the shifted hybrid pendulum-mass-spring-damper system has strangeness index $\mu = 2$ and satisfies Hypothesis 2.33 with $a = 3$ and $d = 4$. The algebraic equations and the difference equations are given by

$$
\begin{aligned}
0 &= x_2^2 + (y_2 - y_1)^2 - L^2,\\
0 &= 2x_2 v_2 + 2(y_2 - y_1)(v_3 - v_1),\\
0 &= 2v_2^2 + 2(v_2 - v_1)^2 - \frac{4}{m}\lambda(x_2^2 + (y_2 - y_1)^2) - 2(y_2 - y_1)\left(\mathrm{g} + \frac{f(\sigma_\tau y_1, \sigma_\tau y_2, \sigma_\tau \lambda)}{M} - \frac{C}{M}v_1 - \frac{K}{M}y_1\right).
\end{aligned}
$$

Let us emphasize that despite the higher index, the algebraic equations do not depend on derivatives of $\sigma_\tau x$. Note that also the Lagrange-multiplier is delayed in (5.27d) such that this example is not included in the specific retarded Hessenberg forms as studied in [13]. ♠

Remark 5.19. Models that feature a similar delay structure as in (5.21) and (5.22) arise in the time-discretization via waveform relaxzation [15, 72, 159] or the analysis of semi-explicit time-integrators [4], see also section 1.1.5. ♣

5.3 Solvability of the hybrid model

In the previous section, we have established that the shifted hybrid numerical-experimental system (5.21) can be solved in the interval $[0,\tau)$ and is regular in the sense of Theorem 2.32, provided that the subsystems satisfy Hypothesis 2.33 and the history function is admissible. The question that remains to be answered is whether a solution exists on time intervals $[0,T)$ with $T>\tau$.

Remark 5.20. For the linear time-invariant case, this is discussed in detail in Chapter 3 for a distributional solution concept and in [50, 52, 53, 97, 173, 211] for the solution concept as defined in Definition 5.6. Results for linear time-varying systems are developed for instance in [98, 99]. Moreover, a special class of nonlinear DDAEs is discussed in [13]. ♣

In view of the method of steps discussed in the previous section, the question that remains to be answered is, which conditions on the subsystems and the history function ensure that the initial condition (5.15b) is consistent for all $i=1,\dots,M$. Unfortunately, the regularity of the subsystems and an admissible history function are not sufficient to establish a solution for $t>\tau$, see for instance, the discussion in Chapter 4 and the following example.

Example 5.21. Consider the regular DDAE

$$\dot{x}(t)=y(t), \qquad 0=x(t)-y(t-1).$$

Applying the method of steps, equation (5.15) yields

$$x_{[i]}=y_{[i-1]} \qquad \text{and} \qquad y_{[i]}=\dot{y}_{[i-1]}. \tag{5.28}$$

For the history function $\phi(t)=\left[\begin{smallmatrix}0\\ t+1\end{smallmatrix}\right]$ we obtain $x_{[1]}(t)=t$ and $y_{[1]}(t)=1$, and we deduce that the history function is admissible. However, the initial value $y_{[1]}(1)=1$ is not consistent for the associated DAE on the interval $[1,2)$. In particular, the solution exists only on the interval $[0,1)$. ♠

The issue in the previous example is, as already outlined in chapter 4, that the equation $z_i=\dot{z}_{i-1}$ results in solutions z_i that become less smooth for increasing i, and possible discontinuities of the form

$$x_{[i-1]}^{(k)}(\tau^-)\neq x_{[i]}^{(k)}(0)$$

are propagated to

$$x_{[i]}^{(k-1)}(\tau^-)\neq x_{[i+1]}^{(k-1)}(0).$$

The discontinuity propagation leads to the classification introduced in Definition 4.9, with a complete characterization presented in Theorem 4.16. Unfortunately, the main tool for the proof of Theorem 4.16 is the Weierstraß canonical form, and thus in general, we cannot expect to have a similar result for the nonlinear DDAE (5.1a). Instead, we use the classification given in [98] and make use of the fact, that in the linear case, this classification provides an upper bound for the classification in Definition 4.9 in the sense of Corollary 4.31.

Definition 5.22. Assume that the DDAE (5.1a) satisfies Hypothesis 2.33 and let

$$\dot{x} = \mathfrak{f}\left(t, x, \sigma_\tau x, \ldots, \sigma_\tau x^{(s)}\right) \tag{5.29}$$

denote the underlying DDE of the DDAE (5.1a) and assume $\frac{\partial \mathfrak{f}}{\partial \sigma_\tau x^{(s)}} \not\equiv 0$. Then (5.1a) is called *retarded, neutral,* or *advanced* if $s = 0$, $s = 1$, or $s \geq 2$ in (5.29).

Lemma 5.23. *Assume that the DDAE* (5.1a) *satisfies Hypothesis 2.33 and let*

$$0 = D\left(t, x, \dot{x}, \sigma_\tau x\right), \tag{5.30a}$$

$$0 = A\left(t, x, \sigma_\tau x, \sigma_\tau \dot{x}, \ldots, \sigma_\tau x^{(s-1)}\right) \tag{5.30b}$$

denote the associated strangeness-free reformulation with the convention that either A does not depend on $\sigma_\tau x^{(k)}$ for any $k \in \mathbb{N}$, or

$$\frac{\partial A}{\partial \sigma_\tau x^{(s-1)}} \not\equiv 0.$$

Then (5.1a) *is retarded, neutral, or advanced, if $\frac{\partial A}{\partial \sigma_\tau x^{(k)}} \equiv 0$ for all $k \in \mathbb{N}$, $s = 1$, or $s = 2$, respectively.*

Proof. The proof follows immediately from rewriting (5.30) as in (2.39) and (2.40). ■

Theorem 5.24. *Suppose that the DDAE* (5.1a) *is sufficiently smooth, has strangeness-index μ, satisfies Hypothesis 2.33 with characteristic values μ, a, d and $\mu+1, a, d$, is not advanced, and the history function $\phi_0 \in \mathscr{C}^1([0,\tau], \mathbb{R}^{n_x})$ is admissible. Then the initial trajectory problem* (5.1) *is solvable.*

Proof. Since the DDAE (5.1a) satisfies Hypothesis 2.33 and is not advanced, Lemma 5.23 implies that the strangeness-free reformulation is of the form

$$0 = D(t, x, \dot{x}, \sigma_\tau x), \qquad\qquad 0 = A(t, x, \sigma_\tau x) \tag{5.31}$$

with the understanding that A may not depend on $\sigma_\tau x$. Applying the method of steps to (5.31) yields the sequence of initial value problems

$$\begin{aligned} 0 &= D\left(t + (i-1)\tau, x_{[i]}, \dot{x}_{[i]}, x_{[i-1]}\right), \\ 0 &= A\left(t + (i-1)\tau, x_{[i]}, x_{[i-1]}\right), \\ x_{[i]}(0) &= x_{[i-1]}(\tau^-). \end{aligned} \tag{5.32}$$

Since the history function is admissible, we can (locally) solve (5.32) for $i = 1$ and by [127, Theorem 4.13] this solution is also a solution of (5.15). Although this solution is of local nature it can be globalized by applying the cited theorem again until we reach the boundary of $\mathscr{M}_\mu$ (cf. [127, Remark 4.14]). If we assume that the solution exists on the time interval $[0,\tau)$ this immediately implies

$$0 = \lim_{t \nearrow \tau} A(t, x_{[1]}(t), x_{[0]}(t)) = A(\tau, x_{[1]}(\tau^-), x_{[0]}(\tau)).$$

Hence, $x_{[1]}(\tau^-)$ is consistent for the DAE (5.15) with $i = 2$. The result follows iteratively by repeating this procedure. ■

Corollary 5.25. *Suppose that the numerical and experimental subsystems* (5.2) *and* (5.3) *both satisfy Hypothesis 2.33 with* $\mu = 0$. *Then for any* $\tau > 0$ *and for any admissible history function* ϕ, *the initial trajectory problem for the shifted hybrid numerical-experimental system* (5.21) *is solvable.*

Proof. Theorem 5.15 ensures that the shifted hybrid system is strangeness-free. Lemma 5.23 thus implies that (5.21) is not advanced. The result follows from Theorem 5.24. ■

Example 5.26. Although the system for the pendulum (1.3) is not strangeness-free, Example 5.18 shows that the shifted hybrid system resulting from coupling the pendulum with the mass-spring-damper system is not advanced. In particular, Theorem 5.24 ensures that the associated initial trajectory problem is solvable. ♠

If the DDAE (5.1a) is advanced, then in general we cannot expect a solution for the ITP (5.1), see for instance Example 1.5 and the results from Chapter 4. However, if the initial trajectory is linked smoothly to the solution, i.e., the initial trajectory satisfies the splicing condition (4.23), then we can expect to establish further results similar to Theorem 4.27. Nevertheless, mimicking the strategy from the proof of Theorem 4.27 in the nonlinear case, requires the use of the underlying DDE (5.29) and thus the implicit function theorem. A detailed analysis of this setting is currently under investigation and subject to further research.

Part II

Structured realization theory

6 Problem setting and background

From a modeling perspective, it is often difficult to describe a physical or chemical system exactly via differential equations and modeling laws usually apply only for ideal settings. Thus, it is desirable to construct a model in an automated fashion directly from data, that may come from some experiment. The data may be in the form of

- a time series, for instance obtained from a numerical simulation or experiment, or
- transfer function evaluations, for instance obtained by a vector network analyzer [11].

The standing assumption in this second part of the thesis is that the data is generated by a dynamical system Σ, exemplified in Figure 6.1. Hereby we do not assume any knowledge about the system, in particular, there is no realization, i.e., a description of the internal dynamics, of the system available.

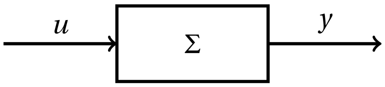


Figure 6.1 – Input-output mapping of a black-box system

Despite the inaccessibility of a description of detailed internal dynamics, there may yet be significant auxiliary information or at least a basic understanding of how the system should behave, allowing one to surmise general structural features of the underlying dynamical system. For example, vibration effects are naturally associated with subsystems that have second-order structure; internal transport or signal propagation will naturally be associated with time delays—see Table 6.1 for further examples.

Example 6.1 (Acoustic transmission, [191]). Consider the acoustic transmission example from section 1.1.4 and suppose we are interested in the acoustic pressure $y(t) = p(\xi_0, t)$ at a fixed point $\xi_0 \in (0, L)$ in the duct (see Figure 1.5), which we view as the output of an abstract yet unknown system that is driven by the input fluid velocity $u(t)$, determined by the acoustic driver positioned at $\xi = 0$. Instead of using the *partial differential equation* (PDE) model (1.8), we simply assume that the output pressure depends linearly on the input velocity in a way that is invariant to translation in

Table 6.1 – Examples for system structures with output mapping $y(t) = Cx(t)$

	state space description	**transfer function**
second-order	$A_1\ddot{x}(t) + A_2\dot{x}(t) + A_3x(t) = Bu(t)$	$C\left(s^2A_1 + sA_2 + A_3\right)^{-1}B$
state delay	$A_1\dot{x}(t) + A_2x(t) + A_3x(t-\tau) = Bu(t)$	$C\left(sA_1 + A_2 + \mathrm{e}^{-\tau s}A_3\right)^{-1}B$
neutral delay	$A_1\dot{x}(t) + A_2x + A_3\dot{x}(t-\tau) = Bu(t)$	$C\left(sA_1 + A_2 + s\mathrm{e}^{-\tau s}A_3\right)^{-1}B$
viscoelastic	$A_1\ddot{x}(t) + \int_0^t h(t-\tau)A_2\dot{x}(\tau)\mathrm{d}\tau + A_3x(t) = Bu(t)$	$C\left(s^2A_1 + s\hat{h}(s)A_2 + A_3\right)^{-1}B$

time, and so the output could be anticipated to involve some superposition of internal states that are lagged in time according to propagation delays related to the distance traveled by the signal. Assuming a uniform sound speed $c > 0$ throughout the duct, we allow for a direct propagation delay $\tau_1 = \xi_0/c$ between the input and output location and a second propagation delay $\tau_2 = (2L-\xi_0)/c$, associated with a reflected signal. A semi-empirical model for the state evolution of a system that has these basic features could have the form

$$A_1x(t) + A_2x(t-\tau_1) + A_3x(t-\tau_2) = bu(t),$$

with an output port map given by $y(t) = c^T x(t)$. The matrices A_1, A_2, and A_3, the port maps associated with the vectors b and c, as well as their dimensions are unknown. ♠

Throughout this part of the thesis we make the assumption that the system Σ in Figure 6.1 is linear, i.e., there exists a *linear time-invariant* (LTI) operator S with $y = Su$. We are thus interested in finding a suitable operator (in the sense of section 1.2), that approximates the data in some suitable norm.

Problem 6.2. *Construct a structured LTI operator $\widetilde{S}$ solely from input/output data such that*

$$\|y - \widetilde{y}\| = \|Su - \widetilde{S}u\| \leq \varepsilon\|u\| \tag{6.1}$$

for all admissible input signals u, a small parameter $\varepsilon \geq 0$, and suitable norms.

In order to solve Problem 6.2 we have to address the question what kind of data we assume available and define precisely, what a *structured LTI operator* is. Recall that LTI systems can be represented either in the time domain or in the frequency domain [6]. The mapping from one domain to the other is given by the Laplace-transform for continuous time systems and the Z-transform for discrete time systems. Moreover, the $\mathscr{L}_\infty$ error in the time domain can be bounded by the $\mathscr{H}_2$ error in the frequency domain via

$$\left\|y - \widetilde{y}\right\|_{\mathscr{L}_\infty} := \sup_{t>0}\|y(t) - \widetilde{y}(t)\|_\infty \leq \|S - \widetilde{S}\|_{\mathscr{H}_2}\|u\|_{\mathscr{L}_2}. \tag{6.2}$$

In fact, for *single-input/single-output* (SISO) systems, the $\mathscr{H}_2$ norm is the $\mathscr{L}_2$-$\mathscr{L}_\infty$ induced norm of the underlying convolution operator, i.e. $\|S - \widetilde{S}\|_{\mathscr{H}_2}$ is the smallest number such that (6.2) holds for all inputs $u \in \mathscr{L}_2$ [20].

It is well-known (cf. [20] and the references therein) that if the solution operator S is the convolution operator of a standard state-space realization, that is (assuming a zero initial condition and no direct feedthrough)

$$(Su)(t) = \int_0^t C\exp(A(t-s))Bu(s)\mathrm{d}s,$$

the $\mathscr{H}_2$ error $\|S-\widetilde{S}\|_{\mathscr{H}_2}$ is minimized if the transfer function of $\widetilde{S}$ interpolates the transfer function of S at the mirror images of the poles of $\widetilde{S}$. Thus our approach to solve Problem 6.2 is to construct $\widetilde{S}$ such that it is an interpolant of S in the frequency domain.

Remark 6.3. If a state-space description of the dynamical process under investigation is known, one could use *model order reduction* (MOR) methods (see the recent surveys and books [6, 8, 17, 28, 29, 106, 177]) to obtain a low-dimensional and cheap-to-evaluate surrogate model. Among the many MOR methods let us mention rational interpolation (formerly known as moment matching) [14, 18] and $\mathscr{H}_2$-optimal interpolation [89] as methods that also interpolate the transfer function. Note that the $\mathscr{H}_2$ optimality conditions for structured problems are much more involved [22, 70] and to our knowledge, there exists no general computational strategy to obtain optimal interpolation points even if the state-space description is available. Preservation of system structures of the state-space description is for instance considered in [18, 57, 58, 78, 133, 155, 199]. We note that almost all of these approaches require an internal description. Notable exceptions are provided in [71, 184, 189, 191]. ♣

Remark 6.4. For some model problems, for instance a circuit that involves a lossless transmission line [39], it is possible to transform a hyperbolic PDE into a delay equation [61, 139] that is — from a computational perspective — much easier to solve. See also section 1.1.4 for a detailed example. Thus even if a state-space description is available, it may be advantageous to choose a different structure for the surrogate model than for the original model. Notice that many of these problems are characterized by slowly decaying Hankel singular values or Kolmogorov n-widths (see [213] for a connection between the two concepts), which prevents classical MOR methods from succeeding and thus requires a special treatment [35, 48, 162, 180]. ♣

The second part of the thesis is organized as follows: First, we precisely state the problem for realizing a *delay differential-algebraic equation* (DDAE) from frequency measurements of a transfer function in section 6.1 and introduce the term system structure, that allows us to not only identify a DDAE from measurements, but a larger system class (cf. Problem 6.6 and Problem 6.9). The framework to obtain an interpolant of the frequency domain data is derived in chapter 7 and can be understood as a generalization of the Loewner framework [150] (see section 6.2 for further details). If we can access only input/output measurements in the time domain, we can estimate frequency data via the *empirical transfer function estimate* (ETFE) [137] or the *least-squares transfer function estimate* (lsTFE) [168]. This approach together with the estimation of unknown parameters in the system structure, for instance the delay time τ, is presented in chapter 8.

Remark 6.5. If the data under consideration is prone to noise one may wish to use a least-squares approach instead of interpolation, such as *vector fitting* [66, 67, 93], or *dynamic mode decomposition* [128, 186, 210]. We consider this, however, a second step and thus postpone this to future work. ♣

Let us mention that most of the content of this part of the thesis is already published in the journal articles [189, 191] and the preprint [77]. The presented results are joint work with Christopher Beattie (Virginia Tech), Elliot Fosong (University of Cambridge), Serkan Gugercin (Virginia Tech), and Philipp Schulze (TU Berlin).

6.1 Problem setting

Recall that a LTI DDAE is given by

$$\begin{aligned} E\dot{x}(t) &= A_1x(t) + A_2x(t-\tau) + Bu(t),\\ y(t) &= Cx(t), \end{aligned} \tag{6.3}$$

and — as before — we call $x(t) \in \mathbb{R}^{n_x}$, $u(t) \in \mathbb{R}^{n_u}$, and $y(t) \in \mathbb{R}^{n_y}$, the *state*, *input*, and *output* of the system (6.3), which we assume to be exponentially bounded. In this case the Laplace transform may be applied to (6.3) and rearranged to $\widehat{y}(s) = H(s)\widehat{u}(s)$ with

$$H(s) = C(sE - A_1 - \exp(-\tau s)A_2)^{-1}B,$$

provided that $\det(sE - A_1 - \exp(-\tau s)A_2)$ is not vanishing identically (i.e., the DDAE (6.3) is delay-regular, cf. Theorem 3.20) and the initial condition

$$x(t) = 0 \qquad \text{for } t \in [-\tau, 0]$$

is satisfied. The function $H : \mathbb{C} \to \mathbb{C}^{n_y \times n_u}$ is called the *transfer function* of (6.3). Since the transfer function characterizes the input-output behavior of (6.3), measurements of H seem appropriate to construct the realization. More precisely, we assume that the following data is available: Suppose we have $2n$ points in the complex plane, which may be interpreted as complex driving frequencies, $\{\mu_1, \ldots, \mu_n\}$ and $\{\sigma_1, \ldots, \sigma_n\}$. In addition to these complex frequencies, we have the so-called *left tangential direction* vectors $\{\ell_1, \ldots, \ell_n\}$ and the *right tangential direction* vectors $\{r_1, \ldots, r_n\}$ where $\ell_i \in \mathbb{R}^p$ and $r_i \in \mathbb{R}^m$ for $i = 1, \ldots, n$. In the SISO case, i.e., $n_u = n_y = 1$, these tangential directions are assigned the value one, i.e., $\ell_i = r_i = 1$. Unlike projection-based model reduction, which requires access to the state space quantities, data-driven interpolatory model reduction only assumes access to the action of the transfer function evaluated at the driving frequencies along the tangential directions, i.e.,

$$\ell_i^T H(\mu_i) = f_i^T \qquad \text{and} \qquad H(\sigma_i)r_i = g_i \qquad \text{for } i = 1, \ldots, n. \tag{6.4}$$

If the direction vectors ℓ_i and ℓ_j are linearly independent, one can allow μ_i to coincide with μ_j, and similarly for the σ_i's. However, for simplicity the only coincidence of interpolation points that we admit will be between left and right interpolation points, i.e., $\mu_i = \sigma_j$. If this is the case for an index pair (i, j), then bitangential derivative data is assumed to be available. Since we assume that each of the two sets $\{\mu_i\}_{i=1}^n$ and $\{\sigma_i\}_{i=1}^n$ consists of n distinct points, if $\mu_i = \sigma_j$ for an index pair (i, j), without loss of generality, we assume $i = j$. Then, the corresponding bitangential derivative data is defined as

$$\ell_i^T H'(\mu_i)r_i = \theta_i,$$

where H' denotes the derivative of H, i.e., $H' := \frac{\mathrm{d}}{\mathrm{d}s}H$. Following [7, 150], we summarize the *interpolation data* as

$$\begin{aligned}
\text{left interpolation data:} \quad & \{(\mu_i, \ell_i, f_i) \mid \mu_i \in \mathbb{C}, \ell_i \in \mathbb{C}^{n_y}, f_i \in \mathbb{C}^{n_u}, i = 1,\dots,n\},\\
\text{right interpolation data:} \quad & \{(\sigma_i, r_i, g_i) \mid \sigma_i \in \mathbb{C}, r_i \in \mathbb{C}^{n_y}, g_i \in \mathbb{C}^{n_u}, i = 1,\dots,n\},\\
\text{bitangential derivative data:} \quad & \{(i, \theta_i) \mid i \in \{1,\dots,n\} \text{ for which } \mu_i = \sigma_i, \theta_i \in \mathbb{C}\},
\end{aligned} \tag{6.5}$$

with the understanding that the last category may be empty if $\{\mu_i\}_{i=1}^n \cap \{\sigma_i\}_{i=1}^n = \varnothing$. Note that in the case $\mu_i = \sigma_i$, the compatibility of the conditions (6.4) requires that $f_i^T r_i = \ell_i^T g_i$. For the ease of presentation in the next sections, we summarize the interpolation data in the matrices

$$\begin{aligned}
\mathscr{M} &:= \operatorname{diag}(\mu_1,\dots,\mu_n) \in \mathbb{C}^{n\times n}, & \mathscr{S} &:= \operatorname{diag}(\sigma_1,\dots,\sigma_n) \in \mathbb{C}^{n\times n},\\
\mathscr{L} &:= \begin{bmatrix} \ell_1 & \dots & \ell_n \end{bmatrix} \in \mathbb{C}^{n_y\times n}, & \mathscr{R} &:= \begin{bmatrix} r_1 & \dots & r_n \end{bmatrix} \in \mathbb{C}^{n_u\times n},\\
\mathscr{F} &:= \begin{bmatrix} f_1 & \dots & f_n \end{bmatrix} \in \mathbb{C}^{n_u\times n}, & \mathscr{G} &:= \begin{bmatrix} g_1 & \dots & g_n \end{bmatrix} \in \mathbb{C}^{n_y\times n}.
\end{aligned} \tag{6.6}$$

The problem that we are interested in solving can thus be formulated as follows.

Problem 6.6 (Realization problem for DDAEs)**.** *Given the interpolation data in* (6.5)*, find matrices* $\widetilde{E}, \widetilde{A}_1, \widetilde{A}_2 \in \mathbb{C}^{n_x\times n_x}$*,* $\widetilde{B} \in \mathbb{C}^{n_x\times n_u}$*, and* $\widetilde{C} \in \mathbb{C}^{n_y\times n_x}$ *and a parameter* $\tau \geq 0$*, such that the transfer function*

$$\widetilde{H}(s) = \widetilde{C}\left(s\widetilde{E} - \widetilde{A}_1 - \exp(-\tau s)\widetilde{A}_2\right)^{-1}\widetilde{B} \tag{6.7}$$

satisfies the interpolation conditions

$$\ell_i^T \widetilde{H}(\mu_i) = f_i^T \qquad \text{and} \qquad \widetilde{H}(\sigma_i) r_i = g_i \qquad \text{for } i = 1,\dots,n.$$

If $\mu_i = \sigma_i$ *for any index* i*, then additionally,*

$$\ell_i^T \widetilde{H}'(\mu_i) r_i = \theta_i$$

is to be satisfied.

Remark 6.7. One may furthermore ask for the transfer function $\widetilde{H}$ in (6.7) to be *real*, i.e., to satisfy

$$\overline{\widetilde{H}(s)} = \widetilde{H}(\overline{s}) \qquad \text{for all } s \in \mathbb{C}, \tag{6.8}$$

where $\overline{z}$ denotes the complex conjugate of $z \in \mathbb{C}$. A sufficient condition to ensure a real transfer function is to ask for the system matrices $\widetilde{E}, \widetilde{A}_1, \widetilde{A}_2, \widetilde{B}, \widetilde{C}$ to be real, which we henceforth refer to as *real realization problem.* ♣

For $\tau = 0$, Problem 6.6 reduces to the task of identifying a linear system in generalized state-space form, a task which is successfully solved via the Loewner framework [150], see the forthcoming section 6.2. In that sense, the particular structure of the transfer function in (6.7) can be seen as a generalization of the Loewner framework.

Depending on the application at hand, it may not be desirable to construct a realization that depends on the past. Instead, one may prescribe a different system structure. For instance, a general *resistor-inductor-capacitor* (RLC) network may be modeled as *differential-algebraic equation* (DAE) with integral term [78], given by

$$A_1\dot{x}(t) + A_2 x(t) + A_3 \int_0^t x(\theta)\mathrm{d}\theta = Bu(t), \qquad y(t) = B^T x(t). \tag{6.9}$$

The transfer function associated with (6.9) is given by

$$H(s) = B^T \left(sA_1 + A_2 + \frac{1}{s} A_3 \right)^{-1} B$$

and we expect better approximation properties of the realization by preserving this form. Further examples for system structures are listed in Table 6.1. Instead of formulating and solving a realization problem for each of these system classes, we are interested in a general scheme that is able to construct a realization for a given system structure.

Remark 6.8. In terms of MOR, which aims at producing a computationally inexpensive surrogate model of a given dynamical system, the preservation of structure in the *reduced order model* (ROM) often allows one to derive a ROM with a smaller state-space dimension n_x, while maintaining comparable or at times even better accuracy than what unstructured reduced models produce, see Section 5 in [18]. Additionally, since the internal structure of models often reflects core phenomenological properties, structured models may behave in ways that remain qualitatively consistent with the phenomena that are being modeled – possibly more so than unstructured models having higher objective fidelity. In contrast to the structured realization problem (cf. Problem 6.6), most structure-preserving MOR techniques are developed in a projection-based context, thus assuming access to internal dynamics in the form of differential equations. For details we refer to [18,57,58,78,133,155,199]. Notable exceptions are provided in [71,184,189,191]. ♣

Although the term *system structure* can have wide-ranging meanings, for our purposes we will understand the term to refer to equivalence classes of systems having realizations associated with a linearly independent function family $\{h_1, h_2, \ldots, h_K\}$ that appear as

$$H(s) = C \left(\sum_{k=1}^{K} h_k(s) A_k \right)^{-1} B, \tag{6.10}$$

where $C \in \mathbb{R}^{n_y \times n_x}$, $A_k \in \mathbb{R}^{n_x \times n_x}$ for $k = 1, \ldots, K$, $B \in \mathbb{R}^{n_x \times n_u}$. We assume in all that follows that the functions $h_k \colon \mathbb{C} \to \mathbb{C}$ are meromorphic. For any given function family, we will refer to associated matrix-valued functions having the form $\sum_{k=1}^{K} h_k(s) A_k$ as an *affine structure*. By standard abuse of notation, we use $H(s)$ to denote either the system itself or the transfer function of the system evaluated at the point $s \in \mathbb{C}$. The two systems $H(s)$ and $\widetilde{H}(s)$ are called *structurally equivalent* if $H(s), \widetilde{H}(s) \in \mathbb{C}^{n_y \times n_u}$ for $s \in \mathbb{C}$ and if they each have the form

$$H(s) = C \left(\sum_{k=1}^{K} h_k(s) A_k \right)^{-1} B \qquad \text{and} \qquad \widetilde{H}(s) = \widetilde{C} \left(\sum_{k=1}^{K} \tilde{h}_k(s) \widetilde{A}_k \right)^{-1} \widetilde{B},$$

with $\operatorname{span}\{h_1, h_2, \ldots, h_K\} \equiv \operatorname{span}\{\tilde{h}_1, \tilde{h}_2, \ldots, \tilde{h}_K\}$. In particular, we allow different state-space dimensions, i.e., for $\widetilde{C} \in \mathbb{R}^{n_y \times n}$, $\widetilde{A}_k \in \mathbb{R}^{n \times n}$, and $\widetilde{B} \in \mathbb{R}^{n \times n_u}$ the integers n_x and n need not be the same. Given an original (full order) system associated with $H(s)$, we aim to construct a structurally equivalent system $\widetilde{H}(s)$ that interpolates the original system at the driving frequencies, yielding to the following generalization of Problem 6.6.

Problem 6.9 (Structured realization problem)**.** *Given the data in* (6.5) *and a system structure associated with the linearly independent function family* $\{h_1,\dots,h_K\}$, *find matrices* $\widetilde{A}_k \in \mathbb{C}^{n_x \times n_x}$, $k=1,\dots,K$, $\widetilde{B} \in \mathbb{C}^{n_x \times n_u}$, *and* $\widetilde{C} \in \mathbb{C}^{n_y \times n_x}$, *such that the transfer function*

$$\widetilde{H}(s) = \widetilde{C}\left(\sum_{k=1}^{K} h_k(s)\widetilde{A}_k\right)^{-1}\widetilde{B} \tag{6.11}$$

satisfies the interpolation conditions

$$\ell_i^T \widetilde{H}(\mu_i) = f_i^T \qquad \textit{and} \qquad \widetilde{H}(\sigma_i) r_i = g_i \qquad \textit{for } i=1,\dots,n. \tag{6.12a}$$

If $\mu_i = \sigma_i$ *for any index* i, *then additionally,*

$$\ell_i^T \widetilde{H}'(\mu_i) r_i = \theta_i \tag{6.12b}$$

is to be satisfied.

Remark 6.10. Comparing Problem 6.6 and Problem 6.9, we observe that the coefficient functions h_k may depend on possibly unknown parameters like the time delay τ, which also need to be identified. This may be done by fitting a realization obtained as solution of Problem 6.9 via least-squares optimization to additional data (see the forthcoming section 8.3). ♣

6.2 The Loewner framework

A special case of Problem 6.6 and Problem 6.9 is the generalized state-space realization problem, which can be obtained by setting $\tau = 0$ in Problem 6.6, i.e., by considering the dynamical system

$$E\dot{x}(t) = A_1 x(t) + Bu(t), \qquad y(t) = Cx(t) \tag{6.13}$$

with associated transfer function $H(s) = C(sE - A_1)^{-1}B$. This problem is successfully solved by the Loewner realization framework introduced in [150], which uses a *Loewner matrix* $\mathbb{L} \in \mathbb{C}^{n \times n}$ and a *shifted Loewner matrix* $\mathbb{L}_\sigma \in \mathbb{C}^{n \times n}$, whose entries $[\mathbb{L}]_{i,j}$ and $\left[\mathbb{L}_\sigma\right]_{i,j}$ for $i,j = 1,\dots,n$ are defined as

$$[\mathbb{L}]_{i,j} := \frac{f_i^T r_j - \ell_i^T g_j}{\mu_i - \sigma_j}, \qquad \text{and} \qquad \left[\mathbb{L}_\sigma\right]_{i,j} := \frac{\mu_i f_i^T r_j - \sigma_j \ell_i^T g_j}{\mu_i - \sigma_j}, \qquad \text{if} \quad \mu_i \neq \sigma_j, \tag{6.14a}$$

$$[\mathbb{L}]_{i,i} := \theta_i, \qquad \text{and} \qquad \left[\mathbb{L}_\sigma\right]_{i,i} := f_i^T r_i + \mu_i \theta_i, \qquad \text{if} \quad \mu_i = \sigma_i. \tag{6.14b}$$

For SISO systems the definition in (6.14) reduces to

$$\begin{aligned}
[\mathbb{L}]_{i,j} &= \frac{H(\mu_i)-H(\sigma_j)}{\mu_i-\sigma_j}, &&\text{and} & \left[\mathbb{L}_\sigma\right]_{i,j} &= \frac{\mu_i H(\mu_i)-\sigma_j H(\sigma_j)}{\mu_i-\sigma_j}, &&\text{if}\quad \mu_i \neq \sigma_j,\\
[\mathbb{L}]_{i,i} &= H'(\mu_i), &&\text{and} & \left[\mathbb{L}_\sigma\right]_{i,i} &= H(\mu_i)+\mu_i H'(\mu_i), &&\text{if}\quad \mu_i = \sigma_i,
\end{aligned}$$

i.e., $\mathbb{L}$ and $\mathbb{L}_\sigma$ are the divided differences matrices corresponding to the transfer functions $H(s)$ and $sH(s)$, respectively.

Remark 6.11. The Loewner matrices satisfy the Sylvester equations

$$\mathscr{M}\mathbb{L}-\mathbb{L}\mathscr{S} = \mathscr{L}^T\mathscr{G}-\mathscr{F}^T\mathscr{R} \qquad \text{and} \qquad \mathscr{M}\mathbb{L}-\mathbb{L}\mathscr{S} = \mathscr{M}\mathscr{L}^T\mathscr{G}-\mathscr{F}^T\mathscr{R}\mathscr{S},$$

with data matrices as defined in (6.6). For further details we refer to [150]. ♣

Theorem 6.12 (Loewner realization [150])**.** *Let the matrices* $\mathbb{L}$ *and* $\mathbb{L}_\sigma$ *be defined as in* (6.14) *and assume that* $\det(\tilde{s}\mathbb{L}-\mathbb{L}_\sigma)\neq 0$ *for all* $\tilde{s}\in\{\mu_i\}_{i=1}^n\cup\{\sigma_i\}_{i=1}^n$. *Then the system*

$$-\mathbb{L}\dot{\tilde{x}}(t) = -\mathbb{L}_\sigma\tilde{x}(t)+\mathscr{F}^T u(t), \qquad \tilde{y}(t)=\mathscr{G}\tilde{x}(t) \tag{6.15}$$

with $\mathscr{F},\mathscr{G}$ *as defined in* (6.6) *is a minimal realization of an interpolant of the data, i.e., its transfer function*

$$\widetilde{H}(s)=\mathscr{G}(\mathbb{L}_\sigma - s\mathbb{L})^{-1}\mathscr{F}^T$$

satisfies the interpolation conditions (6.12).

Thus in view of Problem 6.9, the Loewner realization corresponds to the specific setting

$$K=2, \qquad h_1(s)\equiv 1, \qquad \text{and} \qquad h_2(s)=-s.$$

The condition $\det(\tilde{s}\mathbb{L}-\mathbb{L}_\sigma)\neq 0$ in Theorem 6.12 can be relaxed by means of the truncated *singular value decomposition* (SVD) [6, Remark 3.2.1].

Theorem 6.13 ([150, Theorem 5.1])**.** *Suppose that*

$$\operatorname{rank}\left(\tilde{s}\mathbb{L}-\mathbb{L}_\sigma\right)=\operatorname{rank}\begin{bmatrix}\mathbb{L} & \mathbb{L}_\sigma\end{bmatrix}=\operatorname{rank}\begin{bmatrix}\mathbb{L}\\ \mathbb{L}_\sigma\end{bmatrix}=: r \qquad \textit{for all } \tilde{s}\in\{\mu_i\}\cup\{\sigma_i\}. \tag{6.16}$$

Then a minimal realization of an interpolant of the data is given by the system

$$\begin{aligned}
-Y^*\mathbb{L}X\dot{\tilde{x}}(t) &= -Y^*\mathbb{L}_\sigma X\tilde{x}(t)+Y^*\mathscr{F}^T u(t),\\
\tilde{y}(t) &= \mathscr{G}X\tilde{x}(t),
\end{aligned}\tag{6.17}$$

where $Y\in\mathbb{C}^{n\times r}$ *and* $X\in\mathbb{C}^{n\times r}$ *are computed from the short SVD* $\tilde{s}\mathbb{L}-\mathbb{L}_\sigma = Y\Sigma X^*$ *for some* $\tilde{s}\in\{\mu_i\}\cup\{\sigma_i\}$, *where* $Y\in\mathbb{C}^{n\times r}$ *and* $X\in\mathbb{C}^{n\times r}$ *have orthonormal columns and* $\Sigma\in\mathbb{R}^{r\times r}$ *is diagonal with positive elements.*

Remark 6.14. It should be noted that the pencil $s\mathbb{L} - \mathbb{L}_\sigma$ can be singular (i.e., $\det(s\mathbb{L} - \mathbb{L}_\sigma) \equiv 0$) while at the same time the matrices $\begin{bmatrix} \mathbb{L} & \mathbb{L}_\sigma \end{bmatrix}$ and $\begin{bmatrix} \mathbb{L}^* & \mathbb{L}_\sigma^* \end{bmatrix}$ have full row rank. In this case, the condition (6.16) is not satisfied and thus Theorem 6.13 does not apply. Indeed, the matrices

$$\mathbb{L} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \mathbb{L}_\sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

have no common (left or right) nullspace. But even if the rank condition (6.16) is satisfies and thus Theorem 6.13 can be applied, the pencil in (6.17) may have a high index (cf. Definition 2.12), may be close to a pencil with high index or may even be close to a singular pencil. In this case, a further regularization is required to prevent numerical issues. We refer to [34] for further details. ♣

The Loewner realization framework is an effective and broadly applicable approach for constructing rational approximants directly from interpolation data; it has been extended to parametric systems [10, 115], to realization independent methods for optimal $\mathcal{H}_2$ approximation [19], to bilinear systems [9], and to switched systems [87]. However, the Loewner framework is only capable of producing *rational* approximants and, so in particular, it cannot capture the transcendental character of transfer functions for dynamical systems containing time delays or distributed parameter subsystems that model convection or transport (cf. [63]).

7

Structured interpolatory realizations

In this chapter we provide a solution for Problem 6.6 and the more general Problem 6.9.

7.1 Interpolation conditions

Suppose we are given interpolation data as in (6.5) and for the moment assume that we already have a realization of the form $\widetilde{H}(s) = \widetilde{C}\widetilde{\mathscr{K}}(s)^{-1}\widetilde{B}$. If we can impose conditions on $\widetilde{C}, \widetilde{B}$ and the matrix function $\widetilde{\mathscr{K}}$ such that $\widetilde{H}(s) = \widetilde{C}\widetilde{\mathscr{K}}(s)^{-1}\widetilde{B}$ satisfies the interpolation conditions (6.12), then we can revert the process and use the conditions to construct the realization. The following observation, which corresponds to an equivalent parametrization of the interpolation conditions (6.12), suggests how one might proceed.

Theorem 7.1. *Let $\widetilde{\mathscr{K}} : \mathbb{C} \to \mathbb{C}^{n\times n}$ be a continuously differentiable matrix-valued function, which is nonsingular at $s = \mu_i$ and $s = \sigma_j$ for $i, j = 1, \ldots, n$. The realization $\widetilde{H}(s) = \widetilde{C}\widetilde{\mathscr{K}}(s)^{-1}\widetilde{B}$ satisfies the interpolation conditions* (6.12a) *if and only if*

$$\mathscr{G} = \widetilde{C}P_{\mathscr{G}} \qquad \text{and} \qquad \mathscr{F}^T = P_{\mathscr{F}}^T\widetilde{B}, \tag{7.1}$$

where $P_{\mathscr{G}}, P_{\mathscr{F}} \in \mathbb{C}^{n\times n}$ are two matrices, whose columns $p_{\mathscr{G}}^i := P_{\mathscr{G}}e_i$ and $p_{\mathscr{F}}^i := P_{\mathscr{F}}e_i$, respectively, solve the linear systems

$$\widetilde{\mathscr{K}}(\sigma_i)p_{\mathscr{G}}^i = \widetilde{B}r_i \qquad \text{and} \qquad \widetilde{\mathscr{K}}(\mu_i)^T p_{\mathscr{F}}^i = \widetilde{C}^T\ell_i, \tag{7.2}$$

where e_i is the ith column of the $n \times n$ identity matrix. Moreover, if $\mu_i = \sigma_i$, then $\widetilde{H}$ satisfies in addition the bitangential interpolation condition (6.12b)*, provided that*

$$\left(p_{\mathscr{F}}^i\right)^T \widetilde{\mathscr{K}}'(\mu_i)p_{\mathscr{G}}^i = -\theta_i, \tag{7.3}$$

where $\widetilde{\mathscr{K}}'$ denotes the derivative of $\widetilde{\mathscr{K}}$.

Proof. The transfer function $\widetilde{H}(s) = \widetilde{C}\widetilde{\mathscr{K}}(s)^{-1}\widetilde{B}$ is well-defined at $s = \mu_i$ and $s = \sigma_i$. Assume first that (7.1) and (7.2) are satisfied. Multiplying the first equation in (7.1) by e_i yields $g_i = \widetilde{C}p^i_{\mathscr{G}}$. Then, using the first equation in (7.2) and the fact that $\widetilde{\mathscr{K}}(\sigma_i)$ is nonsingular, one immediately obtains $g_i = \widetilde{H}(\sigma_i)r_i$, i.e., the right tangential interpolation holds. Similarly, using the second expression in (7.1) and the definition of $p^i_{\mathscr{F}}$ in (7.2), we arrive at $f_i^T = \ell_i^T\widetilde{H}(\mu_i)$; thus (6.12a) holds. For the other direction, we observe that if $p^i_{\mathscr{F}}$ and $p^i_{\mathscr{G}}$ are the unique solutions of (7.2), then the interpolation conditions immediately imply (7.1). Moreover, if $\mu_i = \sigma_i$, then (7.3) yields

$$\ell_i^T\widetilde{H}'(\mu_i)r_i = -\ell_i^T\widetilde{C}\widetilde{\mathscr{K}}(\mu_i)^{-1}\widetilde{\mathscr{K}}'(\mu_i)\widetilde{\mathscr{K}}(\sigma_i)^{-1}\widetilde{B}r_i = -\left(p^i_{\mathscr{F}}\right)^T\widetilde{\mathscr{K}}'(\mu_i)p^i_{\mathscr{G}} = \theta_i. \qquad \blacksquare$$

Evidently, in order to satisfy the collected tangent interpolation conditions (6.12a), we can now equivalently require the realization $\widetilde{H}(s)$ to satisfy the conditions of Theorem 7.1. In particular we need $\widetilde{\mathscr{K}}(s)$ to be nonsingular at the driving frequencies $s = \mu_i$ and $s = \sigma_j$. For $\widetilde{\mathscr{K}}(s) = \sum_{k=1}^K h_k(s)\widetilde{A}_k$, the other conditions (7.1) and (7.2) can be rewritten as

$$\mathscr{G} = \widetilde{C}P_{\mathscr{G}}, \qquad \mathscr{F}^T = P_{\mathscr{F}}^T\widetilde{B}, \tag{7.4}$$

$$\sum_{k=1}^K \widetilde{A}_k P_{\mathscr{G}} h_k(\mathscr{S}) = \widetilde{B}\mathscr{R}, \qquad \sum_{k=1}^K h_k(\mathscr{M})P_{\mathscr{F}}^T\widetilde{A}_k = \mathscr{L}^T\widetilde{C}, \tag{7.5}$$

where we set $h_k(\mathscr{M}) := \operatorname{diag}(h_k(\mu_1),\ldots,h_k(\mu_n))$ and $h_k(\mathscr{S}) := \operatorname{diag}(h_k(\sigma_1),\ldots,h_k(\sigma_n))$. To fulfill additionally the bitangential interpolation conditions (6.12b) for the case $\mu_i = \sigma_i$, the third condition of Theorem 7.1 needs to be satisfied.

If the matrices $P_{\mathscr{F}}$ and $P_{\mathscr{G}}$ are nonsingular, then

$$\widetilde{H}(s) = \widetilde{C}\widetilde{\mathscr{K}}(s)^{-1}\widetilde{B} = \mathscr{G}\left(P_{\mathscr{F}}^T\widetilde{\mathscr{K}}(s)P_{\mathscr{G}}\right)^{-1}\mathscr{F}^T,$$

and hence the realization is unique up to the basis transformation described by $P_{\mathscr{F}}$ and $P_{\mathscr{G}}$. In this case, the matrices $\widetilde{B}$ and $\widetilde{C}$ are given directly by the data without further computations and the matrices $P_{\mathscr{F}}$ and $P_{\mathscr{G}}$ capture the non-uniqueness of the realization. In Section 7.3 we will use these matrices to tailor the realization to interpolate additional data. In any case, we view equations (7.4) and (7.5) not as a coupled system but as a staggered process. First, fix matrices $P_{\mathscr{F}}, P_{\mathscr{G}}$ and determine $\widetilde{B}$ and $\widetilde{C}$ from (7.4). In a second step, use this information to solve (7.5). With this viewpoint, i.e., not counting $P_{\mathscr{F}}$ and $P_{\mathscr{G}}$ as unknowns, we have Kn^2 unknowns from the coefficient matrices $\widetilde{A}_k$ and $(n_u + n_y)n$ unknowns from the input and output matrices $\widetilde{B}$ and $\widetilde{C}$, giving a total of $Kn^2 + (n_u + n_y)n$ unknowns. For these unknowns, (7.4) and (7.5) constitute $2n^2 + (n_u + n_y)n$ equations, leaving $(K-2)n^2$ degrees of freedom. In particular, we can expect a unique solution for $K = 2$.

Remark 7.2. There are $(K-2)n^2$ degrees of freedom to solve the structured realization problem, and therefore the $K = 1$ case does not have enough degrees of freedom to guarantee a solution in general. To further examine the case $K = 1$, assume for simplicity that $\widetilde{H}$ is a *single-input/single-output* (SISO) system, i.e., $\widetilde{B} = \widetilde{b} \in \mathbb{R}^n$ and $\widetilde{C}^T = \widetilde{c} \in \mathbb{R}^n$. Then the reduced model has the form $H(s) = \frac{1}{h_1(s)}\widetilde{c}^T\widetilde{A}^{-1}\widetilde{b}$. Therefore, the interpolation conditions yield

$$\widetilde{c}^T\widetilde{A}^{-1}\widetilde{b} = H(\sigma_i)h_1(\sigma_i) \qquad \text{and} \qquad \widetilde{c}^T\widetilde{A}^{-1}\widetilde{b} = H(\mu_i)h_1(\mu_i), \qquad \text{for} \qquad i = 1,\ldots,n. \tag{7.6}$$

Since $\widetilde{c}^T \widetilde{A}^{-1} \widetilde{b}$ is constant, for the interpolation problem in (7.6) to have a solution, we need

$$H(\sigma_i) h_1(\sigma_i) = H(\mu_i) h_1(\mu_i) = \mathsf{c},$$

where c is a constant for $i = 1, \ldots, n$. This clearly will not be the case in general and we cannot expect to have a solution. Interestingly, if this condition holds, a solution can be found easily by setting $\widetilde{A} = 1$, $\widetilde{b} = 1$ and $\widetilde{c} = \mathsf{c}$. Based on these considerations, we will focus on $K \geq 2$ in the remainder of the thesis. ♣

Remark 7.3. The nonsingularity of the matrices $P_{\mathscr{F}}$ and $P_{\mathscr{G}}$ is connected to the controllability and observability of the realization. A SISO system in standard state-space form, i.e., $\widetilde{\mathscr{K}}(s) = sI_n - \widetilde{A}$, $\widetilde{B} = \widetilde{b} \in \mathbb{R}^n$, and $\widetilde{C} = \widetilde{c} \in \mathbb{R}^{1 \times n}$ is called controllable, if $n = \operatorname{rank}\left(\begin{bmatrix} \widetilde{b} & \widetilde{A}\widetilde{b} & \cdots & \widetilde{A}^{n-1}\widetilde{b} \end{bmatrix}\right)$. It is called observable, if $n = \operatorname{rank}\left(\begin{bmatrix} \widetilde{c}^T & \widetilde{A}^T \widetilde{c}^T & \cdots & (\widetilde{A}^T)^{n-1} \widetilde{c}^T \end{bmatrix}\right)$. To establish the connection between the matrices $P_{\mathscr{F}}$ and $P_{\mathscr{G}}$ to these concepts, we observe that for SISO systems in standard state-space form $\widetilde{\mathscr{K}}$ and its pointwise inverse form a set of commutative matrices. Hence we have

$$\begin{aligned} \operatorname{rank}(P_{\mathscr{G}}) &= \operatorname{rank}\left(\begin{bmatrix} \widetilde{\mathscr{K}}(\sigma_1)^{-1}\widetilde{b} & \cdots & \widetilde{\mathscr{K}}(\sigma_n)^{-1}\widetilde{b} \end{bmatrix}\right) \\ &= \operatorname{rank}\left(\begin{bmatrix} \widetilde{\mathscr{K}}(\sigma_1)^{-1}\widetilde{b} & \widetilde{\mathscr{K}}(\sigma_1)^{-1}\widetilde{\mathscr{K}}(\sigma_2)^{-1}\widetilde{b} & \cdots & \left(\prod_{i=1}^{n} \widetilde{\mathscr{K}}(\sigma_i)^{-1}\right)\widetilde{b} \end{bmatrix}\right) \\ &= \operatorname{rank}\left(\begin{bmatrix} \widetilde{b} & \widetilde{A}\widetilde{b} & \cdots & \widetilde{A}^{n-1}\widetilde{b} \end{bmatrix}\right) \end{aligned}$$

such that $P_{\mathscr{G}}$ is nonsingular if and only if the realization is controllable. Similarly, $P_{\mathscr{F}}$ is nonsingular if and only if the realization is observable. ♣

Note that the Loewner pencil with $h_1(s) \equiv 1$ and $h_2(s) = -s$ (cf. section 6.2) satisfies the conditions of Theorem 7.1 with $\widetilde{\mathscr{K}}(s) = \mathbb{L}_\sigma - s\mathbb{L}$, i.e., the Loewner framework is a special case of Theorem 7.1 with matrices $P_{\mathscr{F}}$ and $P_{\mathscr{G}}$ set to the identity. Indeed, for $\mu_i \neq \sigma_j$, the (i, j) component of the Loewner pencil is

$$e_i^T \widetilde{\mathscr{K}}(s) e_j = [\mathbb{L}_\sigma]_{i,j} - s[\mathbb{L}]_{i,j} = \left(\frac{\mu_i - s}{\mu_i - \sigma_j}\right) f_i^T r_j + \left(\frac{s - \sigma_j}{\mu_i - \sigma_j}\right) \ell_i^T g_j,$$

so it immediately follows $e_i^T \widetilde{\mathscr{K}}(\mu_i) = \ell_i^T \mathscr{G} = \ell_i^T \widetilde{C}$ and $\widetilde{\mathscr{K}}(\sigma_j) e_j = \mathscr{F}^T r_j = \widetilde{B} r_j$. Similarly, for the case $\mu_i = \sigma_i$, we obtain

$$e_i^T \widetilde{\mathscr{K}}(\mu_i) = \ell_i^T \widetilde{C}, \qquad \widetilde{\mathscr{K}}(\sigma_i) e_i = \widetilde{B} r_i, \qquad \text{and} \qquad e_i^T \widetilde{\mathscr{K}}'(\mu_i) e_i = -\mathbb{L} = -\theta_i.$$

Following the discussion above Remark 7.2, we can expect a unique solution of (7.5) only for the special case $K = 2$, which we discuss in detail in Section 7.2 and show its close relation to the Loewner framework. If $K \geq 3$, we need a strategy to deal with the remaining degrees of freedom. To this end we propose two approaches, which both provide interpolation of further data while maintaining the dimension of the matrices in the realization. The first approach uses additional interpolation points (Section 7.3.1), while the second one interpolates additional derivative evaluations of the transfer functions (Section 7.3.2).

7.2 Structured Loewner realizations for the case $K = 2$

Setting $P_{\mathscr{F}} = P_{\mathscr{G}} = I_n$ gives $2n^2 + (n_u + n_y)n$ equations in (7.4) and (7.5) for the $Kn^2 + (n_u + n_y)n$ unknowns such that we can expect (under some regularity) a unique solution for the case $K = 2$. In this case $\widetilde{B} = \mathscr{F}^T$, $\widetilde{C} = \mathscr{G}$, and the matrix equations in (7.5) reduce to

$$h_1(\mathscr{M})\widetilde{A}_1 + h_2(\mathscr{M})\widetilde{A}_2 = \mathscr{L}^T\mathscr{G} \qquad \text{and} \qquad \widetilde{A}_1 h_1(\mathscr{S}) + \widetilde{A}_2 h_2(\mathscr{S}) = \mathscr{F}^T\mathscr{R}.$$

To decouple these equations, we multiply the first equation from the right by $h_2(\mathscr{S})$ and the second equation from the left by $h_2(\mathscr{M})$. Subtracting the resulting systems yields the Sylvester-like equation

$$h_2(\mathscr{M})\widetilde{A}_1 h_1(\mathscr{S}) - h_1(\mathscr{M})\widetilde{A}_1 h_2(\mathscr{S}) = h_2(\mathscr{M})\mathscr{F}^T\mathscr{R} - \mathscr{L}^T\mathscr{G} h_2(\mathscr{S}). \tag{7.7}$$

Similarly, we can eliminate $\widetilde{A}_1$ and obtain

$$h_1(\mathscr{M})\widetilde{A}_2 h_2(\mathscr{S}) - h_2(\mathscr{M})\widetilde{A}_2 h_1(\mathscr{S}) = h_1(\mathscr{M})\mathscr{F}^T\mathscr{R} - \mathscr{L}^T\mathscr{G} h_1(\mathscr{S}). \tag{7.8}$$

Remark 7.4. If the desired model is a generalized state space system as in (6.13), i.e., $h_1(s) = s$ and $h_2(s) \equiv -1$, then (7.7) and (7.8) are given by the Sylvester equations

$$\widetilde{A}_1\mathscr{S} - \mathscr{M}\widetilde{A}_1 = \mathscr{F}^T\mathscr{R} - \mathscr{L}^T\mathscr{G} \qquad \text{and} \qquad \widetilde{A}_2\mathscr{S} - \mathscr{M}\widetilde{A}_2 = \mathscr{M}\mathscr{F}^T\mathscr{R} - \mathscr{L}^T\mathscr{G}\mathscr{S}, \tag{7.9}$$

respectively. Up to a sign factor, these are the Sylvester equations that define the Loewner matrix and the shifted Loewner matrix, see 6.11. In particular, if $\sigma_i \neq \mu_j$ for $i, j = 1, \ldots, n$, then $\widetilde{A}_1 = -\mathbb{L}$ and $\widetilde{A}_2 = -\mathbb{L}_\sigma$ are the unique solutions of (7.7) and (7.8) and the Loewner framework is a special case of the general framework presented in this paper. Similarly, the proportional ansatz for the realization of delay systems introduced in [189] is covered by our framework. ♣

Those elements of $\widetilde{A}_1$ and $\widetilde{A}_2$, for which $\mu_i \neq \sigma_j$, may be obtained by multiplying (7.7) and (7.8) from left by e_i^T and from right by e_j yielding

$$[\widetilde{A}_1]_{i,j} = \frac{h_2(\mu_i) f_i^T r_j - \ell_i^T g_j h_2(\sigma_j)}{h_2(\mu_i)h_1(\sigma_j) - h_1(\mu_i)h_2(\sigma_j)}, \quad [\widetilde{A}_2]_{i,j} = \frac{h_1(\mu_i) f_i^T r_j - \ell_i^T g_j h_1(\sigma_j)}{h_1(\mu_i)h_2(\sigma_j) - h_2(\mu_i)h_1(\sigma_j)} \tag{7.10}$$

under the generic assumption that $h_1(\mu_i)h_2(\sigma_j) \neq h_2(\mu_i)h_1(\sigma_j)$. This is satisfied for all possible choices of interpolation points with $\mu_i \neq \sigma_j$ if the functions h_1 and h_2 satisfy the *Haar condition* [60], see also the forthcoming Section 7.3.1. The components for which $\mu_i = \sigma_i$ can be obtained by translating the conditions in Theorem 7.1 to the $K = 2$ case. This yields

$$h_1(\mu_i)[\widetilde{A}_1]_{i,i} + h_2(\mu_i)[\widetilde{A}_2]_{i,i} = \ell_i^T g_i \qquad \text{and} \qquad h_1'(\mu_i)[\widetilde{A}_1]_{i,i} + h_2'(\mu_i)[\widetilde{A}_2]_{i,i} = -\theta_i \tag{7.11}$$

and consequently

$$[\widetilde{A}_1]_{i,i} = \frac{h_2(\mu_i)\theta_i + h_2'(\mu_i)\ell_i^T g_i}{h_2'(\mu_i)h_1(\mu_i) - h_1'(\mu_i)h_2(\mu_i)}, \quad [\widetilde{A}_2]_{i,i} = \frac{h_1(\mu_i)\theta_i + h_1'(\mu_i)\ell_i^T g_i}{h_1'(\mu_i)h_2(\mu_i) - h_2'(\mu_i)h_1(\mu_i)}, \tag{7.12}$$

for the components with $\mu_i = \sigma_i$ under the generic assumption $h_2'(\mu_i)h_1(\mu_i) \neq h_1'(\mu_i)h_2(\mu_i)$. Consequently, we have shown the subsequent result.

Theorem 7.5. *Let $\widetilde{A}_1$ and $\widetilde{A}_2$ be as in* (7.10) *and* (7.12) *where the denominators are assumed nonzero. If*

$$\det\left(h_1(\tilde{s})\widetilde{A}_1 + h_2(\tilde{s})\widetilde{A}_2\right) \neq 0 \qquad \text{for all } \tilde{s} \in \{\mu_i\}_{i=1}^n \cup \{\sigma_i\}_{i=1}^n, \tag{7.13}$$

then the transfer function $H(s) = \mathscr{G}\left(h_1(s)\widetilde{A}_1 + h_2(s)\widetilde{A}_2\right)^{-1}\mathscr{F}^T$ satisfies the interpolation conditions (6.12).

Remark 7.6. For the analysis of assumption (7.13) in Theorem 7.5, we observe that the function

$$\eta\colon \mathbb{C} \to \mathbb{C}, \qquad s \mapsto \det\left(h_1(s)\widetilde{A}_1 + h_2(s)\widetilde{A}_2\right)$$

is meromorphic, since by definition h_1 and h_2 are meromorphic. The identity theorem for holomorphic functions implies that either $\eta \equiv 0$ and hence that $h_1(s)\widetilde{A}_1 + h_2(s)\widetilde{A}_2$ is singular for every $s \in \mathbb{C}$, or that set of zeros of η has no accumulation point and consequently, is a set of measure zero, i.e., the transfer function H in Theorem 7.5 is defined for almost all $s \in \mathbb{C}$. ♣

The matrices $\widetilde{A}_1$ and $\widetilde{A}_2$ have a structure similar to the Loewner matrix and the shifted Loewner matrix. This gives rise to the idea that the result of Theorem 7.5 can be obtained from the standard Loewner framework using transformed data.

Corollary 7.7. *Suppose that $h_2(\mathscr{S})$ and $h_2(\mathscr{M})$ are nonsingular and that the denominators in* (7.10) *and* (7.12) *are nonzero. Construct the Loewner matrix $\mathbb{L}$ and the shifted Loewner matrix $\mathbb{L}_\sigma$ for the transformed data*

$$\begin{aligned} &\textit{left interpolation data:} && \left\{\left(\frac{h_1(\mu_i)}{h_2(\mu_i)}, \frac{\ell_i}{h_2(\mu_i)}, f_i\right) \textit{ for } i = 1,\dots,n\right\}, \\ &\textit{right interpolation data:} && \left\{\left(\frac{h_1(\sigma_i)}{h_2(\sigma_i)}, \frac{r_i}{h_2(\sigma_i)}, g_i\right) \textit{ for } i = 1,\dots,n\right\}, \\ &\textit{bitangential derivative data:} && \left\{\left(i, \frac{h_2(\mu_i)\theta_i + h_2'(\mu_i)\ell_i^T g_i}{h_1'(\mu_i)h_2(\mu_i) - h_2'(\mu_i)h_1(\mu_i)}\right) \textit{ for which } \mu_i = \sigma_i\right\} \end{aligned} \tag{7.14}$$

If $\det(h_2(\tilde{s})\mathbb{L}_\sigma - h_1(\tilde{s})\mathbb{L}) \neq 0$ *for all $\tilde{s} \in \{\mu_i\}_{i=1}^n \cup \{\sigma_i\}_{i=1}^n$, then the transfer function*

$$H(s) = \mathscr{G}(h_2(s)\mathbb{L}_\sigma - h_1(s)\mathbb{L})^{-1}\mathscr{F}^T$$

interpolates the data.

Proof. Simple calculations yield that, when constructing the Loewner pencil with the transformed interpolation data (7.14), the Loewner matrix and the shifted Loewner matrix coincide with $-\widetilde{A}_1$ and $\widetilde{A}_2$ given in (7.10) and (7.12). Theorem 7.5 completes the proof. ■

Corollary 7.7 allows one to transfer many results of the standard Loewner framework to the general framework considered in this subsection. In particular, this allows us to keep the system matrices real if the set of interpolation data is closed under complex conjugation. The details are formulated in Lemma 7.8.

Lemma 7.8. *Let the set of interpolation data be closed under complex conjugation, i.e., there exist unitary matrices $T_{\mathscr{F}}, T_{\mathscr{G}} \in \mathbb{C}^{n\times n}$ with*

$$\begin{aligned} &T_{\mathscr{F}}^{*}\mathscr{M}T_{\mathscr{F}} \in \mathbb{R}^{n\times n}, \quad T_{\mathscr{F}}^{*}\mathscr{L}^{T} \in \mathbb{R}^{n}, \quad T_{\mathscr{F}}^{*}\mathscr{F}^{T} \in \mathbb{R}^{n},\\ &T_{\mathscr{G}}^{*}\mathscr{S}T_{\mathscr{G}} \in \mathbb{R}^{n\times n}, \quad \mathscr{R}T_{\mathscr{G}} \in \mathbb{R}^{n}, \qquad \mathscr{G}T_{\mathscr{G}} \in \mathbb{R}^{n}. \end{aligned}$$

Moreover, assume that the set of θ_i's (for the case $\mu_i = \sigma_i$) is closed under complex conjugation. Then, the realization $(T_{\mathscr{F}}^{*}\widetilde{A}_1 T_{\mathscr{G}}, T_{\mathscr{F}}^{*}\widetilde{A}_2 T_{\mathscr{G}}, T_{\mathscr{F}}^{*}\mathscr{F}^{T}, \mathscr{G}T_{\mathscr{G}})$ with $(\widetilde{A}_1, \widetilde{A}_2, \mathscr{F}^{T}, \mathscr{G})$ from Theorem 7.5 consists of real-valued matrices and interpolates the data.

Proof. First we note that if the set of interpolation data is closed under complex conjugation, so is the set of transformed data in Corollary 7.7. Based on this observation, the proof for the case $\mu_i \neq \sigma_j$ for all $i, j = 1,\dots,n$ simply follows the lines of [11, section 2.4.4.]. This can also be comprehended after multiplying the Sylvester-like equations (7.7) and (7.8) from left by $T_{\mathscr{F}}^{*}$ and from right by $T_{\mathscr{G}}$. Similar reasoning proves the claim for the $\mu_i = \sigma_i$ case. ∎

Example 7.9. A special case of Lemma 7.8 applies when the interpolation data is sorted such that the real values have the highest indices, i.e.,

$$\begin{aligned} \mathscr{M} &= \operatorname{diag}(\mu_1, \overline{\mu_1}, \dots, \mu_{2\ell-1}, \overline{\mu_{2\ell-1}}, \mu_{2\ell+1}, \dots, \mu_n),\\ \mathscr{L} &= \begin{bmatrix} \ell_1 & \overline{\ell_1} & \dots & \ell_{2\ell-1} & \overline{\ell_{2\ell-1}} & \ell_{2\ell+1} & \dots & \ell_n \end{bmatrix},\\ \mathscr{F} &= \begin{bmatrix} f_1 & \overline{f_1} & \dots & f_{2\ell-1} & \overline{f_{2\ell-1}} & f_{2\ell+1} & \dots & f_n \end{bmatrix},\\ \mathscr{S} &= \operatorname{diag}(\sigma_1, \overline{\sigma_1}, \dots, \sigma_{2r-1}, \overline{\sigma_{2r-1}}, \sigma_{2r+1}, \dots, \sigma_n),\\ \mathscr{R} &= \begin{bmatrix} r_1 & \overline{r_1} & \dots & r_{2r-1} & \overline{r_{2r-1}} & r_{2r+1} & \dots & r_n \end{bmatrix},\\ \mathscr{G} &= \begin{bmatrix} g_1 & \overline{g_1} & \dots & g_{2r-1} & \overline{g_{2r-1}} & g_{2r+1} & \dots & g_n \end{bmatrix}. \end{aligned}$$

In this case possible choices for $T_{\mathscr{F}}$ and $T_{\mathscr{G}}$ are given by block diagonal unitary matrices

$$T_{\bullet} = \operatorname{blkdiag}\left(\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -\imath \\ 1 & \imath \end{bmatrix}, \dots, \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -\imath \\ 1 & \imath \end{bmatrix}, 1, \dots, 1\right),$$

where $\bullet \in \{\mathscr{F}, \mathscr{G}\}$. One can also obtain the real realization directly from Theorem 7.1 by choosing $P_{\mathscr{F}}^{T} = T_{\mathscr{F}}^{*}$ and $P_{\mathscr{G}} = T_{\mathscr{G}}$ (see discussion after Theorem 7.1). ♠

Remark 7.10. The result from Corollary 7.7 can (formally) be obtained by rewriting the transfer function

$$H(s) = \widetilde{C}(h_1(s)\widetilde{A}_1 + h_2(s)\widetilde{A}_2)^{-1}\widetilde{B} = \widetilde{C}\left(\frac{h_1(s)}{h_2(s)}\widetilde{A}_1 + \widetilde{A}_2\right)^{-1}\widetilde{B}\frac{1}{h_2(s)}.$$

This corresponds to a similar strategy as in [71]. ♣

Remark 7.11. Similar as in the Loewner framework, we can parameterize the realization with a feedthrough term D. Assuming $\{\mu_i\}_{i=1}^n \cap \{\sigma_i\}_{i=1}^n = \emptyset$, simple calculations lead to the realization

$$H(s) = \widetilde{C}\left(h_1(s)\widetilde{A}_1 + h_2(s)\widetilde{A}_2\right)^{-1}\widetilde{B} + D$$

with $\widetilde{C} = \mathscr{G} - D\mathscr{R}$, $\widetilde{B} = \mathscr{F}^T - \mathscr{L}^T D$,

$$\widetilde{C} = \mathscr{G} - D\mathscr{R}, \qquad \left[\widetilde{A}_1\right]_{i,j} = \frac{h_2(\mu_i)\left(f_i^T - \ell_i^T D\right) r_j - \ell_i^T \left(g_j - D r_j\right) h_2(\sigma_j)}{h_2(\mu_i) h_1(\sigma_j) - h_1(\mu_i) h_2(\sigma_j)},$$

$$\widetilde{B} = \mathscr{F}^T - \mathscr{L}^T D, \qquad \left[\widetilde{A}_2\right]_{i,j} = \frac{h_1(\mu_i)\left(f_i^T - \ell_i^T D\right) r_j - \ell_i^T \left(g_j - D r_j\right) h_1(\sigma_j)}{h_1(\mu_i) h_2(\sigma_j) - h_2(\mu_i) h_1(\sigma_j)},$$

which interpolates the data (6.5) for all matrices $D \in \mathbb{R}^{n_y \times n_u}$. For the special case $h_1(s) = s$ and $h_2(s) \equiv -1$, we recover the results from [150] given by $\widetilde{C} = \mathscr{G} - D\mathscr{R}$, $\widetilde{B} = \mathscr{F}^T - \mathscr{L}^T D$, $\widetilde{A}_1 = -\mathbb{L}$, and $\widetilde{A}_2 = -\mathbb{L}_\sigma - \mathscr{L}^T D \mathscr{R}$. ♣

7.3 Structured realization for the case $K \geq 3$

When $K \geq 3$, the conditions in Theorem 7.1 do not provide enough conditions for the available degrees of freedom (even if $P_{\mathscr{F}}$ and $P_{\mathscr{G}}$ are fixed). Hence, we have some freedom in choosing the matrices $\widetilde{A}_k$ with $k = 1, \ldots, n$. We can exploit these degrees of freedom, for instance, by fitting the transfer function to additional data. For simplicity we assume $\{\mu_i\}_{i=1}^n \cap \{\sigma_i\}_{i=1}^n = \emptyset$ for the remainder of this section.

7.3.1 Interpolation at additional points

In this subsection we focus on fitting the transfer function to additional data or, equivalently, matching the given data with a smaller state space dimension. To this end, we assume that we have $(Q_{\mathscr{F}} - 1)n$ additional left interpolation points and $(Q_{\mathscr{G}} - 1)n$ additional right interpolation points at hand, which we group in sets of n. More precisely, the left interpolation data is grouped into the matrices

$$\begin{gathered} \mathscr{M}_q := \operatorname{diag}(\mu_{q;1}, \mu_{q;2}, \ldots, \mu_{q;n}) \in \mathbb{C}^{n \times n}, \qquad \mathscr{L}_q := \begin{bmatrix} \ell_{q;1} & \ell_{q;2} & \cdots & \ell_{q;n} \end{bmatrix} \in \mathbb{C}^{n_y \times n}, \\ \mathscr{F}_q := \begin{bmatrix} f_{q;1} & f_{q;2} & \cdots & f_{q;n} \end{bmatrix} \in \mathbb{C}^{n_u \times n}, \end{gathered} \tag{7.15a}$$

where $q = 1, \ldots, Q_{\mathscr{F}}$. Here, we set $\mu_{1;i} := \mu_i$, $f_{1;i} := f_i$, and $\ell_{1;i} := \ell_i$, such that we have $\mathscr{M}_1 = \mathscr{M}$, $\mathscr{L}_1 = \mathscr{L}$, and $\mathscr{F}_1 = \mathscr{F}$. Similarly, we introduce for $q = 1, \ldots, Q_{\mathscr{G}}$ the matrices

$$\begin{gathered} \mathscr{S}_q := \operatorname{diag}(\sigma_{q;1}, \sigma_{q;2}, \ldots, \sigma_{q;n}) \in \mathbb{C}^{n \times n}, \qquad \mathscr{R}_q := \begin{bmatrix} r_{q;1} & r_{q;2} & \cdots & r_{q;n} \end{bmatrix} \in \mathbb{C}^{n_u \times n}, \\ \mathscr{G}_q := \begin{bmatrix} g_{q;1} & g_{q;2} & \cdots & g_{q;n} \end{bmatrix} \in \mathbb{C}^{n_y \times n}. \end{gathered} \tag{7.15b}$$

To use the full capacity of the available degrees of freedom, we assume $K = Q_{\mathscr{F}} + Q_{\mathscr{G}}$, with $Q_{\mathscr{F}}, Q_{\mathscr{G}} \geq 1$. The next result, which is a generalization of the $K = 2$ case, gives us the necessary and sufficient conditions that the matrices in the realization $H(s)$ must satisfy to interpolate all prescribed information.

Theorem 7.12. *Let $H(s) = \widetilde{C}\widetilde{\mathscr{K}}(s)^{-1}\widetilde{B}$ with $\widetilde{\mathscr{K}}(s) = \sum_{k=1}^{K} h_k(s)\widetilde{A}_k$ and suppose that $\widetilde{\mathscr{K}}(s)$ is nonsingular for all $\tilde{s} \in \{\mu_{q;i}\}_{q=1}^{Q_{\mathscr{F}}} \cup \{\sigma_{q;i}\}_{q=1}^{Q_{\mathscr{G}}}$ for all $i = 1,\ldots,n$.*

(i) The left interpolation conditions $\ell_{q;i}^T \widetilde{H}(\mu_{q;i}) = f_{q;i}^T$ are satisfied for all $i = 1,\ldots,n$ and all $q = 1,\ldots,Q_{\mathscr{F}}$ if and only if there exist matrices $P_{\mathscr{F},q}$ with $q = 1,\ldots,Q_{\mathscr{F}}$ that satisfy

$$\mathscr{F}_q^T = P_{\mathscr{F},q}^T \widetilde{B} \qquad \text{and} \qquad \sum_{k=1}^{K} h_k(\mathscr{M}_q) P_{\mathscr{F},q}^T \widetilde{A}_k = \mathscr{L}_q^T \widetilde{C}. \tag{7.16}$$

(ii) The right interpolation conditions $\widetilde{H}(\sigma_{q;i}) r_{q;i} = g_{q;i}$ are satisfied for all $i = 1,\ldots,n$ and all $q = 1,\ldots,Q_{\mathscr{G}}$ if and only if there exist matrices $P_{\mathscr{G},q}$ with $q = 1,\ldots,Q_{\mathscr{G}}$ that satisfy

$$\mathscr{G}_q = \widetilde{C} P_{\mathscr{G},q} \qquad \text{and} \qquad \sum_{k=1}^{K} \widetilde{A}_k P_{\mathscr{G},q} h_k(\mathscr{S}_q) = \widetilde{B}\mathscr{R}_q. \tag{7.17}$$

Proof. The result follows directly from Theorem 7.1. For the sake of completeness we give the proof of the first statement again. The second identity in (7.16) implies $\ell_{q;i}^T \widetilde{C} = e_i^T P_{\mathscr{F},q}^T \sum_{k=1}^{K} h_k(\mu_{q;i})\widetilde{A}_k$. Thus, by the first identity and the definition of $\widetilde{H}$ we conclude

$$\ell_{q;i}^T \widetilde{H}(\mu_{q;i}) = e_i^T P_{\mathscr{F},q}^T \widetilde{B} = f_{q;i}^T$$

for $i = 1,\ldots,n$ and $q = 1,\ldots,Q_{\mathscr{F}}$. ■

Evidently, in order to satisfy the interpolation conditions (6.12a) it will be sufficient to require that (7.16) and (7.17) hold simultaneously. This gives us the following strategy to determine the realization matrices $\widetilde{A}_k, \widetilde{B}$, and $\widetilde{C}$. Suppose we can find matrices $P_{\mathscr{F},q}$ and $P_{\mathscr{G},q}$ that satisfy the first identity in (7.16) and (7.17), respectively, i.e., that allow us to fix $\widetilde{B}$ and $\widetilde{C}$. Then we can compute the matrices $\widetilde{A}_k$ as follows. Vectorization [104] of the second identity in (7.16) yields

$$\sum_{k=1}^{K} \left(I_n \otimes h_k(\mathscr{M}_q) P_{\mathscr{F},q}^T\right) \operatorname{vec}(\widetilde{A}_k) = \left(\widetilde{C}^T \otimes I_n\right) \operatorname{vec}(\mathscr{L}_q^T),$$

where ⊗ denotes the Kronecker product and $\operatorname{vec}(X)$ denotes the vector of stacked columns of the matrix X. Similarly, equation (7.17) implies

$$\sum_{k=1}^{K} \left(h_k(\mathscr{S}_q) P_{\mathscr{G},q}^T \otimes I_n\right) \operatorname{vec}(\widetilde{A}_k) = \left(I_n \otimes \widetilde{B}\right) \operatorname{vec}(\mathscr{R}_q).$$

All equations together yield the linear algebraic system $\mathbb{A}\boldsymbol{\alpha} = \boldsymbol{\beta}$ with $\mathbb{A} \in \mathbb{C}^{Kn^2 \times Kn^2}$, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{C}^{Kn^2}$ given by

$$
\begin{gathered}
\mathbb{A} := \left[\begin{array}{ccc}
I_n \otimes h_1\left(\mathscr{M}_1\right) P_{\mathscr{F},1}^T & \cdots & I_n \otimes h_K\left(\mathscr{M}_1\right) P_{\mathscr{F},1}^T \\
\vdots & & \vdots \\
I_n \otimes h_1\left(\mathscr{M}_{Q_{\mathscr{F}}}\right) P_{\mathscr{F},Q_{\mathscr{F}}}^T & \cdots & I_n \otimes h_K\left(\mathscr{M}_{Q_{\mathscr{F}}}\right) P_{\mathscr{F},Q_{\mathscr{F}}}^T \\
\hline
h_1\left(\mathscr{S}_1\right) P_{\mathscr{G},1}^T \otimes I_n & \cdots & h_K\left(\mathscr{S}_1\right) P_{\mathscr{G},1}^T \otimes I_n \\
\vdots & & \vdots \\
h_1\left(\mathscr{S}_{Q_{\mathscr{G}}}\right) P_{\mathscr{G},Q_{\mathscr{G}}}^T \otimes I_n & \cdots & h_K\left(\mathscr{S}_{Q_{\mathscr{G}}}\right) P_{\mathscr{G},Q_{\mathscr{G}}}^T \otimes I_n
\end{array}\right], \\
\boldsymbol{\alpha} := \begin{bmatrix} \operatorname{vec}(\widetilde{A}_1) \\ \vdots \\ \operatorname{vec}(\widetilde{A}_K) \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{\beta} := \left[\begin{array}{c}
\left(\widetilde{C}^T \otimes I_n\right) \operatorname{vec}\left(\mathscr{L}_1^T\right) \\
\vdots \\
\left(\widetilde{C}^T \otimes I_n\right) \operatorname{vec}\left(\mathscr{L}_{Q_{\mathscr{F}}}^T\right) \\
\hline
\left(I_n \otimes \widetilde{B}\right) \operatorname{vec}\left(\mathscr{R}_1\right) \\
\vdots \\
\left(I_n \otimes \widetilde{B}\right) \operatorname{vec}\left(\mathscr{R}_{Q_{\mathscr{G}}}\right)
\end{array}\right].
\end{gathered}
\tag{7.18}
$$

Note that the solution of the linear equation system $\mathbb{A}\boldsymbol{\alpha} = \boldsymbol{\beta}$ depends on $P_{\mathscr{F},q}$ and $P_{\mathscr{G},q}$ and there is some freedom in choosing these matrices. A simple possibility is given by

$$
P_{\mathscr{F},q}^T := \begin{bmatrix} \mathscr{F}_q^T & \star \end{bmatrix}, \qquad P_{\mathscr{G},q} := \begin{bmatrix} \mathscr{G}_q \\ \star \end{bmatrix}, \qquad \widetilde{B} := \begin{bmatrix} I_{n_u} \\ 0 \end{bmatrix}, \qquad \text{and} \qquad \widetilde{C} := \begin{bmatrix} I_{n_y} & 0 \end{bmatrix}, \tag{7.19}
$$

which satisfies the first identity in (7.16) and (7.17) for any choice of $\star$. However, the trivial choice of setting these blocks to zero makes the system matrix $\mathbb{A}$ singular, see also Remark 7.3. Instead, we propose to fill the $\star$ part of the matrices $P_{\mathscr{F},q}$ and $P_{\mathscr{G},q}$ such that $P_{\mathscr{F},q}$ and $P_{\mathscr{G},q}$ are nonsingular assuming that $\mathscr{F}_q$ and $\mathscr{G}_q$ have full row rank. A more specific choice of $\star$ may even lead to real-valued realizations as stated in the following lemma.

Lemma 7.13. *Let each of the interpolation data sets be closed under complex conjugation, i.e., there exist unitary matrices $T_{\mathscr{F},q}, T_{\mathscr{G},q} \in \mathbb{C}^{n\times n}$ with*

$$
\begin{aligned}
&T_{\mathscr{F},q}^* \mathscr{M}_q T_{\mathscr{F},q} \in \mathbb{R}^{n\times n}, \quad T_{\mathscr{F},q}^* \mathscr{L}_q^T \in \mathbb{R}^n, \quad T_{\mathscr{F},q}^* \mathscr{F}_q^T \in \mathbb{R}^n, \quad \textit{for} \quad q = 1,\ldots,Q_{\mathscr{F}}, \\
&T_{\mathscr{G},q}^* \mathscr{S}_q T_{\mathscr{G},q} \in \mathbb{R}^{n\times n}, \quad \mathscr{R}_q T_{\mathscr{G},q} \in \mathbb{R}^n, \quad \mathscr{G}_q T_{\mathscr{G},q} \in \mathbb{R}^n, \quad \textit{for} \quad q = 1,\ldots,Q_{\mathscr{G}}.
\end{aligned}
$$

Moreover, let the matrices $P_{\mathscr{F},q}$ and $P_{\mathscr{G},q}$ be as in (7.19) *with free entries $\star$ chosen such that $T_{\mathscr{F},q}^* P_{\mathscr{F},q}^T \in \mathbb{R}^n$ and $P_{\mathscr{G},q} T_{\mathscr{G},q} \in \mathbb{R}^n$ hold. Then, the matrices $\widetilde{A}_1,\ldots,\widetilde{A}_K$, $\widetilde{B}$, and $\widetilde{C}$ from Theorem 7.12 are real matrices (if they exist).*

Proof. The matrices $\widetilde{B}$ and $\widetilde{C}$ from (7.20) are real-valued. In addition, the second equalities in (7.16) and (7.17) are equivalent to

$$\begin{aligned}\sum_{k=1}^{K} T_{\mathscr{F},q}^{*} h_k(\mathscr{M}_q) T_{\mathscr{F},q} T_{\mathscr{F},q}^{*} P_{\mathscr{F},q}^{T} \widetilde{A}_k &= T_{\mathscr{F},q}^{*} \mathscr{L}_q^T \widetilde{C} \qquad \text{and} \\ \sum_{k=1}^{K} \widetilde{A}_k P_{\mathscr{G},q} T_{\mathscr{G},q} T_{\mathscr{G},q}^{*} h_k(\mathscr{S}_q) T_{\mathscr{G},q} &= \widetilde{B} \mathscr{R}_q T_{\mathscr{G},q}.\end{aligned}$$

Since the matrices $\widetilde{A}_k$ are the solutions of these linear matrix equations and since their coefficient matrices as well as the right hand sides are real-valued, we conclude that the matrices $\widetilde{A}_k$ are also real-valued. ■

To complete the discussion, we analyze the regularity of $\mathbb{A}$ in the SISO case, that is $p = m = 1$. Here, we set

$$P_{\mathscr{F},q} := \operatorname{diag}(\mathscr{F}_q), \qquad P_{\mathscr{G},q} := \operatorname{diag}(\mathscr{G}_q), \qquad \widetilde{B} := \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \text{and} \quad \widetilde{C} := \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}. \tag{7.20}$$

With these settings, the (i, j) components of the second matrix equations in (7.16) and (7.17) read as $f_{q;i} \sum_{k=1}^{K} h_k(\mu_{q;i})[\widetilde{A}_k]_{i,j} = 1$ and $g_{q;j} \sum_{k=1}^{K} h_k(\sigma_{q;j})[\widetilde{A}_k]_{i,j} = 1$, respectively. Putting this into matrix notation yields the linear system

$$\begin{bmatrix} f_{1;i} &&&&& \\ & \ddots &&&& \\ && f_{Q_{\mathscr{F}};i} &&& \\ &&& g_{1;j} && \\ &&&& \ddots & \\ &&&&& g_{Q_{\mathscr{G}};j} \end{bmatrix} \begin{bmatrix} h_1(\mu_{1;i}) & \dots & h_K(\mu_{1;i}) \\ \vdots && \vdots \\ h_1(\mu_{Q_{\mathscr{F}};i}) & \dots & h_K(\mu_{Q_{\mathscr{F}};i}) \\ h_1(\sigma_{1;j}) & \dots & h_K(\sigma_{1;j}) \\ \vdots && \vdots \\ h_1(\sigma_{Q_{\mathscr{G}};j}) & \dots & h_K(\sigma_{Q_{\mathscr{G}};j}) \end{bmatrix} \begin{bmatrix} [\widetilde{A}_1]_{i,j} \\ [\widetilde{A}_2]_{i,j} \\ \vdots \\ [\widetilde{A}_K]_{i,j} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \tag{7.21}$$

where the system matrix is the product of a diagonal matrix and a generalized Vandermonde matrix. We notice that reordering the entries of $\boldsymbol{\alpha}$ in (7.18) yields an orthogonal similarity transformation that decouples (7.18) in smaller systems of the form (7.21). This generalized Vandermonde matrix is also called a *Haar matrix* [60] and is nonsingular if the driving frequencies $\mu_{q;i}$ and $\sigma_{q;j}$ are distinct and the functions h_k satisfy the *Haar condition* [60]. In particular, the Haar condition is satisfied for monomials, and thus relevant for second-order systems (cf. Table 6.1). The diagonal matrix is nonsingular if the driving frequencies $\mu_{q;i}$ and $\sigma_{q;j}$ are distinct from the roots of the original transfer function. In this case, the system in (7.21) has a unique solution for each (i, j) combination and hence, via transformations, we can infer that $\mathbb{A}$ is nonsingular.

We illustrate the construction of the realization with additional data with the following example.

Example 7.14. Given scalars $a_1, a_2, a_3, b, c \in \mathbb{R}$ with $bc \neq 0$, we consider the system

$$\begin{aligned} a_1 \dot{x}(t) &= a_2 x(t) + a_3 x(t-1) + b u(t), \\ y(t) &= c x(t) \end{aligned}$$

with transfer function $H(s) = \frac{cb}{sa_1 - a_2 - \mathrm{e}^{-s} a_3}$. Setting $Q_{\mathscr{F}} = 1$ and $Q_{\mathscr{G}} = 2$, we pick distinct interpolation points $\mu_{1;1} = \mu$, $\sigma_{1;1} = \sigma$, and $\sigma_{2;1} = \lambda$. We choose $\widetilde{B} = 1$ and $\widetilde{C} = 1$ with $P_{\mathscr{F},1} = H(\mu), P_{\mathscr{G},1} = H(\sigma)$, and $P_{\mathscr{G},2} = H(\lambda)$. Then the system in (7.21) reads as

$$\begin{bmatrix} H(\mu) & & \\ & H(\sigma) & \\ & & H(\lambda) \end{bmatrix} \begin{bmatrix} \mu & -1 & -\mathrm{e}^{-\mu} \\ \sigma & -1 & -\mathrm{e}^{-\sigma} \\ \lambda & -1 & -\mathrm{e}^{-\lambda} \end{bmatrix} \begin{bmatrix} \widetilde{A}_1 \\ \widetilde{A}_2 \\ \widetilde{A}_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{7.22}$$

The inverse of the Haar matrix is given by

$$\frac{1}{\mu \mathrm{e}^{\mu}(\mathrm{e}^{\sigma} - \mathrm{e}^{\lambda}) + \sigma \mathrm{e}^{\sigma}(\mathrm{e}^{\lambda} - \mathrm{e}^{\mu}) + \lambda \mathrm{e}^{\lambda}(\mathrm{e}^{\mu} - \mathrm{e}^{\sigma})} \begin{bmatrix} \mathrm{e}^{\mu}(\mathrm{e}^{\sigma} - \mathrm{e}^{\lambda}) & -\mathrm{e}^{\sigma}(\mathrm{e}^{\mu} - \mathrm{e}^{\lambda}) & \mathrm{e}^{\lambda}(\mathrm{e}^{\mu} - \mathrm{e}^{\sigma}) \\ \mathrm{e}^{\mu}(\sigma \mathrm{e}^{\sigma} - \lambda \mathrm{e}^{\lambda}) & -\mathrm{e}^{\sigma}(\mu \mathrm{e}^{\mu} - \lambda \mathrm{e}^{\lambda}) & \mathrm{e}^{\lambda}(\mu \mathrm{e}^{\mu} - \sigma \mathrm{e}^{\sigma}) \\ -\mathrm{e}^{\mu}\mathrm{e}^{\sigma}\mathrm{e}^{\lambda}(\sigma - \lambda) & \mathrm{e}^{\mu}\mathrm{e}^{\sigma}\mathrm{e}^{\lambda}(\mu - \lambda) & -\mathrm{e}^{\mu}\mathrm{e}^{\sigma}\mathrm{e}^{\lambda}(\mu - \sigma) \end{bmatrix}$$

such that the solution of (7.22) is $\begin{bmatrix} \widetilde{A}_1 & \widetilde{A}_2 & \widetilde{A}_3 \end{bmatrix} = \frac{1}{cb} \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}$. In fact, we recover the original transfer function. ♠

Clearly, the realization is real-valued if all quantities in (7.21) are real. If we pick the driving frequencies on the imaginary axis, then in general the Haar matrix will be complex-valued. The following lemma shows how to obtain real-valued realizations based on complex interpolation data if the matrices $P_{\mathscr{F},q}$ and $P_{\mathscr{G},q}$ are chosen as in (7.20).

Lemma 7.15. *Let the interpolation data be closed under complex conjugation and sorted as in Example 7.9 such that the unitary matrices $T_{\mathscr{F}}, T_{\mathscr{G}} \in \mathbb{C}^{n \times n}$ from Example 7.9 satisfy*

$$\begin{aligned} &T_{\mathscr{F}}^* \mathscr{M}_q T_{\mathscr{F}} \in \mathbb{R}^{n \times n}, \quad T_{\mathscr{F}}^* \mathscr{L}_q^T \in \mathbb{R}^n, \quad T_{\mathscr{F}}^* \mathscr{F}_q^T \in \mathbb{R}^n, \quad \textit{for} \quad q = 1, \ldots, Q_{\mathscr{F}}, \\ &T_{\mathscr{G}}^* \mathscr{S}_q T_{\mathscr{G}} \in \mathbb{R}^{n \times n}, \quad \mathscr{R}_q T_{\mathscr{G}} \in \mathbb{R}^n, \quad \mathscr{G}_q T_{\mathscr{G}} \in \mathbb{R}^n, \quad \textit{for} \quad q = 1, \ldots, Q_{\mathscr{G}}. \end{aligned}$$

Moreover, let the matrices $P_{\mathscr{F},q}$ and $P_{\mathscr{G},q}$ be as in (7.20)*. Then, the realization*

$$(T_{\mathscr{F}}^* \widetilde{A}_1 T_{\mathscr{G}}, \ldots, T_{\mathscr{F}}^* \widetilde{A}_K T_{\mathscr{G}}, T_{\mathscr{F}}^* \widetilde{B}, \widetilde{C} T_{\mathscr{G}}),$$

with $(\widetilde{A}_1, \ldots, \widetilde{A}_K, \widetilde{B}, \widetilde{C})$ from Theorem 7.12, consists of real-valued matrices and interpolates the data.

Proof. First note that the state space transformation by the unitary matrices $T_{\mathscr{F}}^*$ and $T_{\mathscr{G}}$ does not change the transfer function and thus the interpolation given by Theorem 7.12 is also valid here. It remains to show that the realization consists of real-valued matrices. Since $\widetilde{B}$ and $\widetilde{C}$ are given in (7.20), it is straightforward to see that $T_{\mathscr{F}}^* \widetilde{B}$ and $\widetilde{C} T_{\mathscr{G}}$ are real-valued. As in the proof of Lemma 7.13, we deduce the realness of $T_{\mathscr{F}}^* \widetilde{A}_k T_{\mathscr{G}}$ by observing that the second equalities in (7.16) and (7.17) are equivalent to

$$\sum_{k=1}^{K} T_{\mathscr{F}}^* h_k(\mathscr{M}_q) T_{\mathscr{F}} T_{\mathscr{F}}^* P_{\mathscr{F},q}^T T_{\mathscr{F}} T_{\mathscr{F}}^* \widetilde{A}_k T_{\mathscr{G}} = T_{\mathscr{F}}^* \mathscr{L}_q^T \widetilde{C} T_{\mathscr{G}} \qquad \text{and}$$

$$\sum_{k=1}^{K} T_{\mathscr{F}}^* \widetilde{A}_k T_{\mathscr{G}} T_{\mathscr{G}}^* P_{\mathscr{G},q} T_{\mathscr{G}} T_{\mathscr{G}}^* h_k(\mathscr{S}_q) T_{\mathscr{G}} = T_{\mathscr{F}}^* \widetilde{B} \mathscr{R}_q T_{\mathscr{G}}.$$

Straightforward computations yield that $T_{\mathscr{F}}^* P_{\mathscr{F},q}^T T_{\mathscr{F}}$ and $T_{\mathscr{G}}^* P_{\mathscr{G},q} T_{\mathscr{G}}$ are real-valued. From these linear matrix equations we can determine the matrices $\widetilde{A}_k$ or equivalently the transformed analogues $T_{\mathscr{F}}^* \widetilde{A}_k T_{\mathscr{G}}$. In the latter case, we observe that the coefficient matrices as well as the right hand sides are real-valued and thus we deduce that the matrices $T_{\mathscr{F}}^* \widetilde{A}_k T_{\mathscr{G}}$ are also real-valued for $k = 1, \dots, K$. ■

7.3.2 Matching derivative data

Hermite interpolation provides a well known and robust approach for polynomial approximation that involves the matching of derivative data. When we seek reduced models that are structurally equivalent to standard first order realizations (that is, when in (6.10) we have $K = 2$, $h_1(s) = s$, and $h_2(s) \equiv -1$) then first order necessary conditions for optimality of the reduced order approximant with respect to the $\mathscr{H}_2$ norm are known and they require that the reduced transfer function $\widetilde{H}(s)$ must be a Hermite interpolant of the original $H(s)$ [89]. Even though these necessary conditions do not extend immediately to more general structured systems as in (6.10), it is known for some special cases such as second order systems with modal damping and port-Hamiltonian systems [22], and for systems with simple delay structures [70, 71], that Hermite interpolation (in a different form than for the rational case) still plays a fundamental role in the necessary optimality conditions. Therefore, if derivative information for the transfer function H is accessible then this motivates finding a structurally equivalent realization $H(s)$ that matches both the evaluation data and the derivative data. We therefore assume that

$$(f_i')^T := \ell_i^T H'(\mu_i) \qquad \text{and} \qquad g_i' := H'(\sigma_i) r_i \qquad \text{for } i = 1, \dots, n \tag{7.23}$$

are available, collected in the matrices

$$\mathscr{F}' = \begin{bmatrix} f_1' & \dots & f_n' \end{bmatrix} \qquad \text{and} \qquad \mathscr{G}' = \begin{bmatrix} g_1' & \dots & g_n' \end{bmatrix}.$$

In this section, we derive conditions such that the transfer function $\widetilde{H}$ interpolates the data (6.5) with $\{\mu_i\}_{i=1}^n \cap \{\sigma_i\}_{i=1}^n = \emptyset$ and in addition satisfies the Hermite interpolation condition (7.23).

Theorem 7.16. *Let $H(s) = \widetilde{C}(\sum_{k=1}^K h_k(s)\widetilde{A}_k)^{-1}\widetilde{B}$ and suppose that $\sum_{k=1}^K h_k(\tilde{s})\widetilde{A}_k$ is nonsingular for all $\tilde{s} \in \{\mu_i\}_{i=1}^n \cup \{\sigma_i\}_{i=1}^n$.*

(i) The left interpolation conditions $\ell_i^T \widetilde{H}(\mu_i) = f_i^T$ and the left Hermite interpolation conditions $\ell_i^T \widetilde{H}'(\mu_i) = (f_i')^T$ are satisfied for $i = 1, \dots, n$ if and only if there exist matrices $P_{\mathscr{F}}$ and $P_{\mathscr{F}'}$ that satisfy

$$\mathscr{F}^T = P_{\mathscr{F}}^T \widetilde{B}, \qquad \sum_{k=1}^K h_k(\mathscr{M}) P_{\mathscr{F}}^T \widetilde{A}_k = \mathscr{L}^T \widetilde{C}, \tag{7.24}$$

$$(\mathscr{F}')^T = (P_{\mathscr{F}'})^T \widetilde{B}, \qquad \sum_{k=1}^K h_k(\mathscr{M}) (P_{\mathscr{F}'})^T \widetilde{A}_k = -\sum_{k=1}^K h_k'(\mathscr{M}) P_{\mathscr{F}}^T \widetilde{A}_k. \tag{7.25}$$

(ii) The right interpolation conditions $\widetilde{H}(\sigma_i) r_i = g_i$ and the right Hermite interpolation conditions $\widetilde{H}'(\sigma_i) r_i = g_i'$ for $i = 1, \dots, n$ are satisfied if and only if there exist matrices $P_{\mathscr{G}}$ and $P_{\mathscr{G}'}$

that satisfy

$$\mathscr{G} = \widetilde{C} P_{\mathscr{G}}, \qquad \sum_{k=1}^{K} \widetilde{A}_k P_{\mathscr{G}} h_k(\mathscr{S}) = \widetilde{B}\mathscr{R}, \tag{7.26}$$

$$\mathscr{G}' = \widetilde{C} P_{\mathscr{G}'}, \qquad \sum_{k=1}^{K} \widetilde{A}_k P_{\mathscr{G}'} h_k(\mathscr{S}) = -\sum_{k=1}^{K} \widetilde{A}_k P_{\mathscr{G}} h_k'(\mathscr{S}). \tag{7.27}$$

Proof. We only prove the first statement; the second statement is proved analogously. We observe that (7.24) resembles the left interpolation conditions from Theorem 7.1. It remains to show that the left Hermite interpolation conditions are equivalent to (7.25). To simplify notation, we introduce, as before, $\widetilde{\mathscr{K}}(s) := \sum_{k=1}^{K} h_k(s)\widetilde{A}_k$. The second identity in (7.24) holds if and only if

$$\ell_i^T \widetilde{C} = e_i^T \mathscr{L}^T \widetilde{C} = e_i^T P_{\mathscr{F}}^T \widetilde{\mathscr{K}}(\mu_i) \qquad \text{for } i = 1, \ldots, n.$$

Similarly, from the second identity in (7.25) we obtain

$$-e_i^T P_{\mathscr{F}}^T \widetilde{\mathscr{K}}'(\mu_i) = -e_i^T \sum_{k=1}^{K} h_k'(\mathscr{M}) P_{\mathscr{F}}^T \widetilde{A}_k = e_i^T \sum_{k=1}^{K} h_k(\mathscr{M}) \left(P_{\mathscr{F}'}\right)^T \widetilde{A}_k = e_i^T \left(P_{\mathscr{F}'}\right)^T \widetilde{\mathscr{K}}(\mu_i).$$

Thus, for $i = 1, \ldots, n$ we have

$$\begin{aligned} \ell_i^T \widetilde{H}'(\mu_i) &= -\ell_i^T \widetilde{C} \widetilde{\mathscr{K}}(\mu_i)^{-1} \widetilde{\mathscr{K}}'(\mu_i) \widetilde{\mathscr{K}}(\mu_i)^{-1} \widetilde{B} = -e_i^T P_{\mathscr{F}}^T \widetilde{\mathscr{K}}'(\mu_i) \widetilde{\mathscr{K}}(\mu_i)^{-1} \widetilde{B} \\ &= e_i^T \left(P_{\mathscr{F}'}\right)^T \widetilde{B} = \left(f_i'\right)^T, \end{aligned}$$

where the last identity is nothing else than the first equality in (7.25). ■

Remark 7.17. Evidently, as the number of functions K determining the structure increases, the number of available degrees of freedom to force interpolation increases as well, and in particular, when $K > 4$ there will be sufficient degrees of freedom available to allow matching of higher order derivatives as well. The calculations involved are very technical and unenlightening, so that we choose not to pursue this thread here. It is worth to note that a combination of interpolation at additional interpolation points (section 7.3.1) and Hermite interpolation is possible as well. ♣

As before, it is sufficient and necessary to satisfy (7.24)-(7.27) simultaneously to satisfy the interpolation conditions (6.12a) and the Hermite interpolation conditions (7.23). After choosing the matrices $P_{\mathscr{F}}$, $P_{\mathscr{F}'}$, $P_{\mathscr{G}}$, and $P_{\mathscr{G}'}$, Theorem 7.16 gives $4n^2$ equations for Kn^2 unknown variables. In particular for $K = 4$, we can expect under some regularity conditions that there is a unique solution for the matrices $\widetilde{A}_k$. For $K = 3$, we can either satisfy the left or the right Hermite interpolation conditions. Similarly as in the previous subsection, we can set

$$P_{\mathscr{F}}^T := \begin{bmatrix} \mathscr{F}^T & \star \end{bmatrix}, \qquad P_{\mathscr{F}'}^T := \begin{bmatrix} (\mathscr{F}')^T & \star \end{bmatrix}, \qquad P_{\mathscr{G}} := \begin{bmatrix} \mathscr{G} \\ \star \end{bmatrix}, \qquad P_{\mathscr{G}'} := \begin{bmatrix} \mathscr{G}' \\ \star \end{bmatrix}, \tag{7.28}$$

yielding $\widetilde{B} = \begin{bmatrix} I_m & 0 \end{bmatrix}^T$ and $\widetilde{C} = \begin{bmatrix} I_p & 0 \end{bmatrix}$. For the sake of completeness, we derive the equivalent of the system (7.18) for Hermite interpolation for $K = 4$. Vectorization of the second equations in

(7.24)-(7.27), respectively, yields the system $\mathbb{A}\boldsymbol{\alpha} = \boldsymbol{\beta}$ with $\mathbb{A} \in \mathbb{C}^{4n^2 \times 4n^2}$ and $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{C}^{4n^2}$ given by

$$\begin{aligned}
\mathbb{A} &:= \begin{bmatrix} I_n \otimes h_1(\mathscr{M}) P_{\mathscr{F}}^T & \cdots & I_n \otimes h_4(\mathscr{M}) P_{\mathscr{F}}^T \\ I_n \otimes \left(h_1(\mathscr{M}) \left(P_{\mathscr{F}'}\right)^T + h_1'(\mathscr{M}) P_{\mathscr{F}}^T\right) & \cdots & I_n \otimes \left(h_4(\mathscr{M}) \left(P_{\mathscr{F}'}\right)^T + h_4'(\mathscr{M}) P_{\mathscr{F}}^T\right) \\ h_1(\mathscr{S}) P_{\mathscr{G}}^T \otimes I_n & \cdots & h_4(\mathscr{S}) P_{\mathscr{G}}^T \otimes I_n \\ \left(h_1(\mathscr{S}) \left(P_{\mathscr{G}'}\right)^T + h_1'(\mathscr{S}) P_{\mathscr{G}}^T\right) \otimes I_n & \cdots & \left(h_4(\mathscr{S}) \left(P_{\mathscr{G}'}\right)^T + h_4'(\mathscr{S}) P_{\mathscr{G}}^T\right) \otimes I_n \end{bmatrix}, \\
\boldsymbol{\alpha} &:= \begin{bmatrix} \operatorname{vec}(\widetilde{A}_1) \\ \operatorname{vec}(\widetilde{A}_2) \\ \operatorname{vec}(\widetilde{A}_3) \\ \operatorname{vec}(\widetilde{A}_4) \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{\beta} := \begin{bmatrix} \left(\widetilde{C}^T \otimes I_n\right) \operatorname{vec}\left(\mathscr{L}^T\right) \\ 0 \\ \left(I_n \otimes \widetilde{B}\right) \operatorname{vec}(\mathscr{R}) \\ 0 \end{bmatrix}.
\end{aligned} \tag{7.29}$$

Remark 7.18. Real-valued realizations that accomplish Hermite interpolation may be obtained in the same manner as in the case of additional interpolation points (cf. Lemmas 7.13 and 7.15). The only additional requirement is that $\mathscr{F}$ and $\mathscr{F}'$ as well as $\mathscr{G}$ and $\mathscr{G}'$ need to have the same number of complex conjugate pairs such that

$$T_{\mathscr{F}}^* \mathscr{F}^T \in \mathbb{R}^n, \quad T_{\mathscr{F}}^* \left(\mathscr{F}'\right)^T \in \mathbb{R}^n, \quad \mathscr{G} T_{\mathscr{G}} \in \mathbb{R}^n, \quad \text{and} \quad \mathscr{G}' T_{\mathscr{G}} \in \mathbb{R}^n.$$

♣

7.4 Truncation of redundant data

Suppose that we have solved the linear system (7.18) (or (7.29) for Hermite interpolation) to obtain the realization $\widetilde{H}(s) = \widetilde{C}\widetilde{\mathscr{K}}(s)^{-1}\widetilde{B}$ with $\widetilde{\mathscr{K}}(s) = \sum_{k=1}^K h_k(s)\widetilde{A}_k$. By construction, the matrices satisfy the equations in Theorem 7.12 (or Theorem 7.16). However, $\widetilde{\mathscr{K}}(s)$ might be (numerically) singular at the driving frequencies $\mu_{q;i}$ and $\sigma_{q;i}$. This is likely to happen if we add more and more data, since at some point the information becomes redundant, and $\widetilde{\mathscr{K}}(s)$ might become ill-conditioned or even singular. Let us emphasize, that in view of the theory for *delay differential-algebraic equations* (DDAEs) developed in the previous chapters, general existence and uniqueness results require that $\widetilde{\mathscr{K}}(s)$ is nonsingular for some $s \in \mathbb{C}$, see for instance Theorem 3.20.

To remove redundant data, we proceed similarly as in Theorem 6.13 and assume that for all $s \in \{\mu_{q;i}\}_{i=1}^n \cup \{\sigma_{q;i}\}_{i=1}^n$ we have

$$\operatorname{rank}\left(\sum_{k=1}^K h_k(s)\widetilde{A}_k\right) = \operatorname{rank}\left(\begin{bmatrix} \widetilde{A}_1 & \cdots & \widetilde{A}_K \end{bmatrix}\right) = \operatorname{rank}\left(\begin{bmatrix} \widetilde{A}_1 \\ \vdots \\ \widetilde{A}_K \end{bmatrix}\right) =: r. \tag{7.30}$$

In this case, there exist unitary matrices $V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$ and $W = \begin{bmatrix} W_1 & W_2 \end{bmatrix} \in \mathbb{C}^{n \times n}$ with $V_1, W_1 \in \mathbb{C}^{n \times r}$ and $V_2, W_2 \in \mathbb{C}^{n \times (n-r)}$ such that

$$\widetilde{A}_k V_2 = 0 \qquad \text{and} \qquad \widetilde{A}_k^* W_2 = 0, \qquad \text{for all} \quad k = 1, \ldots, K. \tag{7.31}$$

Theorem 7.19. *Let the realization $\widetilde{H}(s) = \widetilde{C}(\sum_{k=1}^{K} h_k(s)\widetilde{A}_k)^{-1}\widetilde{B}$ satisfy the equations in Theorem 7.12 with matrices $P_{\mathscr{F},q}$ and $P_{\mathscr{G},q}$. Suppose that the matrices $\widetilde{A}_k$ satisfy the rank assumption* (7.30) *and let $V_1, W_1 \in \mathbb{C}^{n\times r}$ complete V_2 and W_2 in* (7.31) *to unitary matrices. For $k = 1,\ldots,K$ set*

$$\widetilde{A}_{k;r} := W_1^* \widetilde{A}_k V_1, \qquad \widetilde{B}_r := W_1^* \widetilde{B}, \qquad \textit{and} \qquad \widetilde{C}_r := \widetilde{C} V_1.$$

If $\operatorname{span}\{\ell_{q;1},\ldots,\ell_{q;n}\} = \mathbb{C}^{n_y}$ *for all $q = 1,\ldots,Q_{\mathscr{F}}$ and* $\operatorname{span}\{r_{q;1},\ldots,r_{q;n}\} = \mathbb{C}^{n_u}$ *for all $q = 1,\ldots,Q_{\mathscr{G}}$, then the realization $\widetilde{H}_r(s) = \widetilde{C}_r(\sum_{k=1}^{K} h_k(s)\widetilde{A}_{k;r})^{-1}\widetilde{B}_r$ interpolates the data.*

Proof. First, bear in mind that by assumption, the affine structure $\sum_{k=1}^{K} h_k(s)\widetilde{A}_{k;r}$ is nonsingular at the driving frequencies $\mu_{q;i}$ and $\sigma_{q;i}$, and we observe that $\widetilde{A}_k V_1 V_1^* = \widetilde{A}_k$ and $W_1 W_1^* \widetilde{A}_k = \widetilde{A}_k$ hold for $k = 1,\ldots,K$ by construction of V_1 and W_1. Thus, for $q = 1,\ldots,Q_{\mathscr{F}}$

$$\begin{aligned}\sum_{k=1}^{K} h_k(\mathscr{M}_q) P_{\mathscr{F},q}^T W_1 \widetilde{A}_{k;r} &= \left(\sum_{k=1}^{K} h_k(\mathscr{M}_q) P_{\mathscr{F},q}^T W_1 W_1^* \widetilde{A}_k\right) V_1 \\ &= \left(\sum_{k=1}^{K} h_k(\mathscr{M}_q) P_{\mathscr{F},q}^T \widetilde{A}_k\right) V_1 = \mathscr{L}_q^T \widetilde{C}_r,\end{aligned}$$

where the last identity follows from (7.16). Similarly, we obtain for $q = 1,\ldots,Q_{\mathscr{G}}$

$$\sum_{k=1}^{K} \widetilde{A}_{k;r} V_1^* P_{\mathscr{G},q} h_k(\mathscr{S}_q) = W_1^* \sum_{k=1}^{K} \widetilde{A}_k V_1 V_1^* P_{\mathscr{G},q} h_k(\mathscr{S}_q) = W_1^* \sum_{k=1}^{K} \widetilde{A}_k P_{\mathscr{G},q} h_k(\mathscr{S}_q) = \widetilde{B}_r \mathscr{R}_q.$$

Furthermore, we notice

$$\mathscr{L}_q^T \widetilde{C} = \sum_{k=1}^{K} h_k(\mathscr{M}_q) P_{\mathscr{F},q}^T \widetilde{A}_k = \sum_{k=1}^{K} h_k(\mathscr{M}_q) P_{\mathscr{F},q}^T \widetilde{A}_k V_1 V_1^* = \mathscr{L}_q^T \widetilde{C} V_1 V_1^*.$$

Since the columns of $\mathscr{L}_q$ span the whole space $\mathbb{C}^p$, the above identity implies $\widetilde{C} = \widetilde{C} V_1 V_1^*$. With the same reasoning we obtain $\widetilde{B} = W_1 W_1^* \widetilde{B}$. Finally, we have

$$\begin{aligned}\ell_{q;i}^T \widetilde{H}_r(\mu_{q;i}) &= e_i^T \mathscr{L}_q^T \widetilde{C}_r \left(\sum_{k=1}^{K} h_k(\mu_{q;i}) \widetilde{A}_{k;r}\right)^{-1} \widetilde{B}_r \\ &= e_i^T \left(\sum_{k=1}^{K} h_k(\mathscr{M}_q) P_{\mathscr{F},q}^T W_1 \widetilde{A}_{k;r}\right) \left(\sum_{k=1}^{K} h_k(\mu_{q;i}) \widetilde{A}_{k;r}\right)^{-1} \widetilde{B}_r \\ &= e_i^T P_{\mathscr{F},q}^T W_1 \left(\sum_{k=1}^{K} h_k(\mu_{q;i}) \widetilde{A}_{k;r}\right) \left(\sum_{k=1}^{K} h_k(\mu_{q;i}) \widetilde{A}_{k;r}\right)^{-1} \widetilde{B}_r \\ &= e_i^T P_{\mathscr{F},q}^T W_1 W_1^* \widetilde{B} = f_{q;i}^T\end{aligned}$$

for $q = 1,\ldots,Q_{\mathscr{F}}$ and $i = 1,\ldots,n$. The right interpolation conditions follow analogously. ■

Remark 7.20. Instead of performing two rank-revealing decompositions on

$$\begin{bmatrix}\widetilde{A}_1 & \cdots & \widetilde{A}_K\end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix}\widetilde{A}_1^* & \cdots & \widetilde{A}_K^*\end{bmatrix}^*,$$

it is computationally more reasonable (as it is also done in the classical Loewner framework [150]) to pick a driving frequency $\tilde{s} \in \{\mu_{q;i}\}_{i=1}^{n} \cup \{\sigma_{q;i}\}_{i=1}^{n}$ and take the *singular value decomposition* (SVD)

$$\sum_{k=1}^{K} h_k(\tilde{s})\widetilde{A}_k = \begin{bmatrix} W_1 & W_2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} \tag{7.32}$$

with $V_1, W_1 \in \mathbb{C}^{n\times(n-r)}$ and $V_2, W_2 \in \mathbb{C}^{n\times r}$. These matrices satisfy the assumptions of Theorem 7.19, since

$$\ker\left(\begin{bmatrix} \widetilde{A}_1^* & \cdots & \widetilde{A}_K^* \end{bmatrix}^*\right) \subseteq \ker\left(\sum_{k=1}^{K} h_k(s)\widetilde{A}_k\right)$$

and the rank assumption (7.30) implies $\widetilde{A}_k V_2 = 0$ for all $k = 1,\ldots,K$. By the same reasoning, $W_2^* \widetilde{A}_k = 0$ for all $k = 1,\ldots,K$ and hence V_1 and W_1 from (7.32) can be used to truncate the data. $\clubsuit$

Example 7.21. If we pick further distinct interpolation points in Example 7.14, then the realization is given by the matrices

$$\widetilde{A}_1 = \frac{1}{cb}\begin{bmatrix} a_1 & \ldots & a_1 \\ \vdots & & \vdots \\ a_1 & \ldots & a_1 \end{bmatrix}, \quad \widetilde{A}_2 = \frac{1}{cb}\begin{bmatrix} a_2 & \ldots & a_2 \\ \vdots & & \vdots \\ a_2 & \ldots & a_2 \end{bmatrix}, \quad \text{and} \quad \widetilde{A}_3 = \frac{1}{cb}\begin{bmatrix} a_3 & \ldots & a_3 \\ \vdots & & \vdots \\ a_3 & \ldots & a_3 \end{bmatrix}.$$

Clearly, the rank assumption (7.30) is satisfied with $r = 1$. Setting $W_1 = \begin{bmatrix} 1 & 0 & \ldots & 0] \end{bmatrix}$ and $V_1 = W_1^T$ yields the true transfer function. $\spadesuit$

In view of Theorem 7.16, we still need to establish that the Hermite interpolation conditions are satisfied after truncation (the left and right interpolation conditions are satisfied by Theorem 7.19). By the same reasoning as in the proof of Theorem 7.19 we can establish the identity

$$\sum_{k=1}^{K} h_k(\mathscr{M})\left(P_{\mathscr{F}'}\right)^T W_1 \widetilde{A}_{k;r} = -\sum_{k=1}^{K} h_k'(\mathscr{M}) P_{\mathscr{F}}^T W_1 \widetilde{A}_{k;r}$$

and compute

$$\begin{aligned} \ell_i^T \widetilde{H}_r'(\mu_i) &= -e_i^T \mathscr{L}^T \widetilde{C}_r \left(\sum_{k=1}^{K} h_k(\mu_i)\widetilde{A}_{k;r}\right)^{-1} \left(\sum_{k=1}^{K} h_k'(\mu_i)\widetilde{A}_{k;r}\right)\left(\sum_{k=1}^{K} h_k(\mu_i)\widetilde{A}_{k;r}\right)^{-1} \widetilde{B}_r \\ &= -e_i^T P_{\mathscr{F}}^T W_1 \left(\sum_{k=1}^{K} h_k'(\mu_i)\widetilde{A}_{k;r}\right)\left(\sum_{k=1}^{K} h_k(\mu_i)\widetilde{A}_{k;r}\right)^{-1} \widetilde{B}_r \\ &= e_i^T \left(\sum_{k=1}^{K} h_k(\mathscr{M})\left(P_{\mathscr{F}'}\right)^T W_1 \widetilde{A}_{k;r}\right)\left(\sum_{k=1}^{K} h_k(\mu_i)\widetilde{A}_{k;r}\right)^{-1} \widetilde{B}_r \\ &= e_i^T \left(P_{\mathscr{F}'}\right)^T W_1 W_1^* \widetilde{B} = e_i^T \left(P_{\mathscr{F}'}\right)^T \widetilde{B} = \left(f_i'\right)^T \end{aligned}$$

and hence the left Hermite interpolation condition is still satisfied. The proof for the right Hermite interpolation condition proceeds analogously. We summarize the previous discussion in the following theorem.

Theorem 7.22. *Let the realization $\widetilde{H}(s) = \widetilde{C}(\sum_{k=1}^{K} h_k(s)\widetilde{A}_k)^{-1}\widetilde{B}$ satisfy the equations in Theorem 7.16 with matrices $P_{\mathscr{F}}$, $P_{\mathscr{F}'}$, $P_{\mathscr{G}}$, and $P_{\mathscr{G}'}$. Suppose that the matrices $\widetilde{A}_k$ satisfy the rank assumption* (7.30) *and let $W_1, V_1 \in \mathbb{C}^{n\times r}$ be as in Theorem 7.19. If* $\operatorname{span}\{\ell_1,\ldots,\ell_n\} = \mathbb{C}^{n_y}$ *and* $\operatorname{span}\{r_1,\ldots,r_n\} = \mathbb{C}^{n_u}$*, then the realization $\widetilde{H}_r(s) = \widetilde{C}_r(\sum_{k=1}^{K} h_k\widetilde{A}_{k;r})^{-1}\widetilde{B}_r$ interpolates the data and derivative data with $\widetilde{A}_{k;r} := W_1^*\widetilde{A}_k V_1$, $\widetilde{B}_r := W_1^*\widetilde{B}$, and $\widetilde{C}_r := \widetilde{C}V_1$.*

Let us emphasize that the function

$$\eta\colon \mathbb{C}\to\mathbb{C}, \qquad s\mapsto \det\left(\sum_{k=1}^{K} h_k(s)\widetilde{A}_k\right)$$

may be identically zero even if the matrices $\begin{bmatrix}\widetilde{A}_1 & \ldots & \widetilde{A}_K\end{bmatrix}$ and $\begin{bmatrix}\widetilde{A}_1^T & \ldots & \widetilde{A}_K^T\end{bmatrix}^T$ in (7.30) have full rank (cf. Remark 6.14). In this case Theorems 7.19 and 7.22 cannot be used as a post-processing step. Instead, an additional regularization similar to [34] is required. This is subject to further research.

7.5 An algorithm for structured realization

In this section, we synthesize the results of the previous subsections into an algorithmic format, starting with interpolation data (6.5) and an affine structure given via continuously differentiable functions h_k for $k = 1,\ldots,K$. The goal is to construct matrices $\widetilde{A}_1,\ldots,\widetilde{A}_K$, $\widetilde{B}$ and $\widetilde{C}$ such that the realization $\widetilde{H}(s) = \widetilde{C}(\sum_{k=1}^{K} h_k(s)\widetilde{A}_k)^{-1}\widetilde{B}$ associated with the affine structure interpolates the data. We construct realizations as described in the previous subsections, taking advantage of the simplifications available when $K = 2$. Before doing so, a pre-processing step is included if the data set is closed under complex conjugation, which facilitates construction of a real-valued realization where appropriate. Although in principle the transformation to a real-valued realization could be performed after assembling the matrices, it is advisable to enforce this in advance, since numerical rounding errors tend to break the underlying conjugate symmetry and will thus cause complex matrix entries in the realization. A post-processing step may also be necessary to truncate redundancies discovered in the interpolation data. The details are summarized in Algorithm 3.

In practical applications, the interpolation data is often not partitioned in the form (6.5). We therefore present a simplification of Algorithm 3 for the SISO case that can be used without pre-processing of the data. Accordingly, we assume to have the interpolation data

$$\{(\lambda_i, H(\lambda_i)) \in \mathbb{C}^2 \mid i = 1,\ldots,Kn\} \tag{7.33}$$

available. For the sake of notational convenience we assume that we do not interpolate on the real axis, that is $\lambda_i \in \mathbb{C}\setminus\mathbb{R}$ for $i = 1,\ldots,Kn$. To ensure a real realization, we add the complex conjugate data and assume n to be an even number and

$$(\overline{\lambda_{2i-1}}, \overline{H(\lambda_{2i-1})}) = (\lambda_{2i}, H(\lambda_{2i})) \qquad \text{for } i = 1,\ldots,\frac{Kn}{2}. \tag{7.34}$$

Algorithm 3 Structured Realization

Input: Interpolation data (6.5), affine structure $h_1(s), \dots, h_K(s)$ with $K \in \mathbb{N}$.
Output: Matrices $\widetilde{A}_1, \dots, \widetilde{A}_K$, $\widetilde{B}$, and $\widetilde{C}$ such that $\widetilde{H}(s) = \widetilde{C}(\sum_{k=1}^{K} h(s)\widetilde{A}_k)^{-1}\widetilde{B}$ interpolates the data

```
if Data is closed under complex conjugation then          ▷ Keep realization real
    Transform data as in Lemma 7.8, Lemma 7.13, Lemma 7.15, or Remark 7.18
end if

if $K = 2$ then
    Transform data as in (7.14)
    Construct Loewner matrices according to (6.14a) and (6.14b) from the transformed data
    Set $\widetilde{A}_1 = -\mathbb{L}$, $\widetilde{A}_2 = \mathbb{L}_\sigma$, $\widetilde{B} = \mathscr{F}^T$ and $\widetilde{C} = \mathscr{G}$
else
    if derivative data (7.23) is available then
        Construct the matrices $\widetilde{B}, \widetilde{C}, P_{\mathscr{F}}, P_{\mathscr{F}'}, P_{\mathscr{G}}$, and $P_{\mathscr{G}'}$, for example as in (7.28)
        Assemble system (7.29) and solve for $\widetilde{A}_1, \dots, \widetilde{A}_K$
    else
        Partition the data as in (7.15) and pick $n$ accordingly
        Construct the matrices $\widetilde{B}, \widetilde{C}, P_{\mathscr{F},q}$ and $P_{\mathscr{G},q}$ matrices, for example as in (7.19)
        Assemble system (7.18) and solve for $\widetilde{A}_1, \dots, \widetilde{A}_K$
    end if
end if

Compute $r$ as in (7.30)
if $r < n$ then                                          ▷ Truncation of redundant data
    Compute $V_1$ and $W_1$ as in Theorem 7.19
    Set $\widetilde{A}_k := W_1^* \widetilde{A}_k V_1$, $\widetilde{B} := W_1^* \widetilde{B}$, and $\widetilde{C} := \widetilde{C} V_1$
end if
```

For the partitioning of the data (7.33) in the form (6.5) we set $Q_{\mathscr{F}} := \lceil \frac{K}{2} \rceil$, $Q_{\mathscr{G}} = \lfloor \frac{K}{2} \rfloor$ such that $Q_{\mathscr{F}} + Q_{\mathscr{G}} = K$. Moreover, we rename the interpolation data as

$$\mu_{j;i} := \lambda_{2(j-1)n+i}, \qquad f_{j;i} := H(\lambda_{2(j-1)n+i}), \qquad \text{for } j = 1, \dots, Q_{\mathscr{F}},\ i = 1, \dots, n, \tag{7.35a}$$

$$\sigma_{j;i} := \lambda_{(2j-1)n+i}, \qquad g_{j;i} := H(\lambda_{(2j-1)n+i}), \qquad \text{for } j = 1, \dots, Q_{\mathscr{G}},\ i = 1, \dots, n, \tag{7.35b}$$

and define the matrix

$$T = \operatorname{blkdiag}\left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -\imath \\ 1 & \imath \end{bmatrix}, \dots, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -\imath \\ 1 & \imath \end{bmatrix} \right) \in \mathbb{C}^{n \times n}. \tag{7.36}$$

With these preliminaries we can present a simplified version of Algorithm 3 in Algorithm 4.

Algorithm 4 Simplified Structured Realization for SISO systems

```
Input: Interpolation data (7.33), function family $\{h_1,\dots,h_K\}$ with $K \in \mathbb{N}$.
Output: Matrices $\widetilde{A}_1,\dots,\widetilde{A}_K$, $\widetilde{B}$, and $\widetilde{C}$ such that $\widetilde{H}(s) = \widetilde{C}(\sum_{k=1}^{K} h_k(s)\widetilde{A}_k)^{-1}\widetilde{B}$ interpolates the data

Add the complex conjugate data as in (7.34).
Partition the data as in (7.35) with $Q_{\mathscr{F}} := \lceil \frac{K}{2} \rceil$ and $Q_{\mathscr{G}} = \lfloor \frac{K}{2} \rfloor$.
Solve the linear systems (7.21) for $\widetilde{A}_1,\dots,\widetilde{A}_K$.
Set $\widetilde{A}_k := T^* \widetilde{A}_k T$ for $k = 1,\dots,K$, $\widetilde{B} := T^* \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^T$, and $\widetilde{C} := \widetilde{B}^T$ with $T$ as in (7.36).
Pick any $i \in \{1,\dots,Kn\}$ and compute $V_1, W_1$ via the SVD as in (7.32).
Set $\widetilde{A}_k := W_1^* \widetilde{A}_k V_1$, $\widetilde{B} := W_1^* \widetilde{B}$, and $\widetilde{C} := \widetilde{C} V_1$
```

7.6 Connection to structure-preserving interpolatory projections

Although our focus here is on data-driven interpolation, we briefly revisit the structure-preserving interpolatory projection framework introduced in [18] and establish a connection with realizations arising from Theorem 7.5.

Theorem 7.23 (Structure-preserving interpolatory projection [18])**.** *Consider the generalized realization $H(s) = \mathscr{C}(s)\mathscr{K}(s)^{-1}\mathscr{B}(s)$ where both $\mathscr{C}(s) \in \mathbb{C}^{n_y \times n_x}$ and $\mathscr{B}(s) \in \mathbb{C}^{n_x \times n_u}$ are analytic in the right half plane and $\mathscr{K}(s) \in \mathbb{C}^{n_x \times n_x}$ is analytic and has full rank throughout the right half plane. Suppose that the left interpolation points $\{\mu_1,\dots,\mu_n\}$ together with the left tangential directions $\{\ell_1,\dots,\ell_n\}$ and the right interpolation points $\{\sigma_1,\dots,\sigma_n\}$ together with the right tangential directions $\{r_1,\dots,r_n\}$ are given. Define $V \in \mathbb{C}^{n_x \times n}$ and $W \in \mathbb{C}^{n_x \times n}$ as*

$$W = \begin{bmatrix} \mathscr{K}(\mu_1)^{-T}\mathscr{C}(\mu_1)^T\ell_1 & \dots & \mathscr{K}(\mu_n)^{-T}\mathscr{C}(\mu_n)^T\ell_n \end{bmatrix} \qquad \textit{and} \tag{7.37a}$$

$$V = \begin{bmatrix} \mathscr{K}(\sigma_1)^{-1}\mathscr{B}(\sigma_1)r_1 & \dots & \mathscr{K}(\sigma_n)^{-1}\mathscr{B}(\sigma_n)r_n \end{bmatrix}. \tag{7.37b}$$

Define

$$\widetilde{\mathscr{K}}(s) = W^T\mathscr{K}(s)V, \quad \widetilde{\mathscr{B}}(s) = W^T\mathscr{B}(s), \quad \widetilde{\mathscr{C}}(s) = \mathscr{C}(s)V. \tag{7.38}$$

Then the reduced transfer function $\widetilde{H}(s) = \widetilde{\mathscr{C}}(s)\widetilde{\mathscr{K}}(s)^{-1}\widetilde{\mathscr{B}}(s)$ satisfies the interpolation conditions (6.12).

Using $\mathscr{K}(s) = \sum_{k=1}^{K} h_k(s)A_k$, $\mathscr{B}(s) = B$, and $\mathscr{C}(s) = C$ for the affine structure then (7.38) leads to a reduced model with

$$\widetilde{B} = W^TB, \qquad \widetilde{C} = CV, \qquad \text{and} \qquad \widetilde{A}_k = W^TA_kV \quad \text{for } k = 1,\dots,K.$$

The question we want to answer next is how (and if) this projection-based reduced model is connected to the data-driven realization constructed in the previous sections. The next result provides the link.

Proposition 7.24. *The projection matrices W and V introduced in* (7.37)*, based on the matrix functions $\mathscr{K}(s) = \sum_{k=1}^{K} h_k(s) A_k$, $\mathscr{B}(s) = B$, and $\mathscr{C}(s) = C$, satisfy the matrix equations*

$$\sum_{k=1}^{K} h_k(\mathscr{M}) W^T A_k = \mathscr{L}^T C \qquad \text{and} \qquad \sum_{k=1}^{K} A_k V h_k(\mathscr{S}) = B\mathscr{R}, \tag{7.39}$$

as well as

$$\sum_{k=1}^{K} h_k'(\mu_i) \left[W^T A_k V \right]_{i,i} = -\theta_i \tag{7.40}$$

for those $i \in \{1,\ldots,n\}$ with $\mu_i = \sigma_i$.

Proof. Let $w_i = We_i$ and $v_i = Ve_i$ denote the columns of the projection matrices W and V. For $i = 1,\ldots,n$ we have

$$e_i^T \sum_{k=1}^{K} h_k(\mathscr{M}) W^T A_k = w_i^T \sum_{k=1}^{K} h_k(\mu_i) A_k = \ell_i^T C,$$

which proves the first identity. The second identity is proven similarly whereas the third identity follows from the definitions of W and V and from

$$\sum_{k=1}^{K} h_k'(\mu_i)[W^T A_k V]_{i,i} = w_i^T \left(\sum_{k=1}^{K} h_k'(\mu_i) A_k \right) v_i = -\ell_i^T H'(\mu_i) r_i = -\theta_i. \qquad \blacksquare$$

Proposition 7.24 gives a better understanding of the realization of Theorem 7.5 connecting it to the projection-based MOR framework. To make this connection more precise, we will investigate the cases $K = 2$ and $K \geq 3$ separately below.

7.6.1 The case $K = 2$

Using the identities $W^T B = \mathscr{F}^T$ and $CV = \mathscr{G}$, we can rewrite (7.7), using $K = 2$, as

$$h_2(\mathscr{M})\widetilde{A}_1 h_1(\mathscr{S}) - h_1(\mathscr{M})\widetilde{A}_1 h_2(\mathscr{S}) = h_2(\mathscr{M}) W^T B\mathscr{R} - \mathscr{L}^T C V h_2(\mathscr{S}).$$

Substituting the expressions for $B\mathscr{R}$ and $\mathscr{L}^T C$ from (7.39) into the right-hand side implies

$$h_2(\mathscr{M})\widetilde{A}_1 h_1(\mathscr{S}) - h_1(\mathscr{M})\widetilde{A}_1 h_2(\mathscr{S}) = h_2(\mathscr{M}) W^T A_1 V h_1(\mathscr{S}) - h_1(\mathscr{M}) W^T A_1 V h_2(\mathscr{S}),$$

which establishes the relation $\widetilde{A}_1 = W^T A_1 V$ as long as the interpolation sets $\{\mu_i\}_{i=1}^n$ and $\{\sigma_i\}_{i=1}^n$ are disjoint. The identity $\widetilde{A}_2 = W^T A_2 V$ is obtained by using (7.8) instead of (7.7). Thus for $K = 2$, our structured realization approach gives exactly the reduced model one would obtain via projection if the original system matrices were to be available. This equivalence of the projected matrices and the matrices obtained by the realizations is also true if there is an overlapping between the left and right interpolation point sets. This may be comprehended by the observation that the projected matrices also satisfy (7.11) which is clear due to Proposition 7.24.

7.6.2 The case $K \geq 3$

Consider the second-order model $H(s) = C(s^2 A_1 + sA_2 + A_3)^{-1} B$. For simplicity, assume that $H(s)$ is a SISO system. Given the $2n$ interpolation points $\{\mu_1, \ldots, \mu_n\}$ and $\{\sigma_1, \ldots, \sigma_n\}$, one can obtain a projection-based reduced model $\widetilde{H}(s) = \widetilde{C}(s^2 \widetilde{A}_1 + s\widetilde{A}_2 + \widetilde{A}_3)^{-1} \widetilde{B}$ using Theorem 7.23. This reduced model will interpolate H at $2n$ interpolation points. However, $\widetilde{H}(s)$ has $3n$ degrees of freedom[1] and should be able to satisfy $3n$ interpolation conditions. The projection framework cannot achieve this goal. However, our structured realization framework with either additional data as in Section 7.3.1 or Hermite interpolation as in Section 7.3.2 will construct a reduced model that can match this maximum number of interpolation conditions. In other words, for $K \geq 3$, the structured realization cannot be obtained via projection and indeed satisfies more interpolation conditions than the projection framework. Next we give a numerical example illustrating this discussion on a delay example.

Example 7.25. We consider a delay system with the affine structure $h_1(s) = s, h_2(s) \equiv -1$, and $h_3(s) = -\exp(-s)$ and matrices

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \qquad A_2 = A_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \text{and} \qquad c = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

with transfer function $H(s) = c^T(sA_1 - A_2 - e^{-s}A_3)^{-1}b$. We set $Q_{\mathscr{F}} = 1$ and $Q_{\mathscr{G}} = 2$ and pick the driving frequencies $\mu_{1;1} = 0, \sigma_{1;1} = 1$, and $\sigma_{2;1} = -1$. We want to make use of the system (7.21), so we set $\widetilde{B} = 1, \widetilde{C} = 1, P_{\mathscr{F},1} = f_{1;1}, P_{\mathscr{G},1} = g_{1;1}$, and $P_{\mathscr{G},2} = g_{2;1}$. Altogether, the solution of the system (7.21) is given by

$$\begin{bmatrix} \widetilde{A}_1 \\ \widetilde{A}_2 \\ \widetilde{A}_3 \end{bmatrix} = \frac{1}{2 - \mathrm{e} - \frac{1}{\mathrm{e}}} \begin{bmatrix} \mathrm{e} - \frac{1}{\mathrm{e}} + \frac{(1-\frac{1}{\mathrm{e}})^2}{2-\mathrm{e}} + \frac{(\mathrm{e}-1)(\mathrm{e}+2)(\mathrm{e}+3)}{-5-2\mathrm{e}} \\ -\mathrm{e} - \frac{1}{\mathrm{e}} - \frac{1-\frac{1}{\mathrm{e}}}{\mathrm{e}-2} - \frac{(\mathrm{e}+2)(\mathrm{e}+3)}{-5-2\mathrm{e}} \\ 2 + \frac{1-\frac{1}{\mathrm{e}}}{\mathrm{e}-2} + \frac{(\mathrm{e}+2)(\mathrm{e}+3)}{-5-2\mathrm{e}} \end{bmatrix}.$$

Clearly, $\widetilde{A}_2 \neq \widetilde{A}_3$; and hence the realization cannot be obtained via projection. ♠

7.7 Numerical examples

To illustrate the consequences of the preceding theoretical discussion, we compare various structured realizations against the standard Loewner realization framework, using in each case response data as in (6.5). In all the following examples, $H(s)$, $\widetilde{H}_{\mathrm{L}}(s)$, $\widetilde{H}_{\mathrm{A}}(s)$, and $\widetilde{H}_{\mathrm{H}}(s)$ will denote, respectively: the transfer function of the original model, the rational approximation via the standard Loewner realization, the structured realization interpolating at additional points (section 7.3.1), and the structured realization satisfying Hermite interpolation conditions (section 7.3.2). In the following plots, we represent interpolation frequencies with solid vertical lines. Additional driving frequencies used for the structured realization interpolating additional points are highlighted as dashed vertical lines.

[1] A second-order model with $\widetilde{H}(s) = (s\widetilde{C}_1 + \widetilde{C}_2)(s^2\widetilde{A}_1 + s\widetilde{A}_2 + \widetilde{A}_3)^{-1}\widetilde{B}$, where not the state x but a linear combination of the state x and the state velocity $\dot{x}$ is measured, has $4n$ degrees of freedom. But here we do not consider this case.

Table 7.1 – Example 7.26 – $\mathscr{H}_\infty$ errors of the different realizations

n	Loewner	Additional points	Hermite
4	2.342312×10^{-1}	4.496194×10^{-2}	4.011660×10^{-2}
6	2.449003×10^{-1}	5.100268×10^{-2}	4.116856×10^{-2}
8	3.397454×10^{-1}	4.673353×10^{-2}	4.307346×10^{-2}
10	5.561860×10^{-1}	4.454640×10^{-2}	3.694951×10^{-2}

We approximately compute the $\mathscr{H}_\infty$ model reduction errors by performing an extensive sampling of the transfer functions on the imaginary axis. To this end, we extend the interval in which the interpolation points are chosen by five orders of magnitude on both sides and sample the extended interval with 50,000 points. For a more efficient way of computing the $\mathscr{H}_\infty$ norm for the general class of systems considered in this paper, see the recent work [3].

If not otherwise stated, the presented examples are SISO systems. Accordingly, the matrices needed for the realizations corresponding to $\widetilde{H}_\mathrm{A}$ and $\widetilde{H}_\mathrm{H}$ have been chosen as in (7.20) and as the analogue for the Hermite case which is

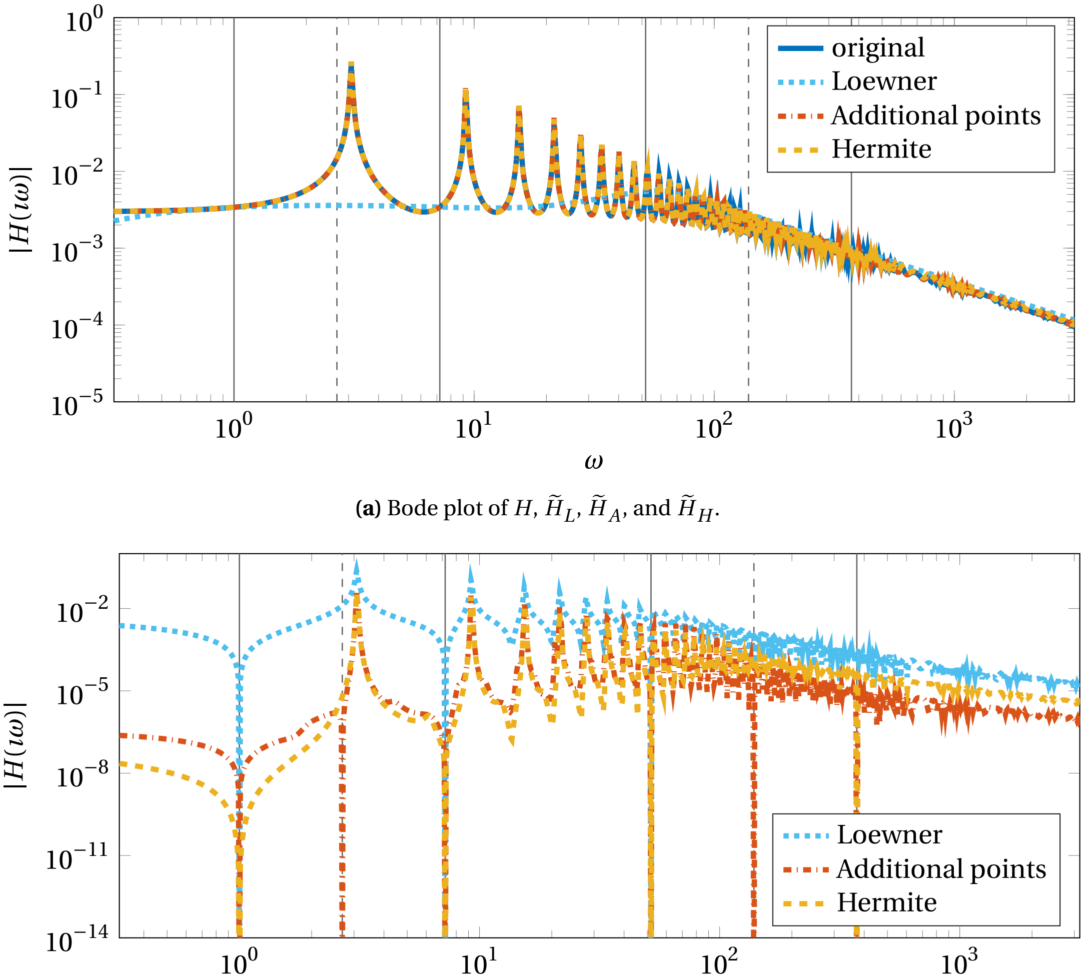
$$P_\mathscr{F} := \operatorname{diag}(\mathscr{F}), \quad P_\mathscr{G} := \operatorname{diag}(\mathscr{G}), \quad P_{\mathscr{F}'} := \operatorname{diag}(\mathscr{F}'), \quad \text{and} \quad P_{\mathscr{G}'} := \operatorname{diag}(\mathscr{G}').$$

Example 7.26. We test our approaches with the delay model from [18] given by the $n_x \times n_x$ matrices

$$A_1 = \nu I_N + T, \quad A_2 = \frac{1}{\tau}\left(\frac{1}{\zeta}+1\right)(T-\nu I_N), \quad A_3 = \frac{1}{\tau}\left(\frac{1}{\zeta}-1\right)(T-\nu I_N),$$

where T is an $n_x \times n_x$ matrix with ones on the sub- and superdiagonal, at the $(1,1)$, and at the (n_x, n_x) position and zeros everywhere else. The functions h_k are given by $h_1(s) = s, h_2(s) \equiv -1$, and $h_3(s) = -\mathrm{e}^{-\tau s}$. We choose $n_x = 500$, $\tau = 1$, $\zeta = 0.01$, and $\nu = 5$. The input matrix $B \in \mathbb{R}^{n_x}$ has ones in the first two components and zeros everywhere else and we choose $C = B^T$. We pick $n = 4$ logarithmically equidistant points on the imaginary axis between $1\imath$ and $100\imath$ (indicated as solid vertical lines in Figure 7.1a) together with their complex conjugates. For the additional points framework (section 7.3.1) we set $Q_\mathscr{F} = 1$ and $Q_\mathscr{G} = 2$, such that we have two additional interpolation points (dashed vertical lines in Figure 7.1a) plus their complex conjugates. The Bode plots of the transfer functions and of the errors are illustrated for the different approaches in Figure 7.1a and Figure 7.1b, respectively. Both of our approaches capture the dynamics of the full model (the graphs are almost on top of that of the original model) and clearly outperform the Loewner realization. This is supported by the $\mathscr{H}_\infty$ errors for the different realizations presented in Table 7.1, given also for other choices of n. Clearly, the choice of the complex driving frequencies μ_i and σ_j is important and should be investigated further, but this is not within the scope of this thesis. ♠

Remark 7.27. As the previous discussion concerning (7.21) indicated, for SISO systems it is possible to permute the (potentially) large-scale system (7.18) into many small uncoupled subsystems. The decoupling of the matrix immediately reduces the computational complexity and gives additional opportunities for parallelism as well. The accuracy for solving these small subsystems depends on



(a) Bode plot of H, $\widetilde{H}_L$, $\widetilde{H}_A$, and $\widetilde{H}_H$.

(b) Bode plot of the absolute errors of $\widetilde{H}_L$, $\widetilde{H}_A$, and $\widetilde{H}_H$.

Figure 7.1 – Example 7.26 – Transfer functions of the different realizations with $n = 4$

Table 7.2 – Example 7.30 – Condition numbers of the linear system (7.18) (κ_{Add}) and the decoupled systems (7.21) (min / max(κ_{AddDec}))

n	κ_{Add}	$\max(\kappa_{\text{AddDec}})$	$\min(\kappa_{\text{AddDec}})$
4	1.665549×10^3	1.665549×10^3	1.279347×10^1
6	3.290936×10^3	2.118390×10^3	7.735186×10^0
8	8.487140×10^3	2.397554×10^3	8.171444×10^0
10	1.526123×10^4	2.957270×10^3	1.026646×10^1

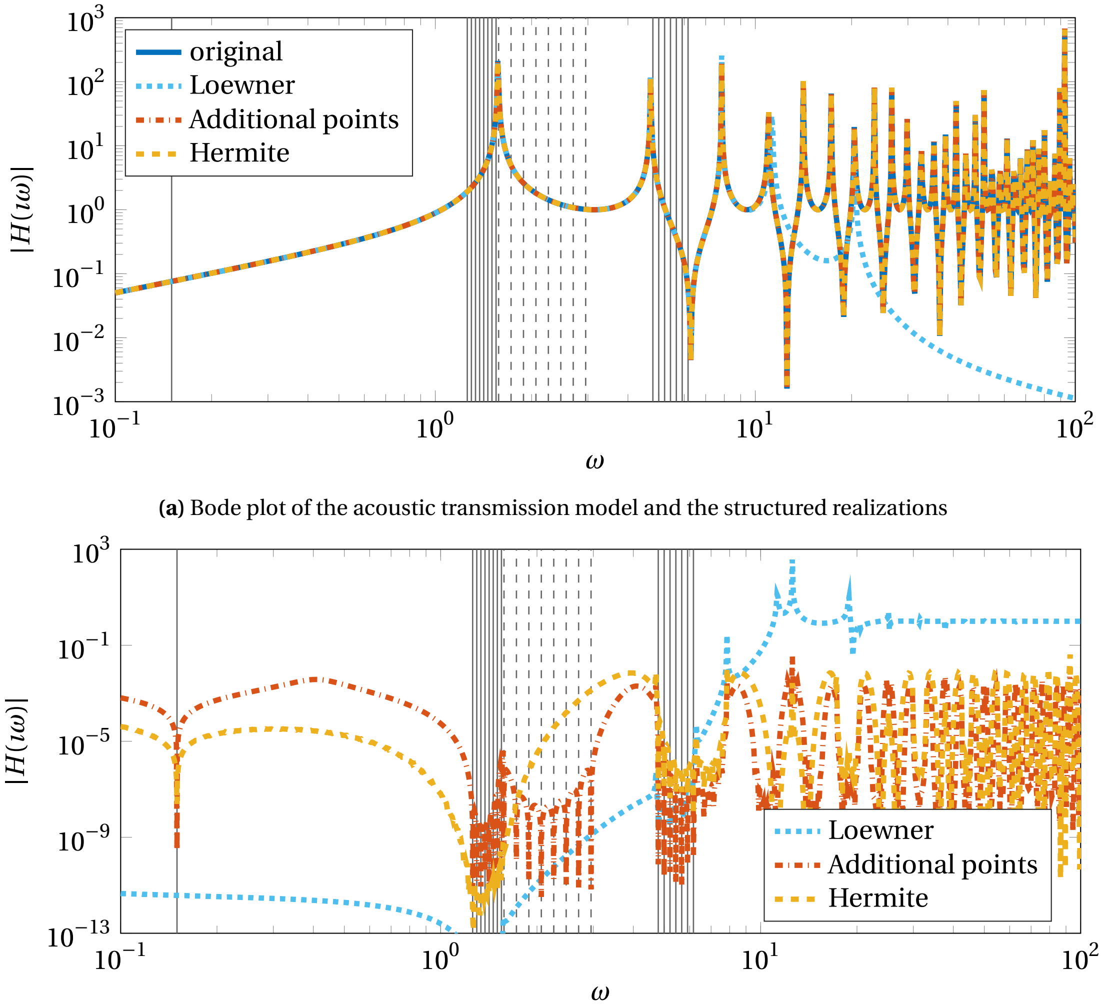
the conditioning of subsystems, which in turn will depend on the choice of interpolation points and the transfer function values (cf. (7.21)). We report in Table 7.2 the minimum and maximum norm-wise matrix condition numbers associated with subsystems produced for different reduction orders in Example 7.26. For comparison, we also include the matrix condition numbers for the aggregate systems (7.18) and note that all are of modest magnitude. Similar observations may be made for the case of Hermite interpolation. ♣

Example 7.28 (Example 6.1 continued)**.** We generate data for this model using a model for acoustic transmission in a duct presented in [63]. Based on a PDE model, the authors of [63] derive an analytic transfer function for this problem: $H(s) = \rho_0 \sinh((L - \xi_0)s/c)/\cosh(Ls/c)$, where ρ_0 is the air density. For our case, we assign parameter values $L = 1$, $\xi_0 = 1/2$, $c = 1$, and $\rho_0 = 1$ and generate data by picking $n = 16$ sampling points on the imaginary axis between $0.1\imath$ and $10\imath$ (see Figure 7.2a). To keep the realization real we add the complex conjugate driving frequencies. We seek structurally equivalent realizations to the hypothesized structure from Example 6.1 that will interpolate this data. The frequency response of the original transfer function $H(s)$ together with the different structurally equivalent realizations is presented in Figure 7.2a. The relative error plot Figure 7.2b shows that structured realizations in this case outperform the Loewner realization by several orders of magnitude. It is noteworthy that the exact transfer function can be written in accordance with the hypothesized structure using matrices $c^T = \begin{bmatrix} 0 & 0 & 0 & \rho_0 \end{bmatrix}$, $b^T = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$ and

$$A_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \end{bmatrix}.$$

♠

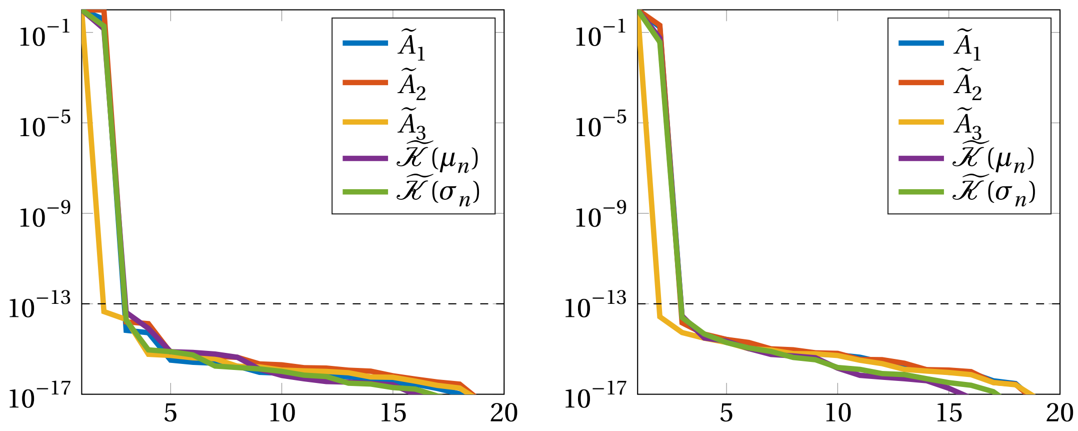
Example 7.29. We consider the transfer function $H(s) = (1 - \exp(-s))/(1 + \exp(-2s))$, which we evaluate on logarithmically spaced points between $\imath$ and $10\imath$ and their complex conjugates. We set $n = 20$ and distribute the points such that they are closed under complex conjugation. Applying Algorithm 3 to the affine structure $h_1(s) = 1$, $h_2(s) = -\exp(-s)$, and $h_3(s) = -\exp(-2s)$ yields numerical rank deficiencies, i.e., the realization suffers from redundant data (see section 7.4). To numerically decide where which data needs to be truncated, we consider the normalized singular values of the matrices, which are presented in Figure 7.3. We set the threshold to 1×10^{-13} (in



(a) Bode plot of the acoustic transmission model and the structured realizations

(b) Relative error plot for the different realizations

Figure 7.2 – Example 7.28 – Bode and relative error plot for $n = 16$



(a) Additional interpolation points

(b) Derivative data

Figure 7.3 – Example 7.29 – Decay of the (normalized) singular values

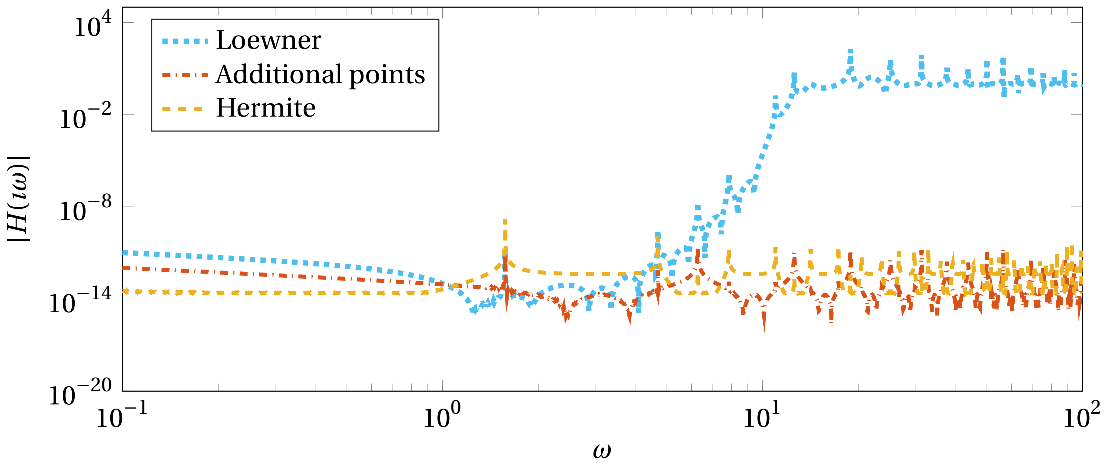


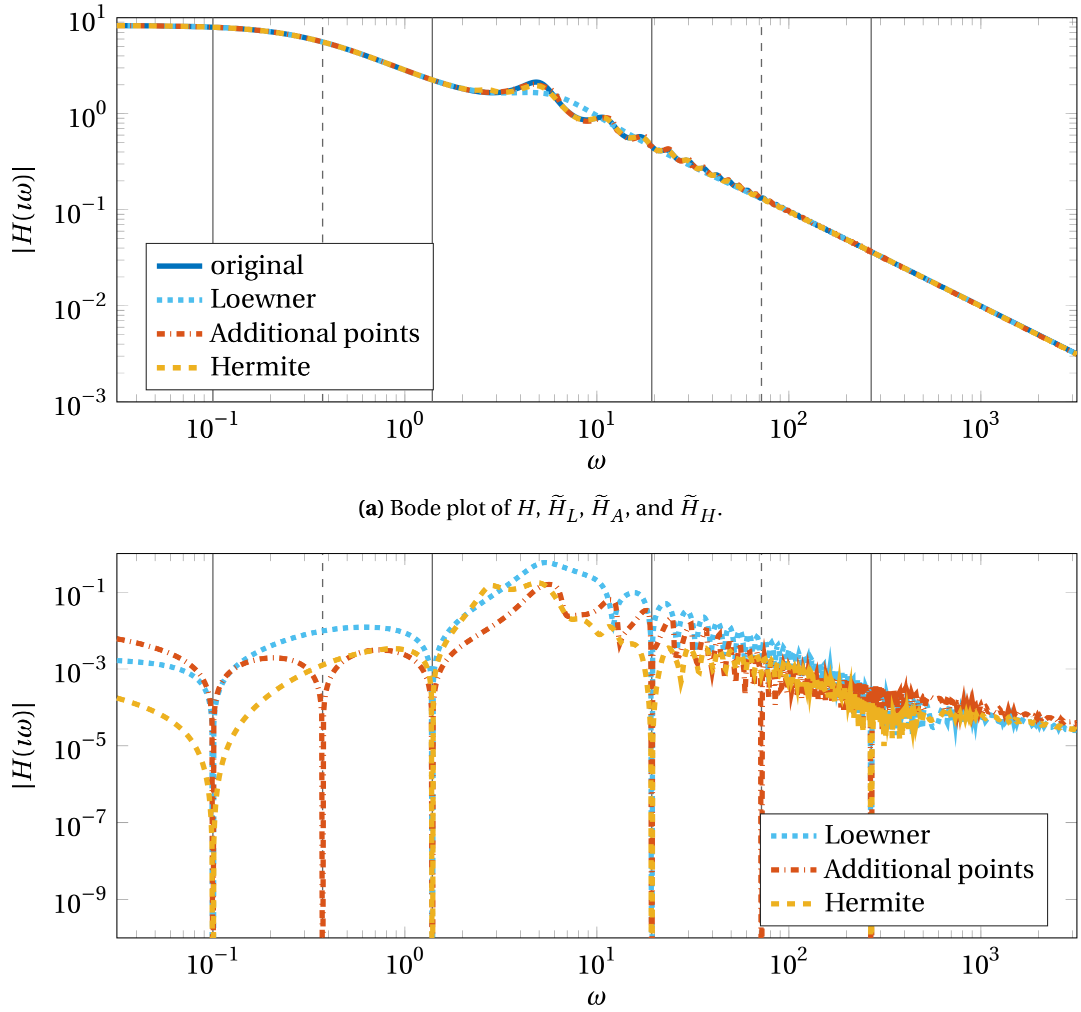
Figure 7.4 – Example 7.29 – Relative error for the different realizations

Figure 7.3 indicated with the dashed line) such that we obtain $r = 2$ for both realizations. We notice that also the Loewner realization results in numerically redundant data and is truncated to dimension $r = 14$. The relative error is displayed in Figure 7.4. ♠

Example 7.30. A heated rod with distributed control and homogeneous Dirichlet boundary conditions, which is cooled by delayed feedback, can be modeled (cf. [40, 158]) as

$$\begin{aligned} \frac{\partial v(\xi,t)}{\partial t} &= \frac{\partial^2 v(\xi,t)}{\partial \xi^2} + a_1(\xi)v(\xi,t) + a_2(\xi)v(\xi,t-1) + u(t) && \text{in } (0,\pi)\times(0,T], \\ v(0,t) &= v(\pi,t) = 0 && \text{in } [0,T]. \end{aligned} \tag{7.41}$$

For the coefficient functions we choose $a_1(\xi) = -2\sin(\xi)$ and $a_2(\xi) = 2\sin(\xi)$. Discretization of (7.41)



(a) Bode plot of H, $\widetilde{H}_L$, $\widetilde{H}_A$, and $\widetilde{H}_H$.

(b) Bode plot of the absolute errors of $\widetilde{H}_L$, $\widetilde{H}_A$, and $\widetilde{H}_H$.

Figure 7.5 – Example 7.30 – Transfer functions of the different realizations with $n = 4$

via centered finite differences with step size $h := \frac{\pi}{N+1}$ yields the system

$$\begin{aligned}\dot{x}(t) &= (L_N + A_{1,N})x(t) + A_{2,N}x(t-1) + Bu(t),\\ y(t) &= Cx(t),\end{aligned}$$

where $L_N \in \mathbb{R}^{n_x \times n_x}$ is the discrete Laplacian and $A_{1,n_x}, A_{2,n_x} \in \mathbb{R}^{n_x \times n_x}$ are discrete approximations of the functions a_1 and a_2, respectively. The input matrix $B \in \mathbb{R}^{n_x}$ is a vector of ones. As output we use the average temperature of the rod, i. e, $C = \frac{1}{\|B\|}B^T$. For our tests we use $n_x = 100$ and $n = 4$ interpolation points on the imaginary axis between $10^{-1}\imath$ and $10^{3}\imath$ together with their complex conjugates. For the realization obtained by interpolating additional data we use the same settings as in Example 7.26. Similarly as in Example 7.26, our approaches are the only ones that capture the qualitative behavior of the original system (cf. Figure 7.5). This is true for all tested numbers of interpolation data n and is further illustrated by the $\mathscr{H}_\infty$ errors listed in Table 7.3. The difference

Table 7.3 – Example 7.30 – $\mathcal{H}_\infty$ errors for the different realizations

n	Loewner	Additional points	Hermite
4	5.863023×10^{-1}	1.596379×10^{-1}	1.751535×10^{-1}
6	7.118732×10^{-1}	4.716281×10^{-1}	7.580182×10^{-2}
8	2.735014×10^{-1}	3.020142×10^{-2}	3.725486×10^{-2}
10	2.110771×10^{-1}	1.796065×10^{-1}	4.085510×10^{-2}

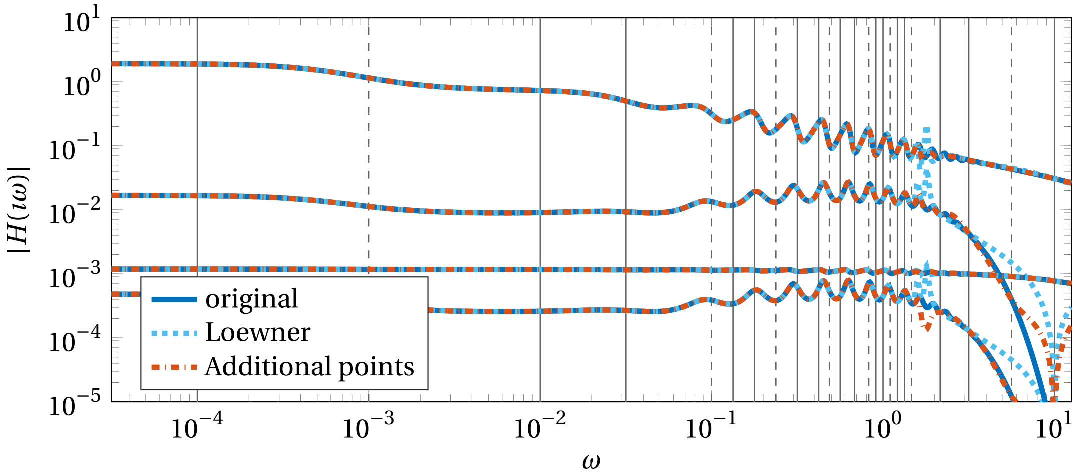


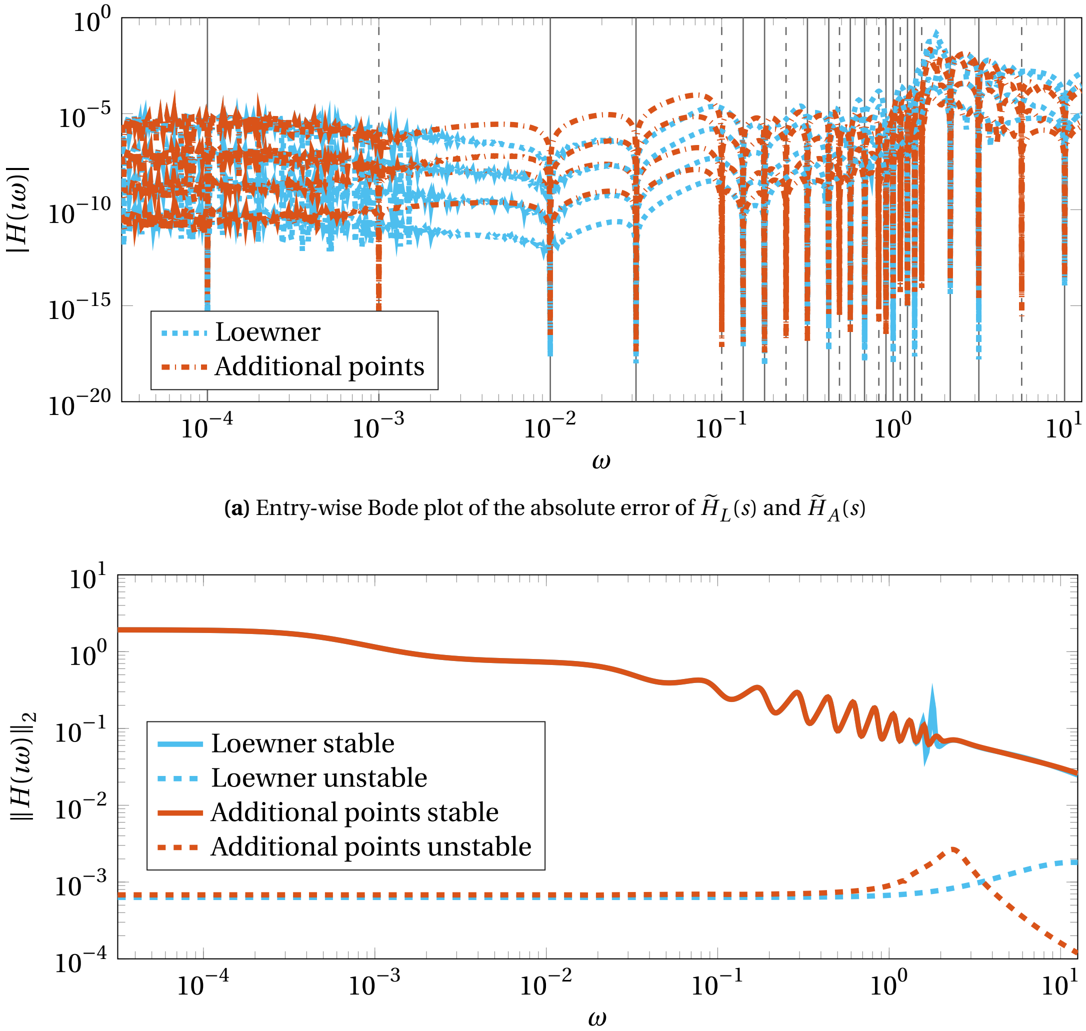
Figure 7.6 – Example 7.31 – Entry-wise Bode plot of H, $\widetilde{H}_L$ and $\widetilde{H}_A$ with $n = 32$

for this example is not as striking as in the two preceding examples, which are much harder to approximate with a rational transfer function of low degree. ♠

Example 7.31. The full model for this example comes from a finite element discretization of a cantilevered Euler-Bernoulli beam [110, § 1.16], resulting in a second order system having the form

$$A_1\ddot{x}(t) + A_2\dot{x}(t) + A_3x(t) = Bu(t), \qquad y(t) = Cx(t).$$

This is a *multiple-input/multiple-output* (MIMO) system ($n_u = 2$ and $n_y = 2$) with $n_x = 800$ internal degrees of freedom. The two input channels represent a point force applied to the state x_N and a distributed force applied to the states x_i with $i \in \{1,2,3,4\}$, i.e., $B = \begin{bmatrix} e_N & \sum_{i=1}^{4} e_i \end{bmatrix}$. The output channels are the displacement history at x_N and x_1, i.e., $C^T = \begin{bmatrix} e_N & e_1 \end{bmatrix}$. The matrices A_1 and A_3 are the mass and the stiffness matrix from the finite element discretization of the beam. As in [23], the damping matrix A_2 is modeled via light proportional damping: $A_2 = \alpha_1 A_1 + \alpha_2 A_3$ with $\alpha_1 = \alpha_2 = 0.05$. The realizations are obtained for 16 complex driving frequencies on the imaginary axis between $10^{-5}\imath$ and $10^{2}\imath$, which we use twice with unit tangential directions $(\ell_i, r_i \in \{e_1, e_2\})$ together with their complex conjugates, giving a total of $n = 32$ interpolation points (see vertical lines in Figure 7.6). The remaining matrices to compute the realization are chosen as in (7.19). Hereby, the $\star$ entries are picked such that the matrices are nonsingular. Since the transfer function of the original model is a rational transfer function, unlike in the previous examples, we expect the



(a) Entry-wise Bode plot of the absolute error of $\widetilde{H}_L(s)$ and $\widetilde{H}_A(s)$

(b) Bode plot of the stable and unstable part of $\widetilde{H}_L(s)$ and $\widetilde{H}_A(s)$

Figure 7.7 – Example 7.31 – Transfer functions of the realizations with $n = 32$.

Loewner realization to perform close to our proposed approach here. This expectation is confirmed by Figure 7.6 where the 2×2 transfer functions of the original model and of the different realizations are plotted entry-wise. The figure shows that both the Loewner realization and the structured realization with additional interpolation points capture the transfer function of the original model for a large frequency range. The accuracy at frequencies higher than one rad/sec could be improved by adding more interpolation points in this frequency region if desired. The error plot in Figure 7.7a shows that for some frequency intervals the Loewner realization is more accurate, while for higher frequencies our approach outperforms the Loewner framework: the maximum error due to $\widetilde{H}_A(s)$ is one order of magnitude smaller than the error due to $\widetilde{H}_L(s)$.

We conclude this example with a remark on the stability of the reduced models. As one expects, stability of the reduced model in the Loewner framework depends on the quality of the interpolation (sampling) points. The Loewner framework does not guarantee a stable reduced model in general. For a better selection of points (in some cases, optimal) one can, for example, combine the Loewner framework with interpolatory $\mathcal{H}_2$ optimal methods as done in [19]. For cases where the Loewner model is unstable, [86] offers various effective post-processing techniques allowing to extract a stable model while not losing much accuracy. One solution is simply to discard the unstable part of the resulting model. Indeed, this choice can be shown to be the best solution in minimizing an $\mathcal{H}_2$-related distance; see, for example, [86, 121, 143] for details. For this beam example, both the Loewner and our approach yield unstable reduced models. Following [86], we check how much the stable and unstable parts of the reduced models contribute to the approximation. For both models, the unstable part has a minor, negligible contribution as illustrated in Figure 7.7b, where the frequency response plots for the stable and unstable parts of the Loewner realization and the structured realization obtained with additional data are displayed. For this example, simply truncating the anti-stable part of the reduced models and taking only the stable part as the approximation causes only a slight loss in accuracy. Indeed, for $\widetilde{H}_A(s)$, while the $\mathcal{L}_\infty$ norm of the anti-stable part is 2.649912×10^{-3}, the $\mathcal{H}_\infty$ norm of the stable part is 1.924082. Computing the poles and the transmission zeros of $\widetilde{H}_A(s)$ shows that the unstable poles are nearly matched by the corresponding transmission zeros as listed in Table 7.4. In particular, it appears that the anti-stable part is due to the fact that the realization is close to a system that is not minimal, i.e., either not controllable or not observable.

Table 7.4 – Example 7.31 – Near pole-zero cancellation for the largest unstable poles

Poles	Zeros
$0.3851 + \imath 1.3934$	$0.3865 + \imath 1.3947$
$0.4512 + \imath 2.3432$	$0.4344 + \imath 2.3501$
$0.5251 - \imath 0.5469$	$0.5249 - \imath 0.5481$
$0.8547 - \imath 2.8547$	$0.8477 - \imath 2.8662$
$1.8070 - \imath 2.7514$	$1.8281 - \imath 2.7535$
$2.3108 + \imath 2.7899$	$2.3440 + \imath 2.7623$

Unlike the case for the Loewner framework, we cannot simply take the stable-part of $\widetilde{H}_A(s)$ as the approximant, since this truncation is performed after conversion to first-order form and destroys the structure we are seeking to retain. For many examples, including the previous ones considered here, no equivalent, generic, finite-dimensional, first-order structure exists. Therefore, one might consider modifying Algorithm 3 so that these near pole-zero cancellations can be detected during the construction and removed without destroying the structure. This is not the focus of this dissertation and is deferred to future work. ♠

8 From time-domain data to structured realizations

If we want to apply the framework from Chapter 7 in form of Algorithm 3 to a more realistic application, we need to provide

(i) the frequency response data in the form (6.5) and

(ii) the affine structure $\sum_{k=1}^{K} h_k(s)\widetilde{A}_k$.

Even if a reasonable affine structure may be known by experts (cf. Example 7.28), this structure might still depend on parameters, that are either unknown or prone to measurements errors, and thus need to be adapted to the model. Moreover, for many applications only measurements in the time domain are possible, such that a direct application of Algorithm 3 or Algorithm 4 is not possible. In this chapter, we present a possibility to resolve these two problems and illustrate the approach in the form of a case study.

8.1 Time-domain data

Instead of assuming access to frequency domain data as required by Algorithm 4, we assume that we only have access to the input and output data in the time domain. Such samples may be obtained by direct measurements or via computer simulation codes. In both cases, the surrogate model must be built from input-output measurements only. Several methods, such as the *signal generator approach* [14], are designed to use time-domain data to obtain frequency measurements. In this work, we use a modification of the *empirical transfer function estimate* (ETFE) method [137] that is presented in [168] and does not assume periodicity of the input and output sequence. Since the main tool of the method from [168] is the solution of a least-squares problem we refer to this approach as *least-squares transfer function estimate* (lsTFE).

We consider the time grid $0 = t_0 < t_1 < \ldots < t_N = t_{\mathrm{f}}$ with $t_j = j\delta_t$ and $N \in \mathbb{N}$ for a given time step size $\delta_t > 0$. Moreover, we assume that measurements of the input and the output at the time grid is available, that is we have access to the data

$$u_j := u(t_j) \in \mathbb{R}^{n_u}, \qquad \text{and} \qquad y_j := y(t_j) \in \mathbb{R}^{n_y} \qquad \text{for } j = 0, 1, \ldots, N. \tag{8.1}$$

For simplicity, we assume in the following that the data under consideration is generated from a *single-input/single-output* (SISO) dynamical system, that is $n_u = n_y = 1$. Moreover, we assume that the system is *bounded-input/bounded-output* (BIBO) stable, i.e., the sequence $(y_j)_{j\in\mathbb{N}}$ is bounded for any bounded sequence $(u_j)_{j\in\mathbb{N}}$ and we make the following crucial assumption for the remainder of this chapter.

> **Assumption 8.1.** *The data in* (8.1) *is generated from a causal, BIBO stable* linear time-invariant *(LTI) system.*

For the case $N = \infty$, Assumption 8.1 guarantees that the output data y_j is obtained via the convolution of the impulse response of the system and the inputs u_j. More precisely, there exist numbers $\mathfrak{h}_i \in \mathbb{R}$ such that

$$y_j = \sum_{i=0}^{j} \mathfrak{h}_i u_{j-i} \qquad \text{for } j \in \mathbb{N}. \tag{8.2}$$

Example 8.2. If the data (8.1) is generated from the discrete-time system

$$\begin{aligned} Ex_{j+1} &= Ax_j + Bu_j, \\ y_j &= Cx_j, \\ x_0 &= 0, \end{aligned}$$

with nonsingular matrix $E \in \mathbb{R}^{n_x \times n_x}$, then $\mathfrak{h}_i = C\left(E^{-1}A\right)^{i-1}\left(E^{-1}B\right)$ for $i > 0$ and $\mathfrak{h}_i = 0$ otherwise. $\spadesuit$

Taking the Z-transforms of $\left(u_j\right)_{j\in\mathbb{N}}$ and $\left(y_j\right)_{j\in\mathbb{N}}$

$$\widehat{u}(z) = \sum_{i=0}^{\infty} u_i z^{-i} \qquad \text{and} \qquad \widehat{y}(z) = \sum_{i=0}^{\infty} y_i z^{-i} \tag{8.3}$$

implies $\widehat{y}(z) = H(z)\widehat{u}(z)$, where H is given by the formal power series

$$H(z) = \sum_{i=0}^{\infty} \mathfrak{h}_i z^{-i}.$$

In practical applications we have $N < \infty$ and thus cannot apply the Z-transform. Instead, we use the *fast Fourier transform* (FFT), which can be interpreted as a special case of the Z-transform. More precisely, we define $q_k := \exp(\frac{2\pi\imath}{N}k)$,

$$\widehat{u}_{k;N} := \sum_{j=0}^{N-1} u_j q_k^{-j}, \qquad \text{and} \qquad \widehat{y}_{k;N} := \sum_{j=0}^{N-1} y_j q_k^{-j}$$

for $k = 0, \ldots, N-1$. Using the index set

$$\mathscr{I} := \left\{ k \in \{0, 1, \ldots, N\} \,\middle|\, |\widehat{u}_{k;N}| > 0 \right\} =: \{k_1, \ldots, k_r\},$$

we can define

$$H_{k;N} := \frac{\widehat{y}_{k;N}}{\widehat{u}_{k;N}} \qquad \text{for } k \in \mathscr{I}$$

as an approximation of the transfer function. This particular way of estimating the transfer function is known as the ETFE [137]. If the sequence $(u_j)_{j\in\mathbb{N}}$ and $(y_j)_{j\in\mathbb{N}}$ are periodic with period N, then $H_{k;N} = H(q_k)$. In practical applications, periodicity cannot always be assumed and thus we pursue a different way here. Following [168] we define the partial sum

$$H_j(z) := \sum_{i=0}^{j} \mathfrak{h}_i z^{-i}.$$

Using the inverse FFT, we observe

$$\begin{aligned} y_j &= \sum_{i=0}^{j} \mathfrak{h}_i u_{j-i} = \sum_{i=0}^{j} \mathfrak{h}_i \left(\frac{1}{N} \sum_{k=0}^{N-1} \widehat{u}_k q_k^{j-i} \right) = \frac{1}{N} \sum_{k=0}^{N-1} \widehat{u}_k H_j(q_k) q_k^j \\ &= \frac{1}{N} \sum_{k\in\mathscr{I}} \widehat{u}_k H_j(q_k) q_k^j. \end{aligned} \tag{8.4}$$

We note that (8.4) provides a direct link between the time domain data y_j and the frequency data $H_j(q_k)$. In addition, we have the following convergence result, which is a generalization of [168, Proposition 3.2].

Theorem 8.3. *Suppose the data* (8.1) *is generated from a dynamical system that satisfies Assumption 8.1. Then*

$$\lim_{j\to\infty} H_j(z) = H(z) \qquad \textit{for all } z \in \mathbb{S} := \{z \in \mathbb{C} \mid |z| = 1\}.$$

Proof. Let $z \in \mathbb{S}$. Then

$$\sum_{i=0}^{j} |\mathfrak{h}_i z^{-i}| = \sum_{i=0}^{j} |\mathfrak{h}_i||z|^{-i} = \sum_{i=0}^{j} |\mathfrak{h}_i|.$$

Assumption 8.1 implies that the system is BIBO stable, which is equivalent to the absolute convergence of the power series $\sum_{i=0}^{\infty} \mathfrak{h}_i$ [226, Chapter 2.8]. Thus the formal power series $H(z)$ converges for every $z \in \mathbb{S}$, which completes the proof. ■

Remark 8.4. If the data is generated from the linear system in Example 8.2, then the rate of convergence depends on the spectral radius of $E^{-1}A$, cf. [168, Proposition 3.2] and [6, Theorem 5.18]. More precisely, let $\rho \geq 0$ denote the spectral radius of $E^{-1}A$, i.e., the modulus of the largest eigenvalue of $E^{-1}A$. Then there exists a constant $c \in \mathbb{R}$ independent of j and ρ such that

$$|H_j(z) - H(z)| \leq c\rho^j$$

for all $z \in \mathbb{S}$. ♣

The relation (8.4) together with the convergence result given by Theorem 8.3 motivates to solve the least-squares problem

$$\operatorname*{arg\,min}_{\hat{H}_{k_i;N}} \sum_{j=j_{\min}}^{N} \left(y_j - \frac{1}{N} \sum_{i=1}^{r} \widehat{u}_{k_i} \hat{H}_{k_i;N} q_{k_i}^j \right), \tag{8.5}$$

with some number $j_{\min} \in \mathbb{N}$ that is chosen in accordance with the expected rate of convergence in Theorem 8.3 and Remark 8.4, cf. [168]. For more details on the choice of $j_{\min}$, we refer to [168, Section 3.6]. The (minimum norm) solution of (8.5) is obtained by computing the Moore–Penrose pseudo-inverse of the matrix

$$F := \frac{1}{N} \begin{bmatrix} \hat{u}_{k_1} q_{k_1}^{j_{\min}} & \dots & \hat{u}_{k_r} q_{k_r}^{j_{\min}} \\ \vdots & \ddots & \vdots \\ \hat{u}_{k_1} q_{k_1}^{N} & \dots & \hat{u}_{k_r} q_{k_r}^{N} \end{bmatrix} \in \mathbb{C}^{(N-j_{\min}+1)\times r},$$

which is given by $F^{\dagger} = \mathscr{V}\Sigma^{-1}\mathscr{U}^{*}$, where $\mathscr{U}\Sigma\mathscr{V}^{*} = F$ denotes the short *singular value decomposition* (SVD) of F. Note that inverting very small but nonzero singular values in Σ poses a numerical problem. Truncating small singular values during the computation of the pseudo-inverse amounts to solving the regularized least-squares problem (cf. [30])

$$\operatorname*{arg\,min}_{\hat{H}\in\mathbb{C}^r} \|F\hat{H} - Y\|_2^2 + \beta\|\hat{H}\|_2^2, \tag{8.6}$$

with

$$\hat{H} = \begin{bmatrix} \hat{H}_{k_1;N} & \dots & \hat{H}_{k_r;N} \end{bmatrix}^T \qquad \text{and} \qquad Y = \begin{bmatrix} y_{j_{\min}} & \dots & y_N \end{bmatrix}^T.$$

Note that the matrix F is dense and, depending on the number r of nonzero Fourier coefficients of the input signal, the numerical solution of (8.6) may become unmanageably expensive. If the user is free to choose the input signal $(u_j)_{j=0}^{N}$ then the numerical issues can be reduced as follows, see also [168]. It is likely that the numerical rank deficiency of F is avoided, if r is small, i.e., if only a small number of the Fourier coefficients of the input sequence $(u_j)_{j=0}^{N}$ is nonzero and $N - j_{\min}$ is large enough. In particular, this ensures that the least-squares problem (8.5) is overdetermined. One way to design a specific input sequence that is sparse in the Fourier domain is to prescribe a set of interpolation points q_{k_i} for $i = 1,\dots,r$ and define

$$u_j := \frac{1}{N}\sum_{i=1}^{r} q_{k_i}^{j}. \tag{8.7}$$

Then, the FFT implies

$$\hat{u}_{k;N} = \sum_{j=0}^{N-1} u_j q_k^{-j} = \frac{1}{N}\sum_{i=1}^{r}\sum_{j=0}^{N-1} \exp\left(\frac{2\pi\imath}{N}(k_i - k)j\right) = \frac{1}{N}\sum_{i=1}^{r} N\delta_{k_i,k},$$

and hence only the Fourier coefficients corresponding to the k_i are nonzero. We note that in this case we do not need to compute the FFT of u_j and F is a generalized Vandermonde matrix.

8.2 Implementation details

In Chapter 7 we have constructed our realizations for continuous-time systems, while the lsTFE approach inSection 8.1 is formulated in the discrete-time setting. To combine both results, we make the following observation: consider the system

$$\begin{aligned}\dot{x}(t) &= A_1 x(t) + Bu(t),\\ y(t) &= Cx(t),\\ x(0) &= x_0\end{aligned}$$

and the control function

$$u(t) = \frac{\delta_t}{t_\mathrm{f}} \sum_{i=1}^{r} \exp\left(2\pi \imath k_i \frac{t}{t_\mathrm{f}}\right). \tag{8.8}$$

Evaluating u at the time grid $t_j = j\delta_t$ with $j \in \{0, 1, \dots, N\}$ reveals

$$u(t_j) = \frac{\delta_t}{N\delta_t} \sum_{i=1}^{r} \exp\left(2\pi \imath k_i \frac{j\delta_t}{N\delta_t}\right) = \frac{1}{N} \sum_{i=1}^{r} \exp\left(\frac{2\pi\imath}{N} k_i j\right) = u_j,$$

i.e., the input signal in (8.8) can be understood as a continuous representation of the discrete input signal in (8.7). For $t > 0$ we have

$$\begin{aligned}y(t) &= C\int_0^t \exp(A_1(t-s))Bu(s)\mathrm{d}s\\ &= \frac{C}{N}\sum_{i=1}^{r}\left(\frac{2\pi\imath}{t_\mathrm{f}}k_i I_{n_x} - A_1\right)^{-1} \exp(A_1 t)\left(\exp\left(\left(\frac{2\pi\imath}{t_\mathrm{f}}k_i I_{n_x} - A_1\right)t\right) - I_{n_x}\right)B.\end{aligned}$$

If we assume that A_1 is asymptotically stable, then for sufficiently large t, we have $\exp(A_1 t) \approx 0$ and hence

$$y(t) \approx \frac{1}{N}\sum_{i=1}^{r} H\left(\frac{2\pi\imath}{t_\mathrm{f}}k_i\right)\exp\left(2\pi\imath k_i \frac{t}{t_\mathrm{f}}\right).$$

A comparison with (8.4) suggests that using the input signal (8.8) in combination with the procedure in Section 8.1 results in an approximation of the transfer function of the continuous-time system at the frequency $\frac{2\pi\imath}{t_\mathrm{f}}k_i$. As a consequence, we can describe frequency bounds $f_\mathrm{min}, f_\mathrm{max} > 0$ and choose $\widetilde{r}$ interpolation points $\widetilde{\lambda}_i$ in the interval $[\imath f_\mathrm{min}, \imath f_\mathrm{max}]$. For a given final time $t_\mathrm{f} = N\delta_t$ and given $i \in \{1, \dots, \widetilde{r}\}$, we can thus compute the number $k_i \in \mathbb{N}$ that minimizes

$$\left|\frac{2\pi\imath}{t_\mathrm{f}}k_i - \widetilde{\lambda}_i\right| = \min_{k\in\mathbb{N}}\left|\frac{2\pi\imath}{t_\mathrm{f}}k - \widetilde{\lambda}_i\right|. \tag{8.9}$$

The transfer function is therefore estimated at the frequencies

$$\lambda_i := \frac{2\pi\imath}{t_\mathrm{f}}k_i \qquad \text{for } i \in \{1, \dots, \widetilde{r}\}. \tag{8.10}$$

Note that for some choices of $\widetilde{\lambda}_i$, we may have $\lambda_i = \lambda_j$ for $i \neq j$. In this case we remove redundant frequencies to obtain r unique frequencies. These frequencies are related to the q_{k_i} via

$$q_{k_i} = \exp(\lambda_i \delta_t). \tag{8.11}$$

We summarize the previous discussion and the results of Section 8.1 in Algorithm 5.

In our examples, we use a logarithmic sampling of the frequency interval $[\imath f_\mathrm{min}, \imath f_\mathrm{max}]$ and pick j_min such that 75 % of the time series is used for the least-squares problem (8.5). For details about the choice of j_min we refer to [168].

Algorithm 5 Least-Squares Transfer Function Estimate

```
Input: $j_{\min}$ and desired interpolation frequencies $\widetilde{\lambda}_i$ $(i = 1,\dots,\widetilde{r})$
Output: actual frequencies $\lambda_i$ $(i = 1,\dots,r)$ together with estimates of the transfer function at these frequencies

Solve the minimization problem (8.9) for $i = 1,\dots,\widetilde{r}$
Remove redundant frequencies to obtain unique frequencies $\lambda_i$ according to (8.10) and corresponding points $q_{k_i}$, cf. (8.11), for $i = 1,\dots,r$
Construct the input signal $u_j$ according to (8.7) and obtain measurements $y_j$
Compute the Fourier coefficients of $u_j$ and assemble the matrix $F$
Solve the regularized minimization problem (8.6)
```

8.3 Estimation of parameters

Comparing Problem 6.9 with the structured realizations in Table 6.1, we observe that the coefficient functions h_k may depend on possibly unknown parameters like the time delay τ, which also need to be identified. Let us thus consider a linearly independent function family

$$h_k : \mathbb{C} \times \mathbb{P} \to \mathbb{C} \qquad \text{for } k = 1,\dots,K$$

with a compact parameter set $\mathbb{P} \subseteq \mathbb{R}^p$. We observe that for any fixed $p \in \mathbb{P}$ we can use Algorithm 3 to obtain a realization

$$\widetilde{H}(s,p) = \widetilde{C}(p)\left(\sum_{k=1}^{K} h_k(s,p)\widetilde{A}_k(p)\right)^{-1}\widetilde{B}(p), \tag{8.12}$$

which interpolates the data for this specific parameter. If further data

$$\left\{\left(\zeta_j, H\left(\zeta_j\right)\right) \in \mathbb{C}^2 \;\middle|\; j = 1,\dots,q\right\} \tag{8.13}$$

are available, in the following referred to as *test data*, we can compute the least-squares mismatch

$$\mathscr{E} : \mathbb{P} \to \mathbb{R}, \qquad p \mapsto \sum_{j=1}^{q} \left\| H(\zeta_j) - \widetilde{H}(\zeta_j, p) \right\|^2 \tag{8.14}$$

between evaluations of the transfer function (8.12) and this data. A simple strategy, as for instance proposed in [189], is to minimize (8.14) over the parameter set $\mathbb{P}$. We note that if an optimal parameter $p^\star \in \mathbb{P}$ is determined, one can add the test data (8.13) to the interpolation data (6.5) and compute a realization that also interpolates the test data via Algorithm 4.

Table 8.1 – Simulation parameters to obtain the transfer function estimates via lsTFE

description	**variable**	**interpolation data**	**test data**
final time	t_f	10000	40
time step	δ_t	5×10^{-3}	1×10^{-5}
frequency sampling interval	$[f_\mathrm{min}, f_\mathrm{max}]$	$[1\times10^{-4}, 1\times10^{0}]$	$[1\times10^{0.3}, 1\times10^{1}]$
requested number of frequency estimates	$\widetilde{r}$	10	6
actual number of frequency estimates	r	8	6

8.4 A case study

To demonstrate the results of this chapter, we revisit the delay example Example 7.26 from the previous chapter. More precisely, we consider the *delay differential-algebraic equation* (DDAE)

$$\begin{aligned} E\dot{x}(t) &= A_1x(t) + A_2x(t-\tau) + Bu(t), && \text{for } t>0,\\ y(t) &= Cx(t), && \text{for } t>0,\\ x(t) &= 0, && \text{for } t\in[-\tau,0], \end{aligned}$$

with $n_x \times n_x$ matrices

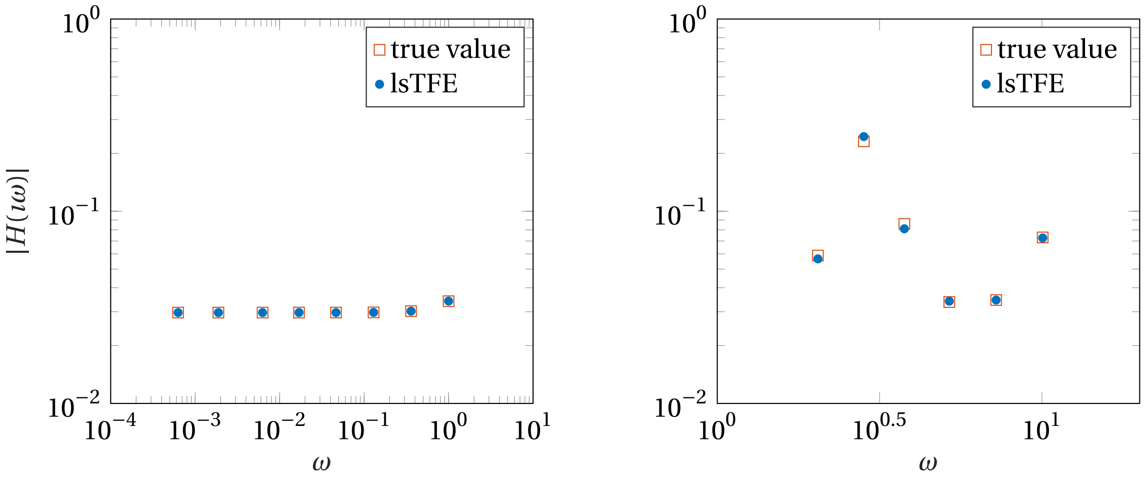
$$E := \nu I_{n_x} + T, \qquad A_1 := \frac{1}{\tau}\left(\frac{1}{\zeta}+1\right)(T-\nu I_{n_x}), \qquad A_2 := \frac{1}{\tau}\left(\frac{1}{\zeta}-1\right)(T-\nu I_{n_x}),$$

where T is an $n_x \times n_x$ matrix with ones on the sub- and superdiagonal, at the $(1,1)$, and at the (n_x, n_x) position and zeros everywhere else. Note that E is nonsingular. We choose $n_x = 12$, $\tau = 1$, $\zeta = 0.01$, and $\nu = 5$. The input matrix $B \in \mathbb{R}^{n_x}$ has ones in the first two components and zeros everywhere else, and we choose $C = B^T$.

We simulate the model twice to obtain estimates of the transfer function: once for setting up the initial model (i.e., for collecting the interpolation data) and once for obtaining additional test data, such that we can estimate the delay via minimizing the least-squares mismatch in (8.14). The simulation parameters are listed in Table 8.1. We use higher frequencies to construct the input function for the test data than for the input function for the interpolation data. These higher frequencies enforce a smaller time step δ_t. In order to have a similar computational cost for both input functions, we therefore adapted the final time t_f. Consequently, Theorem 8.3 suggests that we can expect a better accuracy for the transfer function estimates obtained from the simulation used for the interpolation data.

The resulting transfer function estimates are compared to the true values in Tables 8.2 and 8.3 and visualized in Figure 8.1. In this section, all numerical values are rounded to two decimal places.

The approximation of the transfer function at the lower frequencies (cf. Figure 8.1a) is almost matching the true values. Indeed, the maximum error between the estimates and the true values of the transfer function in the interpolation data set is 9.86×10^{-5}. The approximation for the higher frequencies (cf. Figure 8.1b and Table 8.3) is – as expected – significantly worse. However, even in this frequency range, the approximation is reasonable with a maximum error of 1.47×10^{-2}.



(a) interpolation data

(b) test data

Figure 8.1 – Estimation of the transfer function with lsTFE. The estimates are plotted with blue dots and the true values of the transfer function with red squares.

Table 8.2 – Interpolation data: estimates of the transfer function via lsTFE

frequency ω	**true value** $H(\imath\omega)$	**lsTFE estimate** $\hat{H}_{k_1;N}$	**error**	**norm error**
6.28×10^{-4}	$2.97 \times 10^{-2} + \imath 9.05 \times 10^{-6}$	$2.97 \times 10^{-2} + \imath 9.00 \times 10^{-6}$	2.90×10^{-11} - $\imath 4.71 \times 10^{-8}$	4.71×10^{-8}
1.88×10^{-3}	$2.97 \times 10^{-2} + \imath 2.71 \times 10^{-5}$	$2.97 \times 10^{-2} + \imath 2.70 \times 10^{-5}$	2.61×10^{-10} - $\imath 1.41 \times 10^{-7}$	1.41×10^{-7}
6.28×10^{-3}	$2.97 \times 10^{-2} + \imath 9.05 \times 10^{-5}$	$2.97 \times 10^{-2} + \imath 9.00 \times 10^{-5}$	2.90×10^{-9} - $\imath 4.71 \times 10^{-7}$	4.71×10^{-7}
1.70×10^{-2}	$2.97 \times 10^{-2} + \imath 2.44 \times 10^{-4}$	$2.97 \times 10^{-2} + \imath 2.43 \times 10^{-4}$	2.11×10^{-8} - $\imath 1.27 \times 10^{-6}$	1.27×10^{-6}
4.65×10^{-2}	$2.97 \times 10^{-2} + \imath 6.70 \times 10^{-4}$	$2.97 \times 10^{-2} + \imath 6.66 \times 10^{-4}$	1.59×10^{-7} - $\imath 3.49 \times 10^{-6}$	3.49×10^{-6}
1.29×10^{-1}	$2.97 \times 10^{-2} + \imath 1.87 \times 10^{-3}$	$2.97 \times 10^{-2} + \imath 1.86 \times 10^{-3}$	1.23×10^{-6} - $\imath 9.67 \times 10^{-6}$	9.75×10^{-6}
3.59×10^{-1}	$2.98 \times 10^{-2} + \imath 5.24 \times 10^{-3}$	$2.98 \times 10^{-2} + \imath 5.21 \times 10^{-3}$	9.62×10^{-6} - $\imath 2.62 \times 10^{-5}$	2.79×10^{-5}
1.00×10^{0}	$3.01 \times 10^{-2} + \imath 1.58 \times 10^{-2}$	$3.02 \times 10^{-2} + \imath 1.58 \times 10^{-2}$	8.19×10^{-5} - $\imath 5.50 \times 10^{-5}$	9.86×10^{-5}

Table 8.3 – Test data: estimates of the transfer function via lsTFE

frequency ω	**true value** $H(\imath\omega)$	**lsTFE estimate** $\hat{H}_{k_1;N}$	**error**	**norm error**
2.04×10^{0}	$3.26 \times 10^{-2} + \imath 4.89 \times 10^{-2}$	$3.11 \times 10^{-2} + \imath 4.72 \times 10^{-2}$	-1.57×10^{-3} - $\imath 1.71 \times 10^{-3}$	2.32×10^{-3}
2.83×10^{0}	$6.16 \times 10^{-2} + \imath 2.23 \times 10^{-1}$	$5.93 \times 10^{-2} + \imath 2.37 \times 10^{-1}$	$-2.35 \times 10^{-3} + \imath 1.45 \times 10^{-2}$	1.47×10^{-2}
3.77×10^{0}	2.60×10^{-2} - $\imath 8.19 \times 10^{-2}$	2.06×10^{-2} - $\imath 7.84 \times 10^{-2}$	$-5.43 \times 10^{-3} + \imath 3.54 \times 10^{-3}$	6.48×10^{-3}
5.18×10^{0}	2.79×10^{-2} - $\imath 1.89 \times 10^{-2}$	2.89×10^{-2} - $\imath 1.80 \times 10^{-2}$	$1.01 \times 10^{-3} + \imath 8.97 \times 10^{-4}$	1.35×10^{-3}
7.23×10^{0}	$3.19 \times 10^{-2} + \imath 1.31 \times 10^{-2}$	$3.23 \times 10^{-2} + \imath 1.21 \times 10^{-2}$	3.81×10^{-4} - $\imath 1.00 \times 10^{-3}$	1.07×10^{-3}
1.01×10^{1}	1.88×10^{-2} - $\imath 7.06 \times 10^{-2}$	1.84×10^{-2} - $\imath 7.02 \times 10^{-2}$	$-4.41 \times 10^{-4} + \imath 3.77 \times 10^{-4}$	5.80×10^{-4}

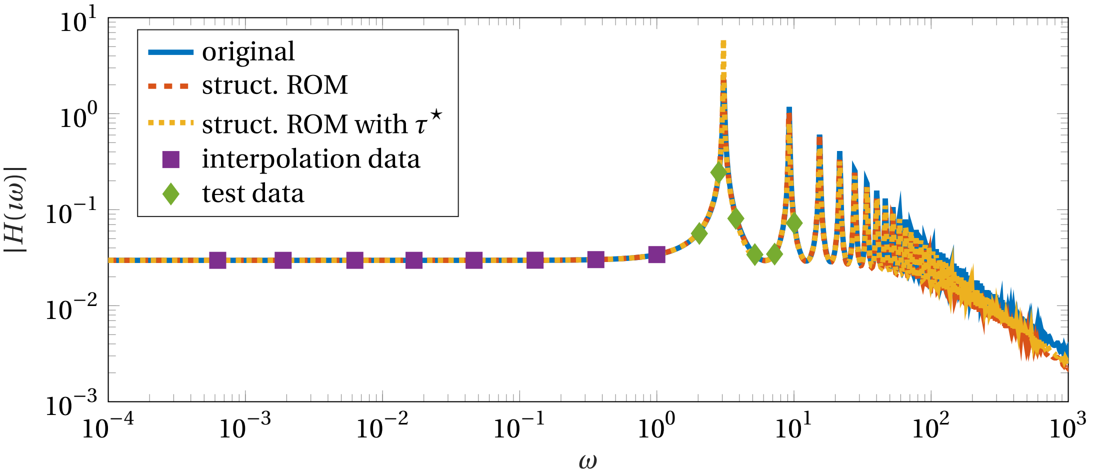


Figure 8.2 – The transfer function of the true model (solid blue line), the realization (dashed red line), and the realization with estimated parameter $\tau^\star$ (dotted yellow line) obtained from the estimated interpolation data (orange squares) and test data (green diamonds).

Before we can apply Algorithm 4, we need to specify the structure via defining the function family $\{h_1,\ldots,h_K\}$. To this end, we first use the actual structure of the original model, i.e., the functions $h_{k\,k=1}^{\,K}$ are given by

$$h_1(s) = s, \qquad h_2(s) \equiv -1, \qquad \text{and} \qquad h_3(s) = -\mathrm{e}^{-s}.$$

Note that the choice for h_3 includes the true value for the delay time $\tau = 1$. To obtain a real realization, we add the complex conjugate data to the estimated transfer function values given in Table 8.2 and choose $Q_{\mathcal{F}} = 2$ and $Q_{\mathcal{G}} = 1$. The transfer function of the obtained realization is depicted as the red dashed line in Figure 8.2. Although we only use approximations of the transfer function, the realization approximates the original model (blue solid line in Figure 8.2) well, even for frequencies larger than the frequencies used to construct the realization.

In a real application, we usually cannot compare the transfer function of the constructed realization with the true transfer function, since we do not know the true model and hence also do not know the transfer function. Instead, it is more reasonable to compare the realization with the true model via simulations in the time domain (see also Problem 6.2). As validation input functions we use

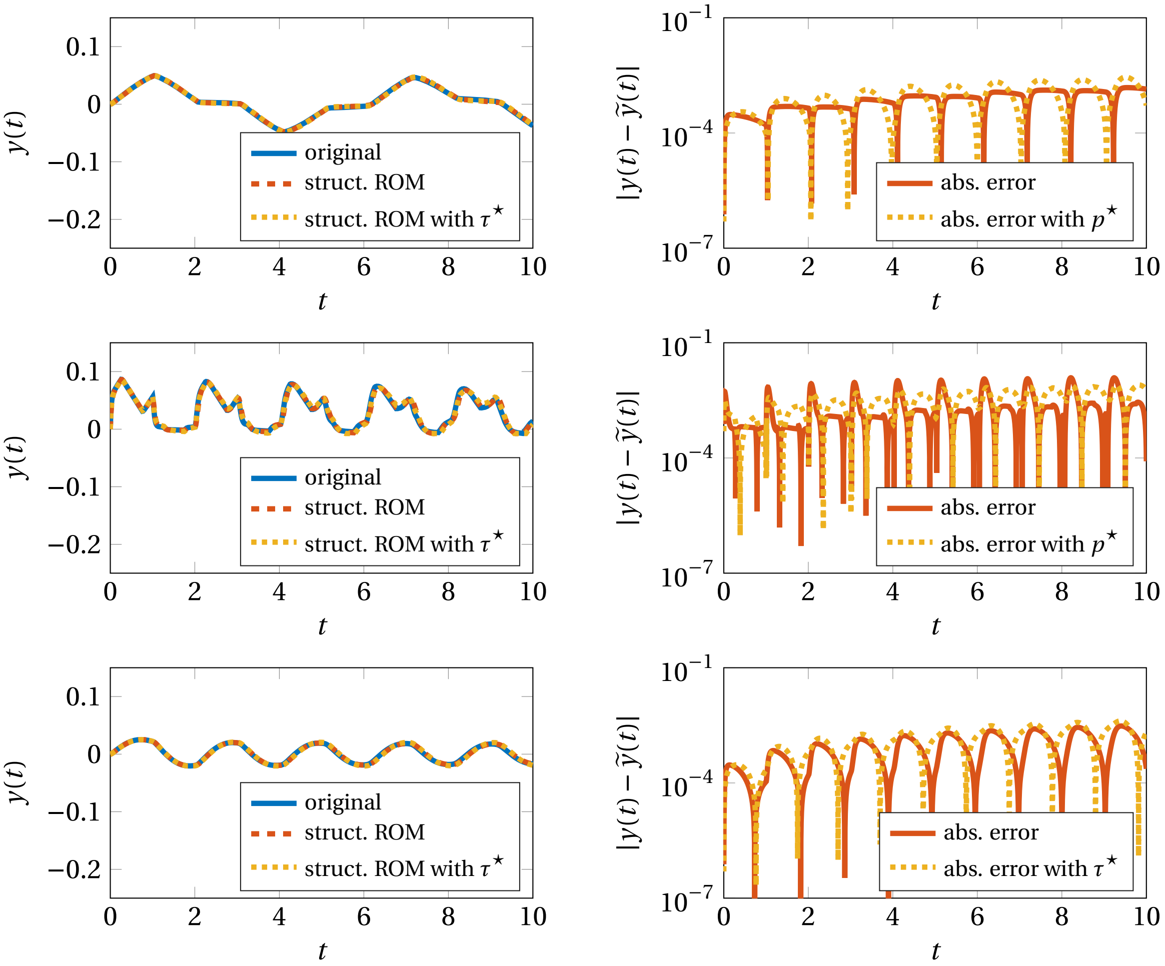
$$u_1(t) = \sin(t), \qquad u_2(t) = 2\left(t - \tfrac{1}{2}\lfloor 2t + \tfrac{1}{2}\rfloor\right)\cdot(-1)^{\lfloor 2t+\frac{1}{2}\rfloor} + 1, \qquad u_3(t) = t\exp(-t^2).$$

The results are presented in Figure 8.3 and Table 8.4. The relative errors given in Table 8.4 indicate slight differences between the accuracies obtained for the three different input signals. Nevertheless, the output trajectories of the realization agree very well with the ones of the original model for all three inputs, as illustrated in Figure 8.3.

As already noted, all previous results have been obtained by exploiting the knowledge of the actual time delay which is equal to one in this case. However, in practical applications we cannot expect to have precise a priori knowledge of the time delay, but rather a rough estimate of it. Thus, in order to

Table 8.4 – Error measurements for the validation inputs with known delay

input signal	$\|u\|_{\mathscr{L}_2}$	$\frac{\|y-\tilde{y}\|_{\mathscr{L}_\infty}}{\|u\|_{\mathscr{L}_2}}$	$\frac{\|y-\tilde{y}\|_{\mathscr{L}_2}}{\|u\|_{\mathscr{L}_2}}$
u_1	2.18×10^{0}	6.96×10^{-4}	1.26×10^{-3}
u_2	3.29×10^{0}	3.73×10^{-3}	3.51×10^{-3}
u_3	3.96×10^{-1}	7.69×10^{-3}	1.01×10^{-2}



(a) time domain simulation

(b) absolute error

Figure 8.3 – Comparison of the output of the original model (blue solid line) with the output of the approximation (red dashed line). Top: input function u_1; middle: input function u_2; bottom: input function u_3.

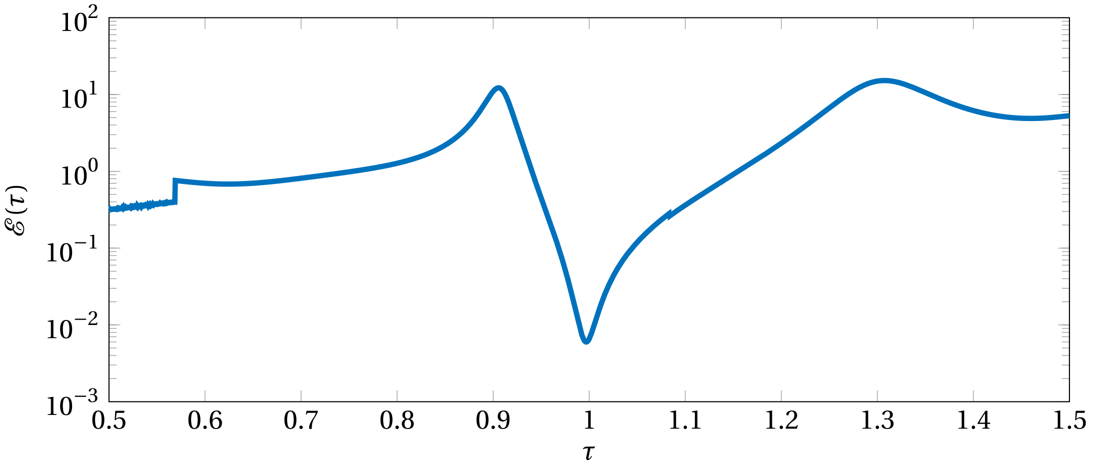


Figure 8.4 – Sampling of the least-squares error (8.14) over the delay time τ.

Table 8.5 – Error measurements for the validation inputs based on the estimated delay

input signal	$\|u\|_{\mathscr{L}_2}$	$\frac{\|y-\tilde{y}\|_{\mathscr{L}_\infty}}{\|u\|_{\mathscr{L}_2}}$	$\frac{\|y-\tilde{y}\|_{\mathscr{L}_2}}{\|u\|_{\mathscr{L}_2}}$
u_1	2.18×10^{0}	1.31×10^{-3}	1.77×10^{-3}
u_2	3.29×10^{0}	2.43×10^{-3}	3.79×10^{-3}
u_3	3.96×10^{-1}	1.43×10^{-2}	1.04×10^{-2}

build the realization from data only, we modify the function h_3 as

$$h_3(s,\tau) = -\mathrm{e}^{-\tau s}$$

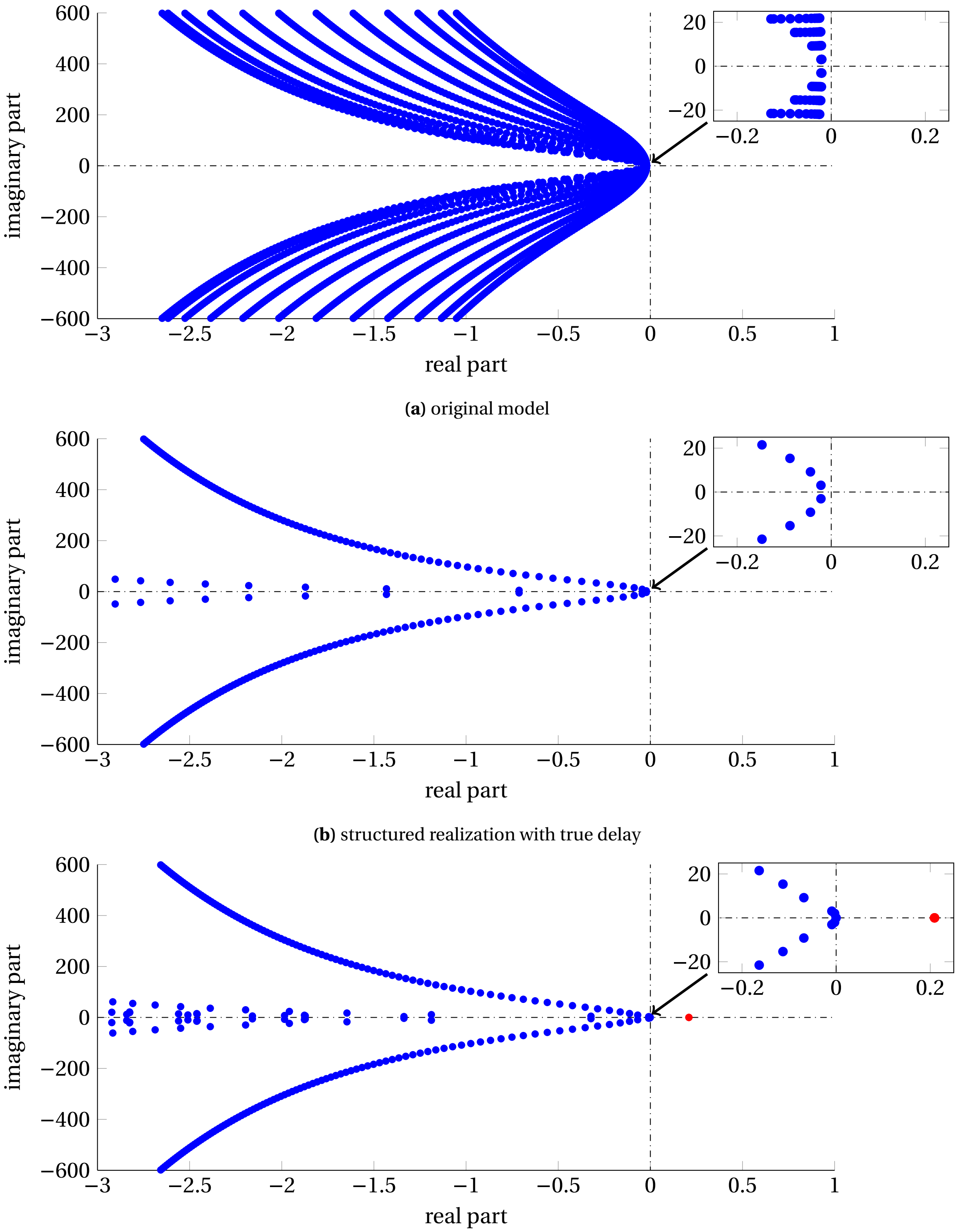
with free parameter τ. As discussed in Section 8.3, we can then use the test data to find an optimal delay time $\tau^\star$. A sampling of the least-squares error (8.14) is provided in Figure 8.4 and reveals a distinct minimum close to the actual time delay $\tau = 1$.

The minimization of the cost function (8.14) is performed using the MATLAB function `fmincon`. As start value we use $\tau = 0.98$, which is obtained from a rough sampling of the cost function. The minimizer determined via `fmincon` is $\tau^\star = 0.996883$ and for this time delay the cost function attains a value of $\mathscr{E}(\tau^\star) = 5.99 \times 10^{-3}$. To simplify the numerical simulations, we use the rounded value $\tau^\star = 0.997$ for the following results. The transfer function of the realization constructed with the estimated parameter $\tau^\star$ and all transfer function estimates (i.e., the interpolation data and the test data), are depicted in Figure 8.2. It is slightly different from the realization with the true parameter, but still approximates the original model well. This statement can be verified by the simulation results with the test inputs u_1, u_2, and u_3, which are presented in Figure 8.3 and Table 8.5.

It is worth to note that there is no guarantee that the realization obtained from Algorithm 3 is stable. In order to investigate the stability of the obtained realizations, we consider the eigenvalues depicted in Figure 8.5, which are computed using the algorithm from [221]. Indeed, the realization obtained from all transfer function estimates and the estimated delay $\tau^\star$ is unstable with one eigenvalue in the right half plane (cf. Figure 8.5c). This is not surprising, since the eigenvalues of the original model (cf. Figure 8.5a) are close to the imaginary axis, such that one can expect that a small perturbation results in an unstable model. Still, the realization constructed only from the interpolation data and with the true value for the delay τ is stable (see Figure 8.5b for the eigenvalues with the largest real part). In contrast to stabilizing post-processing algorithms for rational realizations as offered in [86], a stable–unstable decomposition of a DDAE is not possible in general and thus stability must be enforced during the construction of the realization. This is currently under investigation and subject to further research.

We conclude this case study with a remark about the choice of the interpolation frequencies.

Remark 8.5. In our numerical simulations, we observe that including estimates of the transfer function at smaller frequencies tends to produce less unstable realizations in the sense that fewer eigenvalues are unstable and the real part of the unstable eigenvalues is smaller compared to a realization obtained from estimates of the transfer function at higher frequencies. As an example we refer to the realization obtained only from the interpolation data (with the true delay $\tau = 1$) and the realization obtained from all transfer function estimates (with the estimated delay $\tau = \tau^*$), see Figure 8.5 for the corresponding eigenvalue plots. ♣



(a) original model

(b) structured realization with true delay

(c) structured realization with estimated delay

Figure 8.5 – Eigenvalues of the realization with largest real part

9 Summary and outlook

In the introduction of this thesis we have emphasized the importance of *delay differential-algebraic equations* (DDAEs) in various applications with several examples. They appear, for instance, in chemical reactions, earthquake engineering, feedback control, human blood flow, the realization of transport problems, and the analysis and construction of numerical time-integration methods. Therefore, a rigorous mathematical understanding of DDAEs is essential.

Within the first part of this thesis, we have analyzed the existence and uniqueness of solutions for *initial trajectory problems* (ITPs). Even for *linear time-invariant* (LTI) DDAEs, a distributional solution concept is required. We present a complete analysis in chapter 3. One of the main results is a modified compress-and-shift algorithm. This algorithm constructs a delay-equivalent DDAE that is amenable for the method of steps whenever the DDAE is delay-regular.To establish continuous solutions, we consider the propagation of so-called primary discontinuities in chapter 4. The analysis illustrates two possibilities: Either, the class of DDAEs has to be restricted with a complete characterization given in section 4.1, or the initial trajectories have to satisfy so-called splicing conditions, see section 4.2. Some of these results can be extended to nonlinear DDAEs (cf. chapter 5), which allowed us to establish new existence and uniqueness results for a large class of DDAEs not available yet in the literature.

In the second part of the thesis, we have studied the realization theory for DDAEs. A realization is a dynamical system constructed solely from data and able to mimic the system behavior, not only for the trained control input but also for unknown control functions. In contrast to machine learning, typically, only a small amount of data is available. The data may be available as a trajectory of a measured quantity of the system (the so-called output of the system) in the time domain or measurements of a transfer function in the frequency domain. The construction of our realization in chapter 7 is based on an algebraic characterization of interpolation conditions in the frequency domain. The degrees of freedom that arise if these conditions are applied to DDAEs can be exploited to interpolate additional data while preserving the system matrices' dimension or by satisfying some Hermite interpolation conditions. Our framework extends to further system structures, including second-order systems, fractional systems, and viscoelastic systems. We detail the relation of our realization to the projection-based interpolation of dynamical systems and propose an additional post-processing step to remove data-redundancies. The framework can also be used with time-domain data by extracting frequency information from the given time sequence. The details are presented in Chapter 8.

Although DDAEs appear in various applications, their mathematical understanding is still far from complete with many open problems. Directly linked to this thesis is an extension of the existence and uniqueness results to time-varying and multiple delays that are not commensurate. The solution theory may benefit from a behavior-like approach, which can be particularly useful if the initial trajectory problem is replaced with a boundary trajectory problem. From a modeling perspective, it is essential to know how uncertainties in the system matrices, typically modeled as an additive or multiplicative perturbation, affect the system's properties. A standard question to ask is to find the smallest perturbation in some norm such that a given DDAE is not delay-regular. The norm of the smallest perturbation is referred to as the distance to singularity, or if only a certain class of perturbations is allowed, the structured distance to singularity. One particularly useful structure is the so-called port-Hamiltonian formulation and an extension to DDAEs is a promising research direction. Given the realization of a DDAE from data, a similar research direction is a detailed error analysis if the data is subject to measurement errors. In this case, it may be preferable to use a least-squares approach instead of interpolation. On top, further post-processing steps that ensure stability or passivity of the realization need to be developed.

Bibliography

[1] I. Ahrens and B. Unger. The Pantelides algorithm for delay differential-algebraic equations. *Trans. Math. Appl.*, 4(1):1–36, 2020.

[2] H. Ali and D. Dasgupta. Effects of Time Delays in the Electric Power Grid. In J. Utts and S. Shenoi, editors, *Critical Infrastructure Protection VI*, pages 139–154. Springer Berlin/Heidelberg, 2012.

[3] N. Aliyev, P. Benner, E. Mengi, P. Schwerdtner, and M. Voigt. Large-scale computation of $\mathscr{L}_\infty$-norms by a greedy subspace method. *SIAM J. Matrix Anal. Appl.*, 38(4):1496–1516, 2017.

[4] R. Altmann, R. Maier, and B. Unger. Semi-explicit discretization schemes for weakly-coupled elliptic-parabolic problems. *Math. Comp.*, 2020. In press.

[5] R. Altmann, R. Maier, and B. Unger. A semi-explicit integration scheme for weakly-coupled poroelasticity with nonlinear permeability. 2020. Submitted for publication.

[6] A.C. Antoulas. *Approximation of large-scale dynamical systems.* Advances in Design and Control. SIAM, Philadelphia, PA, 2005.

[7] A.C. Antoulas, C.A. Beattie, and S. Gugercin. Interpolatory model reduction of large-scale dynamical systems. In J. Mohammadpour and K. M. Grigoriadis, editors, *Efficient Modeling and Control of Large-Scale Systems*, pages 3–58. Springer, New York, NY, USA, 2010.

[8] A.C. Antoulas, C.A. Beattie, and S. Güğercin. *Interpolatory Methods for Model Reduction.* SIAM, Philadelphia, PA, 2020.

[9] A.C. Antoulas, I.V. Gosea, and A.C. Ionita. Model Reduction of Bilinear Systems in the Loewner Framework. *SIAM J. Sci. Comput.*, 38(5):B889–B916, 2016.

[10] A.C. Antoulas, A.C. Ionita, and S. Lefteriu. On two-variable rational interpolation. *Linear Algebra Appl.*, 436:2889–2915, 2012.

[11] A.C. Antoulas, S. Lefteriu, and A.C. Ionita. Chapter 8: A tutorial introduction to the Loewner framework for model reduction. In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors, *Model Reduction and Approximation*, pages 335–376. SIAM, 2017.

[12] M. Arnold. DAE aspects of multibody system dynamics. In A. Ilchmann and T. Reis, editors, *Surveys in Differential-Algebraic Equations IV*, pages 41–106. Springer International Publishing, Cham, 2017.

[13] U.M. Ascher and L.R. Petzold. The numerical solution of delay-differential-algebraic equations of retarded and neutral type. *SIAM J. Numer. Anal.*, 32(5):1635–1657, 1995.

[14] A. Astolfi. Model reduction by moment matching for linear and nonlinear systems. *IEEE Trans. Autom. Control*, 55(10):2321–2336, 2010.

[15] Z.Z. Bai and X. Yang. On convergence conditions of waveform relaxation methods for linear differential-algebraic equations. *J. Comput. Appl. Math.*, 235(8):2790–2804, 2011.

[16] C.T.H. Baker, C.A.H. Paul, and H. Tian. Differential algebraic equations with after-effect. *J. Comput. Appl. Math.*, 140(1-2):63–80, 2002.

[17] U. Baur, P. Benner, and L. Feng. Model Order Reduction for Linear and Nonlinear Systems: A System-Theoretic Perspective. *Arch. Comput. Methods Eng.*, 21(4):331–358, 2014.

[18] C. Beattie and S. Gugercin. Interpolatory projection methods for structure-preserving model reduction. *Systems Control Lett.*, 58(3):225–232, 2009.

[19] C. Beattie and S. Gugercin. Realization-independent $\mathscr{H}_2$-approximation. In *Proc. 51st IEEE Conf. Decision Control (CDC)*, pages 4953–4958, Maui, HI, USA, 2012.

[20] C. Beattie and S. Gugercin. Chapter 7: Model reduction by rational interpolation. In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors, *Model Reduction and Approximation*, pages 297–334. SIAM, 2017.

[21] C. Beattie, V. Mehrmann, H. Xu, and H. Zwart. Port-Hamiltonian descriptor systems. *Math. Control. Signals Syst.*, 30(17):1–27, 2018.

[22] C.A. Beattie and P. Benner. $\mathscr{H}_2$-optimality conditions for structured dynamical systems. Preprint MPIMD/14-18, Max Planck Institute Magdeburg, Germany, 2014.

[23] C.A. Beattie and S. Gugercin. Krylov-based model reduction of second-order systems with proportional damping. In *Proc. 44th IEEE Conf. Decision Control (CDC)*, pages 2278–2283, Seville, Spain, 2005.

[24] A. Bellen, N. Guglielmi, and M. Zennaro. On the contractivity and asymptotic stability of systems of delay differential equations of neutral type. *BIT Numer. Math.*, 39(1):1–24, 1999.

[25] A. Bellen, N. Guglielmi, and M. Zennaro. Numerical stability of nonlinear delay differential equations of neutral type. *J. Comput. Appl. Math.*, 125(1-2):251–263, 2000.

[26] A. BELLEN AND M. ZENNARO. *Numerical methods for delay differential equations.* Oxford University Press, 2003.

[27] R. BELLMAN AND K. COOKE. *Differential-difference equations.* Academic Press, New York, 1963.

[28] P. BENNER, A. COHEN, M. OHLBERGER, AND K. WILLCOX. *Model Reduction and Approximation.* SIAM, Philadelphia, PA, Philadelphia, PA, 2017.

[29] P. BENNER, S. GUGERCIN, AND K. WILLCOX. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review*, 57(4):483–531, 2015.

[30] P. BENNER, C. HIMPE, AND T. MITCHELL. On reduced input-output dynamic mode decomposition. *Adv. Comput. Math.*, pages 1–18, 2018.

[31] T. BERGER, A. ILCHMANN, AND S. TRENN. The quasi-Weierstraß form for regular matrix pencils. *Linear Algebra Appl.*, 436(10):4052–4069, 2012.

[32] T. BERGER AND S. TRENN. The quasi-Kronecker form for matrix pencils. *SIAM J. Matrix Anal. & Appl.*, 33(2):336–368, 2012.

[33] J.T. BETTS, S.L. CAMPBELL, AND K. THOMPSON. Lobatto IIIA methods, direct transcription, and DAEs with delays. *Numer. Algorithms*, 69(2):291–300, 2015.

[34] A. BINDER, V. MEHRMANN, A. MIEDLAR, AND P. SCHULZE. A Matlab toolbox for the regularization of descriptor systems arising from generalized realization procedures. Preprint 24-2015, Institut für Mathematik, TU Berlin, Germany, 2016.

[35] F. BLACK, P. SCHULZE, AND B. UNGER. Projection-Based Model Reduction with Dynamically Transformed Modes. *ESAIM: Math. Model. Numer. Anal.*, 54(6):2011–2043, 2020.

[36] G. BOCHAROV AND K. P. HADELER. Structured population models, conservation laws, and delay equations. *J. Differ. Equ.*, 168(1):212–237, 2000.

[37] G.A. BOCHAROV AND F.A RIHAN. Numerical modelling in biosciences using delay differential equations. *Journal of Computational and Applied Mathematics*, 125(1-2):183–199, 2000.

[38] R. BORSCHE, D. KOCOGLU, AND S. TRENN. A distributional solution framework for linear hyperbolic PDEs coupled to switched DAEs. Technical report, Bernoulli Institute for Mathematics, CS and AI, University of Groningen, 2019.

[39] R.K. BRAYTON. Nonlinear oscillations in a distributed network. *Q. Appl. Math.*, 24(4):289–301, 1967.

[40] D. BREDA, S. MASET, AND R. VERMIGLIO. Numerical approximation of characteristic values of partial retarded functional differential equations. *Numer. Math.*, 113(2):181–242, 2009.

[41] D. BREDA, S. MASET, AND R. VERMIGLIO. *Stability of Linear Delay Differential Equations.* Springer New York, 2015.

[42] K.E Brenan, S.L. Campbell, and L.R. Petzold. *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations.* SIAM, 1996.

[43] A. Bunse-Gerstner, R. Byers, V. Mehrmann, and N.K. Nichols. Numerical computation of an analytic singular value decomposition of a matrix valued function. *Numer. Math.*, 60(1):1–39, 1991.

[44] O.S. Bursi. Computational techniques for simulation of monolithic and heterogeneous structural dynamic systems. In O.S. Bursi and D. Wagg, editors, *Modern Testing Techniques for Structural Systems: Dynamics and Control*, pages 1–96. Springer Vienna, 2008.

[45] O.S. Bursi and D. Wagg, editors. *Modern Testing Techniques for Structural Systems.* Springer, Vienna, 2008.

[46] R. Byers, P. Kunkel, and V. Mehrmann. Regularization of linear descriptor systems with variable coefficients. *SIAM Journal on Control and Optimization*, 35(1):117–133, 1997.

[47] R. Byers, V. Mehrmann, and H. Xu. Trimmed linearizations for structured matrix polynomials. *Linear Algebra Appl.*, 429(10):2373–2400, 2008.

[48] N. Cagniart, Y. Maday, and B. Stamm. Model order reduction for problems with large convection effects. In B.N. Chetverushkin, W. Fitzgibbon, Y.A. Kuznetsov, P. Neittaanmäki, J. Periaux, and O. Pironneau, editors, *Contributions to Partial Differential Equations and Applications*, pages 131–150. Springer Cham, 2019.

[49] S. Cambell, A. Ilchmann, V. Mehrmann, and T. Reis, editors. *Applications of Differential-Algebraic Equations: Examples and Benchmarks.* Springer International Publishing, 2019.

[50] S.L. Campbell. Singular linear systems of differential equations with delays. *Appl. Anal.*, 11(2):129–136, 1980.

[51] S.L. Campbell. A general form for solvable linear time varying singular systems of differential equations. *SIAM J. Math. Anal.*, 18(4):1101–1115, 1987.

[52] S.L. Campbell. 2-D (differential-delay) implicit systems. In *13th IMACS World Congr. Comput. Appl. Math.*, pages 1828–1829, 1991.

[53] S.L. Campbell. Nonregular 2D descriptor delay systems. *IMA J. Math. Control Inf.*, 12(1):57–67, 1995.

[54] S.L. Campbell and C.W. Gear. The index of general nonlinear DAEs. *Numer. Math.*, 72(2):173–196, 1995.

[55] S.L. Campbell, P. Kunkel, and V. Mehrmann. Regularization of linear and nonlinear descriptor systems. In L.T. Biegler, S.L. Campbell, and V. Mehrmann, editors, *Control Optim. with Differ. Constraints*, chapter 2. SIAM, Philadelphia, PA, 2012.

[56] S.L. Campbell and V.H. Linh. Stability criteria for differential-algebraic equations with multiple delays and their numerical solutions. *Appl. Math Comput.*, 208(2):397–415, 2009.

[57] Y. Chahlaoui, K.A. Gallivan, A. Vandendorpe, and P. Van Dooren. Model reduction of second-order systems. In P. Benner, V. Mehrmann, and D. C. Sorensen, editors, *Dimension Reduction of Large-Scale Systems*, pages 149–172. Springer, Berlin/Heidelberg, Germany, 2005.

[58] S. Chaturantabut, C. Beattie, and S. Gugercin. Structure-preserving model reduction for nonlinear port-Hamiltonian systems. *SIAM J. Sci. Comput.*, 38(5):B837–B865, 2016.

[59] H. Chen and C. Zhang. Stability analysis of linear multistep and Runge-Kutta methods for neutral multidelay-differential-algebraic systems. *Math. Comput. Model.*, 55(3-4):530–537, 2012.

[60] E.W. Cheney. *Introduction to Approximation Theory.* AMS Chelsea Publishing, New York, NY, USA, second edition, 1982.

[61] K.L. Cooke and D.W. Krumme. Differential-difference equations and nonlinear initial-boundary value problems for linear hyperbolic partial differential equations. *J. Math. Anal. Appl.*, 24(2):372–387, 1968.

[62] D.I. Coroian. Existence and uniqueness for a class of delay differential-algebraic equations. *Bul. ştiinţific al Univ. Baia Mare, Ser. B, Fasc. Mat.*, (2):265–274, 2000.

[63] R. Curtain and K. Morris. Transfer functions of distributed parameter systems: A tutorial. *Automatica*, 45(5):1101–1116, 2009.

[64] S. Dadebo and R. Luus. Optimal control of time-delay systems by dynamic programming. *Optimal Control Applications and Methods*, 13(1):29–41, 1992.

[65] O. Diekmann, M. Gyllenberg, J.A.J. Metz, S. Nakaoka, and A.M. de Roos. Daphnia revisited: local stability and bifurcation theory for physiologically structured population models explained by way of an example. *J. Math. Biol.*, 61:277–318, 2010.

[66] Z. Drmač, S. Gugercin, and C. Beattie. Quadrature-based vector fitting for discretized $\mathcal{H}_2$ approximation. *SIAM J. Sci. Comput.*, 37(2):A625–A652, 2015.

[67] Z. Drmač, S. Gugercin, and C. Beattie. Vector fitting for matrix-valued rational approximation. *SIAM J. Sci. Comput.*, 37(5):A2345–A2379, 2015.

[68] N.H. Du, V.H. Linh, and V. Mehrmann. Robust stability of differential-algebraic equations. In *Surveys in Differential-Algebraic Equations I*, pages 63–95. Springer, Berlin/Heidelberg, 2013.

[69] N.H. Du, V.H. Linh, V. Mehrmann, and D.D. Thuan. Stability and robust stability of linear time-invariant delay differential-algebraic equations. *SIAM J. Matrix Anal. Appl.*, 34(4):1631–1654, 2013.

[70] I. Pontes Duff, S. Gugercin, C. Beattie, C. Poussot-Vassal, and C. Seren. $\mathcal{H}_2$-optimality conditions for reduced time-delay systems of dimension one. *IFAC-PapersOnLine*, 49(10):7–12, 2016. Proc. 13th IFAC Workshop Time Delay Systems (TDS) 2016, Istanbul, Turkey.

[71] I. PONTES DUFF, C. POUSSOT-VASSAL, AND C. SEREN. Realization independent single time-delay dynamical model interpolation and $\mathcal{H}_2$-optimal approximation. In *Proc. 54th IEEE Conf. Decision Control (CDC)*, pages 4662–4667, Osaka, Japan, 2015.

[72] F. EBERT. *On Partitioned Simulation of Electrical Circuits using Dynamic Iteration Methods*. Dissertation, Institut für Mathematik, Technische Universität Berlin, 2008.

[73] K. ENGELBORGHS, T. LUZYANINA, AND D. ROOSE. Numerical bifurcation analysis of delay differential equations using DDE-BIFTOOL. *ACM Trans. Math. Softw.*, 28(1):1–21, 2002.

[74] I.R. EPSTEIN. Delay effects and differential delay equations in chemical kinetics. *International Reviews in Physical Chemistry*, 11(1):135–160, 1992.

[75] T. ERNEUX. *Applied Delay Differential Equations*. Surveys and Tutorials in the Applied Mathematical Sciences. Springer New York, 2009.

[76] L. R. FLETCHER. Regularizability of descriptor systems. *International Journal of Systems Science*, 17(6):843–847, 1986.

[77] E. FOSONG, P. SCHULZE, AND B. UNGER. From time-domain data to low-dimensional structured models. *ArXiv e-print 1902.05112*, 2019.

[78] R.W. FREUND. Structure-preserving model order reduction of RCL circuit equations. In W.H.A. SCHILDERS, H.A. VAN DER VORST, AND J. ROMMES, editors, *Model Order Reduction: Theory, Research Aspects and Applications*, pages 49–73. Springer, Berlin/Heidelberg, Germany, 2008.

[79] E. FRIDMAN. Stability of linear descriptor systems with delay: A Lyapunov-based approach. *J. Math. Anal. Appl.*, 273(1):24–44, 2002.

[80] E. FRIDMAN AND U. SHAKED. A descriptor system approach to H_∞ control of linear time-delay systems. *IEEE Trans. Automat. Contr.*, 47(2):253–270, 2002.

[81] E. FRIDMAN AND U. SHAKED. H_∞-control of linear state-delay descriptor systems: An LMI approach. *Linear Algebra Appl.*, 351-352:271–302, 2002.

[82] F. R. GANTMACHER. *The Theory of Matrices*, volume 2. Chelsea Publishing Company, New York, NY, USA, 1959.

[83] T. GEERTS. 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.

[84] G. GIACOMELLI, A. POLITI, AND S. YANCHUK. Modeling active optical networks. *ArXiv e-print 2004.04793*, 2020.

[85] H. GLUESING-LUERSSEN. *Linear Delay-Differential Systems with Commensurate Delays: An Algebraic Approach*. Springer-Verlag Berlin Heidelberg, Berlin, 2002.

[86] I.V. Gosea and A.C. Antoulas. Stability preserving post-processing methods applied in the Loewner framework. In *Proc. 20th IEEE Workshop Signal Power Integrity (SPI)*, pages 1–4, Turin, Italy, 2016.

[87] I.V. Gosea, M. Petreczky, and A.C. Antoulas. Data-Driven Model Order Reduction of Linear Switched Systems in the Loewner Framework. *SIAM J. Sci. Comput.*, 40(2):B572–B610, 2018.

[88] C.W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind.* Number 1. Pitman, London, 1984.

[89] S. Gugercin, A.C. Antoulas, and C. Beattie. $\mathscr{H}_2$ model reduction for large-scale linear dynamical systems. *SIAM J. Matrix Anal. Appl.*, 30(2):609–638, 2008.

[90] N. Guglielmi and E. Hairer. Computing breaking points in implicit delay differential equations. *Adv. Comput. Math.*, 29(3):229–247, 2008.

[91] N. Guglielmi and E. Hairer. Numerical approaches for state-dependent neutral delay equations with discontinuities. *Math. Comput. Simul.*, 95(1):2–12, 2014.

[92] S. Gumussoy and W. Michiels. Fixed-Order H-Infinity Control for Interconnected Systems Using Delay Differential Algebraic Equations. *SIAM J. Control Optim.*, 49(5):2212–2238, 2011.

[93] B. Gustavsen and A. Semlyen. Rational approximation of frequency domain responses by vector fitting. *IEEE Trans. Power Deliv.*, 14(3):1052–1061, 1999.

[94] P. Ha. *Analysis and Numerical solutions of Delay Differential-Algebraic Equations.* Dissertation, Technische Universität Berlin, 2015.

[95] P. Ha. On the Stability Analysis of Delay Differential-Algebraic Equations. *VNU J. Sci. Math. - Phys.*, 34(2):52–64, 2018.

[96] P. Ha. Spectral Characterizations of Solvability and Stability for Delay Differential-Algebraic Equations. *Acta Math. Vietnamica*, 43(4):715–735, 2018.

[97] P. Ha and V. Mehrmann. Analysis and reformulation of linear delay differential-algebraic equations. *Electron. J. Linear Algebr.*, 23:703–730, 2012.

[98] P. Ha and V. Mehrmann. Analysis and numerical solution of linear delay differential-algebraic equations. *BIT Numer. Math.*, 56(2):633–657, 2016.

[99] P. Ha, V. Mehrmann, and A. Steinbrecher. Analysis of Linear Variable Coefficient Delay Differential-Algebraic Equations. *J. Dynam. Differ. Equations*, 26(4):889–914, 2014.

[100] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princet. Uni. Bull.*, 13:49–52, 1902.

[101] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems.* Springer-Verlag, Berlin, Germany, 2nd edition, 1996.

[102] J.K. HALE AND S.M. VERDUYN LUNEL. *Introduction to Functional Differrential Equations.* Springer-Verlag, New York, 1993.

[103] R. HAUBER. Numerical treatment of retarded differential algebraic equations by collocation methods. *Adv. Comput. Math.*, 7:573–592, 1997.

[104] H.V. HENDERSON AND S.R. SEARLE. The vec-permutation matrix, the vec operator and Kronecker products: a review. *Linear Multilinear Algebra*, 9(4):271–288, 1981.

[105] R.J. HENRY, Z.N. MASOUD, A.H. NAYFEH, AND D.T. MOOK. Cargo pendulation reduction on ship-mounted cranes via boom-luff angle actuation. *J. Vib. Control*, 7(8):1253–1264, 2001.

[106] J.S. HESTHAVEN, G. ROZZA, AND B. STAMM. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations.* Springer Briefs in Mathematics. Springer Cham, Switzerland, 2016.

[107] T. HORIUCHI, M. INOUE, T. KONNO, AND Y. NAMITA. Real-time hybrid experimental system with actuator delay compensation and its application to a piping system with energy absorber. *Earthquake Engineering and Structural Dynamics*, 28:1121–1141, 1999.

[108] X. HU, Y. CONG, AND G.D. HU. Delay-dependent stability of linear multistep methods for DAEs with multiple delays. *Numer. Algorithms*, 79(3):719–739, 2018.

[109] X. HU, Y. CONG, AND G.D. HU. Delay-dependent stability of Runge–Kutta methods for linear delay differential–algebraic equations. *J. Comput. Appl. Math.*, 363(11871330):300–311, 2020.

[110] T.J.R. HUGHES. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis.* Dover Publications Inc., 2012.

[111] A. ILCHMANN AND T. REIS, editors. *Surveys in Differential-Algebraic Equations I.* Springer-Verlag Berlin Heidelberg, 2013.

[112] A. ILCHMANN AND T. REIS, editors. *Surveys in Differential-Algebraic Equations II.* Springer International Publishing, 2015.

[113] A. ILCHMANN AND T. REIS, editors. *Surveys in Differential-Algebraic Equations III.* Springer International Publishing, 2015.

[114] A. ILCHMANN AND T. REIS, editors. *Surveys in Differential-Algebraic Equations IV.* Springer International Publishing, 2017.

[115] A.C. IONITA AND A.C. ANTOULAS. Data-driven parametrized model reduction in the Loewner framework. *SIAM J. Sci. Comput.*, 36(3):A984–A1007, 2014.

[116] S. IWATA AND M. TAKAMATSU. Index Reduction via Unimodular Transformations. *SIAM J. Matrix Anal. Appl.*, 39(3):1135–1151, 2018.

[117] L. JANTSCHER. *Distributionen.* De Gruyter Lehrbuch. Walter de Gruyter, Berlin, New York, 1971.

[118] T. KACZOREK. *Polynomial and Rational Matrices.* Springer, London, 2007.

[119] T. Kalmár-Nagy, G. Stépán, and F.C. Moon. Subcritical Hopf bifurcation in the delay equation model for machine tool vibrations. *Nonlinear Dynamics*, 26(2):121–142, 2001.

[120] I. Karafyllis and M. Krstic. On the relation of delay equations to first-order hyperbolic partial differential equations. *ESAIM - Control. Optim. Calc. Var.*, 20(3):894–923, 2014.

[121] M. Köhler. On the closest stable descriptor system in the respective spaces RH_2 and RH_∞. *Linear Algebra Appl.*, 443:34–49, 2014.

[122] V. Kolmanovskii and A. Myshkis. *Introduction to the Theory and Applications of Functional Differential Equations.* Mathematics and Its Applications. Springer Netherlands, 2013.

[123] A. Kumar, P.D. Christofides, and P. Daoutidis. Singular perturbation modeling of nonlinear processes with nonexplicit time-scale multiplicity. *Chemical Engineering Science*, 53(8):1491–1504, 1998.

[124] P. Kunkel and V. Mehrmann. Regular solutions of nonlinear differential-algebraic equations and their numerical determination. *Numer. Math.*, 79(4):581–600, 1998.

[125] P. Kunkel and V. Mehrmann. Analysis of over-and underdetermined nonlinear differential-algebraic systems with application to nonlinear control problems. *Math. Control. Signals Syst.*, 14(3):233–256, 2001.

[126] P. Kunkel and V. Mehrmann. Index reduction for differential-algebraic equations by minimal extension. *ZAMM Z. Angew. Math. Mech.*, 84(9):579–597, 2004.

[127] P. Kunkel and V. Mehrmann. *Differential-Algebraic Equations. Analysis and Numerical Solution.* European Mathematical Society, Zürich, Switzerland, 2006.

[128] J. Kutz, S. Brunton, B. Brunton, and J. Proctor. *Dynamic Mode Decomposition.* SIAM, Philadelphia, PA, 2016.

[129] Y.N. Kyrychko, K.B. Blyuss, A. Gonzalez-Buelga, S.J. Hogan, and D.J. Wagg. Real-time dynamic substructuring in a coupled oscillator – pendulum system. *Proc. R. Soc. A*, 462:1271–1294, 2006.

[130] K. J. Laidler. A glossary of terms used in chemical kinetics, including reaction dynamics (IUPAC Recommendations 1996). *Pure & Appl. Chem*, 68(1):149–192, 1996.

[131] C. Lainscsek, P. Rowat, L. Schettino, D. Lee, D. Song, C. Letellier, and H. Poizner. Finger tapping movements of Parkinson's disease patients automatically rated using nonlinear delay differential equations. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(1):013119–1 – 013119–13, 2012.

[132] C. Lainscsek, L. Schettino, P. Rowat, E. van Erp, D. Song, and H. Poizner. Nonlinear DDE Analysis of Repetitive Hand Movements in Parkinson's Disease. In V. In, P. Longhini, and A. Palacios, editors, *Applications of Nonlinear Dynamics: Model and Design of Complex Systems*, pages 421–425. Springer Berlin Heidelberg, 2009.

[133] S. LALL, P. KRYSL, AND J.E. MARSDEN. Structure-preserving model reduction for mechanical systems. *Phys. D*, 184(1-4):304–318, 2003.

[134] R. LAMOUR, R. MÄRZ, AND C. TISCHENDORF. *Differential-Algebraic Equations: A Projector Based Analysis*. Springer-Verlag Berlin Heidelberg, 2013.

[135] V.H. LINH AND D.D. THUAN. Spectrum-based robust stability analysis of linear delay differential-algebraic equations. In P. BENNER, M. BOLLHÖFER, D. KRESSNER, C. MEHL, AND T. STYKEL, editors, *Numerical Algebra, Matrix Theory, Differential-Algebraic Equations and Control Theory: Festschrift in Honor of Volker Mehrmann*, pages 533–557. Springer International Publishing, Cham, 2015.

[136] V.H. LINH, N.D. TRUONG, AND M.V. BULATOV. Convergence Analysis of Linear Multistep Methods for a Class of Delay Differential-Algebraic Equations. *Bull. South Ural State Univ. Ser. "Mathematical Model. Program. Comput. Software"*, 11(4):78–93, 2018.

[137] L. LJUNG. On the estimation of transfer functions. *Automatica*, 21(6):677–696, 1985.

[138] H. LOGEMANN. Destabilizing effects of small time delays on feedback-controlled descriptor systems. *Linear Algebra Appl.*, 272(1-3):131–153, 1998.

[139] O. LOPES. Stability and forced oscillations. *J. Math. Anal. Appl.*, 55(3):686–698, 1976.

[140] O. LÜTHJE, S. WOLFF, AND G. PFISTER. Control of chaotic Taylor-Couette flow with time-delayed feedback. *Physical Review Letters*, 86(9):1745–1748, 2001.

[141] T. LUZYANINA AND D. ROOSE. Periodic Solutions of Differential Algebraic Equations With Time Delays: Computation and Stability Analysis. *Int. J. Bifurc. Chaos*, 16(01):67–84, 2006.

[142] D. MACKEY, N. MACKEY, C. MEHL, AND V. MEHRMANN. Vector spaces of linearizations for matrix polynomials. *SIAM J. Matr. Anal. Appl.*, 28(4):971–1004, 2006.

[143] C. MAGRUDER, C.A. BEATTIE, AND S. GUGERCIN. Rational Krylov methods for optimal $\mathscr{H}_2$ model reduction. In *Proc. 49th IEEE Conf. Decision Control (CDC)*, pages 6797–6802, Atlanta, GA, USA, 2010.

[144] R. MÄRZ. Some new results concerning index-3 differential-algebraic equations. *J. Math. Anal. Appl.*, 140(1):177–199, 1989.

[145] R. MÄRZ. The index of linear differential algebraic equations with properly stated leading terms. *Res. Math.*, 42:308–338, 2002.

[146] R. MÄRZ. Characterizing differential algebraic equations without the use of derivative arrays. *Comput. Math. with Appl.*, 50(7):1141–1156, 2005.

[147] R. MÄRZ. Differential-Algebraic Equations from a Functional-Analytic Viewpoint: A Survey. In A. ILCHMANN AND T. REIS, editors, *Surveys in Differential-Algebraic Equations II*, pages 163–285. Springer International Publishing, Cham, 2015.

[148] Z.N. MASOUD AND A.H. NAYFEH. Sway Reduction on Container Cranes Using Delayed Feedback. *Nonlinear Dyn.*, 34(3-4):347–358, 2003.

[149] Z.N. MASOUD, A.H. NAYFEH, AND A. AL-MOUSA. Delayed Position-Feedback Controller for the Reduction of Payload Pendulations of Rotary Cranes. *Journal of Vibration and Control*, 9:257–277, 2003.

[150] A.J. MAYO AND A.C. ANTOULAS. A framework for the solution of the generalized realization problem. *Linear Algebra Appl.*, 425(2-3):634–662, 2007.

[151] G. MAZANTI, I. BOUSSAADA, S.-I. NICULESCU, AND Y. CHITOUR. Effects of Roots of Maximal Multiplicity on the Stability of Some Classes of Delay Differential-Algebraic Systems: The Lossless Propagation Case. *ArXiv e-print 2002.00078*, (2007), 2020.

[152] C. MEHL, V. MEHRMANN, AND M. WOJTYLAK. Linear Algebra Properties of Dissipative Hamiltonian Descriptor Systems. *SIAM Journal on Matrix Analysis and Applications*, 39(3):1489–1519, 2018.

[153] V. MEHRMANN. Index concepts for differential-algebraic equations. In B. ENGQUIST, editor, *Encyclopedia of Applied and Computational Mathematics*, pages 676–681. Springer, Berlin, Germany, 2015.

[154] V. MEHRMANN AND C. SHI. Transformation of high order linear differential-algebraic systems to first order. *Numer. Alg.*, 42:281–307, 2006.

[155] D.G. MEYER AND S. SRINIVASAN. Balancing and model reduction for second-order form linear systems. *IEEE Trans. Automat. Control*, 41(11):1632–1644, 1996.

[156] W. MICHIELS. Spectrum-based stability analysis and stabilisation of systems described by delay differential algebraic equations. *IET Control Theory Appl.*, 5(16):1829–1842, 2011.

[157] W. MICHIELS AND S. GUMUSSOY. Eigenvalue based analysis and controller synthesis for systems described by delay differential algebraic equations. In *IFAC Proc. Vol.*, volume 10, pages 144–149. IFAC, 2012.

[158] W. MICHIELS, E. JARLEBRING, AND K. MEERBERGEN. Krylov-based model order reduction of time-delay systems. *SIAM J. Matrix Anal. Appl.*, 32(4):1399–1421, 2011.

[159] U. MIEKKALA. Dynamic iteration methods applied to linear DAE systems. *J. Comput. Appl. Math.*, 25:133–151, 1989.

[160] J. MILTON, J.L. CABRERA, T. OHIRA, S. TAJIMA, Y. TONOSAKI, C.W. EURICH, AND S.A. CAMPBELL. The time-delayed inverted pendulum: implications for human balance control. *Chaos (Woodbury, N.Y.)*, 19(2), 2009.

[161] S.H. OH AND R. LUUS. Optimal Feedback Control of Time-Delay Systems. *AIChE J.*, 22(1):140–147, 1976.

[162] M. Ohlberger and S. Rave. Nonlinear reduced basis approximation of parameterized evolution equations via the method of freezing. *C. R. Math. Acad. Sci. Paris*, 351(23–24):901–906, 2013.

[163] A. Otto, F. A. Khasawneh, and G. Radons. Position-dependent stability analysis of turning with tool and workpiece compliance. *International Journal of Advanced Manufacturing Technology*, 79(9-12):1453–1463, 2015.

[164] A. Otto and G. Radons. Stability analysis of machine-tool vibrations in the frequency domain. *IFAC-PapersOnLine*, 28(12):328–333, 2015.

[165] A. Otto and G. Radons. Analysis of systems with state-dependent delay and applications in metal cutting. In *24rd Int. Congr. Theor. Appl. Mech.*, Montreal, Canada, 2016.

[166] A. Otto and G. Radons. Transformations from variable delays to constant delays with applications in engineering and biology. In *Time Delay systems. Advances in Delays and Dynamics*, volume 7, pages 169–183. Springer, Cham, 2016.

[167] C.C. Pantelides. The consistent initialization of differential-algebraic systems. *SIAM J. ScI. STAT. Comput.*, 9(2):213–231, 1988.

[168] B. Peherstorfer, S. Gugercin, and K.E. Willcox. Data-driven reduced model construction with time-domain loewner models. *SIAM J. Sci. Comput.*, 39(5):2152–2178, 2017.

[169] M. Peil, M. Jacquot, Y.K. Chembo, L. Larger, and T. Erneux. Routes to chaos and multiple time scale dynamics in broadband bandpass nonlinear delay electro-optic oscillators. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 79(2):1–15, 2009.

[170] S. Petkoski and V.K. Jirsa. Transmission time delays organize the brain network synchronization. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 377(2153), 2019.

[171] L. Petzold. Differential/algebraic equations are not ODE's. *SIAM J. Sci. and Stat. Comput.*, 3(3):367–384, 1982.

[172] J.W. Polderman and J.C. Willems. *Introduction to Mathematical Systems Theory. A Behavioral Approach.* Springer New York, 1998.

[173] L.K. Poppe. The Strangeness Index of a Linear Delay Differential-Algebraic Equation of Retarded Type. *IFAC Proc. Vol.*, 39(10):320–324, 2006.

[174] J.D. Pryce. A simple structural analysis method for DAEs. *BIT Numer. Math.*, 41(2):364–394, 2001.

[175] K. Pyragas. Continuous control of chaos by self-controlling feedback. *Physics Letters A*, 170(6):421–428, 1992.

[176] K. Pyragas and A. Tamaševičius. Experimental control of chaos by delayed self-controlling feedback. *Physics Letters A*, 180(1-2):99–102, 1993.

[177] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations: An Introduction.* UNITEXT. Springer Cham, 2016.

[178] R.J Rabier and W.C. Rheinboldt. Time-dependent linear DAEs with discontinuous inputs. *Linear Algebra Appl.*, 247:1–29, 1996.

[179] S. Reich. On a geometrical interpretation of differential-algebraic equations. *Circuits Syst. Signal Process*, 9(4):367–382, 1990.

[180] J. Reiss, P. Schulze, J. Sesterhenn, and V. Mehrmann. The shifted proper orthogonal decomposition: a mode decomposition for multiple transport phenomena. *SIAM J. Sci. Comput.*, 40(3):A1322–A1344, 2018.

[181] W.C. Rheinboldt. Differential-Algebraic Systems as Differential Equations on Manifolds. *Math. Comput.*, 43(168):473–482, 84.

[182] T. Roose, P.A. Netti, L.L. Munn, Y. Boucher, and R.K. Jain. Solid stress generated by spheroid growth estimated using a linear poroelasticity model. *Microvasc. Res.*, 66(3):204–212, 2003.

[183] A.E. Ruehli. Equivalent Circuit Models for Three-Dimensional Multiconductor Systems. *IEEE Transactions on Microwave Theory and Techniques*, 22(3):216–221, 1974.

[184] G. Scarciotti and A. Astolfi. Model reduction by moment matching for linear time-delay systems. *IFAC Proceedings Volumes*, 47(3):9462–9467, August 2014.

[185] S. Schikora, P. Hövel, H.-J. Wünsche, E. Schöll, and F. Henneberger. All-Optical Noninvasive Control of Unstable Steady States in a Semiconductor Laser. *Physical Review Letters*, 97(21-24):213902, 2006.

[186] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.*, 656:5–28, 2010.

[187] L Scholz. A derivative array approach for linear second order differential-algebraic systems. *Electron. J. Linear Algebr.*, 22:310–347, 2011.

[188] S. Schöps, A. Bartel, M. Günther, E.J.W. ter Maten, and P.C. Müller, editors. *Progress in Differential-Algebraic Equations.* Springer-Verlag Berlin Heidelberg, 2014.

[189] P. Schulze and B. Unger. Data-driven interpolation of dynamical systems with delay. *Systems Control Lett.*, 97:125–131, 2016.

[190] P. Schulze and B. Unger. Model reduction for linear systems with low-rank switching. *SIAM J. Control Optim.*, 56(6):4365–4384, 2018.

[191] P. Schulze, B. Unger, C. Beattie, and S. Gugercin. Data-driven structured realization. *Linear Algebra Appl.*, 537:250 – 286, 2018.

[192] L. Schwartz. *Théorie des Distributions.* Hermann, Paris, 1957, 1959.

[193] P. SCHWERDTNER AND M. VOIGT. Computation of the L_∞-Norm Using Rational Interpolation. *IFAC-PapersOnLine*, 51(25):84–89, 2018.

[194] L.F. SHAMPINE AND P. GAHINET. Delay-differential-algebraic equations in control theory. *Appl. Numer. Math.*, 56(3-4):574–588, 2006.

[195] T.B. SHERIDAN. Space Teleoperation Through Time Delay: Review and Prognosis. *IEEE Transactions on Robotics and Automation*, 9(5):592–606, 1993.

[196] J. SIEBER, K. ENGELBORGHS, T. LUZYANINA, G. SAMAEY, AND D. ROOSE. DDE-BIFTOOL manual - bifurcation analysis of delay differential equations. *ArXiv e-prints 1406.7144v4*, 2016.

[197] B. SIMEON. *Computational Flexible Multibody Dynamics. A Differential-Algebraic Approach.* Springer-Verlag Berlin Heidelberg, 2013.

[198] G. STÉPÁN AND L. KOLLÁR. Balancing with reflex delay. *Mathematical and Computer Modelling*, 31(4-5):199–205, 2000.

[199] T.-J. SU AND R.R. CRAIG JR. Model reduction and control of flexible structures using Krylov vectors. *J. Guid. Control Dynam.*, 14(2):260–267, 1991.

[200] A. TANWANI AND S. TRENN. On observability of switched differential-algebraic equations. In *Proc. IEEE Conf. Decis. Control*, pages 5656–5661, 2010.

[201] H. TIAN, Q. YU, AND J. KUANG. Asymptotic Stability of Linear Neutral Delay Differential-Algebraic Equations and Runge–Kutta Methods. *SIAM J. Numer. Anal.*, 52(1):68–82, 2014.

[202] F. TISSEUR AND K. MEERBERGEN. The Quadratic Eigenvalue Problem. *SIAM Rev.*, 43(2):235–286, 2001.

[203] S. TRENN. *Distributional differential algebraic equations.* PhD thesis, Institut für Mathematik, Technische Universität Ilmenau, Universitätsverlag Ilmenau, Germany, 2009.

[204] S. TRENN. Regularity of distributional differential algebraic equations. *Math. Control Signals Syst.*, 21(3):229–264, 2009.

[205] S. TRENN. Solution concepts for linear DAEs: a survey. In A. ILCHMANN AND T. REIS, editors, *Surveys in Differential-Algebraic Equations I*, chapter 3, pages 137–172. Springer, Berlin/Heidelberg, 1st edition, 2013.

[206] S. TRENN AND B. UNGER. Delay regularity of differential-algebraic equations. In *Proc. 58th IEEE Conf. Decision Control (CDC) 2019, Nice, France*, pages 989–994, 2019.

[207] S. TRENN AND B. UNGER. Feedback regularization of DAEs with delay. *In preparation*, 2020.

[208] S. TRENN AND B. UNGER. Unimodular transformations for DAE initial trajectory problems. *Submitted for publication*, 2020.

[209] S. TRENN AND J.C. WILLEMS. Switched behaviors with impulses - a unifying framework. In *Proc. 51st IEEE Conf. Decis. Control, Maui, USA*, pages 3203–3208, Dec 2012.

[210] J.H. Tu, C.W. Rowley, D.M. Luchtenburg, S.L. Brunton, and J.N. Kutz. On dynamic mode decomposition: Theory and applications. *J. Comput. Dyn.*, 1(2):391–421, 2014.

[211] B. Unger. Discontinuity propagation in delay differential-algebraic equations. *Electron. J. Linear Algebr.*, 34:582–601, 2018.

[212] B. Unger. Delay differential-algebraic equations in real-time dynamic substructuring. *ArXiv e-print 2003.10195*, 2020. Submitted for publication.

[213] B. Unger and S. Gugercin. Kolmogorov n-widths for linear dynamical systems. *Adv. Comput. Math.*, 45(5-6):2273–2286, 2019.

[214] S.M. Verduyn Lunel. Spectral theory for neutral delay equations with applications to control and stabilization. In J. Rosenthal and D.S. Gilliam, editors, *Mathematical Systems Theory in Biology, Communications, Computation, and Finance*, pages 415–467, New York, NY, 2003. Springer New York.

[215] G.C. Verghese, B. Lévy, and T. Kailath. A Generalized State-Space for Singular Systems. *IEEE Trans. Automat. Contr.*, 26(4):811–831, 1981.

[216] M.I. Wallace, J. Sieber, S.A. Neild, D.J. Wagg, and B. Krauskopf. Stability analysis of real-time dynamic substructuring using delay differential equation models. *Earthquake Engineering and Structural Dynamics*, 34:1817–1832, 2005.

[217] L. Weicker, T. Erneux, O. D'Huys, J. Danckaert, M. Jacquot, Y. Chembo, and L. Larger. Slow-fast dynamics of a time-delayed electro-optic oscillator. *Philos. Trans. R. Soc. London A Math. Phys. Eng. Sci.*, 371(1999), 2013.

[218] M.S. Williams and A. Blakeborough. Laboratory testing of structures under dynamic loads: an introductory review. *Philos. Trans. R. Soc. London A Math. Phys. Eng. Sci.*, 359:1651–1669, 2001.

[219] E. Witrant and S.-I. Niculescu. Modeling and control of large convective flows with time-delays. *Mathematics in Engineering, Science and Aerospace*, 1(2):191–205, 2010.

[220] K.-T. Wong. The Eigenvalue Problem $\lambda Tx + Sx$. *J. Differ. Equ.*, 16(2):270–280, 1974.

[221] Z. Wu and W. Michiels. Reliably computing all characteristic roots of delay differential equations in a given right half plane using a spectral method. *J. Comput. Appl. Math.*, 236(9):2499–2514, 2012.

[222] S. Xu, P. Van Dooren, R. Stefan, and J. Lam. Robust Stability and Stabilization for Singular Systems with state delays and parameter uncertainty. *IEEE Trans. Autom. Contr.*, 47(7):1122–1128, 2002.

[223] K. Yamasue and T. Hikihara. Control of microcantilevers in dynamic force microscopy using time delayed feedback. *Review of Scientific Instruments*, 77:053703, 2006.

[224] W. Zhu and L.R. Petzold. Asymptotic stability of linear delay differential-algebraic equations and numerical methods. *Appl. Numer. Math.*, 24:247–264, 1997.

[225] W. ZHU AND L.R. PETZOLD. Asymptotic stability of Hessenberg delay differential-algebraic equations of retarded or neutral type. *Appl. Numer. Math.*, 27(3):309–325, 1998.

[226] R.E. ZIEMER, W.H. TRANTER, AND D.R. FANNIN. *Signals and systems: continuous and discrete*. Prentice Hall, Upper Saddle River, NJ, USA, 4. ed. edition, 1998.