Acta Numerica (2023), pp. 395–515 Printed in the United Kingdom
doi:10.1017/S0962492922000083

Control of port-Hamiltonian differential-algebraic systems and applications

Volker Mehrmann
Institut für Mathematik, Technische Universität Berlin,
Strasse des 17. Juni 136, 10623 Berlin, Germany
E-mail: mehrmann@math.tu-berlin.de

Benjamin Unger
Stuttgart Center for Simulation Science, University of Stuttgart,
Universitätsstrasse 32, 70569 Stuttgart, Germany
E-mail: benjamin.unger@simtech.uni-stuttgart.de

We discuss the modelling framework of port-Hamiltonian descriptor systems and their use in numerical simulation and control. The structure is ideal for automated network-based modelling since it is invariant under power-conserving interconnection, congruence transformations and Galerkin projection. Moreover, stability and passivity properties are easily shown. Condensed forms under orthogonal transformations present easy analysis tools for existence, uniqueness, regularity and numerical methods to check these properties.

After recalling the concepts for general linear and nonlinear descriptor systems, we demonstrate that many difficulties that arise in general descriptor systems can be easily overcome within the port-Hamiltonian framework. The properties of port-Hamiltonian descriptor systems are analysed, and time discretization and numerical linear algebra techniques are discussed. Structure-preserving regularization procedures for descriptor systems are presented to make them suitable for simulation and control. Model reduction techniques that preserve the structure and stabilization and optimal control techniques are discussed.

The properties of port-Hamiltonian descriptor systems and their use in modelling simulation and control methods are illustrated with several examples from different physical domains. The survey concludes with open problems and research topics that deserve further attention.

2020 Mathematics Subject Classification: Primary 65L80
Secondary 37J06

CONTENTS

1. Introduction

Modern key technologies in science and technology require modelling, simulation and optimization (MSO), or control of complex dynamical systems. Most real-world systems are multi-physics systems, combining components from different physical domains, and with different accuracies and scales in the components. To address these requirements, there exist many commercial and open source MSO software packages for simulation and control in all physical domains, e.g. Abaqus,[1] Ansys,[2] COMSOL,[3] Dymola,[4] FEniCS[5] and Simulink.[6] Several of these also have multi-physics components, but all are still very limited when it comes to applications, such as digital twins, which require a cross-domain evolutionary modelling process, the coupling of different domain-specific tools, the incorporation of model hierarchies consisting of coarse and fine discretizations and reduced-order models as well as the incorporation of (optimal) control techniques. The latter point, in particular, requires tools to be open to performing easy and automated model modifications.

Furthermore, flexible compromises between different accuracy levels and computational speed have to be possible to allow uncertainty quantification procedures, as well as error estimates that balance model, discretization, optimization, approximation or round-off errors, combined with sensitivity, stability and robustness measures. Finally, with modern data science tools becoming increasingly powerful,

[1] https://www.3ds.com/products-services/simulia/products/abaqus/
[2] https://www.ansys.com/
[3] https://www.comsol.com
[4] https://www.3ds.com/products-services/catia/products/dymola/
[5] https://fenicsproject.org
[6] https://www.mathworks.com/products/simulink.html

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

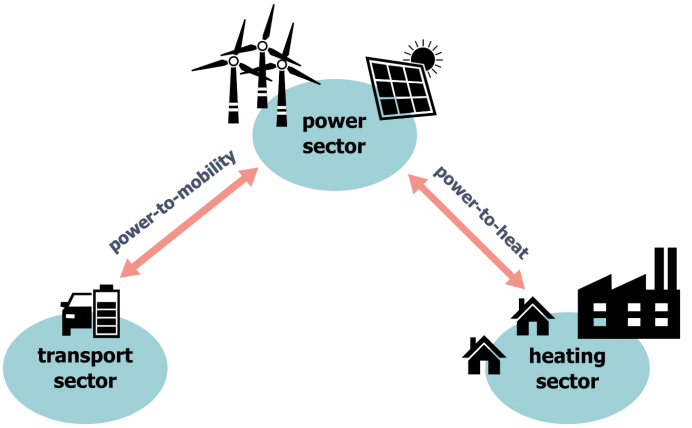

Figure 1.1. Sector coupling and the power-to-X concept.

it is necessary to have models and methods that allow pure data-based approaches and have the flexibility to recycle and re-use components in different applications. On top of all these requirements, the MSO tools cannot be separated from the available computing environments, ranging from process controllers, data, sensor and visualization interfaces, linked up with high-performance and cloud computing facilities.

To address all these challenges in the future and in an increasingly digitized world, a fundamental paradigm shift in MSO is necessary. For every scientific and technological product or process, and the whole life cycle from the design phase to waste recycling, it is necessary to build digital twins with multi-fidelity model hierarchies or catalogues of several models ranging from very fine descriptions, that help to understand the behaviour via detailed and repeated simulations, to very coarse (reduced or surrogate) models, used for real-time control and optimization. Furthermore, the MSO tools should, as far as possible, be open for interaction, automated, and allow the linking of subsystems or numerical methods in a network fashion. They should also allow combination with methods that deal with large sets of real-time data that can and should be employed in a modelling or data assimilation process. Because of all this, it is necessary for mathematical modelling, analysis, numerics, control, optimization, model reduction methods and data science techniques to work hand in hand.

To illustrate these general comments, let us consider a major societal application. In order to reduce global warming, it is necessary to reduce the emissions arising in the production of energy from fossil sources by increasing renewable energy production, such as wind or solar energy. At the same time, it is essential to allow energy-efficient multi-directional sector coupling, such as power-to-heat or power-to-mobility; see Figure 1.1.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

The coupling of energy sectors includes the storage or transformation to another energy carrier (e.g. hydrogen) of superfluous electrical energy, as well the layout and operation of energy transportation networks; see e.g. Brown *et al.* (2018), Conejo, Chen and Constante (2020) and Ramsebner, Haas, Ajanovic and Wietschel (2021).

On the mathematical/computational side, challenges arise because mathematical models of different energy conversion processes and energy transport networks live on very different time scales, such as gas transport networks or electrical power networks. Furthermore, while most energy transport networks are currently operated in a stationary regime, in the future dynamic approaches will be required that allow control and optimization of energy production and transport in real time; see e.g. Bienstock (2015) and Machowski, Lubosny, Bialek and Bumby (2020). These further challenges lead to the following *model class wish list* for a new flexible modelling, simulation, optimization and control framework.

- The model class should allow for automated modelling, in a modularized and network-based fashion, including pure data-based models.
- The mathematical representation should allow the coupling of mathematical models across different scales and physical domains in continuous and discrete time.
- The mathematical models should be close to the real (open or closed) physical system, and easily extendable if further effects have to be considered or when singular limit situations are considered.
- The model class should have nice algebraic, geometric and analytical properties. The models should be easy to analyse concerning existence, uniqueness, robustness, stability, uncertainty, perturbation theory and error analysis.
- The class should be invariant under local variable (coordinate) transformations (in space and time) to understand the local behaviour, for example via local normal forms.
- The model class should allow for simple space and time discretization and model reduction methods, as well as fast solvers for the resulting linear and nonlinear systems of equations.
- The model class should be feasible for simulation, control, optimization and data assimilation.

Can there be such a 'Jack of all trades'? The main goal of this paper is to show that even though many aspects are still under investigation, energy-based modelling via the model class of *dissipative port-Hamiltonian (pH) descriptor systems* has many great features and comes very close to being such a model class. Let us emphasize that the field of pHDAE systems is a highly active research area with many developments currently taking place. In this survey, we thus focus only on selected topics and provide pointers to open problems throughout the paper. In the spirit of the interdisciplinary setting of the port-Hamiltonian framework,

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

we challenge and encourage the mathematical, natural sciences, computer science and engineering communities to apply the conceptual ideas presented in this survey to their applications, improve upon existing results, develop novel theories and software packages, and extend the scope to further application domains.

Structure of the paper

The paper is organized as follows. We first review general nonlinear descriptor systems and its solution theory in Section 2 and associated control theoretical results in Section 3. Dissipative port-Hamiltonian (pH) descriptor systems are introduced in Section 4 and illustrated with several examples from different application domains in Section 5. Condensed forms are presented in Section 6. In Section 7 we start to analyse pH descriptor systems in terms of our model class wish list by discussing the inherent properties of pH descriptor systems. We then turn to structure-preserving model order reduction in Section 8 and discuss time discretization and associated linear system solves in Section 9. We conclude our presentation with a discussion of control methods in Section 10 and a summary (including open problems and future work) in Section 11. We emphasize that within this paper we mainly focus on finite-dimensional problems. Nevertheless, the pH model class can be extended to the infinite-dimensional setting, and we provide a brief discussion and an (incomplete) list of references at the end of Section 11.

Notation

The sets $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{R}$ and $\mathbb{C}$ denote natural numbers, non-negative integers, real numbers and complex numbers, respectively. For a complex number $z \in \mathbb{C}$ we denote its real part as $\operatorname{Re}(z)$. The set of $n \times m$ matrices with values in a field $\mathbb{F}$ is denoted by $\mathbb{F}^{n,m}$. The symbol I is used for the identity matrix, whose dimension is clear from the context. The rank and corank of a matrix $M \in \mathbb{F}^{n,m}$ are denoted by $\operatorname{rank} M$ and $\operatorname{corank} M$, where the latter is defined as

$$\operatorname{corank} M := n - \operatorname{rank} M. \tag{1.1}$$

The transpose of a matrix and the conjugate transpose (if $\mathbb{F} = \mathbb{C}$) are denoted by $M^\top$ and M^H, respectively. To indicate that a matrix $M \in \mathbb{F}^{n,n}$ is positive definite or positive semidefinite, we write $M > 0$ and $M \geq 0$, respectively. The Moore–Penrose inverse of a matrix $M \in \mathbb{F}^{m,n}$, i.e. the unique matrix A satisfying $MAM = M$, $AMA = A$, $(AM)^H = AM$ and $(MA)^H = MA$, is denoted by $A = M^\dagger$. A block diagonal matrix with diagonal blocks $B_1, \ldots, B_k$ is denoted by $\operatorname{diag}(B_1, \ldots, B_k)$, and the span of a list of vectors $v_1, \ldots, v_k$ is denoted by $\operatorname{span}(v_1, \ldots, v_k)$.

The spaces of continuous and k-times continuously differentiable functions (with $k \in \mathbb{N}$) from the time interval $\mathbb{T}$ to some Banach space $\mathcal{X}$ are denoted by $\mathcal{C}(\mathbb{T}, \mathcal{X})$ and $\mathcal{C}^k(\mathbb{T}, \mathcal{X})$, respectively. For a function $f \in \mathcal{C}^1(\mathbb{T}, \mathcal{X})$ we write $\dot{f} := \mathrm{d}/\mathrm{d}t$ to denote the (time) derivative. Similarly, we use the notation $\ddot{f}$ for the second derivative and

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

$f^{(k)}$ for the kth derivative. The Jacobian of a function $f\colon \mathbb{R}^{\ell} \to \mathbb{R}^{n}$ is denoted by $(\partial/\partial z)f(z)$.

Abbreviations

Throughout the paper, we use the following abbreviations.

DAE	differential-algebraic equation
dHDAE	dissipative Hamiltonian differential-algebraic equation
ECRM	effort constraint reduction method
FCRM	flow constraint reduction method
IRKA	iterative rational Krylov algorithm
LTI	linear time-invariant
LTV	linear time-varying
MM	moment matching
MSO	modelling, simulation and optimization
MOR	model order reduction
ODE	ordinary differential equation
PDE	partial differential equation
pH	port-Hamiltonian
pHODE	port-Hamiltonian ordinary differential equation
pHDAE	port-Hamiltonian differential-algebraic equation
ROM	reduced order model

2. The model class of descriptor systems

To allow network-based automated modularized modelling via interconnection, constraint-preserving simulation, optimization, and control of dynamic models, it is common practice in many application domains to use the class of (implicit) control systems, called *descriptor systems* or *differential-algebraic equation (DAE) systems*, of the form

$$F(t, z(t), \dot{z}(t), u(t)) = 0, \tag{2.1a}$$

$$y(t) - G(t, z(t), u(t)) = 0, \tag{2.1b}$$

on some time interval $\mathbb{T} := [t_0, t_{\mathrm{f}}]$ with

$$F\colon \mathbb{T} \times \mathbb{D}_{\mathrm{z}} \times \mathbb{D}_{\dot{\mathrm{z}}} \times \mathbb{D}_{\mathrm{u}} \to \mathbb{R}^{\ell} \quad \text{and} \quad G\colon \mathbb{T} \times \mathbb{D}_{\mathrm{z}} \times \mathbb{D}_{\mathrm{u}} \to \mathbb{R}^{p},$$

with open domains, vector spaces or manifolds $\mathbb{D}_{\mathrm{z}}, \mathbb{D}_{\dot{\mathrm{z}}}, \mathbb{D}_{\mathrm{u}}$. In the finite-dimensional case of real systems, which is predominantly discussed in this paper, we assume for ease of presentation that

$$F\colon \mathbb{T} \times \mathbb{R}^{n} \times \mathbb{R}^{n} \times \mathbb{R}^{m} \to \mathbb{R}^{\ell} \quad \text{and} \quad G\colon \mathbb{T} \times \mathbb{R}^{n} \times \mathbb{R}^{m} \to \mathbb{R}^{p}.$$

We refer to z, u and y as the *state*, *input* and *output*, respectively.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Note that there can be very different roles of inputs in different applications, for example to deal with control actions, interconnections or disturbances, and different roles of outputs, for example for measurements, interconnection or observer design. Also, the models typically have parameters or may have random uncertain components such as unmodelled quantities, uncertainty in parameters, or disturbances.

We remark that most of the results and methods we present also hold for complex systems, but we restrict ourselves to the real case in this survey. In the following, for the sake of a simpler presentation, we will also often omit the time or space argument whenever this is appropriate or clear from the context.

2.1. Solution concept

It is clear that depending on the application, different solution concepts for (2.1) may be necessary; see e.g. Brenan, Campbell and Petzold (1996), Kunkel and Mehrmann (2006) and Lamour, März and Tischendorf (2013).

In the finite-dimensional setting we restrict ourselves to classical function spaces of continuous or continuously differentiable functions. For control problems as in (2.1) we often follow the behaviour framework (e.g. Polderman and Willems 1998) in which a new combined state vector

$$\xi := [z^\top \quad u^\top \quad y^\top]^\top \tag{2.2}$$

is introduced (or $\xi := [z^\top, u^\top]^\top$ if only the state equation (2.1a) is considered). The descriptor system (2.1) is then turned into an under-determined DAE (e.g. Kunkel and Mehrmann 2001), that is, the meaning of the variables is no longer distinguished.

Definition 2.1 (solution concept). Consider the DAE (2.1) on the time interval $\mathbb{T}$ with open subsets $\mathbb{D}_z \subseteq \mathbb{R}^n$, $\mathbb{D}_{\dot z} \subseteq \mathbb{R}^n$, $\mathbb{D}_u \subseteq \mathbb{R}^m$.

(i) Let $u\colon \mathbb{T} \to \mathbb{R}^m$ be a given input. We call a function $z \in \mathcal{C}^1(\mathbb{T}, \mathbb{R}^n)$ a *solution* of the DAE (2.1a) if it satisfies (2.1a) pointwise. It is called a *solution of the initial value problem* (2.1) with initial condition

$$z(t_0) = z_0 \in \mathbb{R}^n \tag{2.3}$$

if it furthermore satisfies (2.3).

(ii) An initial value $z_0 \in \mathbb{R}^n$ is called *consistent* with (2.1a) if the associated initial value problem has at least one solution.

(iii) The control problem (2.1) is called *consistent* if there exists an input $u\colon \mathbb{T} \to \mathbb{R}^m$ such that the resulting DAE (2.1a) has a solution. It is called *regular* if, for every sufficiently smooth input function $u\colon \mathbb{T} \to \mathbb{R}^m$, the corresponding DAE (2.1a) is solvable and the solution is unique for every consistent initial value.

(iv) We call $\xi = [z^\top, u^\top, y^\top]^\top$ a *behaviour solution of the descriptor system* (2.1) if z is a solution of the DAE (2.1a) for this u and ξ satisfies (2.1) pointwise.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Let us emphasize that for DAE systems, typically, not every initial value is consistent. This is due to the fact that in order to deal with algebraic constraints as well as over- and under-determined systems, we allow the Jacobian $(\partial/\partial\dot{\xi})F$ to be singular or even rectangular. We refer to Section 2.2 below for further details. For inconsistent initial values and systems with jumps in the coefficients, one may still obtain a solution using weaker solution concepts; see e.g. Kunkel and Mehrmann (2006), Rabier and Rheinboldt (1996*a*,*b*) or Trenn (2013). However, for ease of presentation, we will not cover these weaker solution concepts in this survey.

2.2. *Solution theory for general nonlinear descriptor systems*

In this subsection we recall the solution theory for general DAE systems

$$\mathcal{F}(t, \xi(t), \dot{\xi}(t)) = 0, \tag{2.4}$$

with $\mathcal{F}\colon \mathbb{T} \times \mathbb{D}_\xi \times \mathbb{D}_{\dot{\xi}} \to \mathbb{R}^L$ and open sets $\mathbb{D}_\xi, \mathbb{D}_{\dot{\xi}} \subseteq \mathbb{R}^N$. Here ξ is the standard state or an extended behaviour vector as in (2.2).

If the Jacobian $(\partial/\partial\dot{\xi})\mathcal{F}$ is not square or singular, then a solution ξ of (2.4), provided such a solution exists, may depend on derivatives of $\mathcal{F}$. This is illustrated in the following example.

Example 2.2. Consider a linear DAE of the form

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{z}_1(t) \\ \dot{z}_2(t) \\ \dot{z}_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_1(t) \\ z_2(t) \\ z_3(t) \end{bmatrix} + \begin{bmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \end{bmatrix}. \tag{2.5}$$

We immediately notice that z_3 does not contribute to the equations and hence can be chosen arbitrarily. Moreover, the third equation dictates $f_3 \equiv 0$, thus detailing that a smooth function $f = [f_1 \quad f_2 \quad f_3]^\top$ is not sufficient for a solution to exist. The second equation yields $z_1(t) = -f_2(t)$, and hence the only valid initial value for z_1 is determined by $-f_2(t_0)$. Substituting $z_1(t) = -f_2(t)$ into the first equation yields

$$z_2(t) = -f_1(t) - \dot{f}_2(t), \tag{2.6}$$

showing that the solution depends on the derivative of f_2. Moreover, we notice that (2.6) constitutes another algebraic equation that is implicitly encoded in (2.5).

The difficulties arising with these differentiations are classified by so-called *index* concepts; see Mehrmann (2015) for a survey. In this paper we mainly make use of the *strangeness index* concept (Kunkel and Mehrmann 2006), which is, roughly speaking, a generalization of the *differentiation index* (Brenan *et al.* 1996) to under- and overdetermined systems. The strangeness index is based on the *derivative array*

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

of level μ (Campbell 1987), defined by

$$\widetilde{\mathcal{F}}_\mu(t,\xi,\eta) := \begin{bmatrix} \mathcal{F}(t,\xi,\dot{\xi}) \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(t,\xi,\dot{\xi}) \\ \vdots \\ \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^\mu \mathcal{F}(t,\xi,\dot{\xi}) \end{bmatrix} \in \mathbb{R}^{(\mu+1)L} \quad \text{with} \quad \eta := \begin{bmatrix} \dot{\xi} \\ \ddot{\xi} \\ \vdots \\ \xi^{(\mu+1)} \end{bmatrix} \in \mathbb{R}^{(\mu+1)N}. \tag{2.7}$$

Since it is *a priori* not clear that the DAE (2.4) is solvable and that the dimension of the solution manifold in terms of the algebraic variables $t, \xi, \ldots, \xi^{(\mu+1)}$ is invariant over time, we need to assume that the set

$$\mathcal{M}_\mu := \left\{(t,\xi,\eta) \in \mathbb{R}^{(\mu+2)N+1} \mid \widetilde{\mathcal{F}}_\mu(t,\xi,\eta) = 0\right\} \tag{2.8}$$

is non-empty and (locally) forms a manifold. For notational convenience we assume that $\mathcal{M}_\mu$ is a manifold of dimension $(\mu+2)N+1-r$. The number r will later correspond to the dimension of the regular part of the DAE. Following Kunkel and Mehrmann (1998), we introduce the Jacobians

$$\mathcal{E}_\mu := \begin{bmatrix} \dfrac{\partial \widetilde{\mathcal{F}}_\mu}{\partial \dot{\xi}} & \cdots & \dfrac{\partial \widetilde{\mathcal{F}}_\mu}{\partial \xi^{(\mu+1)}} \end{bmatrix} \in \mathbb{R}^{(\mu+1)L,(\mu+1)N}, \tag{2.9a}$$

$$\mathcal{A}_\mu := -\begin{bmatrix} \dfrac{\partial \widetilde{\mathcal{F}}_\mu}{\partial \xi} & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{(\mu+1)L,(\mu+1)N}. \tag{2.9b}$$

In the following, we will make some constant rank assumptions, which in turn is the basis for a (local) smooth full-rank decomposition as provided in the next theorem; see Kunkel and Mehrmann (2006, Theorem 4.3).

Theorem 2.3. For open sets $\mathbb{M} \subseteq \mathbb{D} \subseteq \mathbb{R}^k$ let $E \in \mathcal{C}^\mu(\mathbb{D}, \mathbb{R}^{\ell,n})$. Furthermore, assume that $\operatorname{rank} E(z) \equiv r$ for all $z \in \mathbb{M}$. Then, for every $\hat{z} \in \mathbb{M}$, there exists a sufficiently small neighbourhood $\mathbb{V} \subseteq \mathbb{D}$ of $\hat{z}$, and matrix functions $T \in \mathcal{C}^\mu(\mathbb{V}, \mathbb{R}^{n,(n-r)})$ and $Z \in \mathcal{C}^\mu(\mathbb{V}, \mathbb{R}^{\ell,(\ell-r)})$ with pointwise orthonormal columns such that

$$E(z)T(z) = 0 \quad \text{and} \quad Z^\top(z)E(z) = 0 \quad \text{for all } z \in \mathbb{V}.$$

To analyse the nonlinear DAE (2.4) we now make the following assumption, taken from Kunkel and Mehrmann (2001) and presented as in Unger (2020), to filter out the regular part. Note that we use the term *corank* to denote the difference between the size of a matrix and its rank; see also (1.1) for a formal definition.

Assumption 2.4. Assume that the set $\mathcal{M}_\mu$ in (2.8) is a manifold of dimension $(\mu+2)N+1-r$ and the Jacobians defined in (2.9) satisfy

$$\operatorname{rank}[\mathcal{E}_\mu \quad \mathcal{A}_\mu] = r \quad \text{on } \mathcal{M}_\mu. \tag{2.10}$$

Moreover, we have

$$\operatorname{corank}[\mathcal{E}_\mu \quad \mathcal{A}_\mu] - \operatorname{corank}[\mathcal{E}_{\mu-1} \quad \mathcal{A}_{\mu-1}] = v \quad \text{on } \mathcal{M}_\mu \tag{2.11}$$

with the convention that $\operatorname{corank} \partial F_{-1}/\partial \xi = 0$.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

The quantity v in Assumption 2.4 measures the number of equations in the original system that give rise to trivial equations $0 = 0$, that is, it counts the number of redundancies in the system. After the quantification of the regular and redundant parts of the DAE (2.4), we use the next assumption to filter out the algebraic equations.

Assumption 2.5. Suppose the DAE (2.4) satisfies Assumption 2.4. Then the matrix $\mathcal{E}_\mu$ defined in (2.9) satisfies

$$\operatorname{rank} \mathcal{E}_\mu = r - a \quad \text{on } \mathcal{M}_\mu. \tag{2.12}$$

Assumption 2.5 together with Theorem 2.3 ensures (locally) the existence of a smooth matrix function $Z_{\mathrm{a}} \colon \mathcal{M}_\mu \to \mathbb{R}^{(\mu+1)L,a}$ with pointwise maximal rank on $\mathcal{M}_\mu$, that satisfies

$$Z_{\mathrm{a}}^\top \mathcal{E}_\mu = 0 \quad \text{on } \mathcal{M}_\mu. \tag{2.13}$$

The (linearized) algebraic equations are thus encoded in the matrix function

$$Z_{\mathrm{a}}^\top \frac{\partial \widetilde{\mathcal{F}}_\mu}{\partial \xi}. \tag{2.14}$$

To ensure that we are able to solve the algebraic equations for a unknowns requires the matrix in (2.14) to have full rank. This is indeed the case, since (2.13) together with Assumption 2.4 implies that

$$\operatorname{rank} Z_{\mathrm{a}}^\top \mathcal{A}_\mu = \operatorname{rank} Z_{\mathrm{a}} \frac{\partial \widetilde{\mathcal{F}}_\mu}{\partial \xi} = a.$$

Again, Theorem 2.3 implies (locally) the existence of a smooth matrix function $T_{\mathrm{a}} \colon \mathcal{M}_\mu \to \mathbb{R}^{N,N-a}$ with pointwise maximal rank satisfying

$$Z_{\mathrm{a}}^\top \frac{\partial \widetilde{\mathcal{F}}_\mu}{\partial \xi} T_{\mathrm{a}} = 0 \quad \text{on } \mathcal{M}_\mu. \tag{2.15}$$

The remaining differential equations must be contained in the original DAE (2.4) (in contrast to the algebraic equations, which are contained in the derivative array), and thus we make the following additional assumption.

Assumption 2.6. Let Assumptions 2.4 and 2.5 hold and let T_{a} be constructed as in (2.15). Define $d := L - a - v$ and assume that

$$\operatorname{rank} \frac{\partial \mathcal{F}}{\partial \dot{\xi}} T_{\mathrm{a}} = d \quad \text{on } \mathcal{M}_\mu.$$

Once again, we employ Theorem 2.3 to (locally) obtain a smooth matrix function Z_{d} of size $N \times d$ with pointwise maximal rank that satisfies $Z_{\mathrm{d}}^\top (\partial \mathcal{F} / \partial \dot{\xi}) T_{\mathrm{a}} = d$. The matrix function Z_{d} will later be used to filter out the differential equations.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

To summarize the previous discussion, we make the following assumption, which for historical reasons (Kunkel and Mehrmann 2001) and since in the linear case it is actually a theorem, is referred to as a hypothesis. Note that due to the local character of Theorem 2.3, all assumptions hold only in a suitable neighbourhood.

Hypothesis 2.7. There exist integers μ, r, a and v such that $\mathcal{M}_\mu$ defined in (2.8) is non-empty, and such that for every $(t_0, \xi_0, \eta_0) \in \mathcal{M}_\mu$ there exists a (sufficiently small) neighbourhood $\mathcal{U}$ in which the following properties hold.

(i) The set $\mathcal{M}_\mu$ forms a manifold of dimension $(\mu+2)N + 1 - r$.

(ii) We have $\operatorname{rank}[\mathcal{E}_\mu \quad \mathcal{A}_\mu] = r$ on $\mathcal{M}_\mu \cap \mathcal{U}$.

(iii) We have $\operatorname{corank}[\mathcal{E}_\mu \quad \mathcal{A}_\mu] - \operatorname{corank}[\mathcal{E}_{\mu-1} \quad \mathcal{A}_{\mu-1}] = v$ on $\mathcal{M}_\mu \cap \mathcal{U}$ (with the convention $\operatorname{corank}[\mathcal{E}_{-1} \quad \mathcal{A}_{-1}] = 0$).

(iv) We have $\operatorname{rank} \mathcal{E}_\mu = r - a$ on $\mathcal{M}_\mu \cap \mathcal{U}$, such that there exist smooth matrix functions Z_{a} and T_{a} of size $(\mu+1)L \times a$ and $N \times (N-a)$, respectively, and pointwise maximal rank, satisfying $Z_{\mathrm{a}}^\top \mathcal{E}_\mu = 0$, $\operatorname{rank} Z_{\mathrm{a}}^\top \mathcal{A}_\mu = a$ and $Z_{\mathrm{a}}^\top (\partial \tilde{F}_\mu / \partial \xi) T_{\mathrm{a}} = 0$ on $\mathcal{M}_\mu \cap \mathcal{U}$.

(v) We have $\operatorname{rank} (\partial \mathcal{F} / \partial \dot{\xi}) T_{\mathrm{a}} = d := L - a - v$ on $\mathcal{M}_\mu \cap \mathcal{U}$ such that there exists a smooth matrix function Z_{d} of size $N \times d$ and pointwise maximal rank, satisfying $\operatorname{rank} Z_{\mathrm{d}}^\top (\partial \tilde{F} / \partial \dot{\xi}) T_{\mathrm{a}} = d$.

Definition 2.8. Given the DAE (2.4), the smallest value μ such that $\mathcal{F}$ satisfies Hypothesis 2.7 is called the *strangeness index* of (2.4). If $\mu = 0$, then the DAE is called *strangeness-free*.

Remark 2.9. Hypothesis 2.7 is invariant under a large class of equivalence transformations (Kunkel and Mehrmann 2006, Section 4.1), which is why the numbers μ, r, a, v and d are referred to as *characteristic values* for the DAE (2.4).

Following the discussion in Kunkel and Mehrmann (2001), we can use the matrix functions Z_a and Z_d to construct the DAE

$$\widehat{\mathcal{F}}(t, \xi, \dot{\xi}) := \begin{bmatrix} \widehat{\mathcal{F}}_{\mathrm{d}}(t, \xi, \dot{\xi}) \\ \widehat{\mathcal{F}}_{\mathrm{a}}(t, \xi) \end{bmatrix} \tag{2.16}$$

with

$$\hat{F}_{\mathrm{d}}(t, \xi, \dot{\xi}) := (Z_{\mathrm{d}}^\top \mathcal{F})(t, \xi, \dot{\xi}) \quad \text{and} \quad \hat{F}_{\mathrm{a}}(t, \xi) := (Z_{\mathrm{a}}^\top \widetilde{\mathcal{F}}_\mu)(t, \xi).$$

Note that although the matrix functions Z_{a} and Z_{d} depend on derivatives of ξ, it is possible to show (Kunkel and Mehrmann 2001) that the reduced quantities $\widehat{\mathcal{F}}_{\mathrm{a}}$ and $\widehat{\mathcal{F}}_{\mathrm{d}}$ are independent of higher derivatives of ξ. In addition, one can show that (2.16) satisfies Hypothesis 2.7 with characteristic values $\mu = 0$, r, a and v. In particular, (2.16) is strangeness-free.

In the regular case, where $N = L$ and $v = 0$, we can simplify Hypothesis 2.7 as follows; see also Kunkel and Mehrmann (1998).

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Hypothesis 2.10. There exist integers μ and a such that the set $\mathcal{M}_\mu$ defined in (2.8) is non-empty, and such that for every $(t_0, \xi_0, \eta_0) \in \mathcal{M}_\mu$ there exists a (sufficiently small) neighbourhood $\mathcal{U}$ in which the following properties hold.

(i) We have rank $\mathcal{E}_\mu = (\mu+1)N - a$ on $\mathcal{M}_\mu \cap \mathcal{U}$ such that there exists a smooth matrix function Z_{a} of size $(\mu+1)N \times a$ and pointwise maximal rank such that $Z_{\mathrm{a}}^\top \mathcal{E}_\mu = 0$ on $\mathcal{M}_\mu \cap \mathcal{U}$.

(ii) We have rank $Z_{\mathrm{a}}^\top \mathcal{A}_\mu = a$ on $\mathcal{M}_\mu \cap \mathcal{U}$ such that there exists a smooth matrix function T_{a} of size $N \times d$ with $d := N - a$ and pointwise maximal rank, satisfying $Z_{\mathrm{a}}^\top (\partial \tilde{F}_\mu / \partial \xi) T_{\mathrm{a}} = 0$ on $\mathcal{M}_\mu \cap \mathcal{U}$.

(iii) We have rank $(\partial \mathcal{F} / \partial \dot{\xi}) T_{\mathrm{a}} = d := L - a$ on $\mathcal{M}_\mu \cap \mathcal{U}$ such that there exists a smooth matrix function Z_{d} of size $N \times d$ and pointwise maximal rank, satisfying rank $Z_{\mathrm{d}}^\top (\partial \tilde{F} / \partial \dot{\xi}) T_{\mathrm{a}} = d$.

The relation between the original DAE (2.4) and the strangeness-free reformulation (2.16) is given in the following theorem, taken from Kunkel and Mehrmann (2006, Theorems 4.11 and 4.13). For ease of presentation we will focus on the regular case using Hypothesis 2.10, and remark that a similar result is also available for the more general setting described in Hypothesis 2.7; see Kunkel and Mehrmann (2001) for further details.

Theorem 2.11. Let $\mathcal{F}$ as in (2.4) be sufficiently smooth and satisfy Hypothesis 2.10 with characteristic values μ, a and $d := N - a$. Then the following statements hold.

(i) Every sufficiently smooth solution of (2.4) also solves the strangeness-free DAE (2.16).

(ii) Suppose additionally that $\mathcal{F}$ satisfies Hypothesis 2.10 with characteristic values $\mu+1$, a and d. Then, for every $(t_0, \xi_0, \eta_0) \in \mathcal{M}_{\mu+1}$, the strangeness-free problem (2.16) has a unique solution satisfying the initial condition $\xi(t_0) = \xi_0$. Moreover, this solution locally solves the original problem (2.4).

Remark 2.12. The relation of the strangeness index concept as presented above to other index concepts commonly used in the theory of DAEs, such as the *differentiation index* (Campbell and Gear 1995), the *perturbation index* (Hairer and Wanner 1996), the *tractability index* (Lamour *et al.* 2013), the *geometric index* (Rheinboldt 1984, Reich 1990) and the *structural index* (Pantelides 1988, Pryce 2001), is discussed in Mehrmann (2015).

2.3. *Linear time-varying DAE systems*

If the DAE (2.4) is linear time-varying, i.e. of the form

$$E(t)\dot{\xi}(t) = A(t)\xi(t) + f(t), \quad z(t_0) = z_0, \tag{2.17}$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

with smooth matrix functions $E, A\colon \mathbb{T} \to \mathbb{R}^{L,N}$, then the analysis of the previous subsection can be further simplified. In this case the Jacobians (2.9) are given by

$$(\mathcal{E}_\mu)_{i,j} = \binom{i}{j} E^{(i-j)} - \binom{i}{j+1} A^{(i-j-1)}, \quad i,j = 0,\dots,\mu,$$
$$(\mathcal{A}_\mu)_{i,j} = \begin{cases} A^{(i)} & \text{for } i = 0,\dots,\mu,\ j = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Since these matrix functions do not depend on the state variable z or its derivatives, we can get rid of the local character of Hypothesis 2.7 (respectively Hypothesis 2.10) by using the following simplified version of the smooth rank-revealing decomposition (see Theorem 2.3); see e.g. Kunkel and Mehrmann (2006, Theorem 3.9).

Theorem 2.13. Let $E \in \mathcal{C}^\mu(\mathbb{T}, \mathbb{R}^{L,N})$, $\mu \in \mathbb{N}_0 \cup \{\infty\}$, with $\operatorname{rank} E(t) = r$ for all $t \in \mathbb{T}$. Then there exist pointwise orthogonal functions $U \in \mathcal{C}^\mu(\mathbb{T}, \mathbb{R}^{L,L})$ and $V \in \mathcal{C}^\mu(\mathbb{T}, \mathbb{R}^{N,N})$ such that

$$U^\top E V = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$$

with pointwise nonsingular $\Sigma \in \mathcal{C}^\mu(\mathbb{T}, \mathbb{R}^{r,r})$.

In general, we can now proceed as in Hypothesis 2.7 and construct the matrix functions Z_a and Z_d. To avoid checking that $\mathcal{M}_\mu$ is non-empty, we further construct a matrix function Z_v of size $(\mu+1)N \times \nu$ with pointwise maximal rank that filters out equations that do not depend on ξ and its derivatives. If this number of equations is non-zero, we check whether the right-hand side vanishes as well. If this is the case, then we omit these equations. If not, then $\mathcal{M}_\mu = \emptyset$ and the problem has to be regularized; see Kunkel and Mehrmann (2006). Defining

$$g_\mu := \begin{bmatrix} f^\top & \left(\frac{\mathrm{d}}{\mathrm{d}t} f\right)^\top & \cdots & (f^{(\mu)})^\top \end{bmatrix}^\top$$

and

$$\hat{E}_1 := Z_\mathrm{d}^\top E, \quad \hat{A}_1 := Z_\mathrm{d}^\top A, \quad \hat{A}_2 := Z_\mathrm{a}^\top \mathcal{A}_\mu,$$
$$\hat{f}_1 := Z_\mathrm{d}^\top f, \quad \hat{f}_2 := Z_\mathrm{a}^\top g_\mu, \quad \hat{f}_3 := Z_\mathrm{v}^\top g_\mu,$$

we obtain the solution equivalent strangeness-free system

$$\begin{bmatrix} \hat{E}_1(t) \\ 0 \\ 0 \end{bmatrix} \dot{\xi}(t) = \begin{bmatrix} \hat{A}_1(t) \\ \hat{A}_2(t) \\ 0 \end{bmatrix} \xi(t) + \begin{bmatrix} \hat{f}_1(t) \\ \hat{f}_2(t) \\ \hat{f}_3(t) \end{bmatrix}. \tag{2.18}$$

Remark 2.14. In the behaviour case for a control system, where the state variable is given by $\xi = [z^\top \quad u^\top]^\top$, the matrix functions E and A have a block column structure, where the second block column corresponds to the control. Since the constructed coefficients $\hat{A}_1$ and $\hat{A}_2$ are obtained by transformations of the derivative array from the left, the block column structure of A is retained in these matrices.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Moreover, the strangeness-free reformulation does not depend on derivatives of the control u.

If we further allow transformations of the solution space via a pointwise nonsingular matrix function, then we can obtain the following solvability result for the DAE (2.17).

Theorem 2.15. Under some constant rank assumptions the DAE (2.17) is equivalent, in the sense that there is a change of basis in the solution space via a pointwise nonsingular matrix function, to a DAE of the form

$$\begin{aligned} \dot{\xi}_1(t) &= \hat{A}_{13}(t)\xi_3 + \hat{f}_1(t), \\ 0 &= \xi_2(t) + \hat{f}_2(t), \\ 0 &= \hat{f}_3(t), \end{aligned}$$

where $A_{13} \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{d,N-d-a})$ and $\hat{f}_1 \in \mathcal{C}(\mathbb{T}, \mathbb{R}^d)$, $\hat{f}_2 \in \mathcal{C}(\mathbb{T}, \mathbb{R}^a)$, $\hat{f}_3 \in \mathcal{C}(\mathbb{T}, \mathbb{R}^v)$ are determined from g_μ.

(i) If $f \in \mathcal{C}^{\mu+1}(\mathbb{T}, \mathbb{R}^L)$, then (2.17) is solvable if and only if $\hat{f}_3 = 0$.

(ii) An initial value is consistent if and only if in addition the condition

$$\xi_2(t_0) = -\hat{f}_2(t_0)$$

is implied by the initial condition.

(iii) The initial value problem is uniquely solvable if and only if in addition $N - d - a = 0$.

2.4. Linear time-invariant DAE systems

In principle, we can perform the analysis for the linear time-varying case as well in the case of general constant coefficient linear DAE systems

$$E\dot{\xi}(t) = A\xi(t) + f(t), \quad \xi(t_0) = \xi_0, \tag{2.19}$$

with matrices $E, A \in \mathbb{R}^{L,N}$, also referred to *linear time-invariant* (LTI) DAE systems. However, in this setting it is common to work with an equivalence transformation and a corresponding canonical form. For notational convenience, for the next result we also allow complex-valued matrices in (2.19) and work in the field of complex numbers.

We call the matrix pencils $sE_i - A_i$ with $E_i, A_i \in \mathbb{C}^{L,N}$, $i = 1, 2$ *(strongly) equivalent* if there exist nonsingular matrices $S \in \mathbb{C}^{L,L}$ and $T \in \mathbb{C}^{N,N}$ such that

$$S(\lambda E_1 - A_1)T = \lambda E_2 - A_2 \quad \text{for all } \lambda \in \mathbb{C}.$$

In this case we write $\lambda E_1 - A_1 \sim \lambda E_2 - A_2$. The associated canonical form is given by the Kronecker canonical form; see e.g. Gantmacher (1959).

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Theorem 2.16 (Kronecker canonical form). Let $E, A \in \mathbb{C}^{L,N}$. Then

$$\lambda E - A \sim \operatorname{diag}\big(\mathcal{L}_{\epsilon_1}, \dots, \mathcal{L}_{\epsilon_p}, \mathcal{L}_{\eta_1}^\top, \dots, \mathcal{L}_{\eta_q}^\top, \mathcal{J}_{\rho_1}^{\lambda_1}, \dots, \mathcal{J}_{\rho_r}^{\lambda_r}, \mathcal{N}_{\sigma_1}, \dots, \mathcal{N}_{\sigma_s}\big),$$

where the block entries have the following properties.

(i) Every entry $\mathcal{L}_{\epsilon_j}$ is a bidiagonal block of size $\epsilon_j \times (\epsilon_j + 1)$, $\epsilon_j \in \mathbb{N}_0$, of the form

$$\lambda \begin{bmatrix} 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \end{bmatrix}.$$

(ii) Every entry $\mathcal{L}_{\eta_j}^\top$ is a bidiagonal block of size $(\eta_j + 1) \times \eta_j$, $\eta_j \in \mathbb{N}_0$, of the form

$$\lambda \begin{bmatrix} 1 & & \\ 0 & \ddots & \\ & \ddots & 1 \\ & & 0 \end{bmatrix} - \begin{bmatrix} 0 & & \\ 1 & \ddots & \\ & \ddots & 0 \\ & & 1 \end{bmatrix}.$$

(iii) Every entry $\mathcal{J}_{\rho_j}^{\lambda_j}$ is a Jordan block of size $\rho_j \times \rho_j$, $\rho_j \in \mathbb{N}$, $\lambda_j \in \mathbb{C}$, of the form

$$\lambda \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} - \begin{bmatrix} \lambda_j & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_j \end{bmatrix}.$$

(iv) Every entry $\mathcal{N}_{\sigma_j}$ is a nilpotent block of size $\sigma_j \times \sigma_j$, $\sigma_j \in \mathbb{N}$, of the form

$$\lambda \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix} - \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

The Kronecker canonical form is unique up to permutation of the blocks.

Remark 2.17. If the matrices are real-valued and we want to stay within the field of real numbers, then only real-valued transformation matrices S, T may be used. The corresponding canonical form is called the *real Kronecker canonical form*. Here the blocks $\mathcal{J}_{\rho_j}^{\lambda_j}$ with $\lambda_j \in \mathbb{C} \setminus \mathbb{R}$ are in real Jordan canonical form instead, but the other blocks are as in the complex case.

A value $\lambda_0 \in \mathbb{C}$ is called a *(finite) eigenvalue* of $\lambda E - A$ if

$$\operatorname{rank}(\lambda_0 E - A) < \max_{\alpha \in \mathbb{C}} \operatorname{rank}(\alpha E - A).$$

If zero is an eigenvalue of $\lambda A - E$, then $\lambda_0 = \infty$ is said to be an eigenvalue of $\lambda E - A$.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

The blocks $\mathcal{J}_{\rho_j}$ correspond to finite eigenvalues and the blocks $\mathcal{N}_{\sigma_j}$ correspond to the eigenvalue ∞. The size of the largest block $\mathcal{N}_{\sigma_j}$ is called the *(Kronecker) index* ν of the pencil $\lambda E - A$, where, by convention, $\nu = 0$ if E is invertible. A finite eigenvalue is called *semisimple* if the largest Jordan block $\mathcal{J}_{\rho_j}$ associated with this block has $\rho_j = 1$. The pencil $\lambda E - A$ is called *regular* if $N = L$ and $\det(\lambda_0 E - A) \neq 0$ for some $\lambda_0 \in \mathbb{C}$. For regular pencils $\lambda E - A$, the Kronecker canonical form simplifies to the *Weierstrass* canonical form.

Theorem 2.18 (Weierstrass canonical form). Assume that the pencil $\lambda E - A$ with $E, A \in \mathbb{C}^{N,N}$ ($E, A \in \mathbb{R}^{N,N}$) is regular. Then

$$\lambda E - A \sim \lambda \begin{bmatrix} I & 0 \\ 0 & \mathcal{N} \end{bmatrix} - \begin{bmatrix} \mathcal{J} & 0 \\ 0 & I \end{bmatrix}, \tag{2.20}$$

where $\mathcal{J}$ and $\mathcal{N}$ are in Jordan (real Jordan) canonical form and $\mathcal{N}$ is nilpotent.

If the pencil is not regular then there may not exist a solution of (2.19) or it may not be unique (see e.g. Example 2.2), while in the regular case we have the following theorem; see Kunkel and Mehrmann (2006) for the complex case.

Theorem 2.19. Consider a regular matrix pencil $\lambda E - A$ of real square matrices E, A and let S and T be nonsingular matrices which transform (2.19) to its real Weierstrass canonical form (2.20), that is,

$$SET = \begin{bmatrix} I & 0 \\ 0 & \mathcal{N} \end{bmatrix}, \quad SAT = \begin{bmatrix} \mathcal{J} & 0 \\ 0 & I \end{bmatrix}, \quad Sf = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix},$$

where $\mathcal{J}, \mathcal{N}$ are in real Jordan canonical form and $\mathcal{N}$ is nilpotent of nilpotency index ν. Set

$$T^{-1}\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}, \quad T^{-1}\xi_0 = \begin{bmatrix} \xi_{1,0} \\ \xi_{2,0} \end{bmatrix}$$

with analogous partitioning. If $f \in \mathcal{C}^\nu(\mathbb{T}, \mathbb{R}^N)$, then the DAE (2.19) is solvable. An initial value is consistent if and only if

$$\xi_{2,0} = -\sum_{i=0}^{\nu-1} \mathcal{N}^i f_2^{(i)}(t_0).$$

In particular, the set of consistent initial values ξ_0 is non-empty, and every initial value problem with consistent initial condition is uniquely solvable.

Remark 2.20. To clarify the difference between the (Kronecker) index and the strangeness index, observe that if a regular LTI DAE system has (Kronecker) index $\nu > 0$, then its strangeness index is $\mu = \nu - 1$, and if $\nu = 0$, then also $\mu = 0$; see also Mehrmann (2015).

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

3. Control concepts for general DAE systems

In this section we discuss different aspects related to control theory of general DAE systems. Most of our discussion focuses on linear descriptor systems of the form

$$E\dot{z} = Az + Bu, \tag{3.1a}$$

$$y = Cz, \tag{3.1b}$$

with either

- matrices $E, A \in \mathbb{R}^{n,n}$, $B \in \mathbb{R}^{n,m}$, $C \in \mathbb{R}^{p,n}$ for the LTI case, or
- matrix functions $E, A\colon \mathbb{T} \to \mathbb{R}^{n,n}$, $B\colon \mathbb{T} \to \mathbb{R}^{n,m}$, $C\colon \mathbb{T} \to \mathbb{R}^{p,n}$ for the *linear time-varying* (LTV) case.

Remark 3.1. In general, the descriptor system (3.1) may also include a feedthrough term, that is, the output equation (3.1b) is given by

$$y = Cz + Du$$

with a suitable matrix or matrix function D. However, in the DAE context, we can rewrite the descriptor system (3.1) without the feedthrough term, as follows. Consider any decomposition $D = D_{\mathrm{c}}D_{\mathrm{b}}$ and the extended system matrices or matrix functions

$$\hat{E} := \begin{bmatrix} E & 0 \\ 0 & 0 \end{bmatrix}, \quad \hat{A} := \begin{bmatrix} A & 0 \\ 0 & -I \end{bmatrix}, \quad \hat{B} := \begin{bmatrix} B \\ D_{\mathrm{b}} \end{bmatrix}, \quad \hat{C} := \begin{bmatrix} C & D_{\mathrm{c}} \end{bmatrix}.$$

Then the solution of the associated descriptor system contains the solution of the descriptor system with feedthrough term.

3.1. Feedback regularization

As we have seen in Section 2, a DAE system may not be regular, that is, there may not be any initial values such that the initial value problem has a solution, or a solution for a consistent initial value may not be unique. To deal with this situation, we first discuss how to regularize a descriptor system via instantaneous, proportional (linear) state or output feedback, i.e. via feedback laws of the form

$$u = F_1 z + w \quad \text{or} \quad u = F_2 y + w, \tag{3.2}$$

respectively, with suitable matrices or matrix functions F_1 and F_2. After applying such a feedback, the closed-loop system matrices, respectively matrix functions, are given by $\tilde{E} := E$ and

$$\tilde{A}_1 := A + BF_1 \quad \text{and} \quad \tilde{A}_2 := A + BF_2C,$$

respectively.

We start our analysis for the LTI case and recall important conditions for controllability and observability. If the matrix E in (3.1) is nonsingular, then the

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

well-known Hautus lemma (e.g. Dai 1989) asserts that the LTI descriptor system (3.1) is controllable if and only if

$$\operatorname{rank}[\lambda E - A \quad B] = n \quad \text{for all } \lambda \in \mathbb{C}. \tag{3.3}$$

If E is singular, then the situation is more involved and we need the following conditions, taken for instance from Dai (1989) or Bunse-Gerstner, Byers, Mehrmann and Nichols (1999).

Definition 3.2. Consider the LTI descriptor system (3.1) and let S_∞ be a matrix with columns that span the kernel of E.

(i) The system (3.1) is called *controllable at* ∞ or *impulse controllable* if

$$\operatorname{rank}[E \quad AS_\infty \quad B] = n. \tag{3.4}$$

(ii) The system (3.1) is called *strongly controllable* if it is impulse controllable and (3.3) is satisfied.

(iii) The system (3.1) is called *strongly stabilizable* if it is impulse controllable and (3.3) holds for all $\lambda \in \mathbb{C}$ with $\operatorname{Re}(\lambda) \geq 0$.

The corresponding dual conditions with respect to the output equation are given as

$$\operatorname{rank}[\lambda E^\top - A^\top \quad C^\top] = n, \tag{3.5}$$

$$\operatorname{rank}[E^\top \quad A^\top T_\infty \quad C^\top] = n, \tag{3.6}$$

respectively, where T_∞ is a matrix that spans the kernel of $E^\top$.

Definition 3.3. Consider the LTI descriptor system (3.1) and let T_∞ be a matrix with columns that span the kernel of $E^\top$.

(i) The system (3.1) is called *observable at* ∞ or *impulse observable* if condition (3.6) is satisfied.

(ii) The system (3.1) is called *strongly observable* if it is impulse observable and if (3.5) holds for all $\lambda \in \mathbb{C}$.

(iii) The system (3.1) is called *strongly detectable* if it is impulse observable and if (3.5) holds for all $\lambda \in \mathbb{C}$ with $\operatorname{Re}(\lambda) \geq 0$.

A system that satisfies conditions (3.3) and (3.5) is called *minimal*.

Conditions (3.5) and (3.6) are preserved under nonsingular equivalence transformations as well as under state and output feedback. More precisely, if the system satisfies (3.5) and (3.6), then for any nonsingular $U \in \mathbb{R}^{n,n}$, $V \in \mathbb{R}^{n,n}$, and any $F_1 \in \mathbb{R}^{m,n}$ and $F_2 \in \mathbb{R}^{m,p}$, the system with coefficients $(\tilde{E}, \tilde{A}, \tilde{B}, \tilde{C})$ satisfies the same condition for all of the following three choices:

$$\begin{aligned} &\tilde{E} = UEV, \quad &&\tilde{A} = UAV, \quad &&\tilde{B} = UB, \\ &\tilde{E} = E, \quad &&\tilde{A} = A + BF_1, \quad &&\tilde{B} = B, \\ &\tilde{E} = E, \quad &&\tilde{A} = A + BF_2C, \quad &&\tilde{B} = B. \end{aligned}$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Analogous invariance properties hold for (3.5) and (3.6). Further details and properties of LTI DAE systems are discussed in Berger and Reis (2013).

Note, however, that regularity or non-regularity of the pencil and the (Kronecker) index are in general not preserved under state or output feedback, respectively. On the contrary, feedback of the form (3.2) may be used to regularize the system, as detailed in the following theorem taken from Bunse-Gerstner *et al.* (1999).

Theorem 3.4. Consider the LTI descriptor system (3.1).

(i) If (3.1) is impulse controllable, i.e. condition (3.4) is satisfied, then there exists a suitable linear state feedback matrix F_1 such that $\lambda E - (A + BF_1)$ is regular and of (Kronecker) index $\nu \leq 1$.

(ii) If (3.1) is impulse controllable and impulse observable, i.e. conditions (3.4) and (3.6) hold, then there exists a linear output feedback matrix F_2 such that the pencil $\lambda E - (A + BF_2C)$ is regular and of (Kronecker) index $\nu \leq 1$.

Remark 3.5. Although instantaneous feedback is a convenient theoretical approach, it may suffer from the fact that signals have to be measured first, and some calculations have to be carried out, thus resulting in an intrinsically necessary time delay. If this time delay cannot be ignored in the modelling phase, then for some $\tau > 0$ the feedback takes the form

$$u(t) = F_1 z(t - \tau) + w(t) \quad \text{or} \quad u(t) = F_2 y(t - \tau) + w(t),$$

thus rendering the closed-loop system a delay DAE. However, the DAE can be regularized with delayed feedback if and only if it can be regularized with instantaneous feedback; see Trenn and Unger (2019) and Unger (2020) for further details. Nevertheless, we always assume that the feedback delay can be ignored in the modelling phase within this survey.

For the feedback regularization in the LTV and nonlinear case, we follow Campbell, Kunkel and Mehrmann (2012) and use the behaviour approach as introduced in (2.2). In more detail, for the LTV descriptor system (3.1), we form the (matrix) functions

$$\xi := \begin{bmatrix} z \\ u \end{bmatrix}, \quad \mathcal{E} := [E \quad 0], \quad \mathcal{A} := [A \quad B]. \tag{3.7}$$

Ignoring the fact that ξ is composed of parts that may have quite different orders of differentiability, we form the derivative array (2.7) and follow the approach presented in Section 2.3. In more detail, we construct matrices $\widehat{\mathcal{E}}_1$, $\widehat{\mathcal{A}}_1$ and $\widehat{\mathcal{A}}_2$, such that the system

$$\begin{bmatrix} \widehat{\mathcal{E}}_1(t) \\ 0 \\ 0 \end{bmatrix} \dot{\xi}(t) = \begin{bmatrix} \widehat{\mathcal{A}}_1(t) \\ \widehat{\mathcal{A}}_2(t) \\ 0 \end{bmatrix} \xi(t) \tag{3.8}$$

is solution equivalent to (3.1) and strangeness-free. Since the matrix functions

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

are obtained solely by transformations of the derivative array from the left, the partitioning of the matrices as introduced in (3.7) is retained in (3.8); see also Remark 2.14. In particular, the state z and the input function u are not mixed, such that we can rewrite (3.8) as

$$\begin{bmatrix} \hat{E}_1(t) \\ 0 \\ 0 \end{bmatrix} \dot{z}(t) = \begin{bmatrix} \hat{A}_1(t) \\ \hat{A}_2(t) \\ 0 \end{bmatrix} z(t) + \begin{bmatrix} \hat{B}_1(t) \\ \hat{B}_2(t) \\ 0 \end{bmatrix} u(t), \tag{3.9}$$

with $\widehat{\mathcal{E}}_1 = [\hat{E}_1 \quad 0]$, $\widehat{\mathcal{A}}_1 = [\hat{A}_1 \quad \hat{B}_1]$ and $\widehat{\mathcal{A}}_2 = [\hat{A}_2 \quad \hat{B}_2]$.

Remark 3.6. We have constructed (3.8) such that the system is strangeness-free (with respect to the combined state variable ξ). Since (3.9) is simply obtained by rewriting (3.8), it is also strangeness-free with respect to ξ. However, it may not be strangeness-free with respect to the original state variable z (in the sense that we assume u to be given). To distinguish this subtlety in the following, we say that a descriptor system is strangeness-free *as a free system* if it is strangeness-free with respect to z for given input $u \equiv 0$.

To theoretically analyse the regularizability via feedback control, we use the following condensed form; see Kunkel, Mehrmann and Rath (2001).

Theorem 3.7. Consider the LTV descriptor system (3.1) and assume that the corresponding behaviour system defined in (3.7) has a well-defined strangeness index with strangeness-free form (3.8). Then, under some constant rank assumptions, there exist pointwise nonsingular matrix functions $S_z \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{n,n})$, $T_z \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{n,n})$, $S_y \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{p,p})$, $T_u \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{m,m})$ such that setting

$$z = T_z \begin{bmatrix} z_1^\top & z_2^\top & z_3^\top & z_4^\top \end{bmatrix}^\top, \quad u = T_u \begin{bmatrix} u_1^\top & u_2^\top \end{bmatrix}^\top, \quad y = S_y \begin{bmatrix} y_1^\top & y_2^\top \end{bmatrix}^\top,$$

and multiplying (3.9) by appropriate matrix functions from the left, yields a transformed control system of the form

$$\begin{aligned} \dot{z}_1 &= A_{13}(t)z_3 + A_{14}(t)z_4 + B_{12}(t)u_2, && d, && \text{(3.10a)} \\ 0 &= z_2 + B_{22}(t)u_2, && a - \phi, && \text{(3.10b)} \\ 0 &= A_{31}(t)z_1 + u_1, && \phi, && \text{(3.10c)} \\ 0 &= 0, && \nu, && \text{(3.10d)} \\ y_1 &= z_3, && \omega, && \text{(3.10e)} \\ y_2 &= C_{21}(t)z_1 + C_{22}(t)z_2, && p - \omega, && \text{(3.10f)} \end{aligned}$$

where the number at the end of each block equation denotes the number of equations within this block.

Corollary 3.8. Let the assumptions be as in Theorem 3.7. Furthermore, let the quantities ϕ and ω defined in (3.10) be constant. Then the following properties hold.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

(i) The LTV system (3.1) is consistent. Equation (3.10d) describes a redundancy in the system that can be omitted.

(ii) If $\phi = 0$, then for a given input function u, an initial value is consistent if and only if it implies (3.10b). Solutions of the corresponding initial value problem will generally not be unique.

(iii) The system is regular and strangeness-free (as a free system) if and only if $v = \phi = 0$ and $d + a = n$.

Analogous to the constant coefficient case, we can use proportional feedback to modify some of the system properties. However, the following result (Kunkel and Mehrmann 2006, Theorem 3.80) states that some properties stay invariant.

Theorem 3.9. Consider the LTV descriptor system (3.1) and suppose that the assumptions of Theorem 3.7 are satisfied. Then the characteristic values d, a and v are invariant under proportional state feedback and proportional output feedback.

The strangeness index (as a free system) as well as the regularity of the system can, however, be modified by proportional feedback; see Kunkel *et al.* (2001).

Corollary 3.10. Let the assumptions of Corollary 3.8 hold.

(i) There exists a state feedback $u = Fz + w$ such that the closed-loop system

$$E\dot{z} = (A + BF)z + Bw$$

is regular (as a free system) if and only if $v = 0$ and $d + a = n$.

(ii) There exists an output feedback $u = Fy + w$ such that the closed-loop system

$$E\dot{z} = (A + BFC)z + Bw$$

is regular (as a free system) if and only if $v = 0$, $d + a = n$ and $\phi = \omega$.

A similar local result is also available for nonlinear descriptor systems of the form (2.16); see Campbell *et al.* (2012) for details.

3.2. Stability

One of the key questions in control is whether a system can be stabilized via feedback control. In this section we therefore recall the stability theory for *ordinary differential equations* (ODEs) and discuss how these concepts are generalized to DAE systems. The classical stability concepts for ODEs are as follows; see e.g. Hinrichsen and Pritchard (2005). Consider an ODE of the form

$$\dot{z} = f(t, z), \quad t \in \mathbb{T}_\infty = [t_0, \infty] \tag{3.11}$$

and denote the solution satisfying the initial condition $z(t_0) = z_0$ by $z(\cdot; t_0, z_0)$.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Definition 3.11. We classify the solution $z(\cdot; t_0, z_0)$ of (3.11) as follows.

(i) A solution is *stable* if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $\hat{z}_0 \in \mathbb{R}^n$ with $\|\hat{z}_0 - z_0\| < \delta$,
- the initial value problem (3.11) with initial condition $z(t_0) = \hat{z}_0$ is solvable on $\mathbb{T}_\infty$ and
- the solution $z(t; t_0, \hat{z}_0)$ satisfies $\|z(t; t_0, \hat{z}_0) - z(t; t_0, z_0)\| < \varepsilon$ on $\mathbb{T}_\infty$.

(ii) A solution is *asymptotically stable* if it is stable and there exists $\varrho > 0$ such that, for all $\hat{z}_0 \in \mathbb{R}^n$ with $\|\hat{z}_0 - z_0\| < \rho$,
- the initial value problem (3.11) with initial condition $z(t_0) = \hat{z}_0$ is solvable on $\mathbb{T}_\infty$ and
- the solution $z(t; t_0, \hat{z}_0)$ satisfies $\lim_{t\to\infty} \|z(t; t_0, \hat{z}_0) - z(t; t_0, z_0)\| = 0$.

(iii) A solution is *exponentially stable* if it is stable and *exponentially attractive*, that is, if there exist $\delta > 0$, $L > 0$ and $\gamma > 0$ such that, for all $\hat{z}_0 \in \mathbb{R}^n$ with $\|\hat{z}_0 - z_0\| < \delta$,
- the initial value problem (3.11) with initial condition $z(t_0) = \hat{z}_0$ is solvable on $\mathbb{T}_\infty$ and
- the solution satisfies the estimate

$$\|z(t; t_0, \hat{z}_0) - z(t; t_0, z_0)\| < L\mathrm{e}^{-\gamma(t-t_0)} \text{ on } \mathbb{T}_\infty.$$

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

By shifting the arguments we may assume that the reference solution is the trivial solution $z(t; t_0, z_0) = 0$.

Remark 3.12. To analyse the stability of a given ODE systems (finite- or infinite-dimensional) is analytically and computationally very challenging; see e.g. Adrianova (1995), Dieci, Russell and Van Vleck (1997), Dieci and Van Vleck (2002), Hinrichsen and Pritchard (2005) and La Salle (1976).

For DAE systems

$$F(t, z, \dot{z}) = 0, \quad t \in \mathbb{T}_\infty$$

the stability concepts in Definition 3.11 essentially carry over. However, when perturbing a consistent initial value, it may happen that the perturbed initial value is no longer consistent. Then the solution (if we allow discontinuities in the part of the state vector that is not differentiated) has a discontinuous jump that transfers the solution to the constraint manifold. For a strangeness-free DAE this would not be a problem because such a jump does not destroy the stability properties. If, however, the strangeness index is bigger than zero, then, due to the required differentiations, the solution may only exist in the distributional sense; see e.g. Kunkel and Mehrmann (2006), Rabier and Rheinboldt (1996*a*) and Trenn (2013).

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Example 3.13. Consider the homogeneous linear time-invariant DAE from Du, Linh and Mehrmann (2013):

$$\dot{z}_1 = z_2, \quad 0 = -z_1 - \varepsilon z_2.$$

If $\varepsilon > 0$ then the DAE is strangeness-free and has the solution

$$z_1(t) = \mathrm{e}^{-\varepsilon^{-1}t} z_1(0), \quad z_2(t) = -\varepsilon^{-1}\,\mathrm{e}^{-\varepsilon^{-1}t} z_1(0).$$

With a consistent initial value $z_2(0) = -\varepsilon^{-1} z_1(0)$, the solution is asymptotically stable but this limit would not exist for $\varepsilon \to 0$ unless $\varepsilon^{-1} z_1(t)$ is bounded for $t \to 0$. If $\varepsilon = 0$ then the DAE has strangeness index one, and for the solution $z_1 = 0$, $z_2 = \dot{z}_1 = 0$ the initial value $z_1(0)$ is restricted as well. For $z_1(0) = 1$, z_1 exists and is the discontinuous function that jumps from 1 to 0 at $t = 0$ and z_2 would only be representable by a delta distribution. Finally, if $\varepsilon < 0$, then the solution is unstable.

The stability analysis and computational methods for DAE systems, therefore, assume uniquely solvable strangeness-free systems. If the system is not strangeness-free, then we first perform a strangeness-free reformulation, as discussed in Section 2.2.

For a strangeness-free system, a solution of the system is called *stable*, *asymptotically stable*, *(uniformly) exponentially stable*, respectively, if it satisfies the corresponding condition in Definition 3.11 for consistent perturbed initial values $\hat{z}_0$. Then many analytical results and computational methods can be extended to the case of strangeness-free DAE systems; see Kunkel and Mehrmann (2007), Linh and Mehrmann (2009, 2011, 2014) and Linh, Mehrmann and Van Vleck (2011).

For LTI ODE systems

$$\dot{z} = Az, \tag{3.12}$$

with $A \in \mathbb{R}^{n,n}$, it is well known (e.g. Adrianova 1995) that the system is asymptotically (and also uniformly exponentially) stable if all the eigenvalues are in the open left half of the complex plane and stable if all the eigenvalues are in the closed left half-plane and the eigenvalues on the imaginary axis are semisimple, that is, the associated Jordan blocks have size at most one.

The stability analysis can also be carried out via the computation of a Lyapunov function given by $V(z) = \frac{1}{2} z^\top X z$, where for stability $X = X^\top > 0$ is a solution of the *Lyapunov inequality*

$$A^\top X + XA \leq 0, \tag{3.13}$$

and for asymptotic stability it is a positive definite solution of the strict inequality $A^\top X + XA < 0$; see e.g. Kailath (1980).

The spectral characterization of stability for ODE systems can be generalized to LTI DAE systems

$$E\dot{z} = Az, \tag{3.14}$$

with $E, A \in \mathbb{R}^{n,n}$; see e.g. Du *et al.* (2013).

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Theorem 3.14. Consider the DAE (3.14) with a regular pencil $\lambda E - A$ of (Kronecker) index at most one. The trivial solution $z = 0$ then has the following stability properties.

(i) If all finite eigenvalues have non-positive real part and the eigenvalues on the imaginary axis are semisimple, then the trivial solution $z = 0$ is stable.

(ii) If all finite eigenvalues have negative real part, then the trivial solution $z = 0$ is uniformly and thus exponentially and asymptotically stable.

For LTV ordinary differential equations

$$\dot{z} = A(t)z, \tag{3.15}$$

the different stability properties are characterized by means of the fundamental solution $\Phi(\cdot, t_0) \in \mathcal{C}^1(\mathbb{T}, \mathbb{R}^{n,n})$ that satisfies

$$\frac{\partial}{\partial t}\Phi(t, t_0) = A(t)\Phi(t, t_0), \quad \Phi(t_0, t_0) = I_n \tag{3.16}$$

such that $z(t; t_0, z_0) = \Phi(t, t_0)z_0$; see e.g. Adrianova (1995).

Theorem 3.15. Consider the LTV ODE (3.15) with fundamental solution Φ as in (3.16). The trivial solution of the LTV ODE (3.15)

(i) is stable if and only if there exists a constant $L > 0$ with $\|\Phi(t, t_0)\| \leq L$ on $\mathbb{T}$;

(ii) is asymptotically stable if and only if $\|\Phi(t, t_0)\| \to 0$ for $t \to \infty$;

(iii) is exponentially stable if there exist $L > 0$ and $\gamma > 0$ such that $\|\Phi(t, t_0)\| \leq L\, \mathrm{e}^{-\gamma(t-t_0)}$ on $\mathbb{T}$.

To obtain the results that extend this characterization to LTV DAE systems

$$E(t)\dot{z} = A(t)z, \tag{3.17}$$

assume again that the initial value problem associated with (3.17) has a unique solution for every consistent initial value and is strangeness-free. If the system is not strangeness-free then we first perform the transformation to strangeness-free form as in Section 2.2.

For a regular strangeness-free system (3.17) there exist pointwise orthogonal matrix functions $S \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{n,n})$, $T \in \mathcal{C}^1(\mathbb{T}, \mathbb{R}^{n,n})$ such that

$$SET = \begin{bmatrix} E_{11} & 0 \\ 0 & 0 \end{bmatrix}, \quad SAT - SE\dot{T} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{3.18}$$

with E_{11}, A_{22} pointwise nonsingular, and $z = V\left[\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right]$. Under the condition of a bounded matrix function $A_{22}^{-1}A_{21}$, we obtain the algebraic equation $z_2 = -A_{22}^{-1}A_{21}z_1$ and the so-called inherent ODE associated with (3.17) given by

$$\dot{z}_1 = E_{11}^{-1}\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)z_1. \tag{3.19}$$

It is then clear that for the different stability concepts to extend to DAE systems it is necessary that (3.19) satisfies the corresponding stability conditions.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Remark 3.16. For ODE systems there is also a well-known extension of the spectral stability analysis via the computation of Lyapunov, Bohl and Sacker-Sell spectral intervals. These results have been extended to DAE systems; see Berger (2012), Linh and Mehrmann (2009, 2012, 2014) and Linh *et al.* (2011). We will not discuss this topic further, but just mention that it is computationally highly expensive.

For general autonomous nonlinear ODE systems $\dot{z} = f(z)$, the fundamental approach to analyse the stability properties is to compute a Lyapunov function $V(z)$ such that $\dot{V}(z)$ is negative definite in a neighbourhood of the solution z. If such a Lyapunov function exists, then the equilibrium solution $z = 0$ is asymptotically stable; see e.g. La Salle and Lefschetz (1961) and Adrianova (1995). This approach can also be used for general strangeness-free DAE systems by reducing the system to the inherent ODE.

3.3. Stabilization

Since, in physical systems, the stability of a solution is typically a crucial property, it is important to know how a stable system behaves under disturbances or uncertainties in the coefficients and how an unstable system can be stabilized with the help of feedback control strategies. First let us consider this question for LTI control problems of the form (3.1) and ask whether it is possible to achieve stability or asymptotic stability via proportional state or output feedback.

We have seen in Theorem 3.4 that for a strongly stabilizable system there exists an $F \in \mathbb{R}^{m,n}$ such that the pair $(E, A + BF)$ is regular and of (Kronecker) index at most one. Moreover, for a strongly stabilizable and strongly detectable system, there exists an $F \in \mathbb{R}^{m,p}$ such that the pair $(E, A + BFC)$ is regular and of (Kronecker) index at most one. In the construction of stabilizing feedbacks, we can therefore assume that such a (preliminary) state or output feedback has been performed and therefore that the pair (E, A) is regular and of (Kronecker) index at most one.

The calculation of stabilizing feedback control laws can then be performed via an optimal control approach (see also Section 3.5 below), by minimizing the cost functional

$$\mathcal{J}(z,u) = \frac{1}{2}\int_{t_0}^{\infty} \begin{bmatrix} z \\ u \end{bmatrix}^{\top} \begin{bmatrix} W_{\mathrm{z}} & S \\ S^{\top} & W_{\mathrm{u}} \end{bmatrix} \begin{bmatrix} z \\ u \end{bmatrix} \mathrm{d}t \tag{3.20}$$

subject to the constraint (3.1). We could also have used the output function y instead of the state function z by inserting $y = Cz$ and modifying the weights accordingly. The following results, which are based on the Pontryagin maximum principle, are taken from Mehrmann (1991).

Theorem 3.17. Consider the optimal control problem to minimize (3.20) subject to the constraint (3.1) with a pair (E, A) that is regular and of (Kronecker) index at most one. Suppose that a continuous solution $u^{\star}$ to the optimal control problem exists and let $z^{\star}$ be the solution of (3.1) with this input function. Then there exists

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

a Lagrange multiplier function $\lambda \in \mathcal{C}^1(\mathbb{T}, \mathbb{R}^n)$ such that $z^\star$, $u^\star$ and λ satisfy the boundary value problem

$$\begin{bmatrix} 0 & E & 0 \\ -E^\top & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\lambda} \\ \dot{z} \\ \dot{u} \end{bmatrix} = \begin{bmatrix} 0 & A & B \\ A^\top & W_{\mathrm{z}} & S \\ B^\top & S^\top & W_{\mathrm{u}} \end{bmatrix} \begin{bmatrix} \lambda \\ z \\ u \end{bmatrix}, \tag{3.21}$$

with boundary conditions

$$E^\dagger E z(t_0) = z_0, \quad \lim_{t\to\infty} E^\top \lambda(t) = 0, \tag{3.22}$$

where $E^\dagger$ denotes the Moore–Penrose inverse of E.

Theorem 3.18. Suppose that $z^\star$, $u^\star$ and λ satisfy the boundary value problem (3.21), (3.22) and suppose, furthermore, that the matrix

$$\begin{bmatrix} W_{\mathrm{z}} & S \\ S^\top & W_{\mathrm{u}} \end{bmatrix}$$

is positive semidefinite. Then

$$\mathcal{J}(z, u) \geq \mathcal{J}(z^\star, u^\star)$$

for all z and u satisfying (3.1).

The solution of the optimality boundary value problem (3.21) with boundary conditions (3.22) will yield the optimal control u and the corresponding optimal state z. However, in many real-world applications we would like the optimal control to be a state feedback. A sufficient condition for this to hold is that the matrix pencil associated with (3.21) is regular of (Kronecker) index at most one and has no purely imaginary eigenvalue. If the matrix W_{u} is positive definite and (E, A, B) is strongly stabilizable, then this can be guaranteed (Mehrmann 1991) and we can proceed as follows. Recall that we have assumed that the pair (E, A) is regular and of (Kronecker) index at most one. Then the coefficients $E, A, B, W_{\mathrm{z}}, S, W_{\mathrm{u}}$ can be transformed such that (E, A) is in Weierstrass canonical form (2.20), that is,

$$PEQ = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \quad PAQ = \begin{bmatrix} J & 0 \\ 0 & I \end{bmatrix}, \quad PB = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix},$$

with transformed cost function

$$Q^\top W_{\mathrm{z}} Q = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}, \quad Q^\top S = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}.$$

Setting

$$Q^{-1} z =: \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \quad Q^{-1}\lambda =: \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}, \quad Q^{-1} z_0 =: \begin{bmatrix} z_{1,0} \\ z_{2,0} \end{bmatrix}$$

and reordering equations and unknowns, we obtain the transformed boundary value

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

problem

$$\begin{bmatrix} 0 & I & 0 & 0 & 0 \\ -I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\lambda}_1 \\ \dot{z}_1 \\ \dot{\lambda}_2 \\ \dot{z}_2 \\ \dot{u} \end{bmatrix} = \begin{bmatrix} 0 & J & 0 & 0 & B_1 \\ J^\top & W_{11} & 0 & W_{12} & S_1 \\ 0 & 0 & 0 & I & B_2 \\ 0 & W_{12}^\top & I & W_{22} & S_2 \\ B_1^\top & S_1^\top & B_2^\top & S_2^\top & W_{\mathrm{u}} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ z_1 \\ \lambda_2 \\ z_2 \\ u \end{bmatrix}, \tag{3.23}$$

with boundary conditions

$$z_1(t_0) = z_{1,0}, \quad \lim_{t\to\infty} \lambda_1(t) = 0. \tag{3.24}$$

Solving the third and fourth equation in (3.23) gives

$$z_2 = -B_2 u \quad \text{and} \quad \lambda_2 = -W_{12}^\top z_1 - W_{22} z_2 - S_2 u.$$

Inserting these into the other equations gives the reduced optimality system

$$\begin{bmatrix} 0 & I & 0 \\ -I & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\lambda}_1 \\ \dot{z}_1 \\ \dot{u} \end{bmatrix} = \begin{bmatrix} 0 & J & B_1 \\ J^\top & W_{11} & \tilde{S}_1 \\ B_1^\top & \tilde{S}_1^\top & \tilde{W}_{\mathrm{u}} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ z_1 \\ u \end{bmatrix}, \tag{3.25}$$

with $\tilde{S}_1 = S_1 - W_{12}B_2$, $\tilde{W}_{\mathrm{u}} = W_{\mathrm{u}} - S_2^\top B_2 - B_2^\top S_2 + B_2^\top W_{22} B_2$ and boundary conditions (3.24). This is the classical optimality condition associated with the ODE constraint $\dot{z}_1 = Jz_1 + B_1 u$ and the cost matrix

$$\tilde{\mathcal{W}} = \begin{bmatrix} W_{11} & \tilde{S}_1 \\ \tilde{S}_1^\top & \tilde{W}_u \end{bmatrix},$$

for which the standard theory for optimal control with ODEs constraints can be applied; see e.g. Mehrmann (1991).

With an ansatz $\lambda_1 = Xz_1$ with $X = X^\top$, the optimal control takes the form of a state feedback

$$u = Fz_1 = -\tilde{W}_{\mathrm{u}}^{-1}\left(B_1^\top X + \tilde{S}_1^\top\right) z_1.$$

Inserting $\lambda_1 = Xz_1$ into (3.25), we obtain the system

$$\begin{aligned} \dot{z}_1 &= \left(J - B_1 \tilde{W}_{\mathrm{u}}^{-1}\left(\tilde{S}_1 + B_1^\top X\right)\right) z_1, \\ -X\dot{z}_1 &= \left(\left(J - B_1 \tilde{W}_{\mathrm{u}}^{-1} \tilde{S}_1^\top\right)^\top X + W_{11} - \tilde{S}_1 \tilde{W}_{\mathrm{u}}^{-1} \tilde{S}_1^\top\right) z_1, \\ 0 &= \lim_{t\to\infty} Xz_1(t). \end{aligned}$$

A sufficient condition for this system to have a solution (Mehrmann 1991) is that we can find a positive semidefinite solution $X = X^\top$ to the algebraic Riccati equation

$$0 = W_{11} + XJ + J^\top X - \left(B_1^\top X + \tilde{S}_1^\top\right)^\top \tilde{W}_{\mathrm{u}}^{-1}\left(B_1^\top X + \tilde{S}_1^\top\right).$$

If $\tilde{W}_{\mathrm{u}}$ is not invertible then there are further restrictions on the boundary conditions that may defer the solvability of the boundary value problem; see Mehrmann (1991) for details and Section 10.2 below.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

3.4. Passivity

Another important property of control systems is the concept of *passivity*. Let us first introduce a passivity definition for general DAE control systems of the form (2.1) with a state space $\mathcal{Z}$, input space $\mathcal{U}$ and output space $\mathcal{Y}$. See Byrnes, Isidori and Willems (1991) for the definition for ODE systems.

To introduce this definition we consider a positive definite and quadratic *storage function* $\mathcal{H}\colon \mathcal{Z} \to \mathbb{R}$ as well as a *supply function* $\mathcal{S}\colon \mathcal{Y} \times \mathcal{U} \to \mathbb{R}$ satisfying

$$\begin{aligned} \mathcal{S}(0,u) &= 0 \quad \text{for all } u \in \mathcal{U}, \\ \mathcal{S}(y,0) &= 0 \quad \text{for all } y \in \mathcal{Y}. \end{aligned}$$

Definition 3.19. An autonomous DAE of the form (2.1) is called *(strictly) dissipative* with respect to the storage function $\mathcal{H}(z)$ and supply function $\mathcal{S}(y,u)$ if there exists a positive semidefinite (positive definite) function $\Phi\colon \mathcal{X} \to \mathbb{R}$, such that for any $u \in \mathcal{U}$ and for any $t_0 < t_1$ the equation

$$\mathcal{H}(z(t_1)) - \mathcal{H}(z(t_0)) = \int_{t_0}^{t_1} \mathcal{S}(y(s),u(s)) - \Phi(z(s))\,\mathrm{d}s \tag{3.26}$$

holds for all $(y,u) \in \mathcal{Y} \times \mathcal{U}$.

A (strictly) dissipative system is often called *(strictly) passive* if the particular supply rate $S(y,u) = y^\top u$ is used.

Equation (3.26) is called the *storage energy balance equation* and directly implies that the *dissipation inequality*

$$\mathcal{H}(z(t_1)) - \mathcal{H}(z(t_0)) \leq \int_{t_0}^{t_1} \mathcal{S}(y(s),u(s))\,\mathrm{d}s$$

holds for all $(y,u) \in \mathcal{Y} \times \mathcal{U}$. Note that for passive systems the storage energy balance equation is typically called *power balance equation*.

For LTI ODE systems of the form

$$\begin{aligned} \dot{z} &= Az + Bu, \\ y &= Cz + Du, \end{aligned}$$

passivity can be characterized (Willems 1971) via the existence of a positive definite solution $X = X^\top$ of a linear matrix inequality, the *Kalman–Yakubovich–Popov inequality*

$$W(X) := \begin{bmatrix} -XA - A^\top X & C^\top - XB \\ C - B^\top X & D + D^\top \end{bmatrix} \geq 0. \tag{3.27}$$

For strict passivity this inequality has to be strict. Note that (3.27) generalizes the Lyapunov inequality (3.13), which is just the leading block, and hence (strict) passivity directly implies (asymptotic) stability.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

The relationship between passivity and the linear matrix inequality (3.27) has been extended to LTI DAE systems in Reis, Rendel and Voigt (2015) and Reis and Voigt (2015). Since passive systems are closely related to port-Hamiltonian systems, we will come back to this topic in Section 7.4.

3.5. *Optimal control*

An important task in control theory is the solution of optimal control problems that minimize a cost functional subject to an ODE or DAE system. The optimal control theory for general nonlinear DAE systems was presented in Kunkel and Mehrmann (2008). In this section we recall these general results.

Consider the optimal control problem to minimize a cost functional

$$\mathcal{J}(z,u) = \mathcal{M}(t_{\mathrm{f}}) + \int_{t_0}^{t_{\mathrm{f}}} \mathcal{K}(t,z(t),u(t))\,\mathrm{d}t$$

subject to a constraint given by an initial value problem associated with a nonlinear DAE system

$$F(t,z,u,\dot{z}) = 0, \quad z(t_0) = z_0.$$

We can rewrite this problem in the behaviour representation (see the discussion in Section 2.2) with $\xi = [z^\top \quad u^\top]^\top$, and then study the optimization problem

$$\mathcal{J}(\xi) = \mathcal{M}(\xi(t_{\mathrm{f}})) + \int_{t_0}^{t_{\mathrm{f}}} \mathcal{K}(t,\xi(t))\,\mathrm{d}t = \min! \tag{3.28}$$

subject to the constraint

$$F(t,\xi,\dot{\xi}) = 0, \quad [I_n \quad 0]\xi(t_0) = z_0. \tag{3.29}$$

If Hypothesis 2.7 holds, then for this system we have (locally via the implicit function theorem) a strangeness-free reformulation (Kunkel and Mehrmann 2006) given by

$$\begin{aligned} \dot{z}_1 &= \mathcal{L}(t,z_1,u), \quad z_1(t_0) = z_{1,0}, \\ z_2 &= \mathcal{R}(t,z_1,u), \end{aligned} \tag{3.30}$$

and the associated cost function reads

$$\mathcal{J}(z_1,z_2,u) = \mathcal{M}(z_1(t_{\mathrm{f}}),z_2(t_{\mathrm{f}})) + \int_{t_0}^{t_{\mathrm{f}}} \mathcal{K}(t,z_1,z_2,u)\,\mathrm{d}t. \tag{3.31}$$

For this formulation, the necessary optimality conditions in the space

$$\mathbb{W} := \mathcal{C}^1(\mathbb{T},\mathbb{R}^d) \times \mathcal{C}(\mathbb{T},\mathbb{R}^a) \times \mathcal{C}(\mathbb{T},\mathbb{R}^m)$$

are presented in the following theorem. We refer to Kunkel and Mehrmann (2008) for the original presentation and the proof.

Theorem 3.20. Let ξ be a local solution of (3.28) subject to (3.29) in the sense that $(z_1,z_2,u) \in \mathbb{W}$ is a local solution of (3.31) subject to (3.30). Then there exist

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

unique Lagrange multipliers $(\lambda_1, \lambda_2, \gamma) \in \mathbb{W}$ such that $(z_1, z_2, u, \lambda_1, \lambda_2, \gamma)$ solves the boundary value problem

$$\begin{aligned}
\dot{z}_1 &= \mathcal{L}(t, z_1, u), \ z_1(t_0) = z_{1,0},\\
z_2 &= \mathcal{R}(t, z_1, u),\\
\dot{\lambda}_1 &= \frac{\partial}{\partial z_1}\mathcal{K}(t, z_1, z_2, u)^\top - \frac{\partial}{\partial z_1}\mathcal{L}(t, z_1, z_2, u)^\top\lambda_1 - \frac{\partial}{\partial z_1}\mathcal{R}_{z_1}(t, z_1, u)^\top\lambda_1,\\
\lambda_1(t_\mathrm{f}) &= -\frac{\partial}{\partial z_1}\mathcal{M}(z_1(t_\mathrm{f}), z_2(t_\mathrm{f}))^\top\\
0 &= \frac{\partial}{\partial z_2}\mathcal{K}(t, z_1, z_2, u)^\top + \lambda_2,\\
0 &= \frac{\partial}{\partial u}\mathcal{K}(t, z_1, z_2, u)^\top - \frac{\partial}{\partial u}\mathcal{L}(t, z_1, u)^\top\lambda_1 - \frac{\partial}{\partial u}\mathcal{R}(t, z_1, u)^\top\lambda_2,\\
\gamma &= \lambda_1(t_0).
\end{aligned}$$

Theorem 3.20 is a local result based on the implicit function theorem that has to be modified to turn it into a computationally feasible procedure (Kunkel and Mehrmann 2008), and which can be substantially strengthened for the minimization of quadratic cost functionals

$$\mathcal{J}(z, u) = \frac{1}{2}z(t_\mathrm{f})^\top M z(t_\mathrm{f}) + \frac{1}{2}\int_{t_0}^{t_\mathrm{f}} (z^\top W_\mathrm{z} z + 2z^\top S u + u^\top W_\mathrm{u} u)\,\mathrm{d}t, \tag{3.32}$$

with $W_\mathrm{z} = W_\mathrm{z}^\top \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{n,n})$, $W_\mathrm{u} = W_\mathrm{u}^\top \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{m,m})$, $S \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{n,m})$ and $M = M^\top \in \mathbb{R}^{n,n}$, subject to LTV DAE constraints

$$E\dot{z} = Az + Bu + f, \quad z(t_0) = z_0, \tag{3.33}$$

with $E \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{n,n})$, $A \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{n,n})$, $B \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{n,m})$, $f \in \mathcal{C}(\mathbb{T}, \mathbb{R}^n)$, $z_0 \in \mathbb{R}^n$, $u \in \mathbb{U} := \mathcal{C}(\mathbb{T}, \mathbb{R}^m)$ and $f \in \mathcal{C}(\mathbb{T}, \mathbb{R}^n)$. Using the property that for the Moore–Penrose inverse $E^\dagger$ of E we have

$$E\dot{z} = EE^\dagger E\dot{z} = E\frac{\mathrm{d}}{\mathrm{d}t}(E^\dagger E z) - E\frac{\mathrm{d}}{\mathrm{d}t}(E^\dagger E)z,$$

we interpret (3.33) as

$$E\frac{\mathrm{d}}{\mathrm{d}t}(E^\dagger E z) = \left(A + E\frac{\mathrm{d}}{\mathrm{d}t}(E^\dagger E)\right)z + Bu + f, \quad (E^\dagger E z)(t_0) = z_0.$$

This allows the particular solution space (Kunkel and Mehrmann 1996)

$$\mathbb{Z} := \mathcal{C}^1_{E^\dagger E}(\mathbb{T}, \mathbb{R}^n) := \{z \in \mathcal{C}(\mathbb{T}, \mathbb{R}^n) \mid E^\dagger E z \in \mathcal{C}^1(\mathbb{T}, \mathbb{R}^n)\},$$

which ensures that the differentiability of the state variable is only required in this restricted space. Note that with this definition we slightly extend our solution concept from Definition 2.1. Using the solution space $\mathbb{Z}$, Kunkel and Mehrmann (2008) derived the following necessary optimality condition.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Theorem 3.21. Consider the optimal control problem (3.32) subject to (3.33) with a consistent initial condition. Suppose that (3.33) is strangeness-free as a behaviour system and that the range of M is contained in the cokernel of $E(t_\mathrm{f})$. If $(z, u) \in \mathbb{Z} \times \mathbb{U}$ is a solution to this optimal control problem, then there exists a Lagrange multiplier function $\lambda \in \mathbb{Z}$ such that (z, λ, u) satisfy the boundary value problem

$$E\frac{\mathrm{d}}{\mathrm{d}t}(E^\dagger E z) = \left(A + E\frac{\mathrm{d}}{\mathrm{d}t}(E^\dagger E)\right)z + Bu + f, \tag{3.34a}$$

$$E^\top \frac{\mathrm{d}}{\mathrm{d}t}(EE^\dagger \lambda) = W_\mathrm{z} z + Su - (A + EE^\dagger \dot{E})^\top \lambda, \tag{3.34b}$$

$$0 = S^\top z + W_\mathrm{u} u - B^\top \lambda, \tag{3.34c}$$

$$(E^\dagger E z)(t_0) = z_0, \quad (EE^\dagger \lambda)(t_\mathrm{f}) = -E^\dagger(t_\mathrm{f})^\top M z(t_\mathrm{f}). \tag{3.34d}$$

Remark 3.22. If we have an output equation, then we can also consider the cost functional (3.32) with the state replaced by the output. This problem can be treated analogously by inserting the output equation in the cost functional and renaming the coefficients. The same approach with a modified cost functional can also be used when we want to optimally drive the solution to a reference function $\tilde{z}$.

Kunkel and Mehrmann (2011) (see also Backes 2006, Kurina and März 2004) demonstrated that it is generally not possible to drop the assumptions of Theorem 3.21 and instead consider the formal optimality system

$$\begin{aligned} E\dot{z} &= Az + Bu + f, & z(t_0) &= z_0, \\ \frac{\mathrm{d}}{\mathrm{d}t}(E^\top \lambda) &= W_\mathrm{z} z + Su - A^\top \lambda, & (E^\top \lambda)(t_\mathrm{f}) &= -Mz(t_\mathrm{f}), \\ 0 &= S^\top z + W_\mathrm{u} u - B^\top \lambda. \end{aligned} \tag{3.35}$$

However, if this formal optimality system has a unique solution, then the state z and the control u are correct but the optimal Lagrange multiplier may be different. In this case, under some further assumptions, Kunkel and Mehrmann (2011) also showed a sufficient condition in generalizing results from Backes (2006).

Theorem 3.23. Consider the optimal control problem (3.32) subject to (3.33) with a consistent initial condition, and suppose that in the cost functional (3.32) we have that the matrix functions $\begin{bmatrix} W_z & S \\ S^\top & W_\mathrm{u} \end{bmatrix}$ and M are (pointwise) positive semidefinite. If $(z^\star, u^\star, \lambda)$ satisfies the formal optimality system (3.35), then for any (z, u) satisfying (3.33) we have

$$\mathcal{J}(z, u) \geq \mathcal{J}(z^\star, u^\star).$$

Numerically, the solution of the boundary value problem (which is always a DAE) is a challenge, in particular for large-scale problems. For the case when $E = I$ and if W_u is positive definite, which means that the optimality system is strangeness-free, then a classical approach successfully employed in many applications is to resolve

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

(3.34c) for u, and then to decouple the state equation (3.34a) and the adjoint equation (3.34b) via the solution of a Riccati differential equation.

If some further conditions hold, then the Riccati approach can also be carried out for DAE systems; see Kunkel and Mehrmann (2011). If the constraint system with $u = 0$ is regular, strangeness-free, and E has constant rank, then, using Theorem 2.3, there exist pointwise orthogonal $P \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{n,n})$ and $Q \in \mathcal{C}^1(\mathbb{T}, \mathbb{R}^{n,n})$ such that

$$\begin{aligned} \tilde{E} = PEQ = \begin{bmatrix} E_{11} & 0 \\ 0 & 0 \end{bmatrix}, \quad \tilde{A} = PAQ - PE\dot{Q} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \\ \tilde{B} = PB = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad \tilde{W}_z = Q^\top W_z Q = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}, \quad \tilde{f} = Pf = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, \\ \tilde{S} = Q^\top S = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}, \quad z = Q\tilde{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \quad z_0 = Q\tilde{z}_0 = \begin{bmatrix} z_{0,1} \\ z_{0,2} \end{bmatrix}, \end{aligned} \tag{3.36}$$

with $E_{11} \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{d,d})$ and $A_{22} \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{a,a})$ pointwise nonsingular. Constructing the formal optimality system associated with this transformed system and rearranging the equations, the following theorem is proved in Kunkel and Mehrmann (2011).

Theorem 3.24. The DAE in (3.34) is regular and strangeness-free if and only if

$$\hat{W}_u = \begin{bmatrix} 0 & A_{22} & B_2 \\ A_{22}^\top & W_{22} & S_2 \\ B_2^\top & S_2^\top & W_{\mathrm{u}} \end{bmatrix}$$

is pointwise nonsingular, where we used the notation of (3.36).

If $\hat{W}_u$ is pointwise nonsingular, then

$$\begin{bmatrix} -\lambda_2 \\ z_2 \\ u \end{bmatrix} = -\hat{W}_u^{-1} \left(\begin{bmatrix} 0 & A_{21} \\ A_{12}^\top & W_{21} \\ B_1^\top & S_1^\top \end{bmatrix} \begin{bmatrix} -\lambda_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} f_2 \\ 0 \\ 0 \end{bmatrix} \right).$$

The remaining equations can be written as

$$\begin{bmatrix} E_{11}\dot{z}_1 \\ \frac{\mathrm{d}}{\mathrm{d}t}\left(\left(-E_{11}^\top\right)(-\lambda_1)\right) \end{bmatrix} = \begin{bmatrix} 0 & A_{11} \\ A_{11}^\top & W_{11} \end{bmatrix} \begin{bmatrix} -\lambda_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} 0 & A_{12} & B_1 \\ A_{21}^\top & W_{21}^\top & S_1 \end{bmatrix} \begin{bmatrix} -\lambda_2 \\ z_2 \\ u \end{bmatrix} + \begin{bmatrix} f_1 \\ 0 \end{bmatrix}.$$

Defining

$$\begin{aligned} F_1 &:= E_{11}^{-1}\left(A_{11} - [0 \quad A_{12} \quad B_1]\hat{W}_u^{-1}\left[A_{21}^\top \quad W_{21}^\top \quad S_1\right]^\top\right), \\ G_1 &:= E_{11}^{-1}[0 \quad A_{12} \quad B_1]\hat{W}_u^{-1}[0 \quad A_{12} \quad B_1]^\top E_{11}^{-\top}, \\ H_1 &:= W_{11} - \left[A_{21}^\top \quad W_{21}^\top \quad S_1\right]\hat{W}_u^{-1}\left[A_{21}^\top \quad W_{21}^\top \quad S_1\right]^\top, \\ g_1 &:= E_{11}^{-1}\left(f_1 - [0 \quad A_{12} \quad B_1]\hat{W}_u^{-1}\left[f_2^\top \quad 0 \quad 0\right]^\top\right), \\ h_1 &:= -\left[A_{21}^\top \quad W_{21}^\top \quad S_1\right]\hat{W}_u^{-1}\left[f_2^\top \quad 0 \quad 0\right]^\top, \end{aligned}$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

we obtain the boundary value problem

$$\dot{z}_1 = F_1 z_1 + G_1\left(E_{11}^\top \lambda_1\right) + g_1, \qquad z_1(t_0) = z_{0,1}, \tag{3.37a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(E_{11}^\top \lambda_1\right) = H_1 z_1 - F_1^\top \left(E_{11}^\top \lambda_1\right) + h_1, \quad \left(E_{11}^\top \lambda_1\right)(t_\mathrm{f}) = -M_{11} z_1(t_\mathrm{f}). \tag{3.37b}$$

Making the ansatz $E_{11}^\top \lambda_1 = X_{11} z_1 + v_1$, one can solve the two initial value problems

$$\dot{X}_{11} + X_{11}F_1 + F_1^\top X_{11} + X_{11}G_1 X_{11} - H_1 = 0, \quad X_{11}(t_\mathrm{f}) = -M_{11},$$

and

$$\dot{v}_1 + X_{11}G_1 v_1 + F_1^\top v_1 + X_{11} g_1 - h_1 = 0, \quad v_1(t_\mathrm{f}) = 0,$$

to obtain X_{11} and v_1 and to decouple the solution of (3.37). In Kunkel and Mehrmann (2011), a Riccati approach is also obtained directly for the original optimality system (3.34) by the modified ansatz

$$\begin{aligned} \lambda &= XEz + v = XEE^\dagger Ez + v, \\ \frac{\mathrm{d}}{\mathrm{d}t}(EE^\dagger \lambda) &= \frac{\mathrm{d}}{\mathrm{d}t}(EE^\dagger X)Ez + (EE^\dagger X)\dot{E}E^\dagger Ez + (EE^\dagger X)E\frac{\mathrm{d}}{\mathrm{d}t}(E^\dagger Ex) + \frac{\mathrm{d}}{\mathrm{d}t}(E^\dagger Ev), \end{aligned} \tag{3.38}$$

where

$$X \in \mathcal{C}^1_{EE^\dagger}(\mathbb{T}, \mathbb{R}^{n,n}), \quad v \in \mathcal{C}^1_{EE^\dagger}(\mathbb{T}, \mathbb{R}^n)$$

to fit to the solution spaces for z and λ. If W_u is invertible, then introducing the notation

$$F := A - BW_\mathrm{u}^{-1}S^\top, \quad G := BW_\mathrm{u}^{-1}B^\top, \quad H := W - SW_\mathrm{u}^{-1}S^\top$$

yields two initial value problems for the Riccati DAE

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}(E^\top XE) + E^\top XF + F^\top XE + E^\top XGXE - H = 0, \\ (E^\top XE)(t_\mathrm{f}) = -M, \end{aligned} \tag{3.39}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}(E^\top v) + E^\top XGv + F^\top v + E^\top Xf = 0, \quad (E^\top v)(t_\mathrm{f}) = 0.$$

For this to be solvable, we must have $M = E(t_\mathrm{f})^\top \tilde{M} E(t_\mathrm{f})$ with suitable $\tilde{M}$ and $H = E^\top \tilde{H} E$ with suitable $\tilde{H}$.

The major advantage of the ansatz via Riccati equations is that the resulting control can be directly expressed as a feedback control, for example using (3.38), we get

$$u = W_\mathrm{u}^{-1}(B^\top \lambda - S^\top z) = W_\mathrm{u}^{-1}(B^\top XEE^\dagger E - S^\top)z + v.$$

A similar result is also obtained if the cost functional is formulated in terms of the output and using output feedback.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Note that in the LTI DAE case, further results have been obtained that allow the use of efficient numerical techniques for the computation of the solutions to (3.39) via eigenvalue methods; see Mehrmann (1991).

In this section we have recalled several properties for general descriptor systems. In the following sections we study the special class of port-Hamiltonian descriptor systems and show that the structure of the systems directly encodes many desirable properties from our model wish list.

4. Port-Hamiltonian descriptor systems

To fulfil as many points on our wish list as possible, instead of general descriptor systems we will use energy-based modelling within the class of (dissipative) port-Hamiltonian (pH) systems and their generalization to descriptor systems.

4.1. Nonlinear (dissipative) port-Hamiltonian descriptor systems

We start our exposition by introducing the general model class of (dissipative) pH descriptor systems, or pH differential-algebraic equation (pHDAE) systems, introduced in Mehrmann and Morandin (2019).

Definition 4.1 (pH descriptor system, pHDAE). Consider a time interval $\mathbb{T}$, a state space $\mathcal{Z} \subseteq \mathbb{R}^n$ and an extended space $\mathcal{S} := \mathbb{T} \times \mathcal{Z}$. Then a *(dissipative) port-Hamiltonian descriptor system* (pHDAE) is a descriptor system of the form

$$E(t,z)\dot{z} + r(t,z) = (J(t,z) - R(t,z))\eta(t,z) + (G(t,z) - P(t,z))u, \tag{4.1a}$$

$$y = (G(t,z) + P(t,z))^\top \eta(t,z) + (S(t,z) - N(t,z))u, \tag{4.1b}$$

with state $z\colon \mathbb{T} \to \mathcal{Z}$, input $u\colon \mathbb{T} \to \mathbb{R}^m$ and output $y\colon \mathbb{T} \to \mathbb{R}^m$, where

$$r, \eta \in \mathcal{C}(\mathcal{S}, \mathbb{R}^\ell), \qquad E \in \mathcal{C}(\mathcal{S}, \mathbb{R}^{\ell,n}), \qquad J, R \in \mathcal{C}(\mathcal{S}, \mathbb{R}^{\ell,\ell})$$
$$G, P \in \mathcal{C}(\mathcal{S}, \mathbb{R}^{\ell,m}), \quad S, N \in \mathcal{C}(\mathcal{S}, \mathbb{R}^{m,m}),$$

and an associated function $\mathcal{H} \in \mathcal{C}^1(\mathcal{S}, \mathbb{R})$, called the *Hamiltonian* of (4.1). Furthermore, the following properties must hold.

(i) The matrix functions

$$\Gamma := \begin{bmatrix} J & G \\ -G^\top & N \end{bmatrix} \in \mathcal{C}(\mathcal{S}, \mathbb{R}^{(\ell+m),(\ell+m)}), \tag{4.2a}$$

$$W := \begin{bmatrix} R & P \\ P^\top & S \end{bmatrix} \in \mathcal{C}(\mathcal{S}, \mathbb{R}^{(\ell+m),(\ell+m)}), \tag{4.2b}$$

called the *structure matrix* and *dissipation matrix* respectively, satisfy $\Gamma = -\Gamma^\top$ and $W = W^\top \geq 0$ in $\mathcal{S}$.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

(ii) The Hamiltonian satisfies

$$\frac{\partial}{\partial z}\mathcal{H}(t,z) = E^\top(t,z)\eta(t,z) \quad \text{and} \quad \frac{\partial}{\partial t}\mathcal{H}(t,z) = \eta^\top(t,z)r(t,z) \tag{4.3}$$

in $\mathcal{S}$ along any solution of (4.1).

If the pH descriptor system has no inputs and outputs, that is, if $G, P \equiv 0$ and the output equation is omitted, then we refer to (4.1) as a *(dissipative) Hamiltonian differential-algebraic equation (dHDAE)*.

Note that in the literature and also in this survey, the adjective *dissipative* is typically omitted, and we follow this tradition even if the system has a dissipative part R.

Remark 4.2. In many applications the coefficients of pHDAE systems do not explicitly depend on time. This is not really a restriction, since we can always make a system of the form (4.1) autonomous by introducing the combined state $\hat{z} := [z^\top, t]^\top$ and reformulating the pHDAE (4.1) as

$$\begin{bmatrix} E(\hat{z}) & r(\hat{z}) \\ 0 & 1 \end{bmatrix} \dot{\hat{z}} = \begin{bmatrix} J(\hat{z}) - R(\hat{z}) & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \eta(\hat{z}) \\ 0 \end{bmatrix} + \begin{bmatrix} G(\hat{z}) - P(\hat{z}) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ 1 \end{bmatrix}, \tag{4.4a}$$

$$\begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} G(\hat{z}) + P(\hat{z}) & 0 \\ 0 & 1 \end{bmatrix}^\top \begin{bmatrix} \eta(\hat{z}) \\ 0 \end{bmatrix} + \begin{bmatrix} S(\hat{z}) - N(\hat{z}) & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ 1 \end{bmatrix}, \tag{4.4b}$$

which is again a pHDAE, since

$$\frac{\partial}{\partial \hat{z}}\mathcal{H}(\hat{z}) = \begin{bmatrix} \frac{\partial}{\partial z}\mathcal{H}(t,z) \\ \frac{\partial}{\partial t}\mathcal{H}(t,z) \end{bmatrix} = \begin{bmatrix} E(\hat{z}) & r(\hat{z}) \\ 0 & 1 \end{bmatrix}^\top \begin{bmatrix} \eta(\hat{z}) \\ 0 \end{bmatrix}.$$

Remark 4.3. For many properties of pHDAEs that we discuss later, it is sufficient to require the properties (4.2) and (4.3) in Definition 4.1 to hold only along any solution of (4.1), thus further extending the model class. Nevertheless, to simplify the presentation, we work with the definition as presented here.

Remark 4.4. If $E = I_n$ is the identity matrix, $r \equiv 0$, and the coefficients do not explicitly depend on time t, then Definition 4.1 reduces to the well-known classical representation for ODE pH systems, called pHODEs in the following, as for instance presented in van der Schaft and Jeltsema (2014).

Remark 4.5. In the literature, pH systems are often described via a Dirac structure (van der Schaft and Jeltsema 2014), and this approach has also been extended to descriptor systems; see Mehrmann and Morandin (2019), Mehrmann and van der Schaft (2023), van der Schaft (2013), van der Schaft and Maschke (2018, 2020) and Section 7.6 below. In this survey, however, we mostly focus on the dynamical systems point of view, which is prevalent in the simulation and control context. For LTI pHDAEs the exact relationship between these different approaches has recently

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

been characterized in Mehrmann and van der Schaft (2023), where a characterization is also given when a general LTI DAE can be transformed to dHDAE form.

In many applications, additional properties of the Hamiltonian, such as convexity or non-negativity, may strengthen the properties of pHDAEs further; see Section 7. We thus make the following definition.

Definition 4.6. The Hamiltonian for a pHDAE of the form (4.1) is called *non-negative* if

$$\mathcal{H}(t, z(t)) \geq 0 \quad \text{for all } (t, z) \in \mathcal{S} \text{ with } z \text{ being a solution of (4.1).}$$

Although satisfied in many applications, it may seem artificial from a mathematical point of view to require the Hamiltonian to be non-negative. However, this is not a restriction, since any Hamiltonian that is bounded from below can be recast as a non-negative Hamiltonian by adding its infimum along any behaviour solution. In the following we therefore always assume that the Hamiltonian is non-negative.

Remark 4.7. The particular structure of the ports with equal dimensions and the described structure ensure that inputs and outputs are *co-located* or *power-conjugated*. This enables easy power-conserving interconnection of pHDAE systems; see Section 7.2 below. However, in many applications there are specific quantities that can be observed and others that can be used for control, and these are not necessarily co-located or power-conjugated. To allow classical control techniques as well as interconnectability, one can extend the inputs and outputs to obtain a power-conjugated formulation. Even if these variables are not explicitly used, they typically have a physical meaning in the context of supplied energy. We will demonstrate this with examples later on; see e.g. Sections 5.3 and 5.6.

There are different generalizations of Definition 4.1 to infinite-dimensional systems. For example, one can formulate operator pHDAE systems via semigroup theory, introduce formal Dirac structures, or follow a gradient flow approach. In the final section we present an incomplete list of references discussing different aspects of infinite-dimensional pHDAE systems. In this survey we focus mainly on the finite-dimensional case. We assume that a space discretization via Galerkin projection is performed for an infinite-dimensional case. For infinite-dimensional examples we mimic the finite-dimensional properties, which are then preserved under Galerkin projection (see Section 7.1 below). We refer to the examples in Section 5 for further details.

4.2. Linear pHDAE systems

Important special subclasses of the general class of pHDAE systems are LTV and LTI pHDAE systems. Such a general class with a quadratic Hamiltonian was introduced in Beattie, Mehrmann, Xu and Zwart (2018).

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Definition 4.8 (linear pHDAE, quadratic Hamiltonian). A linear time-varying descriptor system of the form

$$E(t)\dot{z} + E(t)K(t)z = (J(t) - R(t))Q(t)z + (G(t) - P(t))u, \tag{4.5a}$$
$$y = (G(t) + P(t))^\top Q(t)z + (S(t) - N(t))u, \tag{4.5b}$$

with

$$E, Q \in \mathcal{C}^1(\mathbb{T}, \mathbb{R}^{\ell,n}), \quad J, R, K \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{\ell,\ell}), \quad G, P \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{\ell,m}),$$
$$S = S^\top, \quad N = -N^\top \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{m,m}),$$

is called a *linear pHDAE with quadratic Hamiltonian*

$$\mathcal{H}\colon \mathbb{T} \times \mathbb{R}^n \to \mathbb{R}, \quad (t, z) \mapsto \frac{1}{2} z^\top Q^\top(t) E(t) z \tag{4.6}$$

if the following properties are satisfied.

(i) The differential operator

$$\mathcal{L} := Q^\top E \frac{\mathrm{d}}{\mathrm{d}t} - (Q^\top J Q - Q^\top E K)\colon \mathcal{C}^1(\mathbb{T}, \mathbb{R}^n) \to \mathcal{C}(\mathbb{T}, \mathbb{R}^n) \tag{4.7}$$

is *skew-adjoint*, that is, we have $Q^\top E \in \mathcal{C}^1(\mathbb{T}, \mathbb{R}^{n,n})$ and for all $t \in \mathbb{T}$

$$Q^\top(t)E(t) = E^\top(t)Q(t) \quad \text{and}$$
$$\frac{\mathrm{d}}{\mathrm{d}t}(Q^\top(t)E(t)) = Q^\top(t)[E(t)K(t) - J(t)Q(t)] + [E(t)K(t) - J(t)Q(t)]^\top Q(t).$$

(ii) The matrix function

$$W := \begin{bmatrix} Q & 0 \\ 0 & I_m \end{bmatrix}^\top \begin{bmatrix} R & P \\ P^\top & S \end{bmatrix} \begin{bmatrix} Q & 0 \\ 0 & I_m \end{bmatrix} \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{(n+m),(n+m)}) \tag{4.8}$$

is positive semidefinite, i.e. $W(t) = W^\top(t) \geq 0$ for all $t \in \mathbb{T}$.

If the Hamiltonian is quadratic, and the coefficients do not depend explicitly on the state z, then Definitions 4.1 and 4.8 are closely related. Starting from Definition 4.8, we may set

$$r(t, z) := E(t)K(t)z \quad \text{and} \quad \eta(t, z) := Q(t)z.$$

Using the skew-adjointness of $\mathcal{L}$ in Definition 4.8(i), we then obtain

$$\frac{\partial}{\partial z}\mathcal{H}(t, z) = E^\top(t)Q(t)z = E^\top \eta(t, z)$$

and

$$\frac{\partial}{\partial t}\mathcal{H}(t, z) = z^\top Q^\top(t)E(t)K(t)z - \frac{1}{2} z^\top Q^\top(t)(J(t) + J^\top(t))Q(t)z$$
$$= \eta^\top(t, z)r(t, z) - \frac{1}{2} z^\top Q^\top(t)(J(t) + J^\top(t))Q(t)z.$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Thus, if additionally J is skew-symmetric, i.e. $J(t) = -J^\top(t)$, then the Hamiltonian satisfies the requirements (4.3) from Definition 4.1. Similarly, we notice that the positive semidefiniteness of the dissipation matrix function in Definition 4.8 is slightly more general than its counterpart in Definition 4.1. On the other hand, the requirement for the matrix functions E and Q to be continuously differentiable is a sufficient condition to obtain a continuously differentiable Hamiltonian as required in Definition 4.1.

Another important special class is that of (LTI) pHDAE systems with quadratic Hamiltonian, which is commonly used in closed-loop and data-based control applications as well as linear stability analysis.

Definition 4.9 (LTI pHDAE, quadratic Hamiltonian). A descriptor system of the form

$$E\dot{z} = (J - R)Qz + (G - P)u, \tag{4.9a}$$

$$y = (G + P)^\top Qz + (S - N)u, \tag{4.9b}$$

with matrices $E, Q \in \mathbb{R}^{\ell,n}$, $J, R \in \mathbb{R}^{\ell,\ell}$, $G, P \in \mathbb{R}^{\ell,m}$ and $S, N \in \mathbb{R}^{m,m}$, is called a *linear time-invariant pHDAE with (quadratic) Hamiltonian*

$$\mathcal{H}\colon \mathbb{R}^n \to \mathbb{R}, \quad z \mapsto \frac{1}{2} z^\top Q^\top E z \tag{4.10}$$

if $E^\top Q = Q^\top E$, and the matrices

$$\Gamma := \begin{bmatrix} J & G \\ -G^\top & N \end{bmatrix} \in \mathbb{R}^{(\ell+m),(\ell+m)},$$

$$W := \begin{bmatrix} Q & 0 \\ 0 & I_m \end{bmatrix}^\top \begin{bmatrix} R & P \\ P^\top & S \end{bmatrix} \begin{bmatrix} Q & 0 \\ 0 & I_m \end{bmatrix} \in \mathbb{R}^{(n+m),(n+m)}$$

satisfy $\Gamma = -\Gamma^\top$ and $W = W^\top \geq 0$.

Remark 4.10. Note that in the LTI or LTV case with quadratic Hamiltonian we will assume throughout the paper that the Hamiltonian is non-negative. This is equivalent to $\mathcal{H}$ being uniformly bounded from below; see Beattie *et al.* (2018). A sufficient condition is obtained by assuming $E^\top Q \geq 0$, which will be used later on.

Having introduced the general modelling concept of (dissipative) pH descriptor systems, we now discuss two modelling simplifications, namely removing the Q factor in linear pHDAE systems and removing the feedthrough term $(S - N)u$.

4.3. Removing the Q factor in linear pHDAE systems

In many applications (time-varying or time-invariant) we have that $\ell = n$ and $Q = I_n$ is the identity matrix in Definitions 4.8 and 4.9. In this case E is the Hessian of the Hamiltonian. This representation has many advantages: all the coefficients appear linearly in (4.9), which greatly simplifies the analysis and also the perturbation theory. In many cases this also leads to a convexification of the representation

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

(Egger 2019, Friedrichs and Lax 1971), and Kurula (2020) illustrates that such a formulation even allows an improved solution theory. In the following we will show how the factor Q can be removed; see Beattie *et al.* (2018) and Mehl, Mehrmann and Wojtylak (2021).

If Q has pointwise full column rank in (4.5), then the state equation can be multiplied with $Q^\top$ from the left, yielding a system with the same solution set given by

$$\begin{aligned} Q^\top E\dot{z} + Q^\top EKz &= Q^\top(J-R)Qz + Q^\top(G-P)u, \\ y &= (G+P)^\top Qz + (S-N)u. \end{aligned}$$

Then, setting $\tilde{E} := Q^\top E$, $\tilde{J} := Q^\top JQ$, $\tilde{R} := Q^\top RQ$, $\tilde{G} := Q^\top G$ and $\tilde{P} := Q^\top P$, the transformed system

$$\begin{aligned} \tilde{E}\dot{z} + \tilde{E}Kz &= (\tilde{J}-\tilde{R})z + (\tilde{G}-\tilde{P})u, \\ y &= (\tilde{G}+\tilde{P})^\top z + (S-N)u \end{aligned}$$

is again a pHDAE, but now has $\tilde{Q} = I_n$ and hence $\tilde{E} = \tilde{E}^\top$.

If Q is not of full rank then the situation is more complex. If Q has constant rank in $\mathbb{T}$, then, using a smooth full-rank decomposition (Theorem 2.13), there exist pointwise orthogonal matrix functions $U\colon \mathbb{T} \to \mathbb{R}^{\ell,\ell}$ and $V\colon \mathbb{T} \to \mathbb{R}^{n,n}$ of the same smoothness as Q such that

$$U^\top QV = \begin{bmatrix} Q_{11} & 0 \\ 0 & 0 \end{bmatrix}, \quad U^\top EV = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix},$$
$$U^\top(J-R)U = \begin{bmatrix} J_{11}-R_{11} & J_{12}-R_{12} \\ J_{21}-R_{21} & J_{22}-R_{22} \end{bmatrix},$$

where the $(1,1)$ block in all three block matrices is square of size $\hat{r} = \operatorname{rank}(Q)$ and Q_{11} is pointwise invertible. Since $Q^\top E = E^\top Q$, we get $Q_{11}^\top E_{11} = E_{11}^\top Q_{11}$ and $E_{12} = 0$, and the transformed system, with $\begin{bmatrix} z_1^\top & z_2^\top \end{bmatrix}^\top = V^\top z$ and $U^\top(G-P) = \begin{bmatrix} G_1-P_1 \\ G_2-P_2 \end{bmatrix}$, is given by

$$\begin{aligned} &\begin{bmatrix} E_{11} & 0 \\ E_{21} & E_{22} \end{bmatrix} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} \\ &\quad = \left(\begin{bmatrix} (J_{11}-R_{11})Q_{11} & 0 \\ (J_{21}-R_{21})Q_{11} & 0 \end{bmatrix} - \begin{bmatrix} E_{11} & 0 \\ E_{21} & E_{22} \end{bmatrix} \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \right) \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} G_1-P_1 \\ G_2-P_2 \end{bmatrix} u, \end{aligned}$$

where z_1 is of size $\hat{r}$ and z_2 of size $n-\hat{r}$. By the pHDAE structure it then follows that $E_{11}K_{12} = 0$ and the resulting subsystem

$$\begin{aligned} E_{11}\dot{z}_1 &= ((J_{11}-R_{11})Q_{11} - E_{11}K_{11})z_1 + (G_1-P_1)u, \\ y &= (G_1+P_1)^\top Q_{11}z_1 + (S-N)u \end{aligned}$$

is a pHDAE with Q_{11} square and nonsingular, which determines z_1 independent of z_2. In particular, we can multiply by $Q_{11}^\top$ as discussed earlier.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

However, for given z_1 and u, the remaining DAE system for z_2,

$$E_{22}\dot{z}_2 = [E_{21} \quad E_{22}]\begin{bmatrix}K_{12}\\K_{22}\end{bmatrix} z_2 + (J_{21} - R_{21})Q_{11}z_1 - [E_{21} \quad E_{22}]\begin{bmatrix}K_{11}\\K_{21}\end{bmatrix} z_1 - E_{21}\dot{z}_1 + (G_2 - P_2)u,$$

has no apparent structure. This is not a problem, since the variable z_2 does not contribute to the Hamiltonian. In fact, further equations, as well as state, input and output variables, can always be added to a pHDAE system if they do not contribute to the Hamiltonian; see also Mehrmann and van der Schaft (2023) for a discussion of the extension of pHDAE systems by equations or removal of variables that do not contribute to the Hamiltonian.

Remark 4.11. When Q is not of full rank, even in the case of pHODE systems the solution can grow unboundedly. Mehl, Mehrmann and Wojtylak (2018) presented the Hamiltonian ODE system

$$\begin{bmatrix}\dot{z}_1\\\dot{z}_2\end{bmatrix} = JQ\begin{bmatrix}z_1\\z_2\end{bmatrix} = \begin{bmatrix}0 & -1\\1 & 0\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}z_1\\z_2\end{bmatrix}, \quad \begin{bmatrix}z_1(0)\\z_2(0)\end{bmatrix} = \begin{bmatrix}z_{1,0}\\z_{2,0}\end{bmatrix}$$

with Hamiltonian $\mathcal{H} = \frac{1}{2}z_1^2$. It has the solution $z_1 = z_{1,0}$, $z_2 = z_{2,0} + tz_{1,0}$ and thus has linear growth and is not stable.

Here the first equation $\dot{z}_1 = 0$ is a pHODE with Hamiltonian $\mathcal{H} = \frac{1}{2}z_1^2$, while the second equation $\dot{z}_2 = z_1$ has no specific structure and z_2 does not contribute to the Hamiltonian.

Remark 4.12. Egger (2019) presented a similar approach generating a representation without a Q factor for nonlinear and even infinite-dimensional evolution equations in a weak formulation. He discussed applications in linear generalized gradient systems, for which, even for $Q = I$, it may be more convenient to use a representation that reverses the roles of E and $J - R$. Note that if in $E\dot{z} = (J - R)z$ both E and $J - R$ are (pointwise) invertible, then, multiplying by E^{-1} and setting $\tilde{z} = (J - R)z$, $(J - R)^{-1} = \tilde{J} - \tilde{R}$, the equivalent new system

$$(\tilde{J} - \tilde{R})\dot{\tilde{z}} = \tilde{E}\tilde{z}$$

has $\tilde{R} \geq 0$ and $\tilde{E} > 0$.

In view of the observations concerning the term Q, we suggest one should avoid introducing a term Q in the representation on the modelling level and instead work with an E in front of the derivative.

4.4. Removing the feedthrough term in linear pHDAE systems

In many pHDAE models there is no feedthrough term $(S - N)u$, and in this case, by the semidefiniteness of the dissipation matrix W, also $P = 0$. If this is not the case, then (under some constant rank assumptions) one can always remove the feedthrough

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

term by extending the state space. However, the simple construction presented in Remark 3.1 may destroy the pH structure. In this subsection we therefore discuss how such an extension is possible while preserving the pHDAE structure.

Consider the linear time-varying or time-invariant pHDAE in (4.9) or (4.5), respectively, and assume that $D := S - N$ has constant rank. Under this assumption, by Theorem 2.13 there exists a pointwise orthogonal matrix function U_D, such that

$$D = U_D \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} U_D^\top,$$

with D_1 pointwise nonsingular. By construction, the symmetric part of D_1 is pointwise positive semidefinite. Setting

$$(G - P)U_D = [G_1 - P_1 \quad G_2 - P_2], \quad U_D^\top u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad U_D^\top y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

with analogous partitioning, the system can be written as

$$E\dot{z} = (J - R)z + (G_1 - P_1)u_1 + (G_2 - P_2)u_2, \tag{4.11a}$$
$$y_1 = (G_1 + P_1)^\top z + D_1 u_1, \tag{4.11b}$$
$$y_2 = (G_2 + P_2)^\top z. \tag{4.11c}$$

Using the positive semidefiniteness of the matrix (function) W in Definitions 4.8 and 4.9, we immediately obtain $P_2 = 0$. Let us introduce the new variable $z_2 := D_1 u_1 + P_1^\top z$ to obtain the extended system

$$\begin{bmatrix} E & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{z} \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} J - R & 0 \\ D_1^{-1} P_1^\top & -D_1^{-1} \end{bmatrix} \begin{bmatrix} z \\ z_2 \end{bmatrix} + \begin{bmatrix} G_1 - P_1 \\ I \end{bmatrix} u_1 + \begin{bmatrix} G_2 \\ 0 \end{bmatrix} u_2,$$
$$y_1 = \begin{bmatrix} G_1^\top & I \end{bmatrix} \begin{bmatrix} z \\ z_2 \end{bmatrix},$$
$$y_2 = G_2^\top z.$$

Note that by this extension the Hamiltonian and the output have not changed: they are just formulated in different variables and the added variables do not contribute to the Hamiltonian. Then, multiplying the state equation by the nonsingular matrix (function) $\left[\begin{smallmatrix} I & P_1 \\ 0 & I \end{smallmatrix}\right]$ from the left, we obtain the extended descriptor system

$$\begin{aligned} \mathcal{E}\dot{\xi} &= (\mathcal{J} - \mathcal{R})\xi + \mathcal{G}u, \\ y &= \mathcal{G}^\top \xi \end{aligned} \tag{4.12}$$

with extended state $\xi = \left[z^\top, z_2^\top\right]^\top$ and matrices

$$\mathcal{E} := \begin{bmatrix} E & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathcal{J} := \begin{bmatrix} J + \frac{1}{2}\left(P_1 D_1^{-1} P_1^\top - \left(P_1 D_1^{-1} P_1^\top\right)^\top\right) & -P_1 D_1^{-\top} \\ D_1^{-1} P_1^\top & -\frac{1}{2}\left(D_1^{-1} - D_1^{-\top}\right) \end{bmatrix},$$
$$\mathcal{G} := \begin{bmatrix} G_1 & G_2 \\ I & 0 \end{bmatrix} U_D^\top, \quad \mathcal{R} := \begin{bmatrix} R - \frac{1}{2}\left(P_1 D_1^{-1} P_1^\top + \left(P_1 D_1^{-1} P_1^\top\right)^\top\right) & 0 \\ 0 & \frac{1}{2}\left(D_1^{-1} + D_1^{-\top}\right) \end{bmatrix}.$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Theorem 4.13. Consider a linear time-varying or constant coefficient pHDAE of the form (4.11) and the extended system (4.12). Then both systems have the same input–output relation and Hamiltonian, and the extended system without feedthrough is again a pHDAE.

Proof. It remains to show that $\mathcal{R}$ is (pointwise) positive semidefinite. Since the positive semidefiniteness of $\mathcal{R}$ is equivalent to the positive semidefiniteness of the symmetric part of

$$\begin{bmatrix} R - P_1 D_1^{-1} P_1^\top & 0 \\ 0 & D_1^{-1} \end{bmatrix},$$

the claim is an immediate consequence of the positive semidefiniteness of the dissipation matrix (4.8) and the Schur complement. Note that due to the special form of the coefficient of the derivative, the changes of basis do not introduce extra derivative terms. □

Thus, in the following we often assume that $S, N = 0$, which then also implies $P = 0$. We should be aware, however, that the matrix D_1 may be ill-conditioned with respect to inversion, so from a numerical point of view the removal of the feedthrough term may be not advisable.

Remark 4.14. The extension of the pHDAE system by algebraic equations and variables that do not contribute to the Hamiltonian is the counterpart to Remark 4.11, where equations that do not contribute to the Hamiltonian can be separated from the system.

5. Applications and examples

In this section we illustrate the generality and wide applicability of the model class of pHDAE systems introduced in Section 4 with several examples from different application areas. For further examples we refer to Rashad, Califano, van der Schaft and Stramigioli (2020) and the references therein.

5.1. Singularly perturbed mechanical systems

DAE models are very often obtained as the limiting situations of singularly perturbed problems, where they are written in such a way that the limits can be taken without changing the system representation. For an illustrative example, consider the simple LTI model of a mass–spring–damper system

$$\begin{bmatrix} \dot{q} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} 0 & 1/m \\ -k & -d/m \end{bmatrix} \begin{bmatrix} q \\ p \end{bmatrix},$$

with mass m, damping constant d, spring constant k and Hamiltonian

$$\mathcal{H}(q, p) = \frac{1}{2} k q^2 + \frac{p^2}{2m}.$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Introducing new variables $v := p/m$ and $w := kq$ (see e.g. Mehrmann and van der Schaft 2023), we can express this system as an LTI pHDAE

$$\begin{bmatrix} 1/k & 0 \\ 0 & m \end{bmatrix} \begin{bmatrix} \dot{w} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & -d \end{bmatrix} \begin{bmatrix} w \\ v \end{bmatrix},$$

with Hamiltonian

$$\mathcal{H}(w, v) = \frac{1}{2k} w^2 + \frac{mv^2}{2}.$$

Then we can take either or both of the limits $m \to 0$ and $k \to \infty$, corresponding to neglecting the mass or replacing the spring connection with a rigid connection. Taking $m \to 0$ gives a DAE system of (Kronecker) index one if the damping d is non-zero, and of (Kronecker) index two if $d = 0$. The limiting Hamiltonian $\mathcal{H}(w, v) = w^2/(2k)$ no longer depends on v. For $k \to \infty$ this is a DAE of (Kronecker) index two, regardless of the value of d, with Hamiltonian $\mathcal{H}(w, v) = mv^2/2$ that no longer depends on w. Taking both limits, this is an algebraic equation with zero Hamiltonian and solution $v = 0$, $w = 0$.

In real-world structural mechanics models, the representing equations are large-scale finite element models (Hughes 2012, Zienkiewicz and Taylor 2005), and it is quite common to take such limits for some masses or stiffness constants or by ignoring some small model components, with possibly very negative effects; see Gräbner *et al.* (2016) and Kannan, Hendry, Higham and Tisseur (2014). See also Sections 5.5 and 5.8.

5.2. *RLC circuit*

An RLC circuit can be modelled as a directed graph with incidence matrix

$$\mathcal{A} = [\mathcal{A}_r \quad \mathcal{A}_c \quad \mathcal{A}_\ell \quad \mathcal{A}_v \quad \mathcal{A}_i]$$

conveniently partitioned into components associated with resistors, capacitors, inductors, voltage sources and current sources; see Freund (2011) for further details. Let V denote the vector of voltages at the nodes (except for the ground node at which the voltage is zero). Furthermore, let I_ℓ, I_v and I_i denote the vectors of currents along the edges for the inductors, voltage sources and current sources, respectively, while V_v and V_i denote the vectors of voltages across the edges for the voltage sources and current sources. Using Kirchhoff's current and voltage law combined with the so-called branch constitutive relations yields a pHDAE (in the spirit of Definition 4.9 with $Q = I_n$, $P = 0$, $S = N = 0$)

$$\begin{bmatrix} \mathcal{A}_c \mathsf{C} \mathcal{A}_c^\top & 0 & 0 \\ 0 & \mathsf{L} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{V} \\ \dot{I}_\ell \\ \dot{I}_v \end{bmatrix} = \begin{bmatrix} -\mathcal{A}_r \mathsf{R}^{-1} \mathcal{A}_r^\top & -\mathcal{A}_\ell & -\mathcal{A}_v \\ \mathcal{A}_\ell & 0 & 0 \\ \mathcal{A}_v & 0 & 0 \end{bmatrix} \begin{bmatrix} V \\ I_\ell \\ I_v \end{bmatrix} + \begin{bmatrix} \mathcal{A}_i & 0 \\ 0 & 0 \\ 0 & -I \end{bmatrix} \begin{bmatrix} -I_i \\ V_v \end{bmatrix},$$

$$\begin{bmatrix} V_i \\ -I_v \end{bmatrix} = \begin{bmatrix} \mathcal{A}_i & 0 \\ 0 & 0 \\ 0 & -I \end{bmatrix}^\top \begin{bmatrix} V \\ I_\ell \\ I_v \end{bmatrix}$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

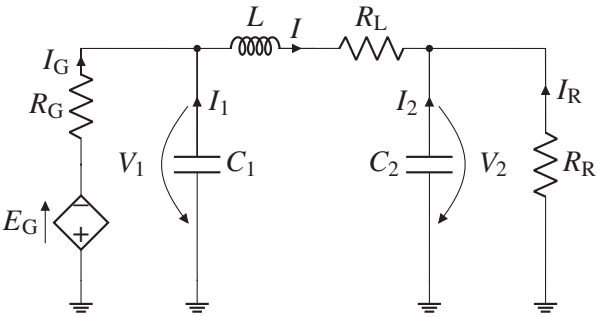

Figure 5.1. Simple DC power network example.

and Hamiltonian

$$\mathcal{H}(V, I_\ell, I_v) = V^\top \mathcal{A}_c \mathsf{C} \mathcal{A}_c^\top V + I_\ell^\top \mathsf{L} I_\ell,$$

which is associated with the stored energy in the capacitors and inductors. Here the positive definite matrices R, C and L are defined via the defining properties of the resistors, capacitors and inductors. Note that in this case we have an LTI pHDAE with a quadratic Hamiltonian, $Q = I$, and no feedthrough term.

For general circuits, Nedialkov, Pryce and Scholz (2022) have recently suggested a new pHDAE formulation that allows, in particular, for a structural index analysis and for more efficient and structurally robust implementations than classical modified nodal analysis. Another recent development is the formulation of dynamic iteration schemes for coupled pHDAE systems and their use in circuit simulation in Günther, Bartel, Jacob and Reis (2021).

5.3. Power networks

A major application of pHDAE modelling arises in power network applications. Consider the following simple model of an electrical circuit in Figure 5.1, which is presented in Mehrmann and Morandin (2019). In this model $L > 0$ is an inductor, $C_1, C_2 > 0$ are capacitors, $R_G, R_L, R_R > 0$ are resistances and E_G a controlled voltage source. This circuit can serve as a surrogate model of a DC generator (E_G,R_G), connected to a load (R_R with a transmission line and given by C_1, C_2, L, R_L). In real-world power networks, one would have a large number of generators (including wind turbines and solar panels) and loads representing customers. With a quadratic Hamiltonian describing the energy stored in the inductor and the two capacitors

$$\mathcal{H}(I, V_1, V_2) = \frac{1}{2} L I^2 + \frac{1}{2} C_1 V_1^2 + \frac{1}{2} C_2 V_2^2, \tag{5.1}$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

a formulation as an LTI pHDAE has the form

$$E\dot{z} = (J - R)z + Gu, \tag{5.2a}$$

$$y = G^\top z, \tag{5.2b}$$

with $z = [I \quad V_1 \quad V_2 \quad I_G \quad I_R]^\top$, $u = E_G$, $y = I_G$, $E = \operatorname{diag}(L, C_1, C_2, 0, 0)$ and $G = e_4 = [0 \quad 0 \quad 0 \quad 1 \quad 0]^\top$,

$$J = \begin{bmatrix} 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} R_L & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & R_G & 0 \\ 0 & 0 & 0 & 0 & R_R \end{bmatrix}.$$

If the generator is shut down (i.e. $E_G = 0$), then the system approaches an equilibrium solution for which $(\mathrm{d}/\mathrm{d}t)\mathcal{H}(z) = 0$, so that $I = I_G = I_R = 0$, and then $z = 0$.

This system can also be considered as a control problem with the control task that a consumer (represented by a resistance R_R) receives a fixed amount of power $P = R_R I_R^2$. This can be achieved by controlling the voltage of the generator E_G, so that the solution converges to the values $I_R = -\sqrt{P/R_R}$, $I = I_G \equiv -I_R$, $V_1 \equiv (R_R + R_L)I_R$, $V_2 \equiv R_R I_R$ and $E_G \equiv -(R_R + R_L + R_G)I_R$.

5.4. *Stokes and Navier–Stokes equation*

A classical example of a partial differential equation which, after proper space discretization, leads to a pHDAE (e.g. Emmrich and Mehrmann 2013) is that of the incompressible or nearly incompressible Navier–Stokes equations, describing the flow of a Newtonian fluid in a domain Ω,

$$\begin{aligned} \frac{\partial v}{\partial_t} - \nu\Delta v + (v \cdot \nabla)v + \nabla p &= f \quad \text{in } \Omega \times \mathbb{T}, \\ \nabla^\top v &= 0 \quad \text{in } \Omega \times \mathbb{T}, \end{aligned}$$

together with suitable initial and boundary conditions; see e.g. Temam (1977). When we linearize around a prescribed stationary vector field v_∞, we obtain the linearized Navier–Stokes equations

$$\begin{aligned} \frac{\partial v}{\partial_t} - \nu\Delta v + (v_\infty \cdot \nabla)v + (v \cdot \nabla)v_\infty + \nabla p &= f \quad \text{in } \Omega \times \mathbb{T}, \\ \nabla^\top v &= 0 \quad \text{in } \Omega \times \mathbb{T}. \end{aligned}$$

If v_∞ is also constant in space, then $(v \cdot \nabla)v_\infty = 0$ and we obtain the Oseen equations. If the term $(v_\infty \cdot \nabla)v$ is also neglected, we obtain the Stokes equation. Performing a finite element discretization in space (e.g. Layton 2008), a Galerkin projection leads to a dHDAE of the form

$$\begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{v} \\ \dot{p} \end{bmatrix} = \left(\begin{bmatrix} A_S & B \\ -B^\top & 0 \end{bmatrix} - \begin{bmatrix} -A_H & 0 \\ 0 & -C \end{bmatrix} \right) \begin{bmatrix} v \\ p \end{bmatrix} + \begin{bmatrix} f \\ 0 \end{bmatrix}, \tag{5.3}$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

where $M = M^\top > 0$ is the mass matrix, $A_S = -A_S^\top$ and $-A_H = -A_H^\top \geq 0$ are the skew-symmetric and symmetric parts of the discretized and linearized convection-diffusion operator, $B^\top$ is the discretized divergence operator, which we assume to be normalized so that it is of full row rank, and $-C = -C^\top > 0$ is a stabilization term, typically of small norm, that is needed for some finite element spaces; see e.g. Rannacher (2000). The variables v and p denote the discretized velocity and pressure, respectively, and f is a forcing or control term.

This becomes a pHDAE by adding an output equation $y = f^\top v$ and an appropriate Hamiltonian; see Altmann and Schulze (2017). Other possible inputs and outputs that are not necessarily co-located can be chosen, for example by different boundary conditions (added to the system via the trace operator and suitable Lagrange multipliers) or measurement points for the velocities or pressures.

5.5. *Multiple-network poroelasticity*

Biot's poroelasticity model for quasi-static deformation (Biot 1941) describes porous materials fully saturated by a viscous fluid. Typical applications include geomechanics (Zoback 2010) and biomedicine (Sobey, Eisenträger, Wirth and Czosnyka 2012). The effect of different fluid compartments can be accounted for by the theory of multiple-network poroelasticity (Bai, Elsworth and Roegiers 1993). For instance, in the investigation of cerebral oedema (Tully and Ventikos 2011), we distinguish different blood cycles (arterial, arteriole/capillary, venous) and cerebrospinal fluid, giving a total of $m = 4$ fluid compartments. The complete model is given by a coupled system of (nonlinear) PDEs

$$-\nabla \cdot (\sigma(u)) + \sum_{i=1}^{m} \nabla(\alpha_i p_i) = f, \tag{5.4a}$$

$$\frac{\partial}{\partial_t}\left(\alpha_i \nabla \cdot u + \frac{1}{M} p_i\right) - \nabla \cdot \left(\frac{\kappa_i(\nabla \cdot u)}{\nu_i} \nabla p_i\right) - \sum_{j \neq i} \beta_{ij}(p_i - p_j) = g_i, \tag{5.4b}$$

with unknown displacements u, unknown pressure variables p_i $(i = 1, \ldots, m)$ for the different fluid compartments, the Biot–Willis fluid–solid coupling coefficients α_i, Biot modulus M, fluid viscosities ν_i, (nonlinear) hydraulic conductivities $\kappa_i = \kappa_i(\nabla \cdot u)$, network transfer coefficients β_{ij}, volume-distributed external forces f and injection g_i. The stress–strain relation is given by

$$\sigma(u) := 2\mu\, \varepsilon(u) + \lambda\, (\nabla \cdot u)\, \mathcal{I}, \quad \varepsilon(u) := \frac{1}{2}\, (\nabla u + (\nabla u)^\top)$$

with the Lamé coefficients μ and λ and the identity tensor $\mathcal{I}$. For simplicity, we consider the system with Dirichlet boundary conditions

$$u = u_\mathrm{b} \quad \text{and} \quad p_i = p_{i,\mathrm{b}} \quad \text{on } \mathbb{T} \times \partial\Omega. \tag{5.4c}$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Following Altmann, Mehrmann and Unger (2021) (see also Egger and Sabouri 2021), a pHDAE of the mixed finite-element discretization of (5.4) is given by

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & K_{\mathrm{u}} & 0 & 0 & 0 \\ 0 & 0 & M_{\mathrm{p}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{w}_h \\ \dot{u}_h \\ \dot{p}_h \\ \dot{\lambda}_{\mathrm{u},h} \\ \dot{\lambda}_{\mathrm{p},h} \end{bmatrix} = \begin{bmatrix} 0 & -K_{\mathrm{u}} & D^\top & B_{\mathrm{u}}^\top & 0 \\ K_{\mathrm{u}} & 0 & 0 & 0 & 0 \\ -D & 0 & -K_{\mathrm{p}}(u_h) & 0 & B_{\mathrm{p}}^\top \\ -B_{\mathrm{u}} & 0 & 0 & 0 & 0 \\ 0 & 0 & -B_{\mathrm{p}} & 0 & 0 \end{bmatrix} \begin{bmatrix} w_h \\ u_h \\ p_h \\ \lambda_{\mathrm{u},h} \\ \lambda_{\mathrm{p},h} \end{bmatrix} + \begin{bmatrix} f_h \\ 0 \\ g_h \\ \dot{u}_{\mathrm{b},h} \\ p_{\mathrm{b},h} \end{bmatrix},$$

with positive definite mass and stiffness matrices M_{p}, K_{u} and $K_{\mathrm{p}}(u_h)$. Let us emphasize that in this representation we may use the boundary conditions as additional inputs (added to the system via the trace operator and suitable Lagrange multipliers) such that the system may be controlled via its boundary.

5.6. *Pressure waves in gas network*

The propagation of pressure waves on acoustic time scales through a network of gas pipelines is modelled in Brouwer, Gasser and Herty (2011) (see also Egger and Kugler 2018, Egger *et al.* 2018) via a linear infinite-dimensional pHDAE system on a finite directed and connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $v \in \mathcal{V}$ and edges $e \in \mathcal{E}$ that correspond to the pipes of the physical network. Let $p^e(t,x)$ denote the pressure and $f^e(t,x)$ the mass flux in pipe e, and consider the equations for the conservation of mass and the balance of momentum

$$a^e \frac{\partial}{\partial t} p^e + \frac{\partial}{\partial x} f^e = 0 \quad \text{on } e \in \mathcal{E}, t > 0,$$
$$b^e \frac{\partial}{\partial t} f^e + \frac{\partial}{\partial x} p^e + d^e q^e = 0 \quad \text{on } e \in \mathcal{E}, t > 0,$$

where the coefficients a^e, b^e encode properties of the fluid and the pipe, and d^e models the damping due to friction at the pipe walls. The coefficients are assumed to be positive and, for ease of presentation, constant on every pipe e. To model the conservation of mass and momentum at the junctions at inner vertices $v \in \mathcal{V}_i$ of the graph, where several pipes $e \in \mathcal{E}(v)$ are connected, we require Kirchhoff's law for the flow as well as continuity of the pressure, that is,

$$\sum\nolimits_{e \in \mathcal{E}(v)} n^e(v) f^e(v) = 0 \qquad \text{for all } v \in \mathcal{V}_i, t > 0,$$
$$p^e(v) = p^{e'}(v) \quad \text{for all } e, e' \in \mathcal{E}(v), v \in \mathcal{V}_i, t > 0.$$

Here $n^e(v) \in \{+1, -1\}$, depending on whether the pipe e starts or ends at the vertex v. The time-dependent quantities $m^e(v)$ and $p^e(v)$ denote the respective functions evaluated at the vertex v. At the boundary vertices $v \in \mathcal{V}_b = \mathcal{V} \setminus \mathcal{V}_i$, we define co-located ports of the network by using the pressure

$$p^e(v) = u_v \quad \text{for } v \in \mathcal{V}_b, e \in \mathcal{E}(v), t > 0$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

as input at $v \in \mathcal{V}_b$, and the mass flux

$$y_v = -n^e(v) f^e(v), \quad v \in \mathcal{V}_b,\ e \in \mathcal{E}(v),\ t > 0$$

as output. We further define initial functions

$$p(0) = p_0, \quad f(0) = f_0 \quad \text{on } \mathcal{E}.$$

Note that typically gas is only inserted at some nodes of the network and extracted at other ends. Thus one could also use different input and output variables at the external nodes that may not necessarily be co-located. Nevertheless, this formulation allows for an easy interconnection while the variables still have a physical interpretation.

In Egger and Kugler (2018) several important properties were shown. These include the existence of unique classical solutions for sufficiently smooth initial data p_0, f_0, as well as global conservation of mass, and that this system has pHDAE structure. Space discretization via a structure-preserving mixed finite element method leads to a block-structured linear time-invariant pHDAE system

$$E\dot{z} = (J - R)z + Gu, \quad z(t_0) = z_0, \tag{5.5a}$$

$$y = G^\top z, \tag{5.5b}$$

with

$$P := 0, \quad S - N := 0, \quad G := \begin{bmatrix} 0 \\ G_2 \\ 0 \end{bmatrix}, \quad z := \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix},$$

$$E := \begin{bmatrix} M_1 & 0 & 0 \\ 0 & M_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad J := \begin{bmatrix} 0 & -J_{12} & 0 \\ J_{12}^\top & 0 & J_{32}^\top \\ 0 & -J_{32} & 0 \end{bmatrix}, \quad R := \begin{bmatrix} 0 & 0 & 0 \\ 0 & R_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Here $z_1 \colon \mathbb{R} \to \mathbb{R}^{n_1}$ represents the discretized pressures and $z_2 \colon \mathbb{R} \to \mathbb{R}^{n_2}$ the discretized fluxes, while $z_3 \colon \mathbb{R} \to \mathbb{R}^{n_3}$ is a Lagrange multiplier vector introduced to penalize the violation of the space-discretized constraints. The coefficients $M_1 = M_1^\top$, $M_2 = M_2^\top$ and $R_{22} = R_{22}^\top$ are positive definite, the matrix J_{32} has full row rank and $\begin{bmatrix} J_{12}^\top & J_{32}^\top \end{bmatrix}$ has full column rank. The discretized Hamiltonian is given by $\mathcal{H}(z) = \frac{1}{2} z^\top E z = \frac{1}{2}(z_1^\top M_1 z_1 + z_2^\top M_2 z_2)$. Note that the Lagrange multiplier does not contribute to the Hamiltonian.

5.7. *Multibody systems*

Another natural class of applications arises in multibody dynamics. Hou (1994) derived the model of a two-dimensional three-link mobile manipulator; see also Bunse-Gerstner *et al.* (1999) for details. After linearizing around a stationary

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

solution, we obtain a control system

$$M\ddot{p} = -D\dot{p} - Kp + Z^\top \lambda + G_1 u,$$
$$0 = Zp,$$

where p is the vector of positions, $Zp = 0$ is the linearized position constraint, its violation is penalized by a Lagrange multiplier vector λ, and $G_1 u$ is the control force applied at the actuators. The mass and stiffness matrices $M = M^\top, K = K^\top$ are positive definite and the damping matrix $D = D^\top$ is positive semidefinite.

This DAE has the first and second time derivative of $Zp = 0$ as hidden algebraic constraints, and it is typically necessary to use a regularization procedure to make the system better suited to numerical simulation and control; see e.g. Eich-Soellner and Führer (1998), Kunkel and Mehrmann (2006), Rabier and Rheinboldt (2000) and Simeon (2013). One possibility is to replace the original constraint with its time derivative $0 = -Z\dot{p}$. By adding a tracking output $y = G_1^\top \dot{p}$ (e.g. Hou and Müller 1994) and transforming to first-order form by introducing

$$z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} := \begin{bmatrix} \dot{p} \\ p \\ \lambda \end{bmatrix},$$

we obtain a linear time-invariant pHDAE system of the form (4.9) with

$$E := \begin{bmatrix} M & 0 & 0 \\ 0 & K & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad R := \begin{bmatrix} D & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad G := \begin{bmatrix} G_1 \\ 0 \\ 0 \end{bmatrix},$$
$$Q := \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}, \quad J := \begin{bmatrix} 0 & -K & Z^\top \\ K & 0 & 0 \\ -Z & 0 & 0 \end{bmatrix}, \quad P := 0, \quad S - N := 0.$$

The quadratic Hamiltonian (4.10) is given by

$$\mathcal{H}(z) = \frac{1}{2} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}^\top \begin{bmatrix} M & 0 \\ 0 & K \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}.$$

Note that the Lagrange multiplier does not contribute to the Hamiltonian.

5.8. *Brake squeal*

Disc brake squeal is a frequent and annoying phenomenon. Gräbner *et al.* (2016) derived a very large finite element model of a brake system (see Figure 5.2) that includes friction as well as circulatory and gyroscopic effects. This has the form

$$M\ddot{q} + \left(C_1 + \frac{\omega_{\text{ref}}}{\omega} C_R + \frac{\omega}{\omega_{\text{ref}}} C_G \right) \dot{q} + \left(K_1 + K_R + \left(\frac{\omega}{\omega_{\text{ref}}} \right)^2 K_G \right) q = f,$$

where q is a vector of finite element coefficients; due to neglecting some small rotational masses, $M = M^\top \geq 0$ is a singular mass matrix; $C_1 = C_1^\top$ models material

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

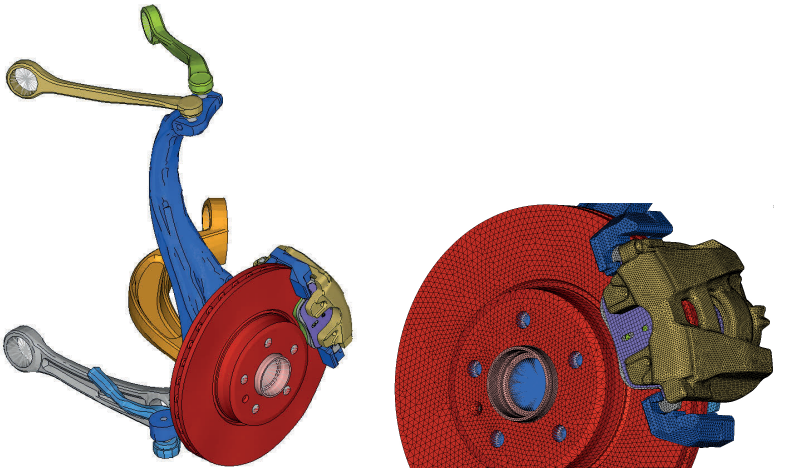

Figure 5.2. Finite element element model of disc brake.

damping; $C_G = -C_G^\top$ models gyroscopic effects; $C_R = C_R^\top$ models friction-induced damping and is typically generated from measurements; $K_1 = K_1^\top \geq 0$ is a stiffness matrix; K_R has no symmetry structure and models circulatory effects; $K_G = K_G^\top$ is a geometric stiffness matrix. One of many parameters is ω, the rotational speed of the disc scaled by a reference velocity ω_{ref}.

Experiments indicate that there is a subcritical Hopf bifurcation (in the parameter ω), when eigenvalues of the associated quadratic parametric eigenvalue problem

$$(\lambda(\omega)^2 M + \lambda(\omega)(G(\omega) + D(\omega)) + (K(\omega) + N(\omega))v(\omega) = 0$$

cross the imaginary axis. Here

$$C_1 + \frac{\omega_{\text{ref}}}{\omega} C_R + \frac{\omega}{\omega_{\text{ref}}} C_G = G(\omega) + D(\omega)$$

and

$$K_1 + K_R + \left(\frac{\omega}{\omega_{\text{ref}}}\right)^2 K_G = K(\omega) + N(\omega)$$

are split into their symmetric and skew-symmetric parts.

By writing the system in first-order formulation, it can be expressed as a perturbed dHDAE system $E\dot{z} = (J - R_{\text{D}})z - R_{\text{N}}z$, with

$$E = \begin{bmatrix} M & 0 \\ 0 & K \end{bmatrix}, \quad J = \begin{bmatrix} -G & -\left(K + \frac{1}{2}N\right) \\ \left(K + \frac{1}{2}N^\top\right) & 0 \end{bmatrix},$$
$$R_{\text{D}} = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}, \quad R_{\text{N}} = \begin{bmatrix} 0 & -\frac{1}{2}N \\ -\frac{1}{2}N^\top & 0 \end{bmatrix}.$$

Instability and squeal arise only from the perturbation term $R_{\text{N}}z$, which is associated with the brake force restricted to the finite element nodes on the brake pad.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Performing a linear stability analysis by solving the eigenvalue problem for an industrial problem, incorporating the perturbation term R_N via a homotopy parameter $E\dot{z} = (J - R_D)z - \alpha R_N z$, $\alpha \in [0, 1]$, Beckesch (2018) determined that for $\alpha = 0$ the spectral abscissa, i.e. the maximal real part of all eigenvalues, is $-5.0462e-06$ and for $\alpha = 0.1$ it is already $2.0336e-05$, that is, the unperturbed problem is already close to a problem with eigenvalues with positive real part. The application task then is to design the brake in such a way (e.g. by including damping devices, so-called shims) that the unperturbed problem $E\dot{z} = (J - R_D)z$ is such that the perturbation $R_N z$ does not lead to eigenvalues in the right half-plane, or at least make sure that they have a small real part.

6. Condensed forms for dHDAE and pHDAE systems

To analyse the solution behaviour of dHDAE or pHDAE systems, it is convenient to study canonical or condensed forms and to reformulate the system by removing high index and redundant parts, as was done for general DAE systems in Section 2.2. But the general condensed forms do not reflect the structure and, in particular, canonical forms such as the Kronecker or Weierstrass form (see Theorems 2.16 and 2.18) are obtained under transformations that may be arbitrarily ill-conditioned.

In this section, to overcome some of these disadvantages, we present condensed forms for LTI and LTV dHDAE and pHDAE systems under (pointwise) orthogonal transformations that preserve the structure. The resulting condensed forms are close to normal forms and display all the important information, but they are typically not canonical.

6.1. Condensed forms for dHDAE systems

We first present a condensed form for LTV dHDAE systems of the form (leaving off the argument t)

$$E\dot{z} = (J - R - EK)z, \quad z(t_0) = z_0, \tag{6.1}$$

which is a special case of the form for pHDAE systems presented in Scholz (2019). In particular, we assume $Q = I_n$ throughout this section; see Section 4.3.

Lemma 6.1. Under some constant rank assumptions, for a dHDAE system of the form (6.1), there exists a pointwise real orthogonal matrix function $Z \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{n,n})$, such that with $\check{z} = Z^\top z$ and multiplying the system by $Z^\top$ from the left, the transformed system

$$\check{E}\dot{\check{z}} = (\check{J} - \check{R} - \check{E}\check{K})\check{z}, \quad \check{z}(t_0) = \check{z}_0, \tag{6.2}$$

with

$$\check{E} := Z^\top E Z, \quad \check{J} := Z^\top J Z, \quad \check{R} := Z^\top R Z, \quad \check{K} := Z^\top K Z - Z^\top \dot{Z},$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

is still a dHDAE system with matrix functions in the block form

$$\check{E} = \begin{bmatrix} E_{11} & E_{12} & 0 & 0 & 0 \\ E_{21} & E_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \check{J} - \check{R} = \begin{bmatrix} J_{11} - R_{11} & J_{12} - R_{12} & J_{13} - R_{13} & J_{14} & 0 \\ J_{21} - R_{21} & J_{22} - R_{22} & J_{23} - R_{23} & 0 & 0 \\ J_{31} - R_{31} & J_{32} - R_{32} & J_{33} - R_{33} & 0 & 0 \\ J_{41} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\check{K} = \begin{bmatrix} K_{11} & K_{12} & 0 & 0 & 0 \\ K_{21} & K_{22} & 0 & 0 & 0 \\ K_{31} & K_{32} & K_{33} & K_{34} & K_{35} \\ K_{41} & K_{42} & K_{43} & K_{44} & K_{45} \\ K_{51} & K_{52} & K_{53} & K_{54} & K_{55} \end{bmatrix} \tag{6.3}$$

and block sizes $n_1 = n_4, n_2, n_3, n_5$. (Note that blocks may be void.) The matrix function $\left[\begin{smallmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{smallmatrix}\right]$ is pointwise symmetric positive definite, the matrix functions $J_{33} - R_{33}$, where $R_{33} \geq 0$ and $J_{41} = -J_{14}^\top$ are pointwise nonsingular, and the block

$$\begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix} \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}$$

is pointwise skew-symmetric.

Proof. We present a constructive proof that is similar to that for the constant coefficient case in Achleitner, Arnold and Mehrmann (2021*b*) and can be directly implemented as a numerical algorithm. The details are presented in Algorithm 1. We make the following remarks on the algorithm. First, the smooth rank-revealing decompositions can be computed using Theorem 2.13. Second, the zero block structure in $\hat{R}$ in Step 2 follows from the positive semidefiniteness of R and the fact that the second and third block rows in $\hat{K}$ do not destroy this structure since the multiplication by $\hat{E}$ puts zeros in these positions. Third, note that the transformed matrix function K in Step 3 does not destroy the structure. The extra zeros in the first two rows and the skew-symmetry of the extra term arising from the change of basis follow from the skew-adjointness of the operator $\check{E}(\mathrm{d}/\mathrm{d}t) - (\check{J} - \check{E}\check{K})$ in Definition 4.8 (for the case $Q = I$). □

From the condensed form we immediately obtain a characterization of existence and uniqueness of solutions.

Corollary 6.2. Consider a dHDAE initial value problem of the form (6.1) in the condensed form (6.3).

(i) The initial value problem (6.1) is uniquely solvable (for consistent initial values) if and only if $n_5 = 0$. If $n_5 \neq 0$, then z_5 can be chosen arbitrarily.

(ii) The solution of the initial value problem is not unique (for consistent initial values) if and only if the matrix functions E, J, R, EK have a common kernel.

(iii) The strangeness index is 0 if and only if $n_1 = n_4 = 0$. Otherwise the strangeness index is 1 if and only if $n_1 = n_4 > 0$.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Algorithm 1 Staircase Algorithm for linear dHDAEs

Input: Pair of dHDAE matrix functions $(E, J - R - EK)$
Output: Matrix function Z and condensed block form (6.3)

Step 1. Assume that E has pointwise constant rank n_1 in $\mathbb{T}$. Then perform a smooth full-rank decomposition

$$E = Z_1 \begin{bmatrix} \tilde{E}_{11} & 0 \\ 0 & 0 \end{bmatrix} Z_1^\top, \tag{6.4}$$

with Z_1 pointwise orthogonal and $\tilde{E}_{11}$ pointwise positive definite of size $\tilde{n}_1 \times \tilde{n}_1$ (or $\tilde{n}_1 = 0$). Set $\tilde{E} := Z_1^\top E Z_1$, $\tilde{J} := Z_1^\top J Z_1$, $\tilde{R} := Z_1^\top R Z_1$ and $\tilde{K} := Z_1^\top (K Z_1 + \dot{Z}_1)$ with

$$\tilde{J} = \begin{bmatrix} \tilde{J}_{11} & -\tilde{J}_{21}^\top \\ \tilde{J}_{21} & \tilde{J}_{22} \end{bmatrix}, \quad \tilde{R} = \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{21}^\top \\ \tilde{R}_{21} & \tilde{R}_{22} \end{bmatrix}, \quad \tilde{K} = \begin{bmatrix} \tilde{K}_{11} & \tilde{K}_{12} \\ \tilde{K}_{21} & \tilde{K}_{22} \end{bmatrix}.$$

Step 2. If $\tilde{n}_1 < n$ then, assuming constant rank of $\tilde{J}_{22} - \tilde{R}_{22}$ in $\mathbb{T}$, apply a full-rank decomposition under pointwise orthogonal congruence,

$$Z_{22}^\top (\tilde{J}_{22} - \tilde{R}_{22}) Z_{22} = \begin{bmatrix} \tilde{\Sigma}_{22} & 0 \\ 0 & 0 \end{bmatrix},$$

with $\tilde{\Sigma}_{22}$ of size $\tilde{n}_2 \times \tilde{n}_2$ pointwise invertible or $\tilde{n}_2 = 0$. Define the matrix functions $Z_2 := \operatorname{diag}(I, Z_{22})$, $\hat{E} := Z_2^\top \tilde{E} Z_2$, $\hat{J} := Z_2^\top \tilde{J} Z_2$, $\hat{R} := Z_2^\top \tilde{R} Z_2$ and $\hat{K} := Z_2^\top \tilde{K} Z_2 - Z_2^\top \dot{Z}_2$ with

$$\hat{J} = \begin{bmatrix} \hat{J}_{11} & -\hat{J}_{21}^\top & -\hat{J}_{31}^\top \\ \hat{J}_{21} & \hat{J}_{22} & 0 \\ \hat{J}_{31} & 0 & 0 \end{bmatrix}, \quad \hat{R} = \begin{bmatrix} \hat{R}_{11} & \hat{R}_{21}^\top & 0 \\ \hat{R}_{21} & \hat{R}_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \hat{K} = \begin{bmatrix} \hat{K}_{11} & \hat{K}_{12} & \hat{K}_{13} \\ \hat{K}_{21} & \hat{K}_{22} & \hat{K}_{23} \\ \hat{K}_{31} & \hat{K}_{32} & \hat{K}_{33} \end{bmatrix},$$

and $\hat{J}_{22} - \hat{R}_{22}$ pointwise nonsingular.

Step 3. If $\tilde{n}_3 := n - \tilde{n}_1 - \tilde{n}_2 > 0$ then, assuming that $\hat{J}_{31}$ has constant rank in $\mathbb{T}$, perform a pointwise full-rank decomposition

$$\hat{J}_{31} = U_{31} \begin{bmatrix} \Sigma_{31} & 0 \\ 0 & 0 \end{bmatrix} V_{31}^\top,$$

with Σ_{31} of size $n_1 \times n_1$ pointwise nonsingular (or $n_1 = 0$). Set

$$Z_3 := \begin{bmatrix} V_{31}^\top & & \\ & I & \\ & & U_{31} \end{bmatrix}.$$

Then $\check{E} := Z_3^\top \hat{E} Z_3$, $\check{J} := Z_3^\top \hat{J} Z_3$, $\check{R} := Z_3^\top \hat{R} Z_3$, $\check{K} = Z_3^\top \hat{K} Z_3 - Z_3^\top \dot{Z}_3$ have the desired form with $n_2 := \tilde{n}_1 - n_1$, $n_3 := \tilde{n}_2$, $n_4 = n_1$, $n_5 := \tilde{n}_3 - n_4$.

Step 4. Set $Z := Z_1 Z_2 Z_3$.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Proof. We prove each item separately.

(i) Note that the last equation can be omitted and the variable z_5 is arbitrary. So the solution is unique if and only if $n_5 = 0$.

(ii) This follows trivially from (i).

(iii) If $n_4 > 0$, then the fourth equation states that $z_1 = 0$ and the first equation yields that z_4 depends via the term $E_{12}\dot{z}_2$ on the derivative of z_2, and hence if $n_4 \neq 0$ then the system has strangeness index 1. Finally, the separated subsystem

$$\begin{bmatrix} E_{22} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{z}_2 \\ \dot{z}_3 \end{bmatrix} = \left(\begin{bmatrix} J_{22} - R_{22} & J_{23} - R_{23} \\ J_{32} - R_{32} & J_{33} - R_{33} \end{bmatrix} - \begin{bmatrix} E_{21}K_{12} + E_{22}K_{22} & 0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} z_2 \\ z_3 \end{bmatrix}$$

has E_{22} positive definite and $J_{33} - R_{33}$ invertible and hence is of strangeness index 0. □

Remark 6.3. In Algorithm 1 we have made several constant rank assumptions. If these are not satisfied, then we can partition the time interval $\mathbb{T}$ into smaller subintervals where the ranks are constants and record the points where these rank changes happen. For a more detailed discussion of such *hybrid DAE systems* with rank changes in the characteristic values; see Hamann and Mehrmann (2008) and Kunkel and Mehrmann (2006, 2018).

In the linear time-invariant case (with $K = 0$, $Q = I$), we have the following corollary.

Corollary 6.4. For every LTI dHDAE system of the form

$$E\dot{z} = (J - R)z, \tag{6.5}$$

with $E, J, R \in \mathbb{R}^{n,n}$, $J = -J^\top$, $R = R^\top \geq 0$ and $E^\top = E \geq 0$, there exists a real orthogonal matrix $Z \in \mathbb{R}^{n,n}$ such that with

$$\check{E} := Z^\top E Z, \quad \check{J} := Z^\top J Z, \quad \check{R} := Z^\top R Z, \quad \check{z} := Z^\top z,$$

the system $\check{E}\dot{\check{z}} = (\check{J} - \check{R})\check{z}$ is still a dHDAE with block matrices

$$\check{E} = \begin{bmatrix} E_{11} & E_{12} & 0 & 0 & 0 \\ E_{21} & E_{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \check{J} - \check{R} = \begin{bmatrix} J_{11} - R_{11} & J_{12} - R_{12} & J_{13} - R_{13} & J_{14} & 0 \\ J_{21} - R_{21} & J_{22} - R_{22} & J_{23} - R_{23} & 0 & 0 \\ J_{31} - R_{31} & J_{32} - R_{32} & J_{33} - R_{33} & 0 & 0 \\ J_{41} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{6.6}$$

and block sizes $n_1 = n_4, n_2, n_3, n_5$. Note that some of the blocks may be void. The matrix $\begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}$ is symmetric positive definite, and the matrices $J_{33} - R_{33}$ and $J_{41} = -J_{14}^\top$ are nonsingular.

Proof. The proof follows directly from the time-varying case and setting $K = 0$. □

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

If we do not restrict ourselves to congruence transformations with orthogonal matrices, then the condensed form (6.6) can be simplified further.

Corollary 6.5. Consider a dHDAE of the form (6.5) with regular matrix pencil $\lambda E - (J - R)$ in staircase form (6.6). Then there exist nonsingular matrices L_1, L_2 such that

$$L_1 \check{E} L_2 = \begin{bmatrix} \hat{E}_{11} & 0 & 0 & 0 \\ 0 & \hat{E}_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad L_1(\check{J} - \check{R})L_2 = \begin{bmatrix} 0 & 0 & 0 & I \\ 0 & \hat{J}_{22} - \hat{R}_{22} & 0 & 0 \\ 0 & 0 & -I & 0 \\ -I & 0 & 0 & 0 \end{bmatrix}. \quad (6.7)$$

The blocks satisfy $\hat{J}_{22} = -\hat{J}_{22}^\top$, $\hat{E}_{11} = \hat{E}_{11}^\top > 0$, $\hat{E}_{22} = \hat{E}_{22} > 0$ and $\hat{R}_{22} = \hat{R}_{22}^\top \geq 0$.

Proof. The proof follows by block Gaussian elimination to create first the block diagonal structure of $\hat{E}$, using the positive definite diagonal block E_{22}. This is followed by block Gaussian elimination using the nonsingular blocks $J_{41} = -J_{14}^\top$ and $J_{33} - R_{33}$ and then scaling these nonsingular blocks. Note that $\hat{E}_{22} = E_{22}$ with E_{22} as in (6.6), and $\hat{J}_{22}$, $\hat{R}_{22}$ are the skew-symmetric and symmetric part of the Schur complement obtained in this way, so the semidefiniteness of $\hat{R}_{22}$ follows from a Schur complement-type argument. See Achleitner *et al.* (2021*b*) for details. □

We also have the corresponding existence and uniqueness result formulated with the (Kronecker) index.

Corollary 6.6. Consider a dHDAE initial value problem of the form (6.5) in the staircase form (6.6).

(i) The initial value problem (6.5) is uniquely solvable (for consistent initial values) if and only if $n_5 = 0$. If $n_5 \neq 0$, then $\tilde{z}_5$ can be chosen arbitrarily.

(ii) The pencil $\lambda E - (J - R)$ is non-regular if and only if the three coefficients E, J, R have a common kernel.

(iii) The (Kronecker) index is zero if and only if $n_1 = n_4 = 0$ and $n_3 = 0$. It is one if and only if $n_1 = n_4 = 0$ and $n_3 > 0$, and otherwise the (Kronecker) index is two if and only if $n_1 = n_4 > 0$.

Proof. The proof is as in the variable coefficient case (with $K = 0$), just observing the different counting between the strangeness index and the Kronecker index; see Mehrmann (2015). □

Remark 6.7. The fact that the dHDAE system is not uniquely solvable if and only if the coefficients have a common nullspace is remarkable. It has significant importance in model evaluation, since it allows us to check this (pointwise) via the singular value decomposition. Moreover, in the constant coefficient case, this allows us to compute the distance to the nearest singular pencil; see Guglielmi and Mehrmann (2022) and Mehl *et al.* (2021). For general pencils, and even for pencils

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

with just the symmetry structure of a dHDAE system, this property does not hold. Consider the pencil $\lambda E - J$ with

$$E = E^\top = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad J = -J^\top = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}.$$

The pencil is not regular, since $\det(\lambda E - J) = 0$ for any $\lambda \in \mathbb{C}$. However, E and J do not have a common nullspace. The problem here is that $E = E^\top$ is not semidefinite. In this case the singularity arises from higher-dimensional singular blocks in the Kronecker form.

Remark 6.8. As the proof of Lemma 6.13 shows, the numerical computation of the condensed forms (6.3) or (6.6) for a given dHDAE system requires a sequence of three (smooth) full-rank decompositions. Unfortunately, these rank decisions may be sensitive under perturbations; see e.g. Byers, Mehrmann and Xu (2007), where the construction of general staircase forms and the challenges are discussed. However, in contrast to general unstructured staircase forms, we see that for dHDAE systems the number of steps is limited to three and often (as in all the examples discussed in Section 5) the first step of performing a full-rank decomposition of E is not necessary.

Note also that the maximal strangeness index is one (the maximal (Kronecker) index of an LTI dHDAE is two), which is of great advantage in iterative solution methods (see Güdücü, Liesen, Mehrmann and Szyld (2022) and Section 9.2), as well as time discretization methods for DAE systems (see Hairer and Wanner (1996), Kunkel and Mehrmann (2006) and Section 9.1).

Furthermore, for general LTI DAE systems it was shown in Mehl *et al.* (2021) when they are equivalent to systems of the form (4.9) and in Mehrmann and van der Schaft (2023) the same result was proved for an extended class of pHDAE systems.

Example 6.9. To illustrate the construction of the condensed forms, consider the dHDAE resulting from the fluid flow example discussed in Section 5.4. In more detail, consider an instationary incompressible fluid flow prescribed in terms of velocity $v\colon \Omega \times \mathbb{T} \to \mathbb{R}^2$ and pressure $p\colon \Omega \times \mathbb{T} \to \mathbb{R}$ on the spatial domain $\Omega = (0,1)^2$ with boundary $\partial\Omega$ for the time period $\mathbb{T} = [0, T]$, that is driven by external forces $f\colon \Omega \times \mathbb{T} \to \mathbb{R}^2$ and has dynamic viscosity $\nu > 0$; see Section 5.4. The system is closed by non-slip boundary conditions and an appropriate initial value v^0 for the velocity. Spatial discretization by a finite difference method on a uniform staggered grid with the semidiscretized velocity $v_h(t) \in \mathbb{R}^{n_v}$ and pressure vectors $p_h(t) \in \mathbb{R}^{n_p}$, $t \in \mathbb{T}$, leads to a pHDAE system for the state $z = \begin{bmatrix} v_h^\top & p_h^\top \end{bmatrix}^\top$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

given by

$$\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{v}_h \\ \dot{p}_h \end{bmatrix} = \left(\begin{bmatrix} A_S & B \\ -B^\top & 0 \end{bmatrix} - \begin{bmatrix} A_H & 0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} v_h \\ p_h \end{bmatrix} + \begin{bmatrix} G_1 \\ 0 \end{bmatrix} u, \qquad y = \begin{bmatrix} G_1^\top & 0 \end{bmatrix} \begin{bmatrix} v_h \\ p_h \end{bmatrix}. \tag{6.8}$$

The matrix $B^\top$ has full row rank if the freedom in the pressure is removed. The initial conditions are $v_h(0) = v_h^0$ and consistently $p_h(0) = p_h^0$. The input u with input matrix $G_1 \in \mathbb{R}^{n_v \times m}$ results from the external forces. System (6.8) is an LTI pHDAE of (Kronecker) index two. To obtain the condensed form we do not have to carry out the first and second step of Algorithm 1 but only Step 3, i.e. the splitting of the discrete divergence operator $B^\top$, by performing a Leray projection, for example, as in Heinkenschloss, Sorensen and Sun (2008), or a full-rank decomposition,

$$B^\top V = \begin{bmatrix} B_1 & 0 \end{bmatrix},$$

with nonsingular matrix B_1 and an orthogonal matrix $V \in \mathbb{R}^{n_p, n_p}$. Performing a congruence transformation, we get a system in condensed form,

$$\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \end{bmatrix} = \left(\begin{bmatrix} J_{11} & J_{12} & B_1 \\ -J_{12}^\top & J_{22} & 0 \\ -B_1^\top & 0 & 0 \end{bmatrix} - \begin{bmatrix} R_{11} & R_{12} & 0 \\ R_{12}^\top & R_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \\ 0 \end{bmatrix} u,$$
$$y = \begin{bmatrix} G_1^\top & G_2^\top & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}.$$

This yields $z_1 = 0$ as B_1 is invertible, and the first equation yields the (hidden) algebraic constraint $B_1 z_3 = (-J_{12} + R_{12}) z_2 - G_1 u$ as well as a consistency condition for the initial value which relates the initial condition for u and z_2 to that for z_3.

Example 6.10. Consider the multiple-network poroelasticity problem discussed in Section 5.5. Ignoring the boundary terms and permuting the rows and columns, we obtain a dHDAE system of the form

$$\begin{bmatrix} K_\mathrm{u} & 0 & 0 \\ 0 & M_\mathrm{p} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & K_\mathrm{u} \\ 0 & -K_\mathrm{p} & -D \\ -K_\mathrm{u} & D^\top & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}, \tag{6.9}$$

with symmetric positive definite matrices M_p, K_u and K_p. Thus Steps 1 and 2 of Algorithm 1 are already done, and we see immediately that (6.9) is a pHDAE of (Kronecker) index two. Notably, if we do not assume that the model is quasi-static, then the $(3, 3)$ block in the matrix on the left-hand side of (6.9) is nonsingular and we end up with an implicit dissipative Hamiltonian ODE that can be viewed as a singular perturbation of (6.9). This illustrates that small perturbations may change the index within the pHDAE framework. Nevertheless, in contrast to general DAEs, the (Kronecker) index may be at most two.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Example 6.11. In the gas network example presented in Section 5.6, and more precisely for the dHDAE obtained by setting $u = 0$ in (5.5), we know directly from the structure what the constraints are, since the system is almost in the form that would be obtained from the staircase algorithm. The first step and second step are already performed, $n_3 = 0$, and it remains to transform the matrix $[0 \quad -J_{32}]$, which has full row rank, by an orthogonal transformation $\hat{U}_J$ to the form $[0 \quad -J_{32}]\hat{U}_J = [\hat{J}_{31} \quad 0]$, with $\hat{J}_{31}$ square and nonsingular. Setting $U_J := \left[\begin{smallmatrix} \hat{U}_J & 0 \\ 0 & I \end{smallmatrix}\right]$ and forming $\hat{E} = U_J^\top E U_J$, $\hat{J} = U_J^\top J U_J$ and $\hat{R} = U_J^\top R U_J$, we get a dHDAE system with $n_1 = n_4 = \operatorname{rank} J_{32}$, $n_3 = n_5 = 0$ and $n_2 = n - 2n_1$.

The simplified construction of the condensed forms for the other examples presented in Sections 5.7 and 5.8 is analogous.

Although for linear time-varying dHDAE systems we get a condensed form under pointwise orthogonal congruence transformations, in contrast to the constant coefficient case, we would need time-varying changes of basis. Such a coordinate transformation requires derivatives of the basis transformation matrices, so a computational method needs to determine a smooth transformation. Such transformations can be determined via methods such as a smooth singular value decomposition (see Theorems 2.3 and 2.13) or a QR decomposition (e.g. Bunse-Gerstner, Byers, Mehrmann and Nichols 1991, Dieci and Eirola 1999, Kunkel and Mehrmann 1991). However, these methods require the solution of matrix differential equations (operating in the orthogonal group). This substantially increases the computational costs.

6.2. Structure-preserving index reduction

Although pointwise condensed forms help a lot in the analysis, they are often not practical in computational techniques. For general DAE systems we therefore proceed differently and use derivative arrays (see Kunkel and Mehrmann (1996, 2006) and Section 2.2) to filter out a strangeness-free system by transformations that act only on the equations and their derivatives and avoid changes of basis. However, this approach will, in general, destroy the dHDAE structure.

An approach that achieves structure preservation on the basis of one smooth change of basis has been proposed by Beattie *et al.* (2018) for the case of pHDAE systems that include the factor Q, using a technique that was introduced for self-adjoint systems in Kunkel, Mehrmann and Scholz (2014) and skew-adjoint systems in Kunkel and Mehrmann (2023). We present this approach here for the case $Q = I$ and assume that the system has a well-defined strangeness index and a unique solution for all consistent initial conditions. Under these assumptions (see Section 2.2) from the (unstructured) derivative array, one can extract an algebraic equation of the form $\hat{A}_2 z = 0$ that contains all the explicit or hidden constraint equations. Then there exists a pointwise orthogonal $T \in \mathcal{C}^1(\mathbb{T}, \mathbb{R}^{n,n})$ such that

$$\hat{A}_2 [T_1 \quad T_2] = [0 \quad \hat{A}_{22}],$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

with $\hat{A}_{22} \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{a,a})$ pointwise nonsingular, so that the columns of the matrix function T_1 span the kernel of $\hat{A}_{22}$.

Using the same proof as in Kunkel *et al.* (2014) for self-adjoint DAE systems (see also Kunkel and Mehrmann 2023), it follows that the matrix function consisting of the first $d = n - a$ rows and columns of $T^\top ET$ is square nonsingular, and therefore positive definite. With this in mind, the original dHDAE can be transformed congruently with T, so that the dHDAE structure is preserved, and with

$$\begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix} = T^\top ET, \quad z = T \begin{bmatrix} z_1 \\ z_2 \end{bmatrix},$$
$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = T^\top (J - R)T - (T^\top ET)(T^\top KT - T^\top \dot{T}),$$

we obtain $A_{22} z_2 = 0$ and hence $z_2 = 0$. Inserting this, the remaining part of the first block equation

$$E_{11} \dot{z}_1 = A_{11} z_1$$

is still a dHDAE but now with E_{11} positive definite, and the Hamiltonian is unchanged since $z_2 = 0$.

Remark 6.12. For nonlinear pHDAE systems satisfying Hypothesis 2.7 with $\mu > 0$, the corresponding local result follows directly via linearization and the implicit function theorem.

6.3. Condensed forms for pHDAE systems

In this section we extend the results presented in the previous section for dHDAE systems to systems with inputs and outputs. For LTI pHDAEs of the form (4.9), the extension of the condensed forms to pHDAE systems was presented in Beattie, Gugercin and Mehrmann (2022*a*). For an LTV pHDAE system (4.5) this result follows by a variation of the condensed form presented in Scholz (2019); see also Kunkel and Mehrmann (2023).

Let us first consider the following result, which allows us to remove the non-uniqueness part, if any exists. Note that we again consider the case $Q = I$, $S - N = 0$ and $P = 0$; see Sections 4.3 and 4.4.

Lemma 6.13. For a pHDAE of the form (4.5) with $Q = I$, $S - N = 0$ and $P = 0$, under some constant rank assumptions, there exists a pointwise orthogonal change of basis $V^\top z =: \tilde{z} = \begin{bmatrix} \tilde{z}_1^\top & \tilde{z}_2^\top & \tilde{z}_3^\top \end{bmatrix}^\top$ such that the system has the form

$$\begin{bmatrix} E_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\tilde{z}}_1 \\ \dot{\tilde{z}}_2 \\ \dot{\tilde{z}}_3 \end{bmatrix} = \begin{bmatrix} J_1 - R_1 - E_1 K_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \\ \tilde{z}_3 \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \\ 0 \end{bmatrix} u, \tag{6.10a}$$

$$y = \begin{bmatrix} G_1^\top & G_2^\top & 0 \end{bmatrix} \begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \\ \tilde{z}_3 \end{bmatrix}, \tag{6.10b}$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

where the system $E_1\dot{z}_1 - (J_1 - R_1 - E_1K_1)z_1 + G_1u$ has a unique solution for every sufficiently often differentiable u, and G_2 has full row rank. Furthermore, the subsystem

$$\begin{bmatrix} E_1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \dot{\tilde{z}}_1 \\ \dot{\tilde{z}}_2 \end{bmatrix} = \begin{bmatrix} J_1 - R_1 - E_1K_1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} u, \tag{6.11a}$$

$$y = \begin{bmatrix} G_1^\top & G_2^\top \end{bmatrix}\begin{bmatrix} \tilde{z}_1 \\ \tilde{z}_2 \end{bmatrix}, \tag{6.11b}$$

obtained by removing the third equation and the variable $\tilde{z}_3$, is still a pHDAE with the same Hamiltonian.

Proof. The proof follows by computing, as in (6.4), a pointwise orthogonal matrix V_1 such that

$$V_1^\top(J - R - EK)V_1 - \left(V_1^\top EV_1\right)V_1^\top\dot{V}_1 = \begin{bmatrix} J_1 - R_1 - E_1K_1 & 0 \\ 0 & 0 \end{bmatrix},$$

$$V_1^\top EV_1 = \begin{bmatrix} E_1 & 0 \\ 0 & 0 \end{bmatrix}, \quad V_1^\top G = \begin{bmatrix} G_1 \\ \tilde{G}_2 \end{bmatrix}.$$

This V_1 exists by the condensed form (6.3), by just combining the first four rows and columns into one block. Then a row compression of $\tilde{G}_2$ via a pointwise orthogonal matrix $\tilde{V}_2$ (assuming constant rank) and a congruence transformation with $V_2 = \operatorname{diag}(I, \tilde{V}_2)$ is performed, so by a congruence transformation with $V = \operatorname{diag}(I, V_2)V_1$ we obtain the zero pattern in (6.10). Updating the output equation accordingly gives the desired form. □

Remark 6.14. In practice, the constant rank assumptions that are required for the derivation of (6.3) can be reduced by performing only the transformation that splits off the common nullspace part of E, J, R, EK.

Lemma 6.13 shows that we can remove redundant equations and variables that do occur in the system. In the following we assume that this reduction has already been performed. The next result presents a condensed form, which extends the form that was obtained in Beattie *et al.* (2022*a*) for LTI pHDAE systems with $K = 0$ and for LTV pHDAE systems in Scholz (2019).

Lemma 6.15. Consider an LTV pHDAE as in (6.11). Then, under some constant rank assumptions, there exists a pointwise orthogonal basis transformation V in the state space and U in the control space, such that in the new variables

$$\hat{z} = \begin{bmatrix} \hat{z}_1^\top & \hat{z}_2^\top & \hat{z}_3^\top & \hat{z}_4^\top & \hat{z}_5^\top & \hat{z}_6^\top \end{bmatrix}^\top = V^\top\begin{bmatrix} \tilde{z}_1^\top & \tilde{z}_2^\top \end{bmatrix}^\top \quad \text{and}$$

$$\hat{u} = \begin{bmatrix} u_1^\top & u_2^\top & u_3^\top \end{bmatrix}^\top = U^\top u,$$

the system has the form

$$\hat{E}\dot{\hat{z}} = (\hat{J} - \hat{R} - \hat{E}\hat{K})\hat{z} + \hat{G}\hat{u},$$

$$\hat{y} = \hat{G}^\top\hat{z},$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

with

$$\hat{E} := \begin{bmatrix} E_{11} & E_{12} & 0 & 0 & 0 & 0 \\ E_{21} & E_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \hat{G} := \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \\ 0 & G_{42} & G_{43} \\ 0 & 0 & G_{53} \\ 0 & 0 & G_{63} \end{bmatrix}, \tag{6.12a}$$

$$\hat{J} - \hat{R} := \begin{bmatrix} J_{11} - R_{11} & J_{12} - R_{12} & J_{13} - R_{13} & J_{14} & J_{15} & 0 \\ J_{21} - R_{21} & J_{22} - R_{22} & J_{23} - R_{23} & J_{24} & 0 & 0 \\ J_{31} - R_{21} & J_{32} - R_{32} & J_{33} - R_{33} & 0 & 0 & 0 \\ J_{41} & J_{42} & 0 & 0 & 0 & 0 \\ J_{51} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{6.12b}$$

$$\hat{K} := \begin{bmatrix} K_{11} & K_{12} & K_{13} & K_{14} & 0 & 0 \\ K_{21} & K_{22} & K_{23} & K_{24} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{6.12c}$$

where E_{22}, $J_{33} - R_{33}$, J_{15}, and G_{42} and G_{63} are pointwise invertible.

Proof. Starting from (6.11), in the first step (similarly as in (6.4)), we determine a pointwise orthogonal matrix function $\tilde{V}_1$ such that

$$\tilde{V}_1^\top E_1 \tilde{V}_1 = \begin{bmatrix} \tilde{E}_{11} & 0 \\ 0 & 0 \end{bmatrix}$$

with $\tilde{E}_{11} > 0$, and then perform a congruence transformation with the matrix function $V_1 = \operatorname{diag}(\tilde{V}_1, I)$, yielding $\tilde{V}_1^\top G_1 = \begin{bmatrix} \tilde{G}_1 \\ \tilde{G}_2 \end{bmatrix}$ and

$$V_1^\top (J - R - EK) V_1 - (V_1^\top E V_1) V_1^\top \dot{V}_1 = \begin{bmatrix} \tilde{J}_{11} - \tilde{R}_{11} - \tilde{E}_{11}\tilde{K}_{11} & \tilde{J}_{12} - \tilde{R}_{12} \\ \tilde{J}_{21} - \tilde{R}_{21} & \tilde{J}_{22} - \tilde{R}_{22} \end{bmatrix}.$$

Next, under the assumption of a constant rank, compute a smooth full-rank decomposition

$$\tilde{V}_2^\top (\tilde{J}_{22} - \tilde{R}_{22}) \tilde{V}_2 = \begin{bmatrix} \hat{J}_{22} - \hat{R}_{22} & 0 \\ 0 & 0 \end{bmatrix},$$

where $\hat{J}_{22} - \hat{R}_{22}$ is invertible and $\hat{R}_{22} \geq 0$. Such a full-rank decomposition exists, since $\tilde{J}_{22} - \tilde{R}_{22}$ has a positive semidefinite symmetric part. Then, defining $V_2 := \operatorname{diag}(I, \tilde{V}_2, I)$ and applying an appropriate congruence transformation with

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

$\hat{V} := V_1 V_2$ yields

$$\hat{V}^\top E \hat{V} = \begin{bmatrix} \tilde{E}_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \hat{V}^\top G = \begin{bmatrix} \hat{G}_1 \\ \hat{G}_2 \\ \hat{G}_3 \\ \hat{G}_4 \end{bmatrix},$$

$$\hat{V}^\top (J - R - EK)\hat{V} = \begin{bmatrix} \hat{J}_{11} - \hat{R}_{11} - \tilde{E}_{11}\tilde{K}_{11} & \hat{J}_{12} - \hat{R}_{12} - \tilde{E}_{11}\tilde{K}_{12} & \hat{J}_{13} - \tilde{E}_{11}\tilde{K}_{13} & 0 \\ \hat{J}_{21} - \hat{R}_{21} & \hat{J}_{22} - \hat{R}_{22} & 0 & 0 \\ \hat{J}_{31} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where $\hat{J}_{22} - \hat{R}_{22}$ is invertible and $\hat{G}_4$ has full row rank. Note that there is no contribution of R in the third block column and row, which is due to the fact that $\hat{V}^\top R \hat{V}$ is pointwise positive semidefinite.

Note also that the terms $\tilde{K}_{12}$ and $\tilde{K}_{13}$ arise due to the fact that the transformation from the right operates in these block columns. Note further that $\hat{G}_4$ has full row rank so it can be transformed by a change of basis to be of the form $[0 \quad \bar{G}_{63}]$ with invertible $\bar{G}_{63}$. Combining this with a smooth full-rank decomposition of the block $\hat{G}_3$, one can perform a smooth transformation

$$\begin{bmatrix} \tilde{V}_3^\top & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \hat{G}_3 \\ \hat{G}_4 \end{bmatrix} U = \begin{bmatrix} 0 & \bar{G}_{42} & \bar{G}_{43} \\ 0 & 0 & \bar{G}_{53} \\ 0 & 0 & \bar{G}_{63} \end{bmatrix},$$

with $\bar{G}_{42}$ and $\bar{G}_{63}$ square and pointwise nonsingular, where the number of rows in $\bar{G}_{63}$ is that of $\hat{G}_4$. Applying an appropriate congruence transformation with $V_3 = \operatorname{diag}(I, I, \tilde{V}_3, I)$, we obtain block matrices

$$\begin{bmatrix} \bar{E}_{11} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \bar{G}_{11} & \bar{G}_{12} & \bar{G}_{13} \\ \bar{G}_{21} & \bar{G}_{22} & \bar{G}_{23} \\ 0 & \bar{G}_{42} & \bar{G}_{43} \\ 0 & 0 & \bar{G}_{53} \\ 0 & 0 & \bar{G}_{63} \end{bmatrix},$$

$$\begin{bmatrix} \bar{J}_{11} - \bar{R}_{11} - \bar{E}_{11}\bar{K}_{11} & \bar{J}_{12} - \bar{R}_{12} - \bar{E}_{11}\bar{K}_{12} & \bar{J}_{13} - \bar{E}_{11}\bar{K}_{13} & \bar{J}_{14} - \bar{E}_{11}\bar{K}_{14} & 0 \\ \bar{J}_{21} - \bar{R}_{21} & \bar{J}_{22} - \bar{R}_{22} & 0 & 0 & 0 \\ \bar{J}_{31} & 0 & 0 & 0 & 0 \\ \bar{J}_{41} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

As a final step we compute a column compression of the full row rank matrix function $\bar{J}_{41}$ and apply an appropriate congruence transformation. This yields the desired form. □

Since in the condensed form (6.12) the blocks J_{51}, G_{42} and G_{63} are pointwise invertible, it follows immediately that $u_3 = 0$, and that $\hat{z}_1 = -G_{53}u_3 = 0$ and $\hat{z}_5$ is

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

uniquely determined by the other variables and their derivatives. These parts are associated with equations for which a regularization is necessary; see Section 6.4.

Example 6.16. The construction in Example 6.11 can also be used to derive the condensed form for the LTI pHDAE, which yields a system of the form (6.12) in which the third, fourth and sixth block rows and the corresponding block columns do not occur. The system has (Kronecker) index two as a free system with $u = 0$.

Example 6.17. The power network from Section 5.3 is already in the condensed form (6.12), where the first, fourth, fifth and sixth block rows and columns do not occur, so the system has (Kronecker) index one as a free system with $u = 0$.

6.4. Regularization via output feedback

Consider a pHDAE system of the form (6.12) and denote the system that is obtained by removing the variables $\hat{z}_1$, $\hat{z}_4$ and the corresponding first and fifth equations by

$$\hat{E}\dot{\hat{z}} = (\hat{J} - \hat{R} - \hat{E}\hat{K})\hat{z} + \hat{G}\,u, \tag{6.13a}$$

$$\hat{y} = \hat{G}^\top \hat{z}. \tag{6.13b}$$

System (6.13) can be viewed as the subsystem that is controllable and observable at ∞ (see Definition 3.3), since we have the following corollary.

Corollary 6.18. For system (6.13) there exists an output feedback

$$u = -\hat{W}\hat{y} + w$$

with $\hat{W} + \hat{W}^\top > 0$, so that the resulting closed-loop system is a pHDAE system

$$\hat{E}\dot{\hat{z}} = (\hat{J} - \hat{R} - \hat{E}\hat{K} - \hat{G}\hat{W}\hat{G}^\top)\hat{z} + \hat{G}w,$$
$$\hat{y} = \hat{G}^\top \hat{z},$$

and is regular and strangeness-free as a free system with $w = 0$.

Proof. This follows directly from the structure of the system, and the fact that $\hat{E}$ is already in a form where the kernel of $\hat{E}$ and $\hat{E}^\top$ can be directly read off. □

Remark 6.19. The condensed form (6.12) is a structured pHDAE version of the condensed forms in Byers, Geerts and Mehrmann (1997*a*) and Byers, Kunkel and Mehrmann (1997*b*), which allow us to remove parts of the system that cannot be made strangeness-free, uniquely solvable or of strangeness index zero ((Kronecker) index at most one) via (output) feedback.

A similar process of removing parts from a system, and that cannot be made strangeness-free by (output) feedback, has been discussed for general nonlinear DAE systems by Campbell *et al.* (2012). The procedure can be applied directly to pHDAE systems.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

7. Properties of pHDAE systems

In this section we discuss several general properties of the model class of pHDAE systems and show why they are a very good candidate for our modelling wish list.

7.1. Invariance under transformations and projection

An essential property of the class of pHDAE systems is the invariance under different equivalence transformations; see Beattie *et al.* (2018) and Mehrmann and Morandin (2019).

Let us begin with general state space-transformations and consider $\tilde{\mathcal{Z}} \subseteq \mathbb{R}^{\tilde{n}}$ and define $\tilde{\mathcal{S}} := \mathbb{T} \times \tilde{\mathcal{Z}}$ with elements $(t, \tilde{z}) \in \tilde{\mathcal{S}}$. Let $\varphi \in \mathcal{C}^1(\tilde{\mathcal{S}}, \mathcal{Z})$ and let $U \in \mathcal{C}(\tilde{\mathcal{S}}, \mathbb{R}^{\ell,\ell})$ be invertible. Define

$$\psi : \tilde{\mathcal{S}} \to \mathcal{S}, \quad (t, \tilde{z}) \mapsto (t, \varphi(t, \tilde{z}))$$

and

$$\begin{aligned}
&\tilde{E} := U^\top (E \circ \psi) \frac{\partial}{\partial z} \varphi, && \tilde{J} := U^\top (J \circ \psi) U, && \tilde{R} := U^\top (R \circ \psi) U, \\
&\tilde{G} := U^\top (G \circ \psi), && \tilde{P} := U^\top (P \circ \psi), && \tilde{\eta} := U^\top (\eta \circ \psi), \\
&\tilde{S} := S \circ \psi, && \tilde{N} := N \circ \psi, && \tilde{r} := U^\top (r \circ \psi) + (E \circ \psi) \frac{\partial}{\partial t} \varphi,
\end{aligned}$$

together with a transformed Hamiltonian

$$\widetilde{\mathcal{H}} := \mathcal{H} \circ \psi. \tag{7.1}$$

We have the following transformation result taken from Mehrmann and Morandin (2019).

Theorem 7.1. Consider a pHDAE (4.1) together with the transformed descriptor system

$$\tilde{E}(t, \tilde{z}) \dot{\tilde{z}} + \tilde{r}(t, \tilde{z}) = (\tilde{J}(t, \tilde{z}) - \tilde{R}(t, \tilde{z})) \tilde{\eta}(t, \tilde{z}) + (\tilde{G}(t, \tilde{z}) - \tilde{P}(t, \tilde{z})) u, \tag{7.2a}$$

$$y = (\tilde{G}(t, \tilde{z}) + \tilde{P}(t, \tilde{z}))^\top \tilde{\eta}(t, \tilde{z}) + (\tilde{S}(t, \tilde{z}) - \tilde{N}(t, \tilde{z})) u, \tag{7.2b}$$

and transformed Hamiltonian $\widetilde{\mathcal{H}}$ in (7.1). Then the following properties hold.

(i) The descriptor system (7.2) is a pHDAE.

(ii) If $\varphi(t, \cdot)$ is a local diffeomorphism for all $t \in \mathbb{T}$, then to any behaviour solution $(\tilde{z}, u, y)$ of (7.2) there corresponds a behaviour solution (z, u, y) of (4.1) with $z(t) = \varphi(t, \tilde{z}(t))$.

(iii) If $\varphi(t, \cdot)$ is a global diffeomorphism for all $t \in \mathbb{T}$, then there is a one-to-one correspondence between a behaviour solution $(\tilde{z}, u, y)$ of (7.2) and a behaviour solution (z, u, y) of (4.1) with $z(t) = \varphi(t, \tilde{z}(t))$.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Proof. The transformed DAE system is obtained from (4.1) by setting $z = \varphi(t, \tilde{z})$, pre-multiplying by $U^\top$ and inserting UU^{-1} in front of z in the first equation. It is clear that if $(\tilde{z}, u, y)$ is a solution of (7.2) then (z, u, y) is a behaviour solution of the original system, and if $\varphi(t, \cdot)$ is a global diffeomorphism then we can apply the inverse transformation to get a behaviour solution $(\tilde{z}, u, y)$ for any behaviour solution (z, u, y) of the original system.

To show that (7.2) is still a pHDAE, we must check the defining conditions. By substitution, we get

$$\begin{aligned}\tilde{W} = \begin{bmatrix} \tilde{R} & \tilde{P} \\ \tilde{P}^\top & \tilde{S} \end{bmatrix} &= \begin{bmatrix} U^\top (R \circ \varphi) U & U^\top (P \circ \varphi) \\ (P \circ \varphi)^\top U & S \circ \varphi \end{bmatrix} \\ &= \begin{bmatrix} U & 0 \\ 0 & I \end{bmatrix}^\top (W \circ \varphi) \begin{bmatrix} U & 0 \\ 0 & I \end{bmatrix} \geq 0,\end{aligned}$$

since positive semidefiniteness is invariant under congruence. We also get

$$\begin{aligned}\frac{\partial}{\partial \tilde{z}} \tilde{\mathcal{H}}(t, \tilde{z}) &= \left(\frac{\partial}{\partial \tilde{z}} \varphi\right)^\top \left(\frac{\partial}{\partial z} \mathcal{H} \circ \varphi\right) = \left(\frac{\partial}{\partial \tilde{z}} \varphi\right)^\top (E^\top z \circ \varphi) \\ &= \left(\frac{\partial}{\partial \tilde{z}} \varphi\right)^\top (E^\top \circ \varphi) U U^{-1} (z \circ \varphi) = \tilde{E}^\top \tilde{z},\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \widetilde{\mathcal{H}}}{\partial t}(t, \tilde{z}) &= \frac{\partial \mathcal{H}(t)}{\partial t} \circ \varphi + \left(\frac{\partial}{\partial z} \mathcal{H} \circ \varphi\right)^\top \frac{\partial \varphi}{\partial t} = z^\top r \circ \varphi + (z^\top E \circ \varphi) \frac{\partial \varphi}{\partial t} \\ &= (z \circ \varphi)^\top \left(r \circ \varphi + (E \circ \varphi) \frac{\partial \varphi}{\partial t}\right) = \tilde{z}^\top U^\top U^{-\top} \tilde{r} = \tilde{z}^\top \tilde{r}. \qquad \square\end{aligned}$$

For linear pHDAE systems with quadratic Hamiltonian, this invariance takes the following form; see Beattie *et al.* (2018).

Theorem 7.2. Consider a linear pHDAE system of the form (4.5) with quadratic Hamiltonian (4.6). Let $U \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{\ell,\ell})$ and $V \in \mathcal{C}^1(\mathbb{T}, \mathbb{R}^{n,n})$ be pointwise nonsingular in $\mathbb{T}$. Then the transformed DAE

$$\begin{aligned}\tilde{E}\dot{\tilde{z}} &= [(\tilde{J} - \tilde{R})\tilde{Q} - \tilde{E}\tilde{K}]\tilde{z} + (\tilde{G} - \tilde{P})u, \\ y &= (\tilde{G} + \tilde{P})^\top \tilde{Q} \tilde{z} + (S + N)u,\end{aligned}$$

with

$$\begin{aligned}&\tilde{E} := U^\top E V, && \tilde{Q} := U^{-1} Q V, && \tilde{J} := U^\top J U, \\ &\tilde{R} := U^\top R U, && \tilde{G} := U^\top G, && \tilde{P} := U^\top P, \\ &\tilde{K} := V^{-1} K V + V^{-1} \dot{V}, && z = V \tilde{z}, &&\end{aligned}$$

is still a pHDAE system with the same quadratic Hamiltonian

$$\tilde{\mathcal{H}}(\tilde{z}) := \frac{1}{2} \tilde{z}^\top \tilde{Q}^\top \tilde{E} \tilde{z} = \mathcal{H}(z).$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Proof. The transformed DAE system is obtained from the original DAE system by setting $z = V\tilde{z}$, pre-multiplying by $U^\top$ and inserting UU^{-1} in front of Q. The transformed operator corresponding to $\mathcal{L}$ in (4.7) is given by

$$\mathcal{L}_V := \tilde{Q}^\top \tilde{E} \frac{\mathrm{d}}{\mathrm{d}t} - \tilde{Q}^\top (\tilde{J}\tilde{Q} - \tilde{E}\tilde{K}).$$

Because

$$\tilde{Q}^\top \tilde{E} = V^\top Q^\top EV, \quad \tilde{Q}^\top \tilde{J}\tilde{Q} = V^\top Q^\top JQV, \quad \tilde{Q}^\top \tilde{E}V^{-1}\dot{V} = V^\top Q^\top E\dot{V},$$

we see that $\mathcal{L}_V$ is again skew-adjoint. It is then straightforward to show that $\tilde{\mathcal{H}}(\tilde{z}) = \mathcal{H}(z)$, and

$$\tilde{W} = \begin{bmatrix} \tilde{Q}^\top \tilde{R}\tilde{Q} & \tilde{Q}^\top \tilde{P} \\ \tilde{P}^\top \tilde{Q} & S \end{bmatrix} = \begin{bmatrix} V & 0 \\ 0 & I \end{bmatrix}^\top W \begin{bmatrix} V & 0 \\ 0 & I \end{bmatrix},$$

where W is as defined in (4.8). Because $W(t)$ is positive semidefinite for all $t \in \mathbb{T}$, so is $\tilde{W}(t)$. □

Note that even if $K = 0$ in an LTV pHDAE system with quadratic Hamiltonian, after the transformation given in Theorem 7.2 the extra term $-\tilde{E}\tilde{K}$ with $\tilde{K} = V^{-1}\dot{V}$ will appear, and if an orthogonal change of basis is carried out in a system with $K = 0$ then $\tilde{K} = V^{-1}\dot{V}$ is skew-symmetric. However, even if $K \neq 0$, Beattie *et al.* (2018) have shown that this term can be removed via a change of basis transformation that does not change the quadratic Hamiltonian.

Lemma 7.3. Consider a pHDAE system

$$\begin{aligned} \tilde{E}\dot{\tilde{z}} &= [(\tilde{J} - \tilde{R})\tilde{Q} - \tilde{E}\tilde{K})]\tilde{z} + (\tilde{G} - \tilde{P})u, \\ y &= (\tilde{G} + \tilde{P})^\top \tilde{Q}\tilde{z} + (S - N)u \end{aligned}$$

with Hamiltonian $\tilde{\mathcal{H}}(\tilde{z}) = \frac{1}{2}\tilde{z}^\top \tilde{Q}^\top \tilde{E}\tilde{z}$ and $\tilde{K} \in \mathcal{C}(\mathbb{T}, \mathbb{R}^{n,n})$. If $V_{\tilde{K}} \in \mathcal{C}^1(\mathbb{T}, \mathbb{R}^{n,n})$ is a pointwise invertible solution of the matrix differential equation $\dot{V} = V\tilde{K}$ with initial condition $V(t_0) = I$, then setting $\tilde{z} = V_K^{-1} z$ and defining

$$E := \tilde{E}V_K^{-1}, \quad Q := \tilde{Q}V_K^{-1}, \quad G := \tilde{G}, \quad J := \tilde{J}, \quad R := \tilde{R}, \quad P := \tilde{P},$$

the system

$$\begin{aligned} E\dot{z} &= (J - R)Qz + (G - P)u, \\ y &= (G + P)^\top Qz + (S - N)u \end{aligned}$$

is again a pHDAE with the same Hamiltonian $\mathcal{H}(z) = \tilde{\mathcal{H}}(\tilde{z})$.

Proof. For a given matrix function $\tilde{K}$, the system $\dot{V} = V\tilde{K}$ always has a solution V_K that is pointwise invertible. The remainder of the proof follows by reversing the proof of Theorem 7.2 with $U = I$ and using $\dot{V}_K V_K^{-1} = -V_K(\mathrm{d}/\mathrm{d}t)(V_K^{-1})$. □

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Another important observation for linear pHDAE systems of the form (4.5) (with $Q = I$) that is important in the context of space discretization and model reduction is that the pHDAE structure is invariant under Galerkin projection.

Corollary 7.4. Consider a pHDAE system of the form (4.5) with quadratic Hamiltonian (4.6) and assume that $K = 0$. If $V \in \mathbb{R}^{n,k}$ for some $k \in \mathbb{N}$, then the projected system in the variable $z = V\tilde{z}$,

$$\begin{aligned} \tilde{E}\dot{\tilde{z}} &= (\tilde{J} - \tilde{R})\tilde{z} + (\tilde{G} - \tilde{P})u, \\ y &= (\tilde{G} + \tilde{P})^\top \tilde{z} + (S - N)u, \end{aligned}$$

with projected matrix functions

$$\tilde{E} := V^\top E V, \quad \tilde{J} := V^\top J V, \quad \tilde{R} := V^\top R V, \quad \tilde{G} := V^\top G, \quad \tilde{P} := V^\top P,$$

is still a pHDAE with projected Hamiltonian $\widetilde{\mathcal{H}}(\tilde{z}) = \frac{1}{2}\tilde{z}^\top \tilde{E}\tilde{z}$.

Remark 7.5. Note that for LTV pHDAE systems with $K \neq 0$, strictly speaking, the invariance under Galerkin projection is violated, unless we incorporate a term $VV^\top$ between E and K so that the projected coefficient K is approximated by $\tilde{K} = V^\top(K - \dot{V})V$.

The discussed invariance of a pHDAE system under transformations allows us to simplify the representation further. These simplifications are discussed in Section 6 for LTI and LTV pHDAE systems.

7.2. Structure-preserving interconnection

Another key property of the pHDAE model class that is particularly important for modularized, network-based modelling across physical domains is that it is preserved under interconnection. To see this, consider two autonomous pHDAEs (see Remark 4.2)

$$\begin{aligned} E_i \dot{z}_i &= (J_i - R_i)\eta_i + (G_i - P_i)u_i, \\ y_i &= (G_i + P_i)^\top \eta_i + (S_i - N_i)u_i, \end{aligned}$$

of the form (4.1) with Hamiltonians $\mathcal{H}_i$, for $i = 1, 2$, and assume that inputs and outputs satisfy a linear interconnection relation

$$\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Then the interconnected system can be written as a pHDAE of the form

$$\begin{aligned} \mathcal{E}\dot{z} &= (\mathcal{J} - \mathcal{R})\eta + \mathcal{G}u, \\ y &= \mathcal{G}^\top \eta, \end{aligned}$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

where

$$z = \begin{bmatrix} z_1^\top & z_2^\top & z_3^\top & z_4^\top & z_5^\top & z_6^\top \end{bmatrix}^\top,$$
$$\eta = \begin{bmatrix} \eta_1^\top & \eta_2^\top & \eta_3^\top & \eta_4^\top & \eta_5^\top & \eta_6^\top & 0 & 0 \end{bmatrix}^\top,$$

with new state variables $z_3 := \eta_3 := u_1$, $z_4 := \eta_4 := u_2$, $z_5 := \eta_5 := y_1$, $z_6 := \eta_6 := y_2$, matrix functions

$$\mathcal{E} = \begin{bmatrix} E_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & E_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathcal{G} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ I & 0 \\ 0 & I \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$\mathcal{J} - \mathcal{R} = \begin{bmatrix} J_1 - R_1 & 0 & G_1 - P_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & J_2 - R_2 & 0 & G_2 - P_2 & 0 & 0 & 0 & 0 \\ -G_1^\top - P_1^\top & 0 & N_1 - S_1 & 0 & I & 0 & -M_{11}^\top & -M_{21}^\top \\ 0 & -G_2 - P_2^\top & 0 & N_2 - S_2 & 0 & I & -M_{12}^\top & -M_{22}^\top \\ 0 & 0 & -I & 0 & 0 & 0 & -L_{11}^\top & -L_{21}^\top \\ 0 & 0 & 0 & -I & 0 & 0 & -L_{12}^\top & -L_{22}^\top \\ 0 & 0 & M_{11} & M_{12} & L_{11} & L_{12} & 0 & 0 \\ 0 & 0 & M_{21} & M_{22} & L_{21} & L_{22} & 0 & 0 \end{bmatrix},$$

and combined Hamiltonian $\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2$. It is clear that the structural conditions of the coefficients are still satisfied.

Unfortunately, due to the extension by extra state variables, the dimension of the state space may substantially increase. However, if we assume that the interconnection is power-preserving (e.g. if $Mu + Ly = 0$ defines a Dirac structure for (y, u)) (see Section 7.6), then index reduction and removal of certain parts (see Section 6) can usually be applied to make the system smaller.

If we restrict ourselves to linear pHDAEs of the form

$$E_i \dot{z}_i = (J_i - R_i) z_i + G_i u_i,$$
$$y_i = G_i^\top z_i$$

(see Sections 4.3 and 4.4), then we can achieve the interconnection in a more condensed form. Assume an ouput feedback of the form $u = Fy + w$ with aggregated variables $u := \begin{bmatrix} u_1^\top & u_2^\top \end{bmatrix}^\top$ and $y := \begin{bmatrix} y_1^\top & y_2^\top \end{bmatrix}^\top$. Define $z := \begin{bmatrix} z_1^\top & z_2^\top \end{bmatrix}^\top$, $J := \operatorname{diag}(J_1, J_2)$, $R := \operatorname{diag}(R_1, R_2)$, $G = \operatorname{diag}(G_1, G_2)$. Then the coupled system has the form

$$E\dot{z} = (J - R + GFG^\top) z + Gw,$$
$$y = G^\top z,$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

which is a pHDAE whenever $R - GF_{\mathrm{sym}}G^\top$ is positive semidefinite, where

$$F_{\mathrm{sym}} := \frac{1}{2}(F + F^\top).$$

A sufficient condition to retain the pH structure is thus to require that F_{sym} is negative semidefinite, corresponding to a potentially dissipative component of the interconnection.

7.3. Storage energy balance equation and dissipation inequality

A key property of pHDAE systems that shows their strong roots in the underlying physical principles is the storage energy balance equation and the associated dissipation inequality; see also Section 3.4.

Theorem 7.6. The pHDAE (4.1) satisfies the storage energy balance equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(t, z(t)) = -\begin{bmatrix}\eta(t,z)\\ u\end{bmatrix}^\top W(t,z)\begin{bmatrix}\eta(t,z)\\ u\end{bmatrix} + y^\top u$$

along any behaviour solution of the pHDAE (4.1).

Proof. Let $[z^\top \quad u^\top \quad y^\top]^\top$ be a behaviour solution of the pHDAE (4.1). Using the structural properties of the pHDAE system, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(t,z) &= \frac{\partial}{\partial t}\mathcal{H}(t,z) + \left(\frac{\partial}{\partial z}\mathcal{H}(t,z)\right)^\top \dot{z} = \eta(t,z)^\top(E(t,z)\dot{z} + r(t,z))\\ &= \eta(t,z)^\top((J(t,z) - R(t,z))z + (B(t,z) + P(t,z) - 2P(t,z))u)\\ &= -\begin{bmatrix}\eta(t,z)\\ u\end{bmatrix}^\top \begin{bmatrix}R(t,z) & P(t,z)\\ P(t,z)^\top & S(t,z)\end{bmatrix}\begin{bmatrix}\eta(t,z)\\ u\end{bmatrix} + y^\top u\\ &= -\begin{bmatrix}\eta(t,z)\\ u\end{bmatrix}^\top W(t,z)\begin{bmatrix}\eta(t,z)\\ u\end{bmatrix} + y^\top u.\end{aligned}$$

□

Using the fact that $W \geq 0$ along any solution of (2.1), we immediately obtain that the pHDAE system satisfies the *dissipation inequality*

$$\mathcal{H}(t_1, z(t_1)) - \mathcal{H}(t_0, z(t_0)) \leq \int_{t_0}^{t_1} y(s)^\top u(s)\,\mathrm{d}s. \tag{7.3}$$

The storage energy balance equation and the dissipation inequality are obvious in all the examples described in Section 5. In all cases the dissipation term R is positive semidefinite. In the disc brake for the unforced system, without applying the brake force (see Section 5.8), this is also the case. The perturbation term which moves eigenvalues to the right half-plane is due to the external force that can be interpreted as a supplied energy via the term $y^\top u$.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

7.4. Stability and passivity of general dHDAE and pHDAE systems

In this subsection we show that for general dHDAE and pHDAE systems the stability analysis is straightforward, since it will turn out that the associated Hamiltonian is a Lyapunov function. To show this, we will use the storage energy balance equation from Theorem 7.6 and for the passivity the dissipation inequality (7.3).

We have already seen in Section 7 that every pHDAE can be easily made autonomous by turning it into the form (4.4a) without changing the Hamiltonian. The dissipation inequality (7.3) then implies that the Hamiltonian $\mathcal{H}$ is locally positive semidefinite in an equilibrium point $z^\star$ and hence $\mathcal{H}$ is a Lyapunov function.

Corollary 7.7. Consider an autonomous dHDAE obtained from the pHDAE (4.1) by setting $u = 0$ and omitting the output equation. If the system is regular and strangeness-free, then the part of the solution that is contributing to the Hamiltonian (see Section 4.3) is stable. Furthermore, in this case a sufficient condition for the system to be asymptotically stable is that $R(t, z) > 0$.

Remark 7.8. Note that in Corollary 7.7 we obtain (asymptotic) stability only for the part of the system that contributes to the Hamiltonian. This is not specific for descriptor pH systems, but already required for pHODEs; see Remark 4.11.

Remark 7.9. For the examples in Section 5 we immediately have asymptotic stability for the circuit (5.2), whereas the multi-body systems in Section 5.7, the gas network problem in Section 5.6, the poroelasticity problem in Section 5.5 and the fluid-dynamics example in Section 5.4 are not strangeness-free, but (asymptotic) stability is obtained after removing the algebraic parts that are associated with a strangeness index that is greater than zero, or (Kronecker) index greater than one. By an appropriate output feedback, the circuit example of Section 5.2 can be made to have (Kronecker) index one.

The dissipation inequality directly implies that strangeness-free pHDAE systems are passive.

Corollary 7.10. Consider an autonomous pHDAE of the form (4.1). If the system is strangeness-free then it is passive. Furthermore, in this case a sufficient condition for the system to be strictly passive is that $W(t, z) > 0$ for all $(t, z) \in \mathbb{T} \times \mathcal{Z}$.

Example 7.11. The disc brake example in Section 5.8 is in general not stable, but it may be stable if the perturbation term is small enough.

7.5. Stability

In this subsection we discuss in more detail the stability of pHDAE systems and, conversely, show that an (asymptotically) stable system can typically be written as a pHDAE system.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Beginning with LTI ODE systems, an immediate consequence of the Lyapunov characterization of stability is the existence of a dHDAE formulation of (asymptotically) stable LTI ODE systems; see Gillis, Mehrmann and Sharma (2018) and Mehrmann and van der Schaft (2023).

Corollary 7.12. Consider the LTI ODE system (3.12) and suppose that $X = X^\top > 0$ is a solution of the Lyapunov inequality (3.13). Setting $E := X, J := -\frac{1}{2}(A^\top X - XA)$, and $R := -\frac{1}{2}(A^\top X + XA)$ implies that

$$E\dot{z} = (J - R)z \tag{7.4}$$

is a dHDAE system with $R \geq 0$. If the Lyapunov inequality is strict, then we have $R > 0$.

Conversely, every dHDAE of the form (7.4) with $E > 0$, $R \geq 0$ ($R > 0$) is (asymptotically) stable.

Proof. The proof follows trivially, since $-2R = A^\top X + XA$. □

Remark 7.13. Note that if we consider an asymptotically stable linear system (3.12) and split $A = J - R$ into its skew-symmetric and symmetric part, then in general we do not have $R \geq 0$. Consider the example

$$A = \begin{bmatrix} 1 & 2\sqrt{7} \\ -2\sqrt{7} & -10 \end{bmatrix}, \quad J = \begin{bmatrix} 0 & 2\sqrt{7} \\ -2\sqrt{7} & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 0 \\ 0 & -10 \end{bmatrix},$$

where the matrix A has eigenvalues -3 and -6, but R is indefinite.

Corollary 7.12 provides the sufficient condition $R > 0$ for asymptotic stability. Nevertheless, a system may also be stable for singular $R \geq 0$, as is shown by the following characterization from Achleitner *et al.* (2021*b*).

Lemma 7.14. Consider the LTI system (3.12) with

$$A = J - R, \quad J = -J^\top, \quad 0 \leq R = R^\top.$$

Then the following conditions are equivalent.

(i) There exists a non-negative integer m_H such that

$$\operatorname{rank}[R \quad JR \quad \cdots \quad J^{m_H}R] = n.$$

(ii) There exists a non-negative integer m_H such that

$$T_{m_H} := \sum_{j=0}^{m_H} J^j R (J^\top)^j > 0.$$

(iii) No eigenvector of J lies in the kernel of R.

(iv) We have $\operatorname{rank}[\lambda I - J \quad R] = n$ for every $\lambda \in \mathbb{C}$, in particular for every eigenvalue λ of J.

Moreover, the smallest possible m_H in (i) and (ii) coincide.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Proof. See Achleitner *et al.* (2021*b*). □

The smallest possible m_H in (i) and (ii) of Lemma 7.14 is called the *hypocoercivity index* of A, and we have the following corollary.

Corollary 7.15. Consider the LTI system (3.12) with

$$A = J - R, \quad J = -J^\top, \quad 0 \leq R = R^\top.$$

Then the system is asymptotically stable if and only if the hypocoercivity index is finite.

Remark 7.16. Achleitner, Arnold and Carlen (2021*a*) showed that if the hypocoercivity index m_H is finite, then for the fundamental solution $\mathrm{e}^{At} \in \mathbb{R}^{n,n}$ of (3.12), the short-time decay in the spectral norm is given by

$$\|\mathrm{e}^{At}\|_2 = 1 - ct^{2m_H+1} + O(t^{2m_H+2}) \quad \text{for } t \to 0+,$$

with a constant $c > 0$.

We now extend these result to the dHDAE setting, where $E \geq 0$ is allowed to be singular. We have the following stability characterizing spectral properties taken from Mehl *et al.* (2018), where an extended result that deals with singular and high index dHDAE systems is also provided.

Theorem 7.17. Consider an LTI dHDAE of the form (7.4) and suppose that the pencil $\lambda E - (J - R)$ is regular and of (Kronecker) index at most one.

(i) If $\lambda_0 \in \mathbb{C}$ is an eigenvalue of $\lambda E - (J - R)$, then $\mathrm{Re}(\lambda_0) \leq 0$.

(ii) If $\omega \in \mathbb{R}$ and $\lambda_0 = \mathrm{i}\omega$ is an eigenvalue of $\lambda E - (J - R)$, then λ_0 is semisimple. Moreover, if the columns of $V \in \mathbb{C}^{n,k}$ form a basis of a regular deflating subspace of $\lambda E - (J - R)$ associated with the eigenvalue λ_0, then $RV = 0$.

Proof.

(i) Let $\lambda_0 \in \mathbb{C}$ be an eigenvalue of $\lambda E - (J - R)$ and let $v \neq 0$ be an eigenvector associated with λ_0. Then we have $\lambda_0 Ev = (J - R)v$ and thus

$$\lambda_0 v^H Ev = v^H Jv - v^H Rv.$$

Considering the real parts of both sides of this equation, we obtain

$$\mathrm{Re}(\lambda_0) v^H Ev = -v^H Rv,$$

where we used the fact that E and R are symmetric and J is skew-symmetric. If $Ev = 0$, then also $(J - R)v = 0$, which would imply that the pencil is singular. Hence we have $Ev \neq 0$, and since E is positive semidefinite, we obtain $v^H Ev > 0$, which finally implies

$$\mathrm{Re}(\lambda_0) = -\frac{v^H Rv}{v^H Ev} \leq 0.$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

(ii) We first prove the 'moreover' part. For this, let the columns of $V \in \mathbb{C}^{n,k}$ form a basis of a regular deflating subspace of $\lambda E - (J - R)$ associated with the eigenvalue $\lambda_0 = \mathrm{i}\omega$, $\omega \in \mathbb{R}$, that is, there exists a matrix $W \in \mathbb{C}^{n,k}$ with full column rank such that $EV = W$ and $(J - R)V = WT$, where $T \in \mathbb{C}^{k,k}$ only has the eigenvalue $\mathrm{i}\omega$. Without loss of generality we may assume that $T = \mathrm{i}\omega I_k + N$ is in Jordan canonical form, where N is strictly upper triangular. Then $V^H(J - R)V = V^H WT$, and taking the Hermitian part on both sides we obtain

$$0 \geq -2V^H RV = V^H WT + T^H W^H V.$$

Since R is positive semidefinite, it remains to show that $V^H RV = 0$, because then we also have $RV = 0$. For this, we show that

$$V^H W = W^H V > 0.$$

This follows, since first $V^H W = V^H EV \geq 0$. If there exists $x \neq 0$ such that $EVx = 0$, then with $y = Vx = y_1 + \mathrm{i}y_2$, y_1, y_2 real, we have $Ey = 0$. This implies that $Ey_1 = 0$ and $y_2 = 0$. Hence

$$y \in \operatorname{span}(\ker E \cup \ker(J - R)) \subseteq \operatorname{span}\left(\bigcup_{\lambda \in \mathbb{S}} \ker(\lambda E - (J - R))\right),$$

with $\mathbb{S} := (\mathbb{C} \cup \{\infty\}) \setminus \{\lambda_0\}$, which contradicts the fact that the columns of V span a regular deflating subspace associated with λ_0.

If M is the inverse of the Hermitian positive definite square root of $V^H W$, then

$$M(V^H WT + T^H V^H W)M = M^{-1}TM + MT^H M^{-1} \leq 0.$$

Moreover,

$$\begin{aligned}\operatorname{trace}(M^{-1}TM + MT^H M^{-1}) &= \operatorname{trace}(M^{-1}TM) + \operatorname{trace}(MT^H M^{-1}) \\ &= \operatorname{trace}(T + T^H) = \operatorname{trace}(N + N^H) = 0,\end{aligned}$$

because N has a zero diagonal. But this implies

$$M^{-1}TM + MT^H M^{-1} = 0$$

and hence also $-2V^H RV = 0$, which finishes the proof of the 'moreover' part.

To show that $\mathrm{i}\omega$ is a semisimple eigenvalue, it remains to show that the matrix $T = \mathrm{i}\omega I_k + N$ is diagonal, i.e. $N = 0$. Since purely imaginary eigenvalues of the system correspond to eigenvectors of the non-dissipative system ($R = 0$) that are in the kernel of R (Mehl, Mehrmann and Sharma 2016), with $RV = 0$ we get $EV = W$ and $JV = WT$, which implies that $V^H JV = V^H WT$. Then

$$M^{-1}TM = MV^H WTM = MV^H JVM$$

implies that T is similar to a matrix which is congruent to J, that is, T is similar to a skew-symmetric matrix, which implies that $N = 0$. Thus, $\mathrm{i}\omega$ is a

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

semisimple eigenvalue of $\lambda E - (J - R)$ and assertion (ii) is proved. See Mehl *et al.* (2018) for further details. □

Analogous to the ODE case, Theorem 7.17 implies that dHDAE systems with regular pencils of (Kronecker) index at most one are stable, but they are not necessarily asymptotically stable. To characterize asymptotic stability, a hypocoercivity index and the corresponding Lyapunov inequality for dHDAE systems is introduced in Achleitner *et al.* (2021*b*) for the DAE case as well. For the precise statement and the proof, we exploit the condensed form from Corollary 6.5 that preserves the dHDAE structure. In more detail, we employ the fact that for any regular dHDAE system of the form (7.4) there exist nonsingular matrices $S, T \in \mathbb{R}^{n,n}$ such that with

$$\hat{E} := SET = \begin{bmatrix} \hat{E}_{11} & 0 & 0 & 0 \\ 0 & \hat{E}_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \hat{J} - \hat{R} = \begin{bmatrix} 0 & 0 & 0 & I \\ 0 & \hat{J}_{22} - \hat{R}_{22} & 0 & 0 \\ 0 & 0 & -I & 0 \\ -I & 0 & 0 & 0 \end{bmatrix}, \tag{7.5}$$

the system $\hat{E}\dot{\hat{z}} = (\hat{J} - \hat{R})\hat{z}$ is still a dHDAE and the matrices $\hat{E}_{11}$, $\hat{E}_{22}$ are symmetric positive definite, $\hat{J}_{22}$ is skew-symmetric and $\hat{R}_{22}$ is symmetric positive semidefinite.

Assume now that the dHDAE system (7.4) is transformed into the form (7.5) with transformed state vector $\hat{z} = \begin{bmatrix} \hat{z}_1^\top & \hat{z}_2^\top & \hat{z}_3^\top & \hat{z}_4^\top \end{bmatrix}^\top$ partitioned according to the block structure; then we immediately obtain that $\hat{z}_1 = 0$, $\hat{z}_3 = 0$ and $\hat{z}_4 = 0$, which gives restrictions in the initial values. Using the fact that $\hat{E}_{22} = \hat{E}_{22}^\top > 0$, the hypocoercivity index of (7.4) is then defined in Achleitner *et al.* (2021*b*) as that of the ODE

$$\dot{\xi}_2 = \hat{E}_{22}^{-1/2}(\hat{J}_{22} - \hat{R}_{22})\hat{E}_{22}^{-1/2}\xi_2, \tag{7.6}$$

with $\xi_2 = \hat{E}_{22}^{1/2}\hat{z}_2$. We have the following characterization of asymptotic stability; see Achleitner *et al.* (2021*b*).

Corollary 7.18. If a dHDAE system of the form (7.4) has a regular pencil $\lambda E - (J - R)$ with (Kronecker) index at most one (i.e. the first block-row and block-column in (7.5) are not present), and non-trivial dynamics with a finite hypocoercivity index, then for every consistent initial condition the solution is asymptotically stable.

Remark 7.19. Using the transformation to the condensed form (7.5) we see that the short-time decay is as for the ODE case. This can be viewed as considering the decay in a seminorm obtained by scaling the solution with the semidefinite matrix E. To see this, let T be the transformation matrix to the form (7.5), and let n_i denote the size of $\hat{z}_i$ for $i = 1, \ldots, 4$. By assuming that the Kronecker index is at most one, we have $n_1 = n_4 = 0$, and the solution $\xi_2(t)$ of (7.6) is asymptotically stable for every initial value $\hat{z}_2(0)$. The solution of the original system is then $z = T\begin{bmatrix} \xi_2^\top & 0 \end{bmatrix}^\top$, hence for every consistent initial value

$$\|z(t)\|^2 = \|T\xi(t)\|^2 \leq \|T\|\|\xi_2(0)\|^2 \, \mathrm{e}^{-2\mu t} \leq \kappa(T)\|\xi(0)\|^2 \, \mathrm{e}^{-2\mu t},$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

where $\mu > 0$ is some exponential decay rate capturing the asymptotic stability of (7.6), and $\kappa(T) = \|T\|\|T^{-1}\|$ is the condition number of T. Note that $\xi(0)$ can be replaced with $z(0)$ by carrying out a transformation of the second component with $\hat{E}_{22}^{1/2}$ and then the coefficient gets an extra factor associated with the condition number of $\hat{E}_{22}$.

The condensed form (7.5) also allows a characterization of (asymptotic) stability via a generalized Lyapunov equation. The following theorem is a simplified and real version of a result in Achleitner *et al.* (2021*b*).

Theorem 7.20. Consider a dHDAE of the form (7.4) with regular matrix pencil $\lambda E - A$ (with $A = J - R$) of Kronecker index at most two and finite hypocoercivity index. Then for every matrix $W \in \mathbb{R}^{n,n}$ the generalized Lyapunov equation

$$E^\top X A + A^\top X E = -E^\top W E \tag{7.7}$$

has a solution. For all solutions X of (7.7), the matrix $E^\top X E$ is unique. Moreover, if W is positive (semi)definite, then every solution X of (7.7) is positive (semi)definite on the image of P_ℓ, the spectral projection onto the left deflating subspace associated with the finite eigenvalues of $\lambda E - A$.

Proof. Due to Theorem 7.17, the eigenvalues are in the closed left half-plane, the eigenvalues on the imaginary axis are semisimple, and the pencil is of Kronecker index at most two. But since the pencil is regular and has a finite hypocoercivity index, its finite spectrum lies in the open left half-plane.

For general linear DAE systems with regular matrix pencil $\lambda E - A$ of (Kronecker) index at most two whose finite eigenvalues lie in the open left half-plane, the result then follows from Stykel (2002). □

Remark 7.21. For LTI DAE systems the characterization of stability via different generalized Lyapunov equations and the relation to pHDAE systems has recently been studied in different contexts, for example in a behaviour context in Gernandt and Haller (2021), via generalized Kalman–Yakubovich–Popov inequalities in Reis *et al.* (2015) and Reis and Voigt (2015), and via linear relations in Gernandt, Haller and Reis (2021).

It is currently under investigation how to extend the results on hypocoercivity to the LTV and nonlinear case to obtain a necessary and sufficient condition for asymptotic stability.

7.6. *Geometric description of pHDAE systems*

Port-Hamiltonian systems are often described through differential geometric structures known as Dirac structures; see e.g. van der Schaft (2000) and van der Schaft and Jeltsema (2014). This viewpoint is extremely helpful in bond-graph representations of dynamical systems (Breedveld 2008, Montbrun-Di Filippo, Delgado, Brie and Paynter 1991, Paynter 1961) and also in understanding limiting situations

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

and the effect of constraints (Mehrmann and van der Schaft 2023, van der Schaft and Maschke 2018, 2020). It is also very important in the construction of time discretization methods; see Section 9.1.

Definition 7.22. Let $\mathcal{F}$ be a Euclidean vector space and $\mathcal{E} := \mathcal{F}^*$ its dual space. Let $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ be a bilinear form on $\mathcal{F} \times \mathcal{E}$ defined via

$$\langle\!\langle (f_1, e_1), (f_2, e_2) \rangle\!\rangle := \langle e_1 \mid f_2 \rangle + \langle e_2 \mid f_1 \rangle,$$

where $\langle \cdot \mid \cdot \rangle$ is the standard duality pairing. Then a linear subspace $\mathcal{D} \subseteq \mathcal{F} \times \mathcal{E}$, such that $\mathcal{D} = \mathcal{D}^{\perp\!\!\!\perp}$ with respect to $\langle\!\langle \cdot, \cdot \rangle\!\rangle$, is called a *Dirac structure* on $\mathcal{F} \times \mathcal{E}$. If $(f, e) \in \mathcal{D}$, then f and e are called *flow* and *effort*, respectively.

In finite dimensions, we have a Dirac structure if $\dim \mathcal{D} = \dim \mathcal{F}$ and

$$\langle e \mid f \rangle = 0 \quad \text{for all } (f, e) \in \mathcal{D}$$

(van der Schaft and Jeltsema 2014).

The concept of a Dirac structure can also be generalized to vector bundles. For this, let $\oplus$ denote the Whitney sum between vector bundles; see e.g. Lang (2012).

Definition 7.23. Consider a state space $\mathcal{Z}$ and a vector bundle $\mathcal{V}$ over $\mathcal{Z}$ with fibres $\mathcal{V}_z$. A *Dirac structure* over $\mathcal{V}$ is a sub-bundle $\mathcal{D} \subseteq \mathcal{V} \oplus \mathcal{V}^*$ such that, for all $z \in \mathcal{Z}$, $\mathcal{D}_z \subseteq \mathcal{V}_z \times \mathcal{V}_z^*$ is a linear Dirac structure.

This definition generalizes the *modulated Dirac structures* (van der Schaft and Jeltsema 2014), where $\mathcal{V} = T\mathcal{Z}$ is the tangent bundle to $\mathcal{Z}$. To associate a Dirac structure to the pHDAE system (4.1), we assume that the system is autonomous (see Remark 4.2) and first show the following lemma, which has been presented in Mehrmann and Morandin (2019).

Lemma 7.24. Consider an autonomous pHDAE system (4.1) with skew-symmetric $J\colon \mathcal{Z} \to \mathcal{L}(\mathcal{V}_z^*, \mathcal{V}_z)$. If $\mathcal{D} \subseteq \mathcal{V} \oplus \mathcal{V}^*$ is a vector sub-bundle with fibres defined by

$$\mathcal{D}_z = \{(f, e) \in \mathcal{V}_z \times \mathcal{V}_z^* \colon f + J(x)e = 0\}, \tag{7.8}$$

then $\mathcal{D}$ is a Dirac structure.

Proof. For generic $(f, e) \in \mathcal{D}_z$ and $(f', e') \in \mathcal{V}_z \times \mathcal{V}_z^*$, we have

$$\begin{aligned} \langle\!\langle (f, e), (f', e') \rangle\!\rangle &= \langle e \mid f' \rangle + \langle e' \mid f \rangle \\ &= \langle e \mid f' \rangle - \langle e' \mid Je \rangle = \langle e \mid f' + Je' \rangle. \end{aligned}$$

We show that $(f', e') \in \mathcal{D}_z$ if and only if $\langle e \mid f' + Je' \rangle = 0$ for all $(f, e) \in \mathcal{D}_z$. If $(f', e') \in \mathcal{D}_z$, then $\langle e \mid f' + Je' \rangle = 0$ holds for any $e \in \mathcal{E}$. If $(f', e') \notin \mathcal{D}_z$, then $f' + Je' \neq 0$ and so there exists $e \in \mathcal{E}$ such that $\langle e \mid f' + Je' \rangle = 1$, but $(f, e) \in \mathcal{D}_z$ with $f = -Je$. □

A Dirac structure for general pHDAE systems is then constructed as follows; see Mehrmann and Morandin (2019).

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Theorem 7.25. Consider an autonomous pHDAE system of the form (4.1). Define the flow fibre $\mathcal{V}_z := \mathcal{F}_z^{\mathrm{s}} \times \mathcal{F}_z^{\mathrm{p}} \times \mathcal{F}_z^{\mathrm{d}}$ for all $z \in \mathcal{Z}$, where

$$\begin{aligned}
\mathcal{F}_z^{\mathrm{s}} &:= E(z)T_z\mathcal{Z} \subseteq \mathbb{R}^{\ell} \text{ is the storage flow fibre,}\\
\mathcal{F}_z^{\mathrm{p}} &:= \mathbb{R}^{m} \text{ is the port flow fibre, and}\\
\mathcal{F}_z^{\mathrm{d}} &:= \mathbb{R}^{\ell+m} \text{ is the dissipation flow fibre.}
\end{aligned}$$

Let us partition $f = (f_{\mathrm{s}}, f_{\mathrm{p}}, f_{\mathrm{d}}) \in \mathcal{V}$ and $e = (e_{\mathrm{s}}, e_{\mathrm{p}}, e_{\mathrm{d}}) \in \mathcal{V}^*$. Then the sub-bundle $\mathcal{D} \subseteq \mathcal{V} \oplus \mathcal{V}^*$ with

$$\mathcal{D}_z = \left\{ (f, e) \in \mathcal{V}_z \times \mathcal{V}_z^* \mid f + \begin{bmatrix} \Gamma(z) & I_{\ell+m} \\ -I_{\ell+m} & 0 \end{bmatrix} e = 0 \right\}$$

is a Dirac structure over $\mathcal{V}$. Furthermore, the system of equations

$$\begin{aligned}
f_{\mathrm{s}} &= -E(z)\dot{z}, & e_{\mathrm{s}} &= \eta(z),\\
f_{\mathrm{p}} &= y, & e_{\mathrm{p}} &= u,\\
e_{\mathrm{d}} &= -W(z)f_d, & (f, e) &\in \mathcal{D}_z
\end{aligned} \tag{7.9}$$

is equivalent to the original pHDAE, and $\langle e \mid f \rangle = 0$ represents the storage energy balance equation.

Proof. Lemma 7.24 implies that $\mathcal{D}$ is a Dirac structure. Writing (4.1) in compact form,

$$\begin{bmatrix} E(z)\dot{z} \\ -y \end{bmatrix} = [(\Gamma(z) - W(z))] \begin{bmatrix} \eta(z) \\ u \end{bmatrix},$$

it follows that $(f, e) \in \mathcal{D}_z$ can be written as

$$\begin{bmatrix} -f_{\mathrm{s}} \\ -f_{\mathrm{p}} \end{bmatrix} = \Gamma(z)e_{\mathrm{s}} + e_{\mathrm{p}}, \quad f_{\mathrm{d}} = (e_{\mathrm{s}}, e_{\mathrm{p}}).$$

Together with the conditions (7.9), this is equivalent to

$$\begin{aligned}
\begin{bmatrix} E(z)\dot{z} \\ -y \end{bmatrix} &= (\Gamma(z) - W(z)) \begin{bmatrix} \eta(z) \\ u \end{bmatrix},\\
f &= \left\{ -E(z)\dot{z},\ y,\ \begin{bmatrix} \eta(z) \\ u \end{bmatrix} \right\},\\
e &= \left\{ \eta(z),\ u,\ -W(z) \begin{bmatrix} \eta(z) \\ u \end{bmatrix} \right\},
\end{aligned}$$

which is exactly the compact form of the pHDAE. Finally, note that the equation $\langle e \mid f \rangle = 0$ can be written as

$$\begin{aligned}
0 &= \langle \eta(z) \mid -E(z)\dot{z} \rangle + \langle u \mid y \rangle + \left\langle -W(z) \begin{bmatrix} \eta(z) \\ u \end{bmatrix} \mid \begin{bmatrix} \eta(z) \\ u \end{bmatrix} \right\rangle\\
&= \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(z) + y^{\top}u - \begin{bmatrix} \eta(z) \\ u \end{bmatrix}^{\top} W(z) \begin{bmatrix} \eta(z) \\ u \end{bmatrix},
\end{aligned}$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

which is the storage energy balance equation. □

Note that if we conversely want to retrieve a pHDAE system from a Dirac structure, then the additional conditions (7.9) and the definition of $\mathcal{H}(z)$ are needed.

Remark 7.26. Since dHDAE systems generalize classical Hamiltonian systems, an immediate question is whether the associated flow has a geometric structure, such as symplecticity or generalized orthogonality when there is no dissipation; see Hairer, Lubich and Wanner (2002). While for pHODE systems this is well established (e.g. Celledoni and Høiseth 2017, van der Schaft and Jeltsema 2014), for LTV dHDAE systems this has only been established recently in Scholz (2019), and in a more general setting in Kunkel and Mehrmann (2023).

Remark 7.27. In the linear time-invariant case, an extension that addresses different Lagrange and Dirac structures and their relation has been introduced in van der Schaft and Maschke (2018). Ignoring the dissipation term as well as the inputs and outputs, the corresponding equation has the form

$$KP\dot{z} = LSz,$$

where $S, P, L, S \in \mathbb{R}^{n,n}$ satisfy $S^\top P = P^\top S$ with $\operatorname{rank}\left[\begin{smallmatrix} S \\ P \end{smallmatrix}\right] = n$, that is, the columns of $\left[\begin{smallmatrix} S \\ P \end{smallmatrix}\right]$ form a *Lagrangian subspace*, and $K^\top L = -L^\top K$ with $\operatorname{rank}\left[\begin{smallmatrix} K \\ L \end{smallmatrix}\right] = n$, that is, the columns of $\left[\begin{smallmatrix} K \\ L \end{smallmatrix}\right]$ are associated with a Dirac structure. A further generalization is discussed in Gernandt *et al.* (2021), and the relation of different representations, including the incorporation of dissipation as well as ports, is discussed in Mehrmann and van der Schaft (2023).

Clearly, if $K = S = I$ then $P = P^\top$ and $S = -S^\top$, and we are in the case of dHDAE systems with $E = E^\top = P$, $S = -S^\top = J$, where the extra condition $E \geq 0$ has to be assumed. If $K = P = I$ then we are in the classical case of pHODE systems with $S = S^\top = Q$ and $L = -L^\top = J$.

Remark 7.28. The representation of pHODE or pHDAE systems via a Dirac structure is an extension of Tellegen's theorem (e.g. Duindam, Macchelli, Stramigioli and Bruyninckx 2009) and it also shows the relation to implicit Lagrange systems (e.g. Yoshimura and Marsden 2006*a*,*b*) as well as gradient systems (e.g. Öttinger and Grmela 1997, Öttinger 2006).

8. Model-order reduction

In this section we discuss structure-preserving *model-order reduction* (MOR) methods for pHDAE systems. The main idea is to replace the potentially high-dimensional pHDAE with a low-dimensional pHDAE surrogate model, such that the output error approximation for a given input is below some given tolerance. A standard approach in the MOR literature (e.g. Antoulas 2005*a*, Quarteroni, Manzoni and Negri 2015, Hesthaven, Rozza and Stamm 2016, Benner, Cohen, Ohlberger and

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Willcox 2017, Antoulas, Beattie and Gugercin 2020) is to construct the surrogate model via Galerkin or Petrov–Galerkin projection. In more detail, for a regular descriptor-system of the form (2.1) (i.e. we assume $\ell = n$ within this section), the projection-based surrogate is given by

$$\hat{F}(t, \hat{z}(t), \dot{\hat{z}}(t), u(t)) = 0, \tag{8.1a}$$

$$\hat{y}(t) - \hat{G}(t, \hat{z}(t), u(t)) = 0, \tag{8.1b}$$

with

$$\hat{F}(t, \hat{z}, \dot{\hat{z}}, u) := V_{\mathrm{r}}^{\top} F(t, V_\ell \hat{z}, V_\ell, \dot{\hat{z}}, u), \tag{8.2a}$$

$$\hat{G}(t, \hat{z}(t), u(t)) := G(t, V_{\mathrm{r}} \hat{z}(t), u(t)) \tag{8.2b}$$

for matrices $V_\ell, V_{\mathrm{r}} \in \mathbb{R}^{n,r}$. The task of MOR is (i) to construct suitable matrices $V_\ell, V_{\mathrm{r}} \in \mathbb{R}^{n,r}$ in a numerically stable way, and (ii) to ensure that $\hat{F}$ and $\hat{G}$ in (8.1) can be evaluated efficiently (without the need to evaluate terms in the full model dimension n). In addition, MOR strives to quantify the error of the *reduced-order model* (ROM) (8.1) and preserve important properties (such as stability or passivity) within the ROM.

For general DAE systems, even if the original system is of (Kronecker) index zero, a Galerkin projection may change the index, the regularity or the stability properties of the free system (with $u = 0$).

Example 8.1. Consider the implicit ODE system

$$\begin{bmatrix} 0 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} \varepsilon & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u.$$

Then with $V_\ell^{\top} := V_{\mathrm{r}}^{\top} := [1 \quad 0]$ we obtain the ROM

$$0 = \varepsilon z_1 + u,$$

which now has (Kronecker) index 1 for $\varepsilon > 0$ and is even singular for $\varepsilon = 0$.

A key advantage of modelling with pHODE and pHDAE systems is that effects as in Example 8.1 do not occur if the structure is not altered. Since pHDAE systems are invariant under Galerkin projection (see Corollary 7.4), the model class is ideal for projection-based discretization and MOR methods. This, together with the invariance under interconnection, allows the construction of model hierarchies ranging from fine models for simulation and parameter studies to very coarse or surrogate models that can be used in control and optimization.

Remark 8.2. MOR for pHODEs is discussed for instance in Afkham and Hesthaven (2017, 2019), Beattie and Gugercin (2011), Borja, Scherpen and Fujimoto (2023), Breiten and Unger (2022), Breiten, Morandin and Schulze (2022*b*), Buchfink, Glas and Haasdonk (2021), Chaturantabut, Beattie and Gugercin (2016), Egger *et al.* (2018), Fujimoto and Kajiura (2007), Gugercin, Polyuga, Beattie and van der Schaft (2009, 2012), Ionescu and Astolfi (2013), Kawano and Scherpen (2018),

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Liljegren-Sailer (2020), Moser and Lohmann (2020), Polyuga and van der Schaft (2011, 2012), Sato and Sato (2018), Scheuermann, Kotyczka and Lohmann (2019), Schulze and Unger (2018), Schwerdtner and Voigt (2020), Wolf, Lohmann, Eid and Kotyczka (2010) and Wu, Hamroun, Le Gorrec and Maschke (2014, 2018). Let us emphasize that for LTI systems, any passive system can be recast as a pH system; see Beattie, Mehrmann and Xu (2022*b*) and Beattie, Mehrmann and Van Dooren (2019). Thus any passivity-preserving MOR method can also be used as a structure-preserving MOR method for LTI pHODEs (with a potentially necessary post-processing step to construct the low-dimensional pH representation). As an example we mention positive-real balanced truncation (Desai and Pal 1984, Guiver and Opmeer 2013, Ionescu and Scherpen 2007, Reis and Stykel 2010*a*,*b*) and interpolation methods (e.g. Antoulas 2005*b*, 2008, Freund 2000, Ionutiu, Rommes and Antoulas 2008, Sorensen 2005).

In the following we focus solely on LTI pHDAE systems, since structure-preserving MOR methods for general pHDAE systems are still under investigation. Early results are available in Liljegren-Sailer (2020) and Schulze (2023). In the following, we discuss different MOR techniques. One important class is that of methods related to the reduction of the underlying Dirac structure and the associated power conservation. These are the effort and flow constraint reduction methods discussed in Section 8.2. Another class includes the recent optimization-based methods of Schwerdtner, Moser, Mehrmann and Voigt (2022) and Moser, Schwerdtner, Mehrmann and Voigt (2022). Like Gillis *et al.* (2018) for dHDAE systems, these use a parametrization of pHDAEs via Cholesky factors of the symmetric semidefinite matrices E, R, the lower half of the skew-symmetric matrix J as well as the matrix G, and perform an optimization to fit the transfer function sampled at a set of sampling points. However, we will not discuss these methods here. Another major class of MOR methods is that of (Galerkin) projection methods that operate in the classical differential equation domain and ensure that the corresponding transfer functions in the frequency domain are well approximated. These methods are the well-known moment matching (Section 8.3) and tangential interpolation (Section 8.4). Before we present these methods, we provide some general considerations in the next subsection.

8.1. General considerations for LTI pHDAE systems

We assume that the pHDAE has been reformulated in such a way that the free system (with $u = 0$) is of (Kronecker) index at most one (see Section 6.4), that $Q = I$ and that the system has no feedthrough term (see Sections 4.3 and 4.4). This means that the system has the form

$$E\dot{z} = (J - R)z + Gu, \tag{8.3a}$$

$$y = G^{\mathsf{T}}z, \tag{8.3b}$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

where the matrix pencil $\lambda E - (J - R)$ is regular and of (Kronecker) index at most one, $E = E^\top \geq 0$, $R = R^\top \geq 0$ and $J = -J^\top$. In view of Corollary 7.4 and the general Petrov–Galerkin projection approach described above, a ROM is constructed by choosing a suitable matrix $V \in \mathbb{R}^{n,r}$, setting $V_{\mathrm{r}} := V_\ell := V$, and constructing the ROM matrices as

$$\hat{E} := V^\top E V \in \mathbb{R}^{r,r}, \quad \hat{J} := V^\top J V \in \mathbb{R}^{r,r}, \tag{8.4a}$$

$$\hat{R} := V^\top R V \in \mathbb{R}^{r,r}, \quad \hat{G} := V^\top G \in \mathbb{R}^{r,m}, \tag{8.4b}$$

such that the structure-preserving surrogate (8.1) for (8.3) is given by

$$\hat{E}\dot{\hat{z}} = (\hat{J} - \hat{R})\hat{z} + \hat{G}u, \tag{8.5a}$$

$$\hat{y} = \hat{G}^\top \hat{z}. \tag{8.5b}$$

In particular, the ROM for the LTI case can be evaluated efficiently and independent of the full model dimension n once the matrices in (8.4) are constructed.

Remark 8.3. Note that the techniques we describe below can also be extended to the case $Q \neq I$. In this case we use a Petrov–Galerkin approach as described in (8.2), i.e. different projection matrices from left and right. In more detail, if Q is nonsingular, then the choice $V_\ell := QV_r$ retains the pHDAE structure in the ROM. This strategy is also prevailing in the context of pHODE systems (e.g. Chaturantabut *et al.* 2016) and is even used to ensure stability preservation in the context of MOR for switched systems (Schulze and Unger 2018).

Remark 8.4. Depending on the application at hand, the system matrices in (8.3) may depend on additional parameters ω. If these parameters are not fixed *a priori* to a specific value, then we want to preserve this parametric dependency in the ROM. A standard assumption in the MOR literature is that the system matrices are available in a parameter-separable form, that is,

$$E(\omega) = \sum_{i=1}^{K} \gamma_i(\omega) E_i, \tag{8.6}$$

with scalar functions γ_i and constant matrices $E_i \in \mathbb{R}^{n,n}$ for $i = 1, \ldots, k$ (and similarly for the other coefficient matrices). In this case the reduced matrices are simply obtained by reducing each E_i separately. If the matrices are not in parameter-separable form, or only with a very large k, then the (discrete) empirical interpolation method can be used instead; see Barrault, Maday, Nguyen and Patera (2004) and Chaturantabut and Sorensen (2010). For more details we refer to Haasdonk (2017).

Remark 8.5. One advantage of the projection-based approach is that besides the pH structure, the Hamiltonian is also approximated with the same ansatz space. Thus the discussed general framework not only preserves the pH-structure but also retains information about the original Hamiltonian. However, reformulating the

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

pH system with a different Hamiltonian may be more amendable for MOR. This is demonstrated in detail in Breiten *et al.* (2022*b*) and Breiten and Unger (2022) for pHODEs and in Breiten and Schulze (2021) for pHDAEs. Similar results are also achieved if the coefficients of the ROM matrices are directly obtained by minimizing a suitable error function; see Schwerdtner and Voigt (2020) for further details.

Although this is not necessary in general, we often also perform another simplification that allows us to clearly separate the dynamical part and the algebraic constraints. These parts have to be treated in a slightly different way and the reduction only takes place in the dynamical equations in order to ensure that the model reduction does not violate the physical principles described by the constraints. For this, let V_0 be an invertible matrix such that

$$V_0^\top E V_0 = \begin{bmatrix} E_{11} & 0 \\ 0 & 0 \end{bmatrix}, \quad V_0^\top (J-R) V_0 = \begin{bmatrix} J_{11}-R_{11} & J_{12}-R_{12} \\ J_{21}-R_{21} & J_{22}-R_{22} \end{bmatrix},$$
$$V_0^\top G = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}, \quad \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = V_0^{-1} z,$$

that is, the transformed system is given by

$$\begin{bmatrix} E_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} J_{11}-R_{11} & J_{12}-R_{12} \\ -J_{12}^\top - R_{12}^\top & J_{22}-R_{22} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} u, \tag{8.7a}$$
$$y = \begin{bmatrix} G_1^\top & G_2^\top \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}. \tag{8.7b}$$

The assumption that (8.3) is of (Kronecker) index at most one implies that $J_{22} - R_{22}$ is nonsingular. The decomposition can be easily obtained by first computing a full-rank factorization of the positive semidefinite matrix E using, for example, a singular value decomposition

$$E = U_0 \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V_0^\top$$

with invertible diagonal matrix Σ, and then forming

$$V_0^\top E V_0 = \begin{bmatrix} E_{11} & 0 \\ 0 & 0 \end{bmatrix} \tag{8.8}$$

with $E_{11} = E_{11}^\top > 0$.

Remark 8.6. If a semi-explicit representation with $E_{11} = I$ is required, then one can compute the Cholesky factorization $E_{11} = L_{11} L_{11}^\top$ and perform another congruence transformation with $\tilde{V}_0 := \operatorname{diag}(L_{11}^{-1}, I)$. This yields, after renaming of

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

the transformed matrices, the equivalent formulation

$$\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} J_{11} - R_{11} & J_{12} - R_{12} \\ J_{21} - R_{12} & J_{22} - R_{22} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} u, \tag{8.9a}$$

$$y = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}^\top \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \tag{8.9b}$$

with $J_{22} - R_{22}$ nonsingular. For many MOR techniques it is essential that the semi-explicit form (8.9) is available. Fortunately, in many applications this can be done directly by exploiting the structure of the equations coming from the physical properties; see the examples in Section 5.

Performing a Laplace transformation for the system (8.3) yields the *transfer function*

$$\mathcal{G}(s) = G^\top (sE - J + R)^{-1} G, \tag{8.10}$$

which can be used to assess the approximation quality of the ROM via the $\mathcal{H}_2$- or $\mathcal{H}_\infty$-norm; see e.g. Antoulas *et al.* (2020). It is important to note that a singular E implies that $\mathcal{G}$ may contain a polynomial term. In general, using the Weierstrass canonical form (see Theorem 2.18), it is easy to see that the (Kronecker) index minus one defines an upper bound for the degree of the polynomial. Measuring the approximation error in the $\mathcal{H}_2$-norm thus requires the polynomial part to be matched exactly, since otherwise the error is unbounded. The situation is analogous for the $\mathcal{H}_\infty$-norm, except that the constant term in the polynomial does not need to be matched exactly.

For simplicity of the presentation, we will only describe the single-input, single-output case, that is, we assume $G \in \mathbb{R}^{n,1}$. All the algorithmic approaches can be easily extended to the multi-input multi-output case.

8.2. *Power conservation based model order reduction*

Two methods that carry out a MOR for the Dirac structure representation are the effort and flow constraint reduction methods that were introduced for standard pHODE systems in Polyuga and van der Schaft (2012) and extended to pHDAE systems in Hauschild, Marheineke and Mehrmann (2019). The basic idea of these approaches is to find a suitable transformation for the dynamic part of the state z_1 that partitions the state into a part associated with the ROM, denoted by $\hat{z}_1$, and a part that does not contribute much to the input–output behaviour of the system, denoted by $\tilde{z}_1$. In more detail, we determine a matrix V_1 with orthonormal columns such that $z_1 = V_1 \begin{bmatrix} \hat{z}_1^\top & \tilde{z}_1^\top \end{bmatrix}^\top$. Then we cut the interconnection between the part of the energy storage port belonging to $\tilde{z}_1$ and the Dirac structure, such that no power is transferred. In this way, the power is exclusively exchanged via the energy storage of $\hat{z}_1$ and the part associated with $\tilde{z}_1$ is omitted.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

In more detail, following the general discussion about Dirac structures in Section 7.6, the relevant constitutive equations in terms of MOR are given by

$$-E\dot{z} = f_{\mathrm{s}}, \quad e_{\mathrm{s}} = z. \tag{8.11}$$

Remark 8.7. Recall that in general the constitutive equation for the effort variable is $e_{\mathrm{s}} = \eta(z)$ in the nonlinear case, and $e_{\mathrm{s}} = Qz$ in the linear case with quadratic Hamiltonian; see Theorem 7.25 for further details. The methods that we will discuss can also be formulated for the more general case (Hauschild *et al.* 2019), but for ease of presentation we proceed here with $Q = I$ (see Section 4.3).

Using the semi-explicit formulation (8.9) and performing a congruence transformation with $V := \mathrm{diag}(V_1, I)$ transforms the constitutive equations (8.11) into the form

$$-\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\hat{z}}_1 \\ \dot{\tilde{z}}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} \hat{f}_{\mathrm{s},1} \\ \tilde{f}_{\mathrm{s},1} \\ f_{\mathrm{s},2} \end{bmatrix}, \quad \begin{bmatrix} \hat{e}_{\mathrm{s},1} \\ \tilde{e}_{\mathrm{s},1} \\ e_{\mathrm{s},2} \end{bmatrix} = \begin{bmatrix} \hat{z}_1 \\ \tilde{z}_1 \\ z_2 \end{bmatrix}. \tag{8.12}$$

For the model reduction we have to identify the part that is influenced by the dissipation (the resistive port). For this we apply a symmetric full-rank decomposition of $V^\top RV$ to compute

$$\begin{bmatrix} \hat{R}_{11} & \hat{\tilde{R}}_{11} & \hat{R}_{12} \\ \hat{\tilde{R}}_{11} & \tilde{R}_{11} & \tilde{R}_{12} \\ \hat{R}_{21} & \tilde{R}_{21} & R_{22} \end{bmatrix} = [Z \quad \hat{Z}] \begin{bmatrix} R_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Z^\top \\ \hat{Z}^\top \end{bmatrix} = ZR_1Z^\top, \tag{8.13}$$

with $0 < R_1 = R_1^\top \in \mathbb{R}^{\ell,\ell}$ and $Z \in \mathbb{R}^{n,\ell}$. Plugging (8.13) into the transformed system and introducing the associated flow and effort variables accordingly, that is,

$$f_{\mathrm{d}} = -R_1 e_{\mathrm{d}}, \quad e_{\mathrm{d}} = Z^\top V^\top V^{-1} z = \begin{bmatrix} \hat{Z}_1^\top & \tilde{Z}_1^\top & Z_2^\top \end{bmatrix} \begin{bmatrix} \hat{e}_{\mathrm{s},1} \\ \tilde{e}_{\mathrm{s},1} \\ e_{\mathrm{s},2} \end{bmatrix},$$

yields a pHDAE with opened resistive port. Inserting the relations (8.12) and introducing the external port variables $(f_{\mathrm{p}}, e_{\mathrm{p}}) = (y, u)$, where

$$y = (VG)^\top (V^\top V^{-1}) z = (VG)^\top e_{\mathrm{s}} = \begin{bmatrix} \hat{G}_1^\top & \tilde{G}_1^\top & G_2^\top \end{bmatrix} \begin{bmatrix} \hat{e}_{\mathrm{s},1} \\ \tilde{e}_{\mathrm{s},1} \\ e_{\mathrm{s},2} \end{bmatrix},$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

we obtain the new representation

$$-\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{f}_{\mathrm{s},1} \\ \tilde{f}_{\mathrm{s},1} \\ f_{\mathrm{s},2} \end{bmatrix} = \begin{bmatrix} \hat{J}_{11} & \tilde{\hat{J}}_{11} & \hat{J}_{12} \\ \hat{\tilde{J}}_{11} & \tilde{J}_{11} & \tilde{J}_{12} \\ \hat{J}_{21} & \tilde{J}_{21} & J_{22} \\ -\hat{G}_1^\top & -\tilde{G}_1^\top & -G_2^\top \\ -\hat{Z}_1^\top & -\tilde{Z}_1^\top & -Z_2^\top \end{bmatrix} \begin{bmatrix} \hat{e}_{\mathrm{s},1} \\ \tilde{e}_{\mathrm{s},1} \\ e_{\mathrm{s},2} \end{bmatrix} + \begin{bmatrix} \hat{Z}_1 \\ \tilde{Z}_1 \\ Z_2 \\ 0 \\ 0 \end{bmatrix} f_{\mathrm{d}} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ I \end{bmatrix} e_{\mathrm{d}} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ I \\ 0 \end{bmatrix} f_{\mathrm{p}} + \begin{bmatrix} \hat{G}_1 \\ \tilde{G}_1 \\ G_2 \\ 0 \\ 0 \end{bmatrix} e_{\mathrm{p}}. \tag{8.14}$$

With these preparations we are now ready to formulate the energy-based MOR methods. The main idea is to cut the interconnection

$$-\dot{\tilde{z}}_1 = \tilde{f}_{\mathrm{s},1}, \quad \tilde{e}_{\mathrm{s},1} = \tilde{z}_1 \tag{8.15}$$

between the energy storage corresponding to $\tilde{z}_1$ and the Dirac structure, in such a way that no energy is transferred. The energy flow through the interconnection (8.15) is set equal to zero by enforcing

$$\tilde{e}_{\mathrm{s},1}^\top \tilde{f}_{\mathrm{s},1} = 0 \quad \text{and} \quad \tilde{z}_1^\top \dot{\tilde{z}}_1 = 0. \tag{8.16}$$

This can be achieved in two canonical choices, leading to two different MOR methods that are discussed in the remainder of this subsection.

In the *effort constraint reduction method* (ECRM), we set $\tilde{e}_{\mathrm{s},1} = 0$, which implies $\tilde{z}_1 = 0$. This choice thus immediately yields (8.16). The reduced Dirac structure is obtained by inserting this relation and removing the second row in (8.14). This yields the reduced pHDAE model

$$\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\hat{z}}_1 \\ \dot{z}_2 \end{bmatrix} = \left(\begin{bmatrix} \hat{J}_{11} & \hat{J}_{12} \\ \hat{J}_{21} & J_{22} \end{bmatrix} - \begin{bmatrix} \hat{R}_{11} & \hat{R}_{12} \\ \hat{R}_{21} & R_{22} \end{bmatrix} \right) \begin{bmatrix} \hat{z}_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} \hat{G}_1 \\ G_2 \end{bmatrix} u, \tag{8.17a}$$

$$y = \begin{bmatrix} \hat{G}_1^\top & G_2^\top \end{bmatrix} \begin{bmatrix} \hat{z}_1 \\ z_2 \end{bmatrix}, \tag{8.17b}$$

It remains to show that (8.17) is indeed port-Hamiltonian, which is easily established with Corollary 7.4, since (8.17) can be constructed via Galerkin projection.

Remark 8.8. Note that the ROM (8.17) is obtained by standard truncation, as is common in balancing type methods; see for instance Gugercin and Antoulas (2004). The situation is different if the pHDAE (8.9) features a Q-term that is not identical to the identity. In this case a simple truncation may destroy the pH structure. Nevertheless, we can proceed as above and rewrite the reduced Dirac structure (obtained by setting $\tilde{e}_{s,1} = 0$ and removing the second block row) as a pHDAE. We refer to Hauschild *et al.* (2019) for further details.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

In the *flow constraint reduction method* (FCRM), the energy transfer between the energy-storing elements and the Dirac structure is cut by setting $\tilde{f}_{s,1} = 0$, which implies $\dot{\tilde{z}}_1 = 0$, and thus also (8.16). Thus $\tilde{z}_1$ is constant and can in particular be chosen as $\tilde{z}_1 = 0$. The second row in (8.14) is then an algebraic equation, which can be resolved for $\tilde{e}_{s,1}$ if $\tilde{J}_{11}$ is invertible, that is,

$$\tilde{e}_{s,1} = -\tilde{J}_{11}^{-1}\big(\hat{\tilde{J}}_{11}\hat{e}_{s,1} + \tilde{J}_{12}e_{s,2} + \tilde{Z}_1 f_d + \tilde{G}_1 e_p\big). \tag{8.18}$$

Substituting (8.18) into (8.14) and removing the second block row yields

$$\begin{aligned} -\begin{bmatrix} I & 0 \\ 0 & I \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \hat{f}_{s,1} \\ f_{s,2} \end{bmatrix} &= \begin{bmatrix} \hat{J}_{11} - \hat{\tilde{J}}_{11}\tilde{J}_{11}^{-1}\hat{\tilde{J}}_{11} & \hat{J}_{12} - \hat{\tilde{J}}_{11}\tilde{J}_{11}^{-1}\tilde{J}_{12} \\ \hat{J}_{21} - \tilde{J}_{21}\tilde{J}_{11}^{-1}\hat{\tilde{J}}_{11} & J_{22} - \tilde{J}_{21}\tilde{J}_{11}^{-1}\tilde{J}_{12} \\ -\hat{G}_1^\top + \tilde{G}_1^\top\tilde{J}_{11}^{-1}\hat{\tilde{J}}_{11} & -G_2^\top + \tilde{G}_1^\top\tilde{J}_{11}^{-1}\tilde{J}_{11} \\ -\hat{Z}_1^\top + \tilde{Z}_1^\top\tilde{J}_{11}^{-1}\hat{\tilde{J}}_{11} & -Z_2^\top + \tilde{Z}_1^\top\tilde{J}_{11}^{-1}\tilde{J}_{12} \end{bmatrix}\begin{bmatrix} \hat{e}_{s,1} \\ e_{s,2} \end{bmatrix} \\ &\quad + \begin{bmatrix} \hat{Z}_1 - \hat{\tilde{J}}_{11}\tilde{J}_{11}^{-1}\tilde{Z}_1 \\ Z_2 - \tilde{J}_{21}\tilde{J}_{11}^{-1}\tilde{Z}_1 \\ \tilde{G}_1^\top\tilde{J}_{11}^{-1}\tilde{Z}_1 \\ \tilde{Z}_1^\top\tilde{J}_{11}^{-1}\tilde{Z}_1 \end{bmatrix} f_d + \begin{bmatrix} 0 \\ 0 \\ 0 \\ I \end{bmatrix} e_d + \begin{bmatrix} 0 \\ 0 \\ I \\ 0 \end{bmatrix} f_p + \begin{bmatrix} \hat{G}_1 - \hat{\tilde{J}}_{11}\tilde{J}_{11}^{-1}\tilde{G}_1 \\ G_2 - \tilde{J}_{21}\tilde{J}_{11}^{-1}\tilde{G}_1 \\ \tilde{G}_1^\top\tilde{J}_{11}^{-1}\tilde{G}_1 \\ \tilde{Z}_1^\top\tilde{J}_{11}^{-1}\tilde{G}_1 \end{bmatrix} e_p. \end{aligned} \tag{8.19}$$

The resulting ROM then is again a pHDAE system, but due to the elimination it now has a feedthrough term (see the third block row in (8.19)). We do not present the technical formulas here; for details we refer to Hauschild *et al.* (2019). In contrast to the ECRM, we immediately conclude that the ROM obtained by the FCRM is not obtained via projection.

The reduced models obtained by the ECRM and FCRM have similar properties but also major differences. Both methods have the same number of reduced states. The ROM in the FCRM has an extra feedthrough term and requires the skew-symmetric matrix $\tilde{J}_{11}$ to be invertible, which is impossible if it is a square matrix of odd size. If $\tilde{J}_{11}$ is singular then the procedure has to be modified, but a (rather technical) construction is possible to deal with this case.

The question that remains to be answered is how to choose the coordinates $\hat{z}_1$ and $\tilde{z}_2$ in an optimal way, which in general is an open problem. Instead, we present a balancing-inspired algorithm to perform the separation, which of course can also be used to compute a (numerically) minimal realization for the pHODE. The details are presented in Algorithm 2; see also Hauschild *et al.* (2019). We emphasize that the resulting pHDAE is not balanced in the classical sense but only inspired by standard balancing; see Breiten *et al.* (2022*b*) and Borja *et al.* (2023) for other pH structure-preserving balancing approaches.

8.3. Moment matching

The *moment matching* (MM) method derives the ROM using a Galerkin projection in such a way that the leading coefficients of the Taylor series expansion of the transfer

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Algorithm 2 Structure-preserving balancing for pHODEs

Input: pHDAE (8.9)
Output: Balanced-like pHDAE (8.12)

Step 1. Set $A_{11} := J_{11} - R_{11}$.
Step 2. Compute solutions $\mathcal{P}_{11}$, $\mathcal{O}_{11}$ of the equations

$$A_{11}\mathcal{P}_{11}\mathcal{P}_{11}^\top + \mathcal{P}_{11}\mathcal{P}_{11}^\top A_{11}^\top + G_1 G_1^\top = 0,$$
$$A_{11}^\top\mathcal{O}_{11}\mathcal{O}_{11}^\top + \mathcal{O}_{11}\mathcal{O}_{11}^\top A_{11} + G_1 G_1^\top = 0.$$

Step 3. Compute the singular value decomposition $U\Sigma W^\top = \mathcal{P}_{11}^\top\mathcal{O}_{11}$, and a QR decomposition $V_1\mathcal{R} = PU$.
Step 4. Partition $V_1 = \begin{bmatrix}\hat{V}_1 & \hat{V}_2\end{bmatrix}$ and perform a congruence transformation with $V := \mathrm{diag}(V_1, I)$ to obtain the form (8.12).

function $\widehat{\mathcal{G}}$ at a given shift parameter $s_0 \in \mathbb{C} \cup \{\infty\}$ of the reduced systems match those of the full-order system $\mathcal{G}$ at s_0. For details of the MM methods for LTI DAE systems, we refer to Freund (2005) for $s_0 \in \mathbb{C}$ and to Benner and Sokolov (2006) for $s_0 = \infty$. The adaptation of these methods for pHODE systems was developed in Polyuga and van der Schaft (2010, 2011). Since the projection space that maps the original to the reduced problem is typically a Krylov subspace constructed by using an Arnoldi or Lanczos iteration (e.g. Bai 2002, Freund 2000, Grimme 1997), the resulting MOR method is applicable to large-scale systems and numerically stable.

For a shift $\sigma_0 \in \mathbb{C}$, a formal expansion of the transfer function $\mathcal{G}$ around s_0 (Antoulas 2005*a*) leads to

$$\mathcal{G}(s) = \sum_{i=0}^{\infty} m_i(\sigma_0 - s)^i. \tag{8.20}$$

The generalized moments m_i can be written as $m_i = G^\top v_i$, with vectors v_i that are determined recursively by solving the linear systems

$$(\sigma_0 E - J + R)v_0 = G, \tag{8.21a}$$
$$(\sigma_0 E - J + R)v_i = Ev_{i-1}, \quad i \geq 1, \tag{8.21b}$$

and employing the Arnoldi process (Saad 2003) to generate an orthogonal basis for this Krylov subspace $\mathcal{V} = \mathrm{span}\{v_0, \ldots, v_{r-1}\}$. Let the columns of V denote this orthonormal basis and construct the matrices for the ROM as in (8.4). It is well known that in this way the moments are matched up to level r; see Freund (2005) and Benner and Sokolov (2006). To ensure that the algebraic constraints are preserved in the ROM, we exploit the semi-explicit form (8.9) and construct the projection matrix only for the dynamic part, as in the following result taken from Hauschild *et al.* (2019).

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Theorem 8.9. Consider the pHDAE (8.7). For given shift $\sigma_0 \in \mathbb{C}$ compute the vectors v_i for $i = 0, \ldots, r-1$ as in (8.21), and construct a matrix $\begin{bmatrix} V_1^\top & V_2^\top \end{bmatrix}^\top$, partitioned accordingly to (8.7) with orthonormal columns such that

$$\operatorname{span} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \operatorname{span}\{v_0, \ldots, v_{r-1}\}.$$

Then the ROM

$$\begin{bmatrix} V_1^\top E_{11} V & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\hat{z}}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} V_1^\top (J_{11} - R_{11}) V_1 & V_1^\top (J_{12} - R_{12}) \\ (J_{21} - R_{21}) V_1 & J_{22} - R_{22} \end{bmatrix} \begin{bmatrix} \hat{z}_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} V_1^\top G_1 \\ G_2 \end{bmatrix} u,$$
$$\hat{y} = \begin{bmatrix} G_1 V_1 \\ G_2 \end{bmatrix} \begin{bmatrix} \hat{z}_1 \\ z_2 \end{bmatrix} \tag{8.22}$$

retains the pH structure and matches the first r moments and the polynomial part of the transfer function.

Remark 8.10. Theorem 8.9 presents a seemingly easy solution to structure-preserving MOR of pHDAE systems of (Kronecker) index one. Nevertheless, this may not be the maximal reduction that is possible, because redundant algebraic conditions cannot be removed; see Mehrmann and Stykel (2005) for further details.

8.4. Tangential interpolation

A fourth and very successful MOR method for LTI ODE systems is the tangential interpolation method; see Antoulas *et al.* (2020) for the general theory and application. In contrast to moment matching, the transfer function and its derivatives are not interpolated at a single point but rather at multiple points. If the system has multiple inputs and outputs, the interpolation is typically only enforced along so-called tangential directions. The main motivation for this approach is the fact that an $\mathcal{H}_2$-reduced model interpolates the full-order model at several interpolation points along tangential directions; see Antoulas *et al.* (2020). For different classes of LTI pHDAE systems, the method was introduced in detail in Beattie *et al.* (2022*a*). We discuss the method for pHDAE systems of the form (8.3).

As in the previous section, we work with single-input single-output systems to ease the presentation, i.e. we assume $m = 1$. All results can be extended to the multi-input multi-output case. For a prescribed set of interpolation frequencies $\sigma_1, \ldots, \sigma_r \in \mathbb{C}$, the goal is to construct a reduced pHDAE system whose transfer function interpolates the transfer function of the original model at the prescribed frequency points, that is, we want

$$\mathcal{G}(\sigma_i) = \widehat{\mathcal{G}}(\sigma_i) \quad \text{for } i = 1, \ldots, r. \tag{8.23}$$

Following the moment matching approach from the previous subsection, we immediately obtain the following result for LTI pHDAE systems of (Kronecker) index one.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Theorem 8.11. Consider a pHDAE (8.7) with (Kronecker) index one. For given interpolation points $\{\sigma_1, \dots, \sigma_r\} \subseteq \mathbb{C}$, construct a matrix $\begin{bmatrix} V_1^\top & V_2^\top \end{bmatrix}^\top \in \mathbb{C}^{n,r}$, partitioned accordingly, that satisfies

$$\operatorname{span}\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \operatorname{span}\{(\sigma_1 E - J + R)^{-1} G, \dots, (\sigma_r E - J + R)^{-1} G\}.$$

Then the ROM (8.22) retains the pH structure, interpolates the original model at the interpolation points, and matches the polynomial part.

As discussed in the previous subsection (see Remark 8.10), the construction in Theorem 8.11 suffers from the fact that possible redundant algebraic equations are not removed. We thus present an alternative approach in the next theorem. Again, the main idea is to construct the ROM via Galerkin projection such that the interpolation conditions (8.23) are satisfied. Since, in general, such a ROM will not match the polynomial part of the transfer function, we follow a strategy from Mayo and Antoulas (2007) (see also Gugercin, Stykel and Wyatt 2013) and modify the feedthrough term without violating the interpolation conditions. The corresponding result for pHDAE systems from Beattie *et al.* (2022*a*) is presented in the following theorem.

Theorem 8.12. Consider a pHDAE (8.7) with (Kronecker) index at most one. For given interpolation points $\{\sigma_1, \dots, \sigma_r\} \subseteq \mathbb{C}$, construct a matrix $V := \begin{bmatrix} V_1^\top & V_2^\top \end{bmatrix}^\top \in \mathbb{C}^{n,r}$, partitioned accordingly, as in (8.11). Define the matrices

$$\begin{aligned}
&\hat{E} := V_1^\top E_{11} V_1, && \hat{D} := -G_2^\top (J_{22} - R_{22})^{-1} G_2, && \hat{B} := V^\top G + \mathbb{1}\hat{D}, \\
&\hat{C} := G^\top V + \hat{D}\mathbb{1}^\top, && \hat{A} := V^\top (J - R) V - \mathbb{1}\hat{D}\mathbb{1}^\top, && \hat{J} := \frac{1}{2}(\hat{A} - \hat{A}^\top), \\
&\hat{R} := -\frac{1}{2}(\hat{A} + \hat{A}^\top), && \hat{P} := \frac{1}{2}(\hat{C}^\top - \hat{B}), && \hat{G} := \frac{1}{2}(\hat{C}^\top + B), \\
&\hat{S} := \frac{1}{2}(\hat{D} + \hat{D}^\top), && \hat{N} := -\frac{1}{2}(\hat{D} - \hat{D}^\top),
\end{aligned}$$

with $\mathbb{1} := [1 \quad \cdots \quad 1]^\top \in \mathbb{R}^r$. Then the ROM

$$\hat{E}\dot{\hat{z}} = (\hat{J} - \hat{R})z + (\hat{G} - \hat{P})u, \tag{8.24a}$$

$$\hat{y} = (\hat{G} + \hat{P})^\top z + (\hat{S} - \hat{N})u \tag{8.24b}$$

satisfies the interpolation conditions (8.23) and matches the polynomial part of $\mathcal{G}$. If, in addition, the matrix $\begin{bmatrix} \hat{R} & \hat{P} \\ \hat{P}^\top & \hat{S} \end{bmatrix}$ is positive semidefinite, then (8.24) is a pHDAE system.

Remark 8.13. In general, the projection matrix V in Theorems 8.11 and 8.12 is complex, and thus the matrices in the ROM are also complex-valued. Nevertheless, if the interpolation points are closed under complex conjugation, then a state-space transformation can be used to find a real-valued realization. In practice, this can

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

be done *a priori* by choosing V appropriately. For details, we refer to Antoulas *et al.* (2020). A similar approach also applies to the MM approach discussed in Theorem 8.9.

Remark 8.14. It is possible to extend these results to the pHDAE systems of (Kronecker) index two. Such a generalization is discussed in detail in Beattie *et al.* (2022*a*).

The crucial question that remains to be answered is the choice of the interpolation points $\sigma_1, \dots, \sigma_r$. It is well known (e.g. Antoulas *et al.* 2020) that an $\mathcal{H}_2$-optimal reduced model interpolates the transfer function of the full-order model at the mirror images of the poles of the ROM. In more detail, assume that the poles $\lambda_i \in \mathbb{C}$ for $i = 1, \dots, r$ of $\widehat{\mathcal{G}}$ are semisimple. If $\widehat{\mathcal{G}}$ is an $\mathcal{H}_2$-optimal approximation, then

$$\mathcal{G}(-\lambda_i) = \widehat{\mathcal{G}}(-\lambda_i) \quad \text{and} \quad \mathcal{G}'(-\lambda_i) = \widehat{\mathcal{G}}'(-\lambda_i) \quad \text{for } i = 1, \dots, r, \tag{8.25}$$

where $\mathcal{G}'$ denotes the derivative with respect to s. Since these poles are not known *a priori*, they cannot be used as interpolation points in Theorems 8.11 and 8.12. Instead, Gugercin, Antoulas and Beattie (2008) proposed a fixed-point iteration to resolve this problem, which is known as the *iterative rational Krylov algorithm* (IRKA). The main idea is to construct a ROM via Theorem 8.11 or Theorem 8.12, compute the poles of the transfer function, and use its mirror images as the next set of interpolation points. This is repeated until convergence. For general unstructured descriptor systems, an Hermite interpolant can be constructed as in Theorems 8.11 and 8.12; see e.g. Gugercin *et al.* (2013). If, however, we preserve the pH-structure as in Theorems 8.11 and 8.12, then in general only a subset of the interpolation conditions (8.25) is satisfied, and hence the resulting ROM may not be optimal with respect to the $\mathcal{H}_2$-norm. Indeed, as our forthcoming numerical examples show (see Section 8.5), the approximation quality can be significantly improved if a different Hamiltonian is used.

One way to achieve such a reformulation with a Hamiltonian that is particularly amendable for MOR is to adapt the strategy for passive ODE systems discussed in Breiten and Unger (2022) to the pHDAE setting as follows. First, consider only the differential part of the pHDAE (8.7), i.e. the implicit pHODE

$$E_{11}\dot{z}_1 = (J_{11} - R_{11})z_1 + G_1 u, \tag{8.26a}$$

$$y = G_1^\top z_1. \tag{8.26b}$$

If (8.26) is not (numerically) minimal, compute a structure-preserving minimal realization, for instance via Algorithm 2 or via the method described in Breiten *et al.* (2022*b*). For the sake of notation, we assume that this step has already been done, that is, we assume that (8.26) is already (numerically) minimal. Second, set $A_{11} := (J_{11} - R_{11})E_{11}^{-1}$, and compute the minimizing solution $X_{11} = X_{11}^\top > 0$ of the algebraic Riccati equation

$$-A_{11}^\top X_{11} - X_{11}A_{11} - \big(G_{11}^\top - X_{11}G_{11}\big)D_{11}^{-1}\big(G_{11} - G_{11}^\top X_{11}\big) = 0.$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Then construct the matrices

$$\tilde{E}_{11} := X_{11}^{-1}, \quad \tilde{J}_{11} := \frac{1}{2}\big(A_{11}X_{11}^{-1} - X_{11}^{-1}A_{11}^{\top}\big),$$
$$\tilde{R}_{11} := -\frac{1}{2}\big(A_{11}X_{11}^{-1} + X_{11}^{-1}A_{11}^{\top}\big),$$

and perform structure-preserving MOR for the system

$$\begin{bmatrix} \tilde{E}_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} \tilde{J}_{11} - \tilde{R}_{11} & J_{12} - R_{12} \\ -J_{12}^{\top} - R_{12}^{\top} & J_{22} - R_{22} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} u, \tag{8.27a}$$

$$y = \begin{bmatrix} G_1^{\top} & G_2^{\top} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \tag{8.27b}$$

provided that the matrix

$$\begin{bmatrix} \tilde{R}_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

is positive semidefinite. Alternatively, one may recast the (Kronecker) index one pHDAE as a pHODE with modified feedthrough term.

Remark 8.15. Interestingly, a (generalized) state-space realization is not necessary to construct an interpolatory ROM. As demonstrated in Mayo and Antoulas (2007) and Antoulas, Lefteriu and Ionita (2017), the ROM can be constructed solely from the interpolation points σ_i and associated measurements of the transfer function $\mathcal{G}$ and its derivative. Generalizations to models with structure are proposed in Schulze, Unger, Beattie and Gugercin (2018), for instance. First attempts to use frequency measurements to construct a low-dimensional pHDAE, i.e. to use interpolation or least-squares approaches as a structure-inducing system identification framework, are presented in Antoulas *et al.* (2017), Benner, Goyal and Van Dooren (2020), Cherifi, Mehrmann and Hariche (2019), Schwerdtner and Voigt (2020), Schwerdtner (2021) and Schwerdtner and Voigt (2021). First approaches that work with time-domain data are presented in Sharma, Wang and Kramer (2022) and Morandin, Nicodemus and Unger (2022). Methods to analyse whether the available data are generated from a passive system are presented in Romer, Montenbruck and Allgöwer (2017), Romer, Berberich, Köhler and Allgöwer (2019) and van Waarde, Camlibel, Rapisarda and Trentelman (2022).

8.5. Numerical examples

To illustrate the performance of the discussed pHDAE structure-preserving MOR methods, we present a numerical example using a multibody system as described in Section 5.7. The reported $\mathcal{H}_2$-norms are computed as in Stykel (2006), using the M-M.E.S.S. Toolbox; see Saak, Köhler and Benner (2021).

A holonomically constrained mass–spring–damper system is a multibody problem that describes the one-dimensional dynamics of g connected mass points in terms of their positions $q\colon \mathbb{T} \to \mathbb{R}^g$, velocities $v\colon \mathbb{T} \to \mathbb{R}^g$ and a Lagrange multiplier

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

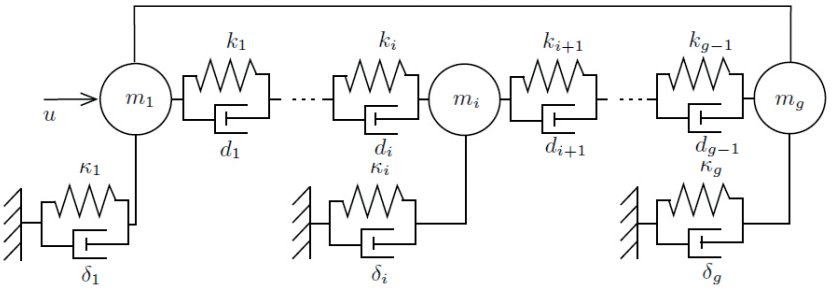

Figure 8.1. Damped mass–spring system with holonomic constraint, taken from Mehrmann and Stykel (2005).

$\lambda\colon \mathbb{T} \to \mathbb{R}$; see Figure 8.1. Here the ith mass of the weight m_i is connected to the $(i+1)$th mass by a spring and a damper with constants $k_i > 0$ and $d_i > 0$, and also connected to the ground by a spring and a damper with the constants $\kappa_i > 0$ and $\delta_i > 0$, respectively. Furthermore, the first and last mass points are connected by a rigid bar. The vibrations are driven by an external force $u\colon \mathbb{T} \to \mathbb{R}$, the control input, acting on the first mass point. The resulting system has a mass matrix $M = \operatorname{diag}(m_1, \ldots, m_g)$, symmetric positive definite tridiagonal stiffness and damping matrices K, $D \in \mathbb{R}^{g,g}$, a constraint matrix $C = [1 \quad 0 \quad \cdots \quad 0 \quad -1] \in \mathbb{R}^{1,g}$, and an input matrix $\tilde{G} = e_1 \in \mathbb{R}^{g,1}$. Since K and D are symmetric positive definite, the problem can be formulated as a pHDAE of (Kronecker) index two by replacing the algebraic constraint $Cq = 0$ with its first derivative $Cv = 0$, yielding the pHDAE

$$\begin{bmatrix} K & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{q} \\ \dot{v} \\ \dot{\lambda} \end{bmatrix} = \left(\begin{bmatrix} 0 & K & 0 \\ -K & 0 & -C^\top \\ 0 & C & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} q \\ v \\ \lambda \end{bmatrix} + \begin{bmatrix} 0 \\ \tilde{G} \\ 0 \end{bmatrix} u \tag{8.28}$$

with $z := [q^\top \quad v^\top \quad \lambda^\top]^\top$ when adding an associated output equation $y = G^\top z$.

The structure of the equations allows an easy construction of the condensed form required for the MOR methods. Performing a full-rank decomposition of C as $CV = [C_1 \quad 0]$ with C_1 invertible and an orthogonal matrix V, a congruence transformation yields the system

$$\begin{bmatrix} K & 0 & 0 & 0 \\ 0 & M_{11} & M_{12} & 0 \\ 0 & M_{12}^\top & M_{22} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{q} \\ \dot{v}_1 \\ \dot{v}_2 \\ \dot{\lambda} \end{bmatrix} = \begin{bmatrix} 0 & K_1 & K_2 & 0 \\ -K_1^\top & -D_{11} & -D_{12} & -C_1^\top \\ -K_2^\top & -D_{12}^\top & -D_{22} & 0 \\ 0 & C_1 & 0 & 0 \end{bmatrix} \begin{bmatrix} q \\ v_1 \\ v_2 \\ \lambda \end{bmatrix} + \begin{bmatrix} 0 \\ G_1 \\ G_2 \\ 0 \end{bmatrix} u.$$

The last equation $C_1 v_1 = 0$ implies that $v_1 = 0$. Differentiating this equation and inserting it into the second equation yields the hidden constraint for the Lagrange multiplier

$$C_1{}^T \lambda = -M_{12}\dot{v}_2 - K_1^\top q - D_{12} v_2 + G_1 u,$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Table 8.1. Parameters for holonomically constrained mass–spring–damper system.

Parameter	g	m_i	k_i	d_i	κ_i	δ_i	κ_1	κ_g	δ_1	δ_g
Value	1000	100	2	5	2	5	4	4	10	10

which imposes a consistency condition for the initial value. The underlying pHODE of size $n_1 = 2(g-1)$ together with the output equation are given by

$$\begin{bmatrix} K & 0 \\ 0 & M_{22} \end{bmatrix} \begin{bmatrix} \dot{q} \\ \dot{v}_2 \end{bmatrix} = \left(\begin{bmatrix} 0 & K_2 \\ -K_2^\top & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & D_{22} \end{bmatrix} \right) \begin{bmatrix} q \\ v_2 \end{bmatrix} + \begin{bmatrix} 0 \\ G_2 \end{bmatrix} u, \tag{8.29a}$$

$$y = \begin{bmatrix} 0 & G_2^\top \end{bmatrix} \begin{bmatrix} q \\ v_2 \end{bmatrix}. \tag{8.29b}$$

If we permute the columns and rows such that the pHODE (8.29) is in the leading blocks, then we can also mimic the MOR strategies from the previous subsections for the pHDAE systems of (Kronecker) index two, by only reducing the pHODE (8.29). Note that the lower-right 2×2 block matrix has no skew-symmetric contribution and hence the relevant submatrix for the FCRM is singular. We thus exclude the FCRM in the following.

For our numerical example, we choose a setting similar to that of Hauschild *et al.* (2019), with parameters as listed in Table 8.1. To compute a ROM with the different methods, we pick $r \in \mathbb{N}$ and reduce only the pHODE (8.29) while the algebraic part is not reduced, knowing that, in general, this is not optimal (see Remark 8.10). Whenever we report a reduced dimension r, this means that we have to add the number of algebraic equations to the dimension of the ROM. A frequency sweep for the different ROMs with $r = 10$ is presented in Figure 8.2 and relative $\mathcal{H}_2$-norms for different values of r are reported in Figure 8.3. For ROMs with $r = 10$ (see Figure 8.2), we observe that MM with shift $s_0 = \infty$ and $s_0 = 10^{-10}$ yields outstanding approximations (errors of order $O(10^{-15})$) for high and low frequencies, respectively. In contrast, the ECRM and rational interpolation with interpolation points computed via IRKA, as described in Section 8.4, provide a uniformly good approximation quality of order $O(10^{-3})$, independently of the chosen frequency. As discussed earlier, the structure-preserving variant of IRKA cannot satisfy all the necessary optimality conditions. To improve the situation, we also present the error for tangential interpolation via structure-preserving IRKA with a modified Hamiltonian, denoted by IRKA (mod. $\mathcal{H}$) in the figures, as described in (8.27), which yields a significant improvement over the original formulation. This can also be seen in the relative $\mathcal{H}_2$-errors displayed in Figure 8.3, where for $r \geq 6$ the IRKA-reduced pHDAE with modified Hamiltonian yields an approximation that is at least one order of magnitude better.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

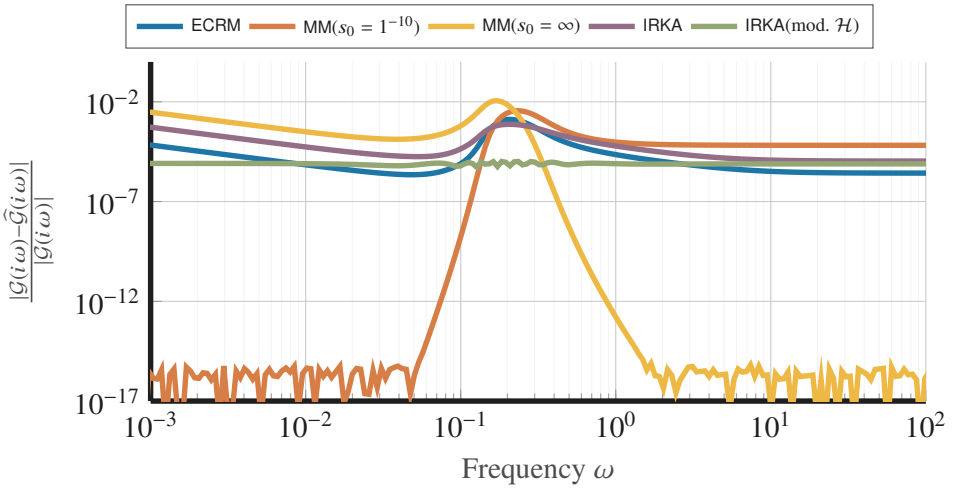

Figure 8.2. Relative errors of reduced transfer functions with $r = 10$ plotted over the frequency for the mass–spring–damper system in the formulation (8.28).

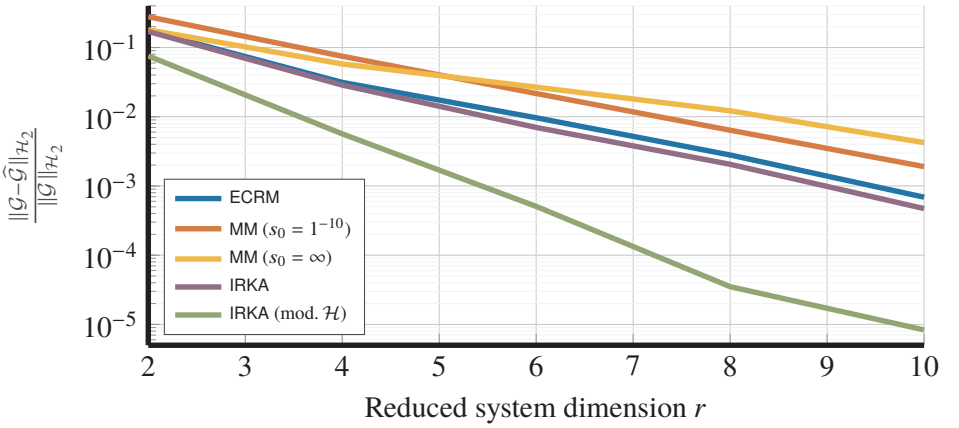

Figure 8.3. Relative $\mathcal{H}_2$ errors of the ROMs for the mass–spring–damper system (8.28) for different reduced orders.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

9. Temporal discretization and linear system solvers

In this section we discuss the time discretization of dHDAE and pHDAE systems and the associated linear system solves. The numerical solution of initial and boundary value problems for general DAE systems of the form (2.1a) has been an important research topic; see the monographs by Brenan *et al.* (1996), Hairer, Lubich and Roche (1989), Hairer and Wanner (1996) and Lamour *et al.* (2013). Following the approach discussed in Kunkel and Mehrmann (2006, Chapter 6), we may assume that the DAE is represented at every time-step in one of the strangeness-free forms (2.16) or (2.18), and then it has been shown that many implicit one-step and multi-step methods can be applied and have the same convergence order as for ODE systems.

9.1. *Time discretization for pHDAEs*

Most of the classical time discretization techniques do not respect a given dHDAE or pHDAE structure in such a way that the time-discretized system still satisfies a storage energy balance equation or dissipation inequality. To analyse for which approaches this is guaranteed is an active research area that is proceeding in different directions. One approach that is currently under investigation is that of partitioned Runge–Kutta methods adapted to the pHDAE structure. Another very promising approach is the discretization of pHDAE systems in such a way that the time-discrete system satisfies a discrete version of the storage energy balance equation; see Celledoni and Høiseth (2017), Kotyczka and Lefèvre (2018) and Kotyczka, Maschke and Lefèvre (2018) for pHODE systems, and Mehrmann and Morandin (2019) for pHDAE systems. Another class of methods, particularly for non-dissipative ODE methods, is based on energy-preserving geometric integration; see e.g. Celledoni *et al.* (2009), Hairer *et al.* (2002) and Quispel and McLaren (2008). The analysis and comparison of these techniques is a topic in its own right, so we only briefly discuss the approach based on collocation methods for pHDAEs of the form (4.1) in Mehrmann and Morandin (2019).

Consider an autonomous pHDAE of the form (4.1) with a given input function $u\colon \mathbb{T} \to \mathbb{R}^m$, a consistent initial value $z(t_0) = z_0$, and suppose that we want to approximate the solution in a time interval $(t_0, t_\mathrm{f} = t_0 + \tau)$ by a polynomial $\tilde{z}(t)$ of degree at most s. For a collocation method, the polynomial $\tilde{z}(t)$ is chosen such that $\tilde{z}(t_i) = z(t_i)$ satisfies the pHDAE (4.1) in the s collocation points $t_i = t_0 + \tau\gamma_i$ with $\gamma_i \in [0, 1]$ for $i = 1, \ldots, s$.

Let ℓ_i denote the ith Lagrange interpolation polynomial with respect to the nodes $\gamma_1, \ldots, \gamma_s$, that is,

$$\ell_i(t) := \prod_{\substack{j=1 \\ j\neq i}}^{s} \frac{t - \gamma_j}{\gamma_i - \gamma_j}.$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Then for collocation methods we require that

$$\dot{\tilde{z}}(t_0 + t\tau) = \sum_{i=1}^{s} \dot{z}_i \ell_i(t), \quad \tilde{z}(t_0 + t\tau) = z_0 + \tau \sum_{j=1}^{s} \dot{z}_j \int_0^t \ell_j(\sigma)\, \mathrm{d}\sigma$$

for the values $\dot{z}_i := \dot{\tilde{z}}(t_i)$, and also

$$z_i := \tilde{z}(t_i) = z_0 + \tau \sum_{j=1}^{s} \alpha_{ij} \dot{z}_j, \quad z_f := \tilde{z}(t_\mathrm{f}) = z_0 + \tau \sum_{j=1}^{s} \beta_j \dot{z}_j,$$

where the coefficients

$$\alpha_{ij} := \int_0^{\gamma_i} \ell_j(\sigma)\, \mathrm{d}\sigma \quad \text{and} \quad \beta_j := \int_0^1 \ell_j(\sigma)\, \mathrm{d}\sigma, \quad i, j = 1 \ldots s$$

are the coefficients of the Butcher diagram of the associated Runge–Kutta method; see Hairer *et al.* (2002).

To preserve the pHDAE structure we use the Dirac structure $\mathcal{D}_z$, as described in (7.8), associated with the pHDAE (4.1), and define the Dirac structure associated with the time discretization

$$\mathcal{D}_{z_i} = \left\{ (f^i, e^i) \in \mathcal{V}_{z_i} \times \mathcal{V}_{z_i}^* \mid f^i + \begin{bmatrix} \Gamma(z_i) & I_{\ell+m} \\ -I_{\ell+m} & 0 \end{bmatrix} e^i = 0 \right\}, \quad i = 1, \ldots, s$$

with $f^i = \left[(f_\mathrm{s}^i)^\top \quad y_i^\top \quad (f_\mathrm{d}^i)^\top \right]^\top$ and $e^i = \left[(e_\mathrm{s}^i)^\top, u_i^\top, (e_\mathrm{d}^i)^\top \right]^\top$, where

$$\begin{aligned} f_\mathrm{s}^i &= -E(z_i)\dot{z}_i, \quad & e_\mathrm{s}^i &= \eta(z_i), \\ e_\mathrm{d}^i &= -W(z_i) f_\mathrm{d}^i, \quad & u_i &= u(z_i). \end{aligned}$$

In this way we obtain a system that is equivalent to the approach of applying the collocation method and then computing discrete inputs and outputs u_i, y_i, for $i = 1, \ldots, s$. This is done by introducing the associated collocation flows, efforts, input and output as

$$\begin{aligned} \tilde{f}_\mathrm{s}(t_0 + \tau t) &= \sum_{i=1}^{s} f_\mathrm{s}^i \ell_i(t), \quad & \tilde{e}_\mathrm{s}(t_0 + \tau t) &= \sum_{i=1}^{s} e_\mathrm{s}^i \ell_i(t), \\ \tilde{f}_\mathrm{d}(t_0 + \tau t) &= \sum_{i=1}^{s} f_\mathrm{d}^i \ell_i(t), \quad & \tilde{e}_\mathrm{d}(t_0 + \tau t) &= \sum_{i=1}^{s} e_\mathrm{d}^i \ell_i(t), \\ \tilde{y}(t_0 + \tau t) &= \sum_{i=1}^{s} y_i \ell_i(t), \quad & \tilde{u}(t_0 + \tau t) &= \sum_{i=1}^{s} u_i \ell_i(t). \end{aligned}$$

Thus the discrete values are in $\mathcal{D}_{\tilde{z}}$ in all collocation points t_i and the discretization preserves the structure. To see this, let us consider the evolution of the Hamiltonian $\mathcal{H}$ along the collocation polynomial $\tilde{z}(t)$. For this, let $\tilde{\mathcal{H}}(t) := \mathcal{H}(\tilde{z}(t))$. Then we

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

have

$$\tilde{\mathcal{H}}(t) - \tilde{\mathcal{H}}(t_0) = \int_{t_0}^{t} \dot{\tilde{\mathcal{H}}}(\sigma)\,\mathrm{d}\sigma,$$

and at the collocation points the storage power balance equation is satisfied, that is,

$$\dot{\tilde{\mathcal{H}}}(t_i) = \frac{\partial}{\partial z_i}\tilde{\mathcal{H}}^\top \dot{z}_i = \eta(z_i)^\top E(z_i)\dot{z}_i = -\left\langle e_{\mathrm{s}}^i \mid f_{\mathrm{s}}^i \right\rangle = \left\langle e_{\mathrm{d}}^i \mid f_{\mathrm{d}}^i \right\rangle + \langle y_i \mid u_i \rangle,$$

for $i = 1, \ldots, s$. Applying the quadrature rule associated with the collocation method to evaluate the integral, we get

$$\begin{aligned}\tilde{\mathcal{H}}(t_{\mathrm{f}}) - \tilde{\mathcal{H}}(t_0) &= \tau \sum_{j=1}^{s} \beta_j \dot{\tilde{\mathcal{H}}}(t_j) + O(\tau^{p+1}) = -\tau \sum_{j=1}^{s} \beta_j \left\langle e_{\mathrm{s}}^j \mid f_{\mathrm{s}}^j \right\rangle + O(\tau^{p+1}) \\ &= \tau \sum_{j=1}^{s} \beta_j \left\langle e_{\mathrm{d}}^j \mid f_{\mathrm{d}}^j \right\rangle + h \sum_{j=1}^{s} \beta_j \langle y_j \mid u_j \rangle + (\tau^{p+1}),\end{aligned}$$

where $p \in \mathbb{N}$ is the approximation order of the quadrature rule. With the same argument we get

$$\tau \sum_{j=1}^{s} \beta_j \left\langle e_{\mathrm{d}}^j \mid f_{\mathrm{d}}^j \right\rangle = \int_{t_0}^{t_{\mathrm{f}}} \left\langle \tilde{e}_{\mathrm{d}}(\sigma) \mid \tilde{f}_{\mathrm{d}}(\sigma) \right\rangle \mathrm{d}\sigma + O(\tau^{p+1}), \tag{9.1a}$$

$$\tau \sum_{j=1}^{s} \beta_j \langle y_j \mid u_j \rangle = \int_{t_0}^{t_{\mathrm{f}}} \langle \tilde{y}(\sigma) \mid \tilde{u}(\sigma) \rangle \,\mathrm{d}\sigma + O(\tau^{p+1}), \tag{9.1b}$$

and hence

$$\tilde{\mathcal{H}}(t_{\mathrm{f}}) - \tilde{\mathcal{H}}(t_0) = \int_{t_0}^{t_{\mathrm{f}}} \left(\langle \tilde{e}_{\mathrm{d}}(\sigma) \mid \tilde{f}_{\mathrm{d}}(\sigma) \rangle + \langle \tilde{y}(\sigma) \mid \tilde{u}(\sigma) \rangle \right) \mathrm{d}\sigma + O(\tau^{p+1}).$$

If $p \geq 2s-2$, then (9.1a) and (9.1b) are actually exact, and if $\beta_j \geq 0$ for $j = 1, \ldots, s$, as is the case for many collocation methods, we have

$$\tau \sum_{j=1}^{s} \left\langle e_{\mathrm{d}}^j \mid f_{\mathrm{d}}^j \right\rangle \leq 0,$$

and thus the discrete system satisfies the same qualitative behaviour as the continuous problem.

If the Hamiltonian $\mathcal{H}$ is quadratic, that is,

$$\mathcal{H}(z) = \frac{1}{2} z^\top E z + v^\top z + c,$$

for some $E = E^\top \in \mathbb{R}^{n,n}$, $v \in \mathbb{R}^n$ and $c \in \mathbb{R}$, then we find that $\tilde{\mathcal{H}} = \mathcal{H}(\tilde{z})$ and $\dot{\tilde{\mathcal{H}}}$ are polynomials of degree $2s$ and degree $2s-1$, respectively. Using the well-known fact (e.g. Hairer and Wanner 1996) that the maximum degree of exactness for quadrature rules with s nodes is $2s-1$, and that it is attained only with Gauss–Legendre

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

collocation methods, it follows that for these methods the integration of $\dot{\tilde{\mathcal{H}}}$ is exact, that is,

$$\begin{aligned}\tilde{\mathcal{H}}(t_{\mathrm{f}}) - \tilde{\mathcal{H}}(t_0) &= \tau \sum_{j=1}^{s} \beta_j \left\langle e_{\mathrm{d}}^j \mid f_{\mathrm{d}}^j \right\rangle + \tau \sum_{j=1}^{s} \beta_j \langle y_j \mid u_j \rangle \\ &= \int_{t_0}^{t_{\mathrm{f}}} \left(\langle \tilde{e}_{\mathrm{d}}(s) \mid \tilde{f}_{\mathrm{d}}(s) \rangle + \langle \tilde{y}(\sigma) \mid \tilde{u}(\sigma) \rangle \right) \mathrm{d}\sigma.\end{aligned}$$

Furthermore, since for these methods we have $\beta_j \geq 0$, it follows that the dissipation term is always non-positive, and we obtain the discrete version of the dissipation inequality

$$\tilde{\mathcal{H}}(t_{\mathrm{f}}) - \tilde{\mathcal{H}}(t_0) \leq \tau \sum_{j=1}^{s} \beta_j \langle y_j \mid u_j \rangle = \int_{t_0}^{t_{\mathrm{f}}} \langle \tilde{y}(\sigma) \mid \tilde{u}(\sigma) \rangle \, \mathrm{d}\sigma,$$

and hence for quadratic Hamiltonians the pHDAE structure is fully preserved.

Example 9.1. Consider the numerical solution of the strangeness-free dHDAE system, given by the power network presented in Section 5.3. We use the artificial constants $E_G = 0$, $L = 2$, $C_1 = 0.01$, $C_2 = 0.02$, $R_{\mathrm{L}} = 0.1$, $R_{\mathrm{G}} = 6$ and $R_{\mathrm{R}} = 3$; see Mehrmann and Morandin (2019). For the time integration we choose $\tau = 0.01$ and $t_{\mathrm{f}} = 1$ and use the implicit midpoint rule, i.e. the Gauss–Legendre collocation method with $s = 1$ stages and order $p = 2$. The numerical result for the consistent initial value

$$z_0 = \sqrt{\frac{10}{R_{\mathrm{R}}}} \begin{bmatrix} 1 & -R_{\mathrm{R}} - R_{\mathrm{L}} & -R_{\mathrm{R}} & -\dfrac{R_{\mathrm{R}} - R_{\mathrm{L}}}{R_{\mathrm{R}}} & -1 \end{bmatrix}^{\top}$$

is presented in Figure 9.1. We observe that after an initial phase the state converges to zero and the Hamiltonian decreases monotonically.

9.2. Linear system solvers

In every step of an implicit time discretization method for a finite-dimensional pHDAE system (directly for the LTV case or in each step of the nonlinear solve in the general case), it is necessary to solve linear algebraic systems of the form

$$Ax = (E + \tau(R - J))x = b, \tag{9.2}$$

where τ is the time-step size. The matrix A can be split into its symmetric and skew-symmetric part $A = H + S$, where $H = \frac{1}{2}((E + \tau R)^{\top} + (E + \tau R)) \geq 0$ and $S = \frac{1}{2}((E + \tau R)^{\top} - (E + \tau R))$. An analogous linear system structure occurs in discretized linear time-varying and nonlinear pHDAE systems, in the construction of reduced models (Egger *et al.* 2018), and by multiplying some equations by -1 in optimization methods; see also Güdücü *et al.* (2022) and Manuoglu and Mehrmann (2019).

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

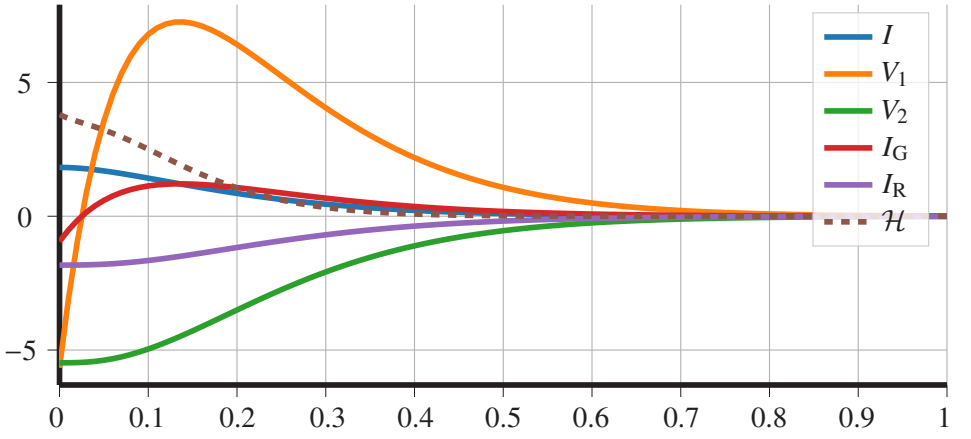

Figure 9.1. Evolution of state components (solid lines) and Hamiltonian (dashed).

The fact that the symmetric part H is positive semidefinite can be exploited in direct or iterative solution methods. In the small and medium scale case we can make use of the following staircase form (Achleitner *et al.* 2021*b*), which we present here in the real case.

Lemma 9.2. Consider $A = H+S \in \mathbb{R}^{n,n}$, where $H = H^\top \geq 0$ and $S = -S^\top$. Then there exist a real orthogonal matrix $U \in \mathbb{R}^{n,n}$, and integers $n_1 \geq n_2 \geq \cdots \geq n_{r-1} \geq 0$ and $n_r \geq 0$, such that

$$U^\top H U = \begin{bmatrix} H_{11} & 0 \\ 0 & 0 \end{bmatrix}, \quad U^\top S U = \begin{bmatrix} S_{11} & S_{12} & & & 0 \\ S_{21} & S_{22} & \ddots & & 0 \\ & \ddots & \ddots & S_{r-2,r-1} & \vdots \\ & & S_{r-1,r-2} & S_{r-1,r-1} & 0 \\ 0 & \cdots & \cdots & 0 & S_{r,r} \end{bmatrix}, \tag{9.3}$$

where $H_{11} = H_{11}^\top \in \mathbb{R}^{n_1,n_1}$ is positive definite, $S_{ii} = -S_{ii}^\top \in \mathbb{R}^{n_i,n_i}$ for $i = 1, \ldots, r$, and $S_{i,i-1} = -S_{i-1,i}^\top = [\Sigma_{i,i-1} \ 0] \in \mathbb{R}^{n_i,n_{i-1}}$ with $\Sigma_{i,i-1}$ being nonsingular for $i = 2, \ldots, r-1$.

Proof. We present the proof for completeness; see also Güdücü *et al.* (2022). The result is trivial when H is nonsingular (and thus positive definite), since in this case it holds with $U = I$, $r = 2$, $n_1 = n$ and $n_2 = 0$.

Let $H = H^\top \geq 0$ be singular. We consider a full-rank decomposition of H with a real orthogonal U such that

$$U_1^\top H U_1 = \begin{bmatrix} \hat{H}_{11} & 0 \\ 0 & 0 \end{bmatrix},$$

where we assume that $\hat{H}_{11} = \hat{H}_{11}^\top \in \mathbb{R}^{n_1,n_1}$, with $0 \leq n_1 \leq n$, is void or positive definite. Applying the same orthogonal congruence transformation to S gives the

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

matrix

$$\hat{S} = U_1^\top S U_1 = \begin{bmatrix} \hat{S}_{11} & \hat{S}_{12} \\ \hat{S}_{21} & \hat{S}_{22} \end{bmatrix}, \tag{9.4}$$

where $\hat{S}_{11} \in \mathbb{R}^{n_1,n_1}$ and $\hat{S}_{21} = -\hat{S}_{12}^\top$, since S is skew-symmetric. If $\hat{H}_{11}$ is void or $\hat{S}_{21} = 0$, then the proof is complete. Otherwise, let

$$\hat{S}_{21} = W_2 \begin{bmatrix} \Sigma_{21} & 0 \\ 0 & 0 \end{bmatrix} V_2^\top$$

be a singular value decomposition, where Σ_{21} is nonsingular (and diagonal), and $W_2 \in \mathbb{R}^{n_1,n_1}$, $V_2 \in \mathbb{R}^{n-n_1,n-n_1}$ are real orthogonal. We define

$$U_2 := \operatorname{diag}(V_2, W_2) \in \mathbb{R}^{n,n}$$

and form

$$U_2^\top U_1^\top H U_1 U_2 = \begin{bmatrix} V_2^\top \hat{H}_{11} V_2 & 0 \\ 0 & 0 \end{bmatrix},$$

where $V_2^\top \hat{H}_{11} V_2 \in \mathbb{R}^{n_1,n_1}$ is void or symmetric positive definite, and

$$U_2^\top U_1^\top S U_1 U_2 = \begin{bmatrix} V_2^\top \hat{S}_{11} V_2 & V_2^\top \hat{S}_{12} W_2 \\ W_2^\top \hat{S}_{21} V_2 & W_2^\top \hat{S}_{22} W_2 \end{bmatrix} = \begin{bmatrix} \tilde{S}_{11} & \tilde{S}_{12} & 0 \\ \tilde{S}_{21} & \tilde{S}_{22} & \tilde{S}_{23} \\ 0 & \tilde{S}_{32} & \tilde{S}_{33} \end{bmatrix}$$

with $\tilde{S}_{21} = [\Sigma_{21}\ 0]$. If $\tilde{S}_{32} = 0$ or is void, then the proof is again complete. Otherwise we continue inductively, and after finitely many steps we obtain a decomposition of the required form. □

When the staircase form (9.3) has been computed, then the transformed linear system $(U^\top A U)(U^\top x) = U^\top b$ can be solved using block Gaussian elimination.

Lemma 9.3. Consider the matrix $U^\top A U$ in (9.3). Then there exist invertible lower and upper block bidiagonal matrices L_s, R_s, respectively, such that

$$L_s U^\top A U R_s = \begin{bmatrix} H_{11} + S_{11} & & & & \\ & \mathcal{S}_1 & & & \\ & & \ddots & & \\ & & & \mathcal{S}_{r-2} & \\ & & & & \mathcal{S}_{r,r} \end{bmatrix},$$

with Schur complements $\mathcal{S}_1, \ldots, \mathcal{S}_{r-2}$ that have positive definite symmetric parts.

Proof. A constructive proof via a sequence of block Gaussian elimination steps is given in Güdücü *et al.* (2022). It relies on the fact that in every step (except the last one) the Schur complement has a symmetric part which is positive definite. □

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Table 9.1. Brake squeal problem. Run times and iteration numbers.

Method	$\tau = 0.001$ time [s]	$\tau = 0.001$ # of iterations	$\tau = 0.0001$ time [s]	$\tau = 0.0001$ # of iterations
Widlund	3.54	85	0.72	16
GMRES	31.10	65	10.51	13

The properties of the linear system resulting from the dHDAE structure also have an immediate advantage in the context of iterative methods. If the symmetric part is positive definite, then Widlund (1978) suggested solving, instead of $Ax = b$, the equivalent system

$$(I + K)x = \hat{b}, \quad \text{where } K = H^{-1}S,\ \hat{b} = H^{-1}b. \tag{9.5}$$

This transformation is a left preconditioning of the original system with its positive definite symmetric part, which defines the H-inner product

$$\langle x, y\rangle_H = y^\top Hx.$$

This implies that one can construct optimal Krylov subspace methods based on three-term recurrences for the system (9.5); see Rapoport (1978) and Widlund (1978). If the symmetric part is semidefinite but singular, then we have to identify the nullspace, which is actually easy in many applications. The advantages of this approach, and the fact that we obtain a rigorous convergence analysis and optimality conditions, are discussed in detail in Güdücü *et al.* (2022) and illustrated by several numerical examples, including those discussed in Section 5.

Example 9.4. The finite element model of the disc brake discussed in Section 5.8 leads to a second-order DAE of the form

$$M\ddot{p} + D\dot{p} + Kp = f,$$

with p the coefficient vector of displacements of the structure, with frequency-dependent mass matrix $M = M^\top > 0$, damping matrix $D = D^\top \geq 0$ and stiffness matrix $K = K^\top > 0$ (Gräbner *et al.* 2016) evaluated for $\omega_{\text{ref}} = 500$. For $f = 0$, after a first-order reformulation and discretization with the implicit mid-point rule, we obtain a linear system with $n = 9338$ and a positive definite symmetric part.

As shown in Table 9.1 (Güdücü *et al.* 2022) a preconditioned GMRES method (preconditioned with the inverse of the symmetric part), despite using fewer iterations, takes a significantly longer time than the method in Widlund (1978), due to the full recurrences in the algorithm compared to three-term recurrences in Widlund's method. This effect becomes even more pronounced for smaller step sizes.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

10. Control methods for pHDAE systems

One of the main advantages to introducing pH descriptor systems is its direct base in control theory. In this section we therefore discuss classical control applications for pHDAE systems.

Consider a linear pHDAE system as in Definition 4.8, and a linear output feedback $u = F(t)y$. Then we can write the system in behaviour form by introducing $\xi = [z^\top \quad u^\top \quad y^\top]^\top$ and new block matrices

$$\mathcal{E} := \operatorname{diag}(E, 0, 0), \quad \mathcal{Q} := \operatorname{diag}(Q, I, I), \quad \mathcal{K} := \operatorname{diag}(K, 0, 0),$$

as well as

$$\mathcal{J} = \begin{bmatrix} J & G & 0 \\ -G^\top & N & I \\ 0 & -I & \frac{1}{2}(F - F^\top) \end{bmatrix}, \quad \mathcal{R} = \begin{bmatrix} R & P & 0 \\ P^\top & S & 0 \\ 0 & 0 & -\frac{1}{2}(F + F^\top) \end{bmatrix}, \tag{10.1}$$

which gives the closed-loop descriptor system

$$\mathcal{E}\dot{\xi} = ((\mathcal{J} - \mathcal{R})\mathcal{Q} - \mathcal{E}\mathcal{K})\xi. \tag{10.2}$$

This is a dHDAE if and only if $-\frac{1}{2}(F + F^\top) \geq 0$ pointwise.

Analogously, for the general nonlinear pHDAE structure in Definition 4.1, we introduce

$$\xi := [z^\top \quad u^\top \quad y^\top]^\top, \quad \tilde{\eta} := [\eta^\top \quad u^\top \quad y^\top]^\top, \quad \tilde{r} := [r^\top \quad 0 \quad 0]^\top,$$

the Hamiltonian $\tilde{\mathcal{H}}(t, \xi) := \mathcal{H}(t, z)$ and matrix functions $\mathcal{E} := \operatorname{diag}(E, 0, 0)$, and $\mathcal{J}$ and $\mathcal{R}$ as defined in (10.1). This gives the system in behaviour form

$$\mathcal{E}\dot{\xi} + \tilde{r} = (\mathcal{J} - \mathcal{R})\tilde{\eta}(\xi) \tag{10.3}$$

satisfying $(\partial/\partial\xi)\tilde{\mathcal{H}} = \mathcal{E}^\top\tilde{\eta}$ and $(\partial/\partial t)\tilde{\mathcal{H}} = \tilde{\eta}^\top\tilde{r}$ pointwise, which is a pHDAE if and only if $\frac{1}{2}(F + F^\top) \leq 0$ pointwise.

In this way, using a parametrization via the output feedback matrix (function) F, we can introduce control methods via the dHDAE systems (10.2) or (10.3), respectively.

10.1. Robust stabilization/passivation

We have seen in Section 7.4 that for dHDAE systems, stability and passivity can be easily characterized, while for asymptotic stability or strict passivity, in general, we have only sufficient conditions.

Considering a linear pHDAE system of the form (4.5) with $Q = I$ that has no feedthrough term (see Section 4.4), and using a linear output feedback $u = Fy + w$, we obtain the closed-loop system

$$\begin{aligned} E\dot{z} + EKz &= (J - R + GFG^\top)z + Gw, \\ y &= G^\top z. \end{aligned}$$

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

If $F = F_{\mathrm{H}} + F_{\mathrm{S}}$ is chosen to have a negative semidefinite symmetric part F_H, then the new dissipation coefficient becomes $\hat{R} = R - GF_{\mathrm{H}}G^{\top}$ and the skew-symmetric part becomes $J - GF_{\mathrm{S}}G^{\top}$. If $\hat{R}$ is positive definite then the system is asymptotically stable. The same approach can also be applied in the general nonlinear case, if an output feedback leads to a positive definite $R(t, z)$ in (4.1). We only have this sufficient condition to guarantee asymptotic stability; see Corollary 7.7. At present, finding a necessary and sufficient condition guaranteeing that a pHDAE system can be made asymptotically stable by output feedback is a problem under investigation.

The situation is much better understood in the case of LTI pHDAE systems, where we have a characterization via the hypocoercivity index being finite in Corollary 7.15. Then, in view of Theorem 7.17(ii), we have different options to obtain asymptotic stability.

Again, if we can achieve $\hat{R} = R - GF_{\mathrm{H}}G^{\top} > 0$, then we have asymptotic stability. But we can also use the skew part F_{S} to change the eigenvectors of the pair $(E, J + GF_{\mathrm{S}}G^{\top})$ in such a way that no eigenvector is in the kernel of $R - GF_{\mathrm{H}}G^{\top}$, or we can use a combination of both. It is clear from the classical theory of unstructured DAE systems (see Section 3.3) that if the system is strongly stabilizable and strongly detectable, then such an output feedback exists and can be computed by ignoring the structure and solving an optimal control problem.

In constructing an output feedback via optimal control methods, we have the freedom to choose the cost functional (3.20) and, furthermore, we also have some freedom in choosing the pHDAE representation, which is not unique (see e.g. the discussion at the end of Section 8.4). This flexibility can be used to make the resulting closed-loop system maximally robust against perturbations, which for general LTI control systems has recently been an important research topic; see e.g. Mehrmann and Xu (2000) and the references therein. For pHODE systems, this topic has recently been of great importance by introducing measures such as the distance to instability for the robustness of pHODE representations (Aliyev, Mehrmann and Mengi 2020) and their optimization (Gillis *et al.* 2018, Gillis and Sharma 2017). For pHDAE systems, this is currently an active research topic.

The analogous question arises in the context of passivity. We see that a regular strangeness-free pHDAE system is passive but in general not necessarily strictly passive, since W in (4.8) is only assumed to be positive semidefinite. To obtain strict passivity, it is necessary to consider the system in the formulation with feedthrough term, and similarly we can analyse how to obtain robust representations as is done for pHODE systems in Bankmann, Mehrmann, Nesterov and Van Dooren (2020), Beattie *et al.* (2019) and Mehrmann and Dooren (2020). For pHDAE systems, this is again an active research topic.

10.2. Optimal control

Due to the many interesting properties of pHDAE systems, one may investigate whether some extra advantages can be obtained also in the context of optimal control

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

problems. Clearly one can just use the general results in Section 3.5 and obtain the same optimality conditions. However, it has been observed in two recent papers (Philipp *et al.* 2021, Faulwasser *et al.* 2022) that some surprising results arise for optimal control problems with LTI pHODE and pHDAE, when as a very special cost functional the supplied energy is minimized, that is,

$$\frac{1}{2}z(t_{\mathrm{f}})^{\top}Mz(t_{\mathrm{f}}) + \frac{1}{2}\int_{t_0}^{t_{\mathrm{f}}} 2y^{\top}u\,\mathrm{d}t = \frac{1}{2}z(t_{\mathrm{f}})^{\top}Mz(t_{\mathrm{f}}) + \frac{1}{2}\int_{t_0}^{t_{\mathrm{f}}} 2z^{\top}G^{\top}u\,\mathrm{d}t, \quad (10.4)$$

subject to the constraint

$$E\dot{z} = (J-R)z + Gu, \quad E^{\dagger}Ez(t_0) = z_0 \quad (10.5)$$

and the output equation is $y = G^{\top}z$. Note that in the cost functional (3.32) we then have $W_{\mathrm{z}} = 0$, $W_{\mathrm{u}} = 0$ and $S = G^{\top}$. Since we are in the LTI case, we can insert the data into the optimality system (3.34) and obtain the following result.

Corollary 10.1. Consider the optimal control problem to minimize (10.4) subject to the constraint (10.5). Assume that the pair $(E, J-R)$ is regular and of (Kronecker) index at most one (as a free system with $u \equiv 0$) and that M is in cokernel E. If $(z,u) \in \mathbb{Z}\times\mathbb{U}$ is a solution to this optimal control problem, then there exists a Lagrange multiplier $\lambda \in \mathcal{C}^1_{E^{\dagger}E}(\mathbb{T},\mathbb{R}^n)$, such that (z,λ,u) satisfy the boundary value problem

$$\begin{bmatrix} 0 & E & 0 \\ -E^{\top} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix}\dot{\lambda}\\ \dot{z}\\ \dot{u}\end{bmatrix} = \begin{bmatrix} 0 & J-R & G \\ (J-R)^{\top} & 0 & G \\ G^{\top} & G^{\top} & 0 \end{bmatrix}\begin{bmatrix}\lambda\\ z\\ u\end{bmatrix}, \quad (10.6)$$

with boundary conditions

$$E^{\dagger}Ez(t_0) = z_0, \quad EE^{\dagger}\lambda(t_{\mathrm{f}}) = -(E^{\dagger})^{\top}Mz(t_{\mathrm{f}}).$$

Use a full-rank decomposition $E = U_E\left[\begin{smallmatrix} E_{11} & 0 \\ 0 & 0\end{smallmatrix}\right]U_E^{\top}$ with $E_{11} = E_{11}^{\top} > 0$ and transform the other coefficients accordingly as

$$U_E^{\top}(J-R)U_E = \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22}\end{bmatrix}, \quad U_E^{\top}MU_E = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22}\end{bmatrix},$$

$$U_E^{\top}G = \begin{bmatrix} G_1 \\ G_2\end{bmatrix}, \quad \begin{bmatrix}\hat{z}_1\\ \hat{z}_2\end{bmatrix} = U_E^{\top}z, \quad \begin{bmatrix}\hat{\lambda}_1\\ \hat{\lambda}_2\end{bmatrix} = U_E^{\top}\lambda, \quad \begin{bmatrix}\hat{z}_{1,0}\\ \hat{z}_{2,0}\end{bmatrix} = U_E^{\top}z_0.$$

After some permutations we can express (10.6) in the form

$$\begin{bmatrix} 0 & E_{11} & 0 & 0 & 0 \\ -E_{11}^{\top} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix}\dot{\hat{\lambda}}_1\\ \dot{\hat{z}}_1\\ \dot{\hat{\lambda}}_2\\ \dot{\hat{z}}_2\\ \dot{u}\end{bmatrix} = \begin{bmatrix} 0 & L_{11} & 0 & L_{12} & G_1 \\ L_{11}^{\top} & 0 & L_{21}^{\top} & 0 & G_1 \\ 0 & L_{21} & 0 & L_{22} & G_2 \\ L_{12}^{\top} & 0 & L_{22}^{\top} & 0 & G_2 \\ G_1^{\top} & G_1^{\top} & G_2^{\top} & G_2^{\top} & 0 \end{bmatrix}\begin{bmatrix}\hat{\lambda}_1\\ \hat{z}_1\\ \hat{\lambda}_2\\ \hat{z}_2\\ u\end{bmatrix}, \quad (10.7)$$

with boundary conditions $\hat{z}_1(t_0) = \hat{z}_{1,0}$ and $\hat{\lambda}_1(t_{\mathrm{f}}) = -E_{11}^{-1}M_{11}\hat{z}_1(t_{\mathrm{f}})$, and consistency

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

condition $\hat{z}_{2,0} = 0$. Here we have used the condition that $M = M^\top$ is in cokernel E, which implies that $M_{12} = 0$, $M_{22} = 0$.

For the structured matrix pencil associated with (10.7) there exists a condensed form under real orthogonal congruence transformations, which was introduced in Byers *et al.* (2007), and from which the spectral properties, the (Kronecker) index and the regularity can be read off. If this pencil is regular, then we directly obtain that the system has (Kronecker) index one if and only if

$$\hat{W}_u = \begin{bmatrix} 0 & J_{22} - R_{22} & G_2 \\ -J_{22} - R_{22} & 0 & G_2 \\ G_2^\top & G_2^\top & 0 \end{bmatrix}$$

is invertible; see Section 2.3. To simplify the algebraic equations, we can perform a congruence transformation with the orthogonal matrix

$$U_S = \begin{bmatrix} \frac{1}{\sqrt{2}} I & -\frac{1}{\sqrt{2}} I & 0 \\ \frac{1}{\sqrt{2}} I & \frac{1}{\sqrt{2}} I & 0 \\ 0 & 0 & I \end{bmatrix},$$

that is, we multiply the system by $U_S^\top$ from the left and set

$$\begin{bmatrix} \tilde{\lambda}_2 \\ \tilde{z}_2 \\ u \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}}(\hat{z}_2 + \hat{\lambda}_2) \\ \frac{1}{\sqrt{2}}(\hat{z}_2 - \hat{\lambda}_2) \\ u \end{bmatrix} = U_S^\top \begin{bmatrix} \hat{\lambda}_2 \\ \hat{z}_2 \\ u \end{bmatrix}.$$

Then we get

$$U_S^\top \hat{W}_{\mathrm{u}} U_S = \begin{bmatrix} -R_{22} & J_{22} & G_2 \\ -J_{22} & R_{22} & 0 \\ G_2^\top & 0 & 0 \end{bmatrix}.$$

Clearly, for this to be invertible we need G_2 to have full column rank, which implies that u is fixed as a linear combination of $\hat{\lambda}_2$ and $\hat{z}_2$. Considering the application examples in Section 5, where we typically have $G_2 = 0$, we cannot expect the DAE associated with the optimality system to be of (Kronecker) index one. Thus we are in the case of a singular control problem; see e.g. Bryson and Ho (2018).

The case when $\hat{W}_{\mathrm{u}}$ is not invertible has been analysed in Schaller *et al.* (2021) for the pHODE case and in Faulwasser *et al.* (2022) for the pHDAE case, where $\mu E - (J - R)$ is regular and of (Kronecker) index at most one. In these papers it is assumed that the image of the matrix G does not intersect with the kernel of R, and that even though this is a singular control problem, the optimal solution is still a feedback control that can be obtained via the solution of a Riccati equation. We refer to Aronna (2018) for the analysis of such problems. The extension of this analysis to the LTV pHDAE case and the case when this assumption is not valid are currently under investigation.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

11. Summary and open problems

This survey paper discusses the model class of port-Hamiltonian descriptor systems (differential-algebraic systems) for numerical simulation and control. We have demonstrated that this model class has many nice properties.

- It allows for automated modelling in a modularized and network-based fashion.
- It allows the coupling of mathematical models across different scales and physical domains.
- It incorporates the properties of the real (open or closed) physical system.
- It has nice algebraic, geometric and analytical properties, and allows analysis concerning existence, uniqueness, robustness, stability and passivity.
- It is invariant under local variable (coordinate) transformations, which leads to local canonical and condensed forms.
- Furthermore, it allows for structure-preserving (space–time) discretization and model reduction methods as well as fast solvers for the resulting linear and nonlinear systems of equations.

Despite the many promising results already available for dHDAE systems and pHDAEs, there are still many open problems that are either under investigation or pose challenging problems to be tackled in the future. In the following, we present an incomplete list to stimulate further research.

Many of the control theoretical concepts presented within this paper rely on instantaneous feedback. While this is a convenient theoretical approach, it is not always possible to realize in applications, where the states or outputs first have to be measured, the control action computed and then fed back into the system, thus resulting in a necessary intrinsic time delay; see also Remark 3.5. Although some initial results for pHODEs with delays are available in the literature (e.g. Schiffer, Fridman, Ortega and Raisch 2016, Breiten, Hinsen and Unger 2022*a*) a general model class for time-delayed pHDAEs is not yet available, and the results presented in this paper have to be extended to the time delay case.

In terms of MOR, the impact of the Hamiltonian on the approximation quality (Breiten *et al.* 2022*b*, Breiten and Unger 2022) needs to be investigated further, particularly with an emphasis on nonlinear MOR methods. Besides, one should consider the optimal approximation of pHDAEs as a multi-objective optimization problem, where the minimization of the approximation error for the input–output dynamic and the Hamiltonian are the two objective functions. Structure-preserving balancing methods are still under investigation; see Breiten and Schulze (2021) for some initial results for LTI systems. Another open problem in structure-preserving MOR is the construction of optimal projection spaces, in the sense that they minimize the Kolmogorov n-widths, respectively the Hankel singular values; see Unger and Gugercin (2019). The first gradient-based optimization procedures are discussed

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

for pHODEs in Moser and Lohmann (2020), Sato and Sato (2018) and Schwerdtner and Voigt (2020). If the n-widths do not decay rapidly, then one cannot expect accurate ROMs with small dimension, and current efforts in the reduction of transport-dominated phenomena (see e.g. Black, Schulze and Unger (2020) and the references therein) need to be adapted to the pH framework.

A topic closely related to MOR is identifying a pHDAE realization from measurements. In view of large data sets, modern artificial intelligence approaches and automated machine learning methods used within the digital twin paradigm, this is an important topic requiring further research, with only a few available results for LTI systems; see Remark 8.15.

It is an open problem to derive necessary and sufficient conditions under which output feedback can make a general nonlinear pHDAE system asymptotically stable. A natural research question in this context would be to extend the concept of hypocoercivity to the LTV and nonlinear cases.

The characterization of distance measures for general dHDAE systems – such as the distance from an asymptotically stable dHDAE system to a system that is only stable, or the distance of a strictly passive pHDAE system to the nearest system that is only passive – is an important research topic because it is a requirement for the design of real-world systems, e.g. Example 5.8. Even if such a characterization is available, then we need computational methods to compute these distances. In the large-scale setting, this is a challenging issue that can only be achieved by a clever combination with model reduction techniques; see Aliyev *et al.* (2020) for a first attempt.

Another interesting research topic is the exact characterization of the relationship between passivity, positive realness and the port-Hamiltonian structure for pHDAEs (see Beattie *et al.* (2019, 2022*b*) for the LTI pHODE case), as well as the characterization via Kalman–Yakobovich–Popov inequalities, as has been done for general LTI DAE systems in Reis *et al.* (2015) and Reis and Voigt (2015). This has recently been achieved in Cherifi, Gernandt and Hinsen (2022*a*) for continuous-time pHDAE systems but is open for LTV pHDAE systems.

A research field that has so far not received much attention is that of discrete-time pH descriptor systems. The primary research in this direction arises from discretizations of pHDAE systems; see Kotyczka and Lefèvre (2018) and Mehrmann and Morandin (2019) for some recent approaches, and Achleitner, Arnold and Mehrmann (2023) for hypocontractivity of discrete-time systems. However, since discrete-time systems arise not only from discretization but also from sampling or realization, it is an open question how to properly define discrete-time pH systems in a general way. This also concerns the stability and passivity analysis.

Since, in many cases, mathematical models are obtained from data via measurements, and the resulting parameters, as well as model coefficients (including the inputs and outputs), are only available with some uncertainty or randomness, it is an open question how to include such uncertainties adequately within the pHDAE framework and also to study robust control methods that deal with such

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

uncertainties; see Breiten *et al.* (2022*b*) and Breiten and Schulze (2021) for the modification of classical balanced truncation methods and Karsai (2022) for the modification of robust control methods in the pHODE case. Another research area is the associated perturbation theory and the error analysis of the relevant numerical methods.

As we have seen, the area of optimal control for pHDAE systems also requires a lot of further research. This concerns the optimal use of the structure for nonlinear systems, the choice of an appropriate cost functional, and also structure-preserving numerical methods, in particular for large-scale problems.

Finally, we would like to discuss a topic that has been touched upon only briefly in this survey: the extension of pH modelling to partial differential equations that has been pursued in many different directions in recent years. One might take the approach of replacing the coefficient matrices with linear differential operators, or follow a differential geometric approach via the extension of Lagrange or Dirac structures, or formulate all classical partial differential equations from physics in the areas of elasticity, electromagnetism, fluid dynamics, structural mechanics, geomechanics, poroelasticity, gas or water transport, to name a few directions, in a structure that resembles the pHDAE structure via the given symmetries or differential forms. These efforts have been the topic of many recent research papers on modelling, numerical methods, optimization and control. Discussing these developments would be a considerable survey in its own right, mainly since the field is growing immensely fast. An essential question that is wide open is how to incorporate the boundary appropriately into the structure, and how to obtain well-posed partial differential equations so that on the one hand they can be treated as controls and on the other hand they can also be used for the interconnection of subsystems. Another important topic that is actively pursued is the appropriate space–time discretization methods, such as finite element or finite volume approaches that preserve the structure.

Instead of discussing this further, we present the following incomplete list of references, which describes many different research directions pursued: Altmann *et al.* (2021), Altmann and Schulze (2017), Aoues, Cardoso-Ribeiro, Matignon and Alazard (2017), Baaiu *et al.* (2009), Bansal *et al.* (2021), Cardoso-Ribeiro, Matignon and Pommier-Budinger (2017), Cherifi, Mehrmann and Schulze (2022*b*), Duindam *et al.* (2009), Egger (2019), Egger and Kugler (2018), Egger *et al.* (2018), Ennsbrunner and Schlacher (2005), Gay-Balmaz and Yoshimura (2018, 2019), Grmela and Öttinger (1997), Jacob and Zwart (2012), Kotyczka (2019), Kotyczka *et al.* (2018), Kurula, Zwart, van der Schaft and Behrndt (2010), Le Gorrec, Zwart and Maschke (2005), Macchelli and Maschke (2009), Macchelli, Melchiorri and Bassi (2005), Macchelli, van der Schaft and Melchiorri (2004*a*,*b*), Matignon and Hélie (2013), Moses Badlyan and Zimmer (2018), Moses Badlyan, Maschke, Beattie and Mehrmann (2018), Öttinger (2006), Öttinger and Grmela (1997), Ramirez Estay (2019), van der Schaft and Maschke (2002), Schöberl and Schlacher (2017), Serhani, Matignon and Haine (2019), Villegas (2007) and Yoshimura and Marsden (2006*a*,*b*).

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

Acknowledgements

The work of V. Mehrmann was supported by the following institutions and programmes: Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) CRC 910 'Control of self-organizing nonlinear systems: Theoretical methods and concepts of application', project no. 163436311; CRC TRR 154 'Mathematical modeling, simulation and optimization at the example of gas networks', project no. 239904186; priority programmes SPP 1984 'Hybrid and multimodal energy systems: System theoretical methods for the transformation and operation of complex networks', project no. 361092219, and SPP 1897, 'Calm, smooth and smart: Novel approaches for influencing vibrations by means of deliberately introduced dissipation', project no. 273845692; DFG Excellence Cluster 2046 Math$^+$, project no. 390685689; Investitionsbank Berlin via the ProFIT program 'Elektrische Antriebe' within the Werner von Siemens Center, Berlin; the BMBF (German Ministry of Education and Research) via the Project EiFer.

B. Unger acknowledges funding from the DFG under Germany's Excellence Strategy – EXC 2075 – 390740016 and is grateful for support by the Stuttgart Center for Simulation Science (SimTech).

References

F. Achleitner, A. Arnold and E. A. Carlen (2021*a*), The hypocoercivity index for the short time behavior of linear time-invariant ODE systems. Available at arXiv:2109.10784.

F. Achleitner, A. Arnold and V. Mehrmann (2021*b*), Hypocoercivity and controllability in linear semi-dissipative ODEs and DAEs, *ZAMM Z. Angew. Math. Mech*. Available at doi:10.1002/zamm.202100171.

F. Achleitner, A. Arnold and V. Mehrmann (2023), Hypocoercivity and hypocontractivity concepts for linear dynamical systems, *Electron. J. Linear Algebra* **9**, 33–61.

L. Y. Adrianova (1995), *Introduction to Linear Systems of Differential Equations*, Vol. 146 of Translations of Mathematical Monographs, American Mathematical Society.

B. M. Afkham and J. S. Hesthaven (2017), Structure preserving model reduction of parametric Hamiltonian systems, *SIAM J. Sci. Comput.* **39**, A2616–A2644.

B. M. Afkham and J. S. Hesthaven (2019), Structure-preserving model-reduction of dissipative Hamiltonian systems, *J. Sci. Comput.* **81**, 3–21.

N. Aliyev, V. Mehrmann and E. Mengi (2020), Computation of stability radii for large-scale dissipative Hamiltonian systems, *Adv. Comput. Math.* **46**, 1–28.

R. Altmann and P. Schulze (2017), A port-Hamiltonian formulation of the Navier–Stokes equations for reactive flows, *Systems Control Lett.* **100**, 51–55.

R. Altmann, V. Mehrmann and B. Unger (2021), Port-Hamiltonian formulations of poroelastic network models, *Math. Comput. Model. Dyn. Sys.* **27**, 429–452.

A. C. Antoulas (2005*a*), *Approximation of Large-Scale Dynamical Systems*, Advances in Design and Control, Society for Industrial and Applied Mathematics.

A. C. Antoulas (2005*b*), A new result on passivity preserving model reduction, *Systems Control Lett.* **54**, 361–374.

A. C. Antoulas (2008), On the construction of passive models from frequency response data, *Automatisierungstechnik* **56**, 447–452.

A. C. Antoulas, C. A. Beattie and S. Gugercin (2020), *Interpolatory Methods for Model Reduction*, Computational Science and Engineering, Society for Industrial and Applied Mathematics.

A. C. Antoulas, S. Lefteriu and A. C. Ionita (2017), Chapter 8: A tutorial introduction to the Loewner framework for model reduction, in *Model Reduction and Approximation* (P. Benner *et al.*, eds), Society for Industrial and Applied Mathematics, pp. 335–376.

S. Aoues, F. L. Cardoso-Ribeiro, D. Matignon and D. Alazard (2017), Modeling and control of a rotating flexible spacecraft: A port-Hamiltonian approach, *IEEE Trans. Control Sys. Tech.* **27**, 355–362.

M. S. Aronna (2018), Second order necessary and sufficient optimality conditions for singular solutions of partially-affine control problems, *Discrete Contin. Dyn. Syst.* **11**, 1179–1199.

A. Baaiu, F. Couenne, D. Eberard, C. Jallut, Y. Legorrec, L. Lefèvre and B. Maschke (2009), Port-based modelling of mass transport phenomena, *Math. Comput. Model. Dyn. Sys.* **15**, 233–254.

A. Backes (2006), Optimale Steuerung der linearen DAE im Fall Index 2. PhD thesis, Humboldt-Universität zu Berlin.

M. Bai, D. Elsworth and J.-C. Roegiers (1993), Multiporosity/multipermeability approach to the simulation of naturally fractured reservoirs, *Water Resour. Res.* **29**, 1621–1633.

Z. Bai (2002), Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems, *Appl. Numer. Math.* **43**, 9–44.

D. Bankmann, V. Mehrmann, Y. Nesterov and P. Van Dooren (2020), Computation of the analytic center of the solution set of the linear matrix inequality arising in continuous- and discrete-time passivity analysis, *Vietnam J. Math.* **48**, 633–660.

H. Bansal, P. Schulze, M. H. Abbasi, H. Zwart, L. Iapichino, W. H. A. Schilders and N. Wouw (2021), Port-Hamiltonian formulation of two-phase flow models, *Systems Control Lett.* **149**, 104881.

M. Barrault, Y. Maday, N. C. Nguyen and A. T. Patera (2004), An 'empirical interpolation' method: Application to efficient reduced-basis discretization of partial differential equations, *C.R. Math. Acad. Sci. Paris* **339**, 667–672.

C. Beattie and S. Gugercin (2011), Structure-preserving model reduction for nonlinear port-Hamiltonian systems, in *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC 2011)*, pp. 6564–6569.

C. Beattie, S. Gugercin and V. Mehrmann (2022*a*), Structure-preserving interpolatory model reduction for port-Hamiltonian differential-algebraic systems, in *Realization and Model Reduction of Dynamical Systems* (C. Beattie *et al.*, eds), Springer, pp. 235–254.

C. Beattie, V. Mehrmann and P. Van Dooren (2019), Robust port-Hamiltonian representations of passive systems, *Automatica* **100**, 182–186.

C. Beattie, V. Mehrmann and H. Xu (2022*b*), Port-Hamiltonian realizations of linear time invariant systems. Available at arXiv:2201.05355.

C. Beattie, V. Mehrmann, H. Xu and H. Zwart (2018), Port-Hamiltonian descriptor systems, *Math. Control Signals Systems* **30**, 1–27.

A. Beckesch (2018), Pfadverfolgung für Finite-Elemente-Modelle parametrischer mechanischer Systeme. Master's thesis, Technische Universität Berlin.

P. Benner and V. I. Sokolov (2006), Partial realization of descriptor systems, *Systems Control Lett.* **55**, 929–938.

P. Benner, A. Cohen, M. Ohlberger and K. Willcox (2017), *Model Reduction and Approximation*, Computational Science and Engineering, Society for Industrial and Applied Mathematics.
P. Benner, P. Goyal and P. Van Dooren (2020), Identification of port-Hamiltonian systems from frequency response data, *Systems Control Lett.* **143**, 104741.
T. Berger (2012), Bohl exponent for time-varying linear differential-algebraic equations, *Internat. J. Control* **85**, 1433–1451.
T. Berger and T. Reis (2013), Controllability of linear differential-algebraic equations: A survey, in *Surveys in Differential-Algebraic Equations I* (A. Ilchmann and T. Reis, eds), Differential-Algebraic Equations Forum, Springer, pp. 1–61.
D. Bienstock (2015), *Electrical Transmission System Cascades and Vulnerability: An Operations Research Viewpoint*, MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics.
M. A. Biot (1941), General theory of three-dimensional consolidation, *J. Appl. Phys.* **12**, 155–164.
F. Black, P. Schulze and B. Unger (2020), Projection-based model reduction with dynamically transformed modes, *ESAIM Math. Model. Numer. Anal.* **54**, 2011–2043.
P. Borja, J. M. A. Scherpen and K. Fujimoto (2023), Extended balancing of continuous LTI systems: A structure-preserving approach, *IEEE Trans. Automat. Control* **68**, 257–271.
P. C. Breedveld (2008), Modeling and simulation of dynamic systems using bond graphs, in *Control Systems, Robotics and Automation: Modeling and System Identification I* (H. Unbehauen, ed.), EOLSS Publishers/UNESCO, pp. 128–173.
T. Breiten and P. Schulze (2021), Structure-preserving linear quadratic Gaussian balanced truncation for port-Hamiltonian descriptor systems. Available at arXiv:2111.05065.
T. Breiten and B. Unger (2022), Passivity preserving model reduction via spectral factorization, *Automatica* **142**, 110368.
T. Breiten, D. Hinsen and B. Unger (2022*a*), Towards a modeling class for port-Hamiltonian systems with time-delay. Available at arXiv:2211.10687.
T. Breiten, R. Morandin and P. Schulze (2022*b*), Error bounds for port-Hamiltonian model and controller reduction based on system balancing, *Comput. Math. Appl.* **116**, 100–115.
K. E. Brenan, S. L. Campbell and L. R. Petzold (1996), *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, Society for Industrial and Applied Mathematics.
J. Brouwer, I. Gasser and M. Herty (2011), Gas pipeline models revisited: Model hierarchies, non-isothermal models and simulations of networks, *Multiscale Model. Simul.* **9**, 601–623.
T. Brown, D. Schlachtberger, A. Kies, S. Schramm and M. Greiner (2018), Synergies of sector coupling and transmission reinforcement in a cost-optimised, highly renewable European energy system, *Energy* **160**, 720–739.
A. E. Bryson and Y.-C. Ho (2018), *Applied Optimal Control: Optimization, Estimation, and Control*, Routledge.
P. Buchfink, S. Glas and B. Haasdonk (2021), Symplectic model reduction of Hamiltonian systems on nonlinear manifolds. Available at arXiv:2112.10815.
A. Bunse-Gerstner, R. Byers, V. Mehrmann and N. K. Nichols (1991), Numerical computation of an analytic singular value decomposition of a matrix valued function, *Numer. Math.* **60**, 1–39.
A. Bunse-Gerstner, R. Byers, V. Mehrmann and N. K. Nichols (1999), Feedback design for regularizing descriptor systems, *Linear Algebra Appl.* **299**, 119–151.

R. Byers, T. Geerts and V. Mehrmann (1997*a*), Descriptor systems without controllability at infinity, *SIAM J. Control Optim.* **35**, 462–479.
R. Byers, P. Kunkel and V. Mehrmann (1997*b*), Regularization of linear descriptor systems with variable coefficients, *SIAM J. Control Optim.* **35**, 117–133.
R. Byers, V. Mehrmann and H. Xu (2007), A structured staircase algorithm for skew-symmetric/symmetric pencils, *Electron. Trans. Numer. Anal.* **26**, 1–13.
C. I. Byrnes, A. Isidori and J. C. Willems (1991), Passivity, feedback equivalence, and the global stabilization of minimum phase nonlinear systems, *IEEE Trans. Automat. Control* **36**, 1228–1240.
S. L. Campbell (1987), A general form for solvable linear time varying singular systems of differential equations, *SIAM J. Math. Anal.* **18**, 1101–1115.
S. L. Campbell and C. W. Gear (1995), The index of general nonlinear DAEs, *Numer. Math.* **72**, 173–196.
S. L. Campbell, P. Kunkel and V. Mehrmann (2012), Regularization of linear and nonlinear descriptor systems, in *Control and Optimization with Differential-Algebraic Constraints* (L. T. Biegler, S. L. Campbell and V. Mehrmann, eds), Vol. 23 of Advances in Design and Control, Society for Industrial and Applied Mathematics, pp. 17–36.
F. L. Cardoso-Ribeiro, D. Matignon and V. Pommier-Budinger (2017), A port-Hamiltonian model of liquid sloshing in moving containers and application to a fluid-structure system, *J. Fluids Structures* **69**, 402–427.
E. Celledoni and E. H. Høiseth (2017), Energy-preserving and passivity-consistent numerical discretization of port-Hamiltonian systems. Available at arXiv:1706.08621.
E. Celledoni, R. I. McLachlan, D. I. McLaren, B. Owren, G. R. W. Quispel and W. M. Wright (2009), Energy-preserving Runge–Kutta methods, *ESAIM Math. Model. Numer. Anal.* **43**, 645–649.
S. Chaturantabut and D. Sorensen (2010), Nonlinear model reduction via discrete empirical interpolation, *SIAM J. Sci. Comput.* **32**, 2737–2764.
S. Chaturantabut, C. A. Beattie and S. Gugercin (2016), Structure-preserving model reduction for nonlinear port-Hamiltonian systems, *SIAM J. Sci. Comput.* **38**, B837–B865.
K. Cherifi, H. Gernandt and D. Hinsen (2022*a*), The difference between port-Hamiltonian, passive and positive real descriptor systems. Available at arXiv:2204.04990.
K. Cherifi, V. Mehrmann and K. Hariche (2019), Numerical methods to compute a minimal realization of a port-Hamiltonian system. Available at arXiv:1903.07042.
K. Cherifi, V. Mehrmann and P. Schulze (2022*b*), Simulations in a digital twin of an electrical machine. Available at arXiv:2207.02171.
A. J. Conejo, S. Chen and G. E. Constante (2020), Operations and long-term expansion planning of natural-gas and power systems: A market perspective, *Proc. IEEE* **108**, 1541–1557.
L. Dai (1989), *Singular Control Systems*, Vol. 118 of Lecture Notes in Control and Information Sciences, Springer.
U. B. Desai and D. Pal (1984), A transformation approach to stochastic model reduction, *IEEE Trans. Automat. Control* **29**, 1097–1100.
L. Dieci and T. Eirola (1999), On smooth decompositions of matrices, *SIAM J. Matrix Anal. Appl.* **20**, 800–819.
L. Dieci and E. S. Van Vleck (2002), Lyapunov and other spectra: A survey, in *Collected Lectures on the Preservation of Stability under Discretization (Fort Collins, CO, 2001)* (D. Estep and S. Tavener, eds), Society for Industrial and Applied Mathematics, pp. 197–218.

L. Dieci, R. D. Russell and E. S. Van Vleck (1997), On the computation of Lyapunov exponents for continuous dynamical systems, *SIAM J. Numer. Anal.* **34**, 402–423.
N. H. Du, V. H. Linh and V. Mehrmann (2013), Robust stability of differential-algebraic equations, in *Surveys in Differential-Algebraic Equations I* (A. Ilchmann and T. Reis, eds), Springer, pp. 63–95.
V. Duindam, A. Macchelli, S. Stramigioli and H. Bruyninckx (2009), *Modeling and Control of Complex Physical Systems: The Port-Hamiltonian Approach*, Springer.
H. Egger (2019), Structure preserving approximation of dissipative evolution problems, *Numer. Math.* **143**, 85–106.
H. Egger and T. Kugler (2018), Damped wave systems on networks: Exponential stability and uniform approximations, *Numer. Math.* **138**, 839–867.
H. Egger and M. Sabouri (2021), On the structure preserving high-order approximation of quasistatic poroelasticity, *Math. Comput. Simul.* **189**, 237–252.
H. Egger, T. Kugler, B. Liljegren-Sailer, N. Marheineke and V. Mehrmann (2018), On structure preserving model reduction for damped wave propagation in transport networks, *SIAM J. Sci. Comput.* **40**, A331–A365.
E. Eich-Soellner and C. Führer (1998), *Numerical Methods in Multibody Dynamics*, Vieweg and Teubner.
E. Emmrich and V. Mehrmann (2013), Operator differential-algebraic equations arising in fluid dynamics, *Comput. Methods Appl. Math* **13**, 443–470.
H. Ennsbrunner and K. Schlacher (2005), On the geometrical representation and interconnection of infinite dimensional port controlled Hamiltonian systems, in *44th IEEE Conference on Decision and Control (CDC 2005)*, IEEE, pp. 5263–5268.
T. Faulwasser, B. Maschke, F. Philipp, M. Schaller and K. Worthmann (2022), Optimal control of port-Hamiltonian descriptor systems with minimal energy supply, *SIAM J. Control Optim.* **60**, 2132–2158.
R. W. Freund (2000), Krylov-subspace methods for reduced-order modeling in circuit simulation, *J. Comput. Appl. Math.* **123**, 395–421.
R. W. Freund (2005), Padé-type model reduction of second-order and higher-order linear dynamical systems, in *Dimension Reduction of Large-Scale Systems* (P. Benner, D. C. Sorensen and V. Mehrmann, eds), Springer, pp. 191–223.
R. W. Freund (2011), The SPRIM algorithm for structure-preserving order reduction of general RLC circuits, in *Model Reduction for Circuit Simulation* (P. Benner *et al.*, eds), Springer, pp. 25–52.
K. O. Friedrichs and P. D. Lax (1971), Systems of conservation equations with a convex extension, *Proc. Nat. Acad. Sci. USA* **68**, 1686–1688.
K. Fujimoto and H. Kajiura (2007), Balanced realization and model reduction of port-Hamiltonian systems, in *2007 American Control Conference*, pp. 930–934.
F. R. Gantmacher (1959), *The Theory of Matrices*, Vol. 2, Chelsea.
F. Gay-Balmaz and H. Yoshimura (2018), Dirac structures in nonequilibrium thermodynamics, *J. Math. Phys.* **59**, 012701.
F. Gay-Balmaz and H. Yoshimura (2019), From Lagranian mechanics to nonequilibrium thermodynamics: A variational perspective, *Entropy* **21**, 8.
H. Gernandt and F. E. Haller (2021), On the stability of port-Hamiltonian descriptor systems, *IFAC-PapersOnLine* **54**(19), 137–142.
H. Gernandt, F. E. Haller and E. Reis (2021), A linear relation approach to port-Hamiltonian differential-algebraic equations, *SIAM J. Matrix Anal. Appl.* **42**, 1011–1044.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

N. Gillis and P. Sharma (2017), On computing the distance to stability for matrices using linear dissipative Hamiltonian systems, *Automatica* **85**, 113–121.
N. Gillis, V. Mehrmann and P. Sharma (2018), Computing the nearest stable matrix pairs, *Numer. Linear Algebra Appl.* **25**, e2153.
N. Gräbner, V. Mehrmann, S. Quraishi, C. Schröder and U. von Wagner (2016), Numerical methods for parametric model reduction in the simulation of disc brake squeal, *ZAMM Z. Angew. Math. Mech.* **96**, 1388–1405.
E. J. Grimme (1997), Krylov projection methods for model reduction. PhD thesis, University of Illinois, Urbana-Champaign.
M. Grmela and H. C. Öttinger (1997), Dynamics and thermodynamics of complex fluids, I: Development of a general formalism, *Phys. Rev. E* **56**, 6620–6632.
C. Güdücü, J. Liesen, V. Mehrmann and D. Szyld (2022), On non-Hermitian positive (semi)definite linear algebraic systems arising from dissipative Hamiltonian DAEs, *SIAM J. Sci. Comput.* **44**, A2871–A2894.
S. Gugercin and A. C. Antoulas (2004), A survey of model reduction by balanced truncation and some new results, *Internat. J. Control* **77**, 748–766.
S. Gugercin, A. C. Antoulas and C. A. Beattie (2008), $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems, *SIAM J. Matrix Anal. Appl.* **30**, 609–638.
S. Gugercin, R. V. Polyuga, C. Beattie and A. van der Schaft (2009), Interpolation-based $\mathcal{H}_2$ model reduction for port-Hamiltonian systems, in *48th IEEE Conference on Decision and Control (CDC 2009)*, pp. 5362–5369.
S. Gugercin, R. V. Polyuga, C. Beattie and A. van der Schaft (2012), Structure-preserving tangential interpolation for model reduction of port-Hamiltonian systems, *Automatica* **48**, 1963–1974.
S. Gugercin, T. Stykel and S. Wyatt (2013), Model reduction of descriptor systems by interpolatory projection methods, *SIAM J. Sci. Comput.* **35**, B1010–B1033.
N. Guglielmi and V. Mehrmann (2022), Computation of the nearest structured matrix triplet with common null space, *Electron. Trans. Numer. Anal.* **55**, 508–531.
C. Guiver and M. R. Opmeer (2013), Error bounds in the gap metric for dissipative balanced approximations, *Linear Algebra Appl.* **439**, 3659–3698.
M. Günther, A. Bartel, B. Jacob and T. Reis (2021), Dynamic iteration schemes and port-Hamiltonian formulation in coupled differential-algebraic equation circuit simulation, *Int J. Circ. Theor. Appl.* **49**, 430–452.
B. Haasdonk (2017), Chapter 2: Reduced basis methods for parametrized PDEs: A tutorial introduction for stationary and instationary problems, in *Model Reduction and Approximation* (P. Benner *et al.*, eds), Society for Industrial and Applied Mathematics, pp. 65–136.
E. Hairer and G. Wanner (1996), *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, second edition, Springer.
E. Hairer, C. Lubich and M. Roche (1989), *The Numerical Solution of Differential-Algebraic Systems by Runge–Kutta Methods*, Springer.
E. Hairer, C. Lubich and G. Wanner (2002), *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, Springer.
P. Hamann and V. Mehrmann (2008), Numerical solution of hybrid systems of differential-algebraic equations, *Comput. Meth. Appl. Mech. Eng.* **197**, 693–705.
S.-A. Hauschild, N. Marheineke and V. Mehrmann (2019), Model reduction techniques for linear constant coefficient port-Hamiltonian differential-algebraic systems, *Control Cybernet.* **48**, 125–152.

M. Heinkenschloss, D. C. Sorensen and K. Sun (2008), Balanced truncation model reduction for a class of descriptor systems with application to the Oseen equations, *SIAM J. Sci. Comput.* **30**, 1038–1063.

J. S. Hesthaven, G. Rozza and B. Stamm (2016), *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*, Springer Briefs in Mathematics, Springer.

D. Hinrichsen and A. J. Pritchard (2005), *Mathematical Systems Theory I: Modelling, State Space Analysis, Stability and Robustness*, Springer.

M. Hou (1994), A three-link planar manipulator model. Technical report, Sicherheitstechnische Regelungs- und Meßtechnik, Bergische Universität–GH Wuppertal, Germany.

M. Hou and P. C. Müller (1994), *LQ* and tracking control of descriptor systems with application to constrained manipulator. Technical report, Sicherheitstechnische Regelungs- und Meßtechnik, Universität Wuppertal.

T. J. R. Hughes (2012), *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*, Courier.

T. C. Ionescu and A. Astolfi (2013), Families of moment matching based, structure preserving approximations for linear port Hamiltonian systems, *Automatica* **49**, 2424–2434.

T. C. Ionescu and J. M. A. Scherpen (2007), Positive real balancing for nonlinear systems, in *Scientific Computing in Electrical Engineering* (G. Ciuprina and D. Ioan, eds), Vol. 11 of Mathematics in Industry, Springer, pp. 153–159.

R. Ionutiu, J. Rommes and A. C. Antoulas (2008), Passivity-preserving model reduction using dominant spectral-zero interpolation, *IEEE Trans. Computer-Aided Design of Integr. Circ. Syst.* **27**, 2250–2263.

B. Jacob and H. Zwart (2012), *Linear Port-Hamiltonian Systems on Infinite-Dimensional Spaces*, Vol. 223 of Operator Theory: Advances and Applications, Birkhäuser.

T. Kailath (1980), *Linear Systems*, Vol. 156 of Information and System Sciences Series, Prentice-Hall.

R. Kannan, S. Hendry, N. J. Higham and F. Tisseur (2014), Detecting the causes of ill-conditioning in structural finite element models, *Computers & Structures* **133**, 79–89.

A. Karsai (2022), Structure-preserving control of port-Hamiltonian systems. Master's thesis, Technische Universität Berlin.

Y. Kawano and J. M. A. Scherpen (2018), Structure preserving truncation of nonlinear port Hamiltonian systems, *IEEE Trans. Automat. Control* **63**, 4286–4293.

P. Kotyczka (2019), Numerical methods for distributed parameter port-Hamiltonian systems. structure-preserving approaches for simulation and control. Habilitation thesis, TU Munich.

P. Kotyczka and L. Lefèvre (2018), Discrete-time port-Hamiltonian systems based on Gauss–Legendre collocation, *IFAC-PapersOnLine* **51**(3), 125–130.

P. Kotyczka, B. Maschke and L. Lefèvre (2018), Weak form of Stokes–Dirac structures and geometric discretization of port-Hamiltonian systems, *J. Comput. Phys.* **361**, 442–476.

P. Kunkel and V. Mehrmann (1991), Smooth factorizations of matrix valued functions and their derivatives, *Numer. Math.* **60**, 115–131.

P. Kunkel and V. Mehrmann (1996), A new class of discretization methods for the solution of linear differential-algebraic equations with variable coefficients, *SIAM J. Numer. Anal.* **33**, 1941–1961.

P. Kunkel and V. Mehrmann (1998), Regular solutions of nonlinear differential-algebraic equations and their numerical determination, *Numer. Math.* **79**, 581–600.

P. Kunkel and V. Mehrmann (2001), Analysis of over- and underdetermined nonlinear differential-algebraic systems with application to nonlinear control problems, *Math. Control Signals Systems* **14**, 233–256.
P. Kunkel and V. Mehrmann (2006), *Differential-Algebraic Equations: Analysis and Numerical Solution*, European Mathematical Society.
P. Kunkel and V. Mehrmann (2007), Stability properties of differential-algebraic equations and spin-stabilized discretizations, *Electron. Trans. Numer. Anal.* **26**, 385–420.
P. Kunkel and V. Mehrmann (2008), Optimal control for unstructured nonlinear differential-algebraic equations of arbitrary index, *Math. Control Signals Systems* **20**, 227–269.
P. Kunkel and V. Mehrmann (2011), Formal adjoints of linear DAE operators and their role in optimal control, *Electron. J. Linear Algebra* **22**, 672–693.
P. Kunkel and V. Mehrmann (2018), Regular solutions of DAE hybrid systems and regularization techniques, *BIT Numer. Math.* **58**, 1049–1077.
P. Kunkel and V. Mehrmann (2023), Local and global canonical forms for differential-algebraic equations with symmetries, *Vietnam J. Math.* **51**, 177–198.
P. Kunkel, V. Mehrmann and W. Rath (2001), Analysis and numerical solution of control problems in descriptor form, *Math. Control Signals Systems* **14**, 29–61.
P. Kunkel, V. Mehrmann and L. Scholz (2014), Self-adjoint differential-algebraic equations, *Math. Control Signals Systems* **26**, 47–76.
G. A. Kurina and R. März (2004), On linear-quadratic optimal control problems for time-varying descriptor systems, *SIAM J. Control Optim.* **42**, 2062–2077.
M. Kurula (2020), Well-posedness of time-varying linear systems, *IEEE Trans. Automat. Control* **65**, 4075–4089.
M. Kurula, H. Zwart, A. van der Schaft and J. Behrndt (2010), Dirac structures and their composition on Hilbert spaces, *J. Math. Anal. Appl.* **372**, 402–422.
J. La Salle and S. Lefschetz (1961), *Stability by Liapunov's Direct Method*, Mathematics in Science and Engineering, Academic Press.
J. P. La Salle (1976), *The Stability of Dynamical Systems*, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics.
R. Lamour, R. März and C. Tischendorf (2013), *Differential-Algebraic Equations: A Projector Based Analysis*, Differential-Algebraic Equations Forum, Springer.
S. Lang (2012), *Differential and Riemannian Manifolds*, Vol. 160 of Graduate Texts in Mathematics, Springer.
W. Layton (2008), *Introduction to the Numerical Analysis of Incompressible Viscous Flows*, Computational Science and Engineering, Society for Industrial and Applied Mathematics.
Y. Le Gorrec, H. Zwart and B. Maschke (2005), Dirac structures and boundary control systems associated with skew-symmetric differential operators, *SIAM J. Control Optim.* **44**, 1864–1892.
B. Liljegren-Sailer (2020), On port-Hamiltonian modeling and structure-preserving model reduction. PhD thesis, Universität Trier.
V. H. Linh and V. Mehrmann (2009), Lyapunov, Bohl and Sacker–Sell spectral intervals for differential-algebraic equations, *J. Dynam. Differ. Equations* **21**, 153–194.
V. H. Linh and V. Mehrmann (2011), Spectral analysis for linear differential-algebraic equations, *Discrete Contin. Dyn. Syst.* **2011**, 991–1000.
V. H. Linh and V. Mehrmann (2012), Spectra and leading directions for linear DAEs, in *Control and Optimization with Differential-Algebraic Constraints* (L. T. Biegler, S. L. Campbell and V. Mehrmann, eds), Vol. 23 of Advances in Design and Control, Society for Industrial and Applied Mathematics, pp. 59–78.

V. H. Linh and V. Mehrmann (2014), Efficient integration of strangeness-free non-stiff differential-algebraic equations by half-explicit methods, *J. Comput. Appl. Math.* **262**, 346–360.

V. H. Linh, V. Mehrmann and E. S. Van Vleck (2011), *QR* methods and error analysis for computing Lyapunov and Sacker–Sell spectral intervals for linear differential-algebraic equations, *Adv. Comput. Math.* **35**, 281–322.

A. Macchelli and B. Maschke (2009), Chapter 4: Infinite-dimensional port-Hamiltonian systems, in *Modeling and Control of Complex Physical Systems: The Port-Hamiltonian Approach* (V. Duindam *et al.*, eds), Springer, pp. 211–272.

A. Macchelli, C. Melchiorri and L. Bassi (2005), Port-based modelling and control of the Mindlin plate, in *44th IEEE Conference on Decision and Control (CDC 2005)*, pp. 5989–5994.

A. Macchelli, A. van der Schaft and C. Melchiorri (2004*a*), Port Hamiltonian formulation of infinite dimensional systems, I: Modeling, in *43rd IEEE Conference on Decision and Control (CDC 2004)*, Vol. 4, pp. 3762–3767.

A. Macchelli, A. van der Schaft and C. Melchiorri (2004*b*), Port Hamiltonian formulation of infinite dimensional systems, II: Boundary control by interconnection, in *43rd IEEE Conference on Decision and Control (CDC 2004)*, Vol. 4, pp. 3768–3773.

J. Machowski, Z. Lubosny, J. W. Bialek and J. R. Bumby (2020), *Power System Dynamics: Stability and Control*, Wiley.

M. Manuoglu and V. Mehrmann (2019), A robust iterative scheme for symmetric indefinite systems, *SIAM J. Sci. Comput.* **41**, A1733–A1752.

D. Matignon and T. Hélie (2013), A class of damping models preserving eigenspaces for linear conservative port-Hamiltonian systems, *Eur. J. Control* **19**, 486–494.

A. J. Mayo and A. C. Antoulas (2007), A framework for the solution of the generalized realization problem, *Linear Algebra Appl.* **425**, 634–662.

C. Mehl, V. Mehrmann and P. Sharma (2016), Stability radii for linear Hamiltonian systems with dissipation under structure-preserving perturbations, *SIAM J. Matrix Anal. Appl.* **37**, 1625–1654.

C. Mehl, V. Mehrmann and M. Wojtylak (2018), Linear algebra properties of dissipative Hamiltonian descriptor systems, *SIAM J. Matrix Anal. Appl.* **39**, 1489–1519.

C. Mehl, V. Mehrmann and M. Wojtylak (2021), Distance problems for dissipative Hamiltonian systems and related matrix polynomials, *Linear Algebra Appl.* pp. 335–366.

V. Mehrmann (1991), *The Autonomous Linear Quadratic Control Problem, Theory and Numerical Solution*, Vol. 163 of Lecture Notes in Control and Information Sciences, Springer.

V. Mehrmann (2015), Index concepts for differential-algebraic equations, in *Encyclopedia of Applied and Computational Mathematics*, Springer, pp. 676–681.

V. Mehrmann and P. V. Dooren (2020), Optimal robustness of port-Hamiltonian systems, *SIAM J. Matrix Anal. Appl.* **41**, 134–151.

V. Mehrmann and R. Morandin (2019), Structure-preserving discretization for port-Hamiltonian descriptor systems, in *58th IEEE Conference on Decision and Control (CDC 2019)*, pp. 6863–6868.

V. Mehrmann and T. Stykel (2005), Balanced truncation model reduction for large-scale systems in descriptor form, in *Dimension Reduction of Large-Scale Systems* (P. Benner, D. C. Sorensen and V. Mehrmann, eds), Springer, pp. 83–116.

V. Mehrmann and A. van der Schaft (2023), Differential–algebraic systems with dissipative Hamiltonian structure, *Math. Control Signals Systems*. Available at doi:10.1007/s00498-023-00349-2.

V. Mehrmann and H. Xu (2000), Numerical methods in control, *J. Comput. Appl. Math.* **123**, 371–394.

J. Montbrun-Di Filippo, M. Delgado, C. Brie and H. M. Paynter (1991), A survey of bond graphs: Theory, applications and programs, *J. Franklin Institute* **328**, 565–606.

R. Morandin, J. Nicodemus and B. Unger (2022), Port-Hamiltonian dynamic mode decomposition. Available at arXiv:2204.13474.

T. Moser and B. Lohmann (2020), A new Riemannian framework for efficient $\mathcal{H}_2$-optimal model reduction of port-Hamiltonian systems, in *59th IEEE Conference on Decision and Control (CDC 2020)*, pp. 5043–5049.

T. Moser, P. Schwerdtner, V. Mehrmann and M. Voigt (2022), Structure-preserving model order reduction for index two port-Hamiltonian descriptor systems. Available at arXiv:2206.03942.

A. Moses Badlyan and C. Zimmer (2018), Operator-GENERIC formulation of thermodynamics of irreversible processes. Available at arXiv:1807.09822.

A. Moses Badlyan, B. Maschke, C. Beattie and V. Mehrmann (2018), Open physical systems: From GENERIC to port-Hamiltonian systems, in *23rd International Symposium on Mathematical Theory of Systems and Networks*, pp. 204–211.

N. Nedialkov, J. D. Pryce and L. Scholz (2022), An energy-based, always index ≤ 1 and structurally amenable electrical circuit model, *SIAM J. Sci. Comput.* **44**, B1122–B1147.

H. C. Öttinger (2006), Nonequilibrium thermodynamics for open systems, *Phys. Rev. E* **73**, 036126.

H. C. Öttinger and M. Grmela (1997), Dynamics and thermodynamics of complex fluids, II: Illustrations of a general formalism, *Phys. Rev. E* **56**, 6633–6655.

C. C. Pantelides (1988), The consistent initialization of differential-algebraic systems, *SIAM J. Sci. Statist. Comput.* **9**, 213–231.

H. M. Paynter (1961), *Analysis and Design of Engineering Systems*, MIT Press.

F. Philipp, M. Schaller, T. Faulwasser, B. Maschke and K. Worthmann (2021), Minimizing the energy supply of infinite-dimensional linear port-Hamiltonian systems, *IFAC-PapersOnLine* **54**(19), 155–160.

J. W. Polderman and J. C. Willems (1998), *Introduction to Mathematical Systems Theory: A Behavioural Approach*, Texts in Applied Mathematics, Springer.

R. V. Polyuga and A. van der Schaft (2010), Structure preserving model reduction of port-Hamiltonian systems by moment matching at infinity, *Automatica* **46**, 665–672.

R. V. Polyuga and A. van der Schaft (2011), Structure-preserving moment matching for port-Hamiltonian systems: Arnoldi and Lanczos, *IEEE Trans. Automat. Control* **56**, 1458–1462.

R. V. Polyuga and A. van der Schaft (2012), Effort- and flow-constraint reduction methods for structure preserving model reduction of port-Hamiltonian systems, *Systems Control Lett.* **61**, 412–421.

J. D. Pryce (2001), A simple structural analysis method for DAEs, *BIT Numer. Math.* **41**, 364–394.

A. Quarteroni, A. Manzoni and F. Negri (2015), *Reduced Basis Methods for Partial Differential Equations: An Introduction*, UNITEXT, Springer International Publishing.

G. R. W. Quispel and D. I. McLaren (2008), A new class of energy-preserving numerical integration methods, *J. Phys. A* **41**, 045206.

P. J. Rabier and W. C. Rheinboldt (1996*a*), Classical and generalized solutions of time-dependent linear differential-algebraic equations, *Linear Algebra Appl.* **245**, 259–293.

P. J. Rabier and W. C. Rheinboldt (1996*b*), Time-dependent linear DAEs with discontinuous inputs, *Linear Algebra Appl.* **247**, 1–29.

P. J. Rabier and W. C. Rheinboldt (2000), *Nonholonomic Motion of Rigid Mechanical Systems from a DAE Viewpoint*, Society for Industrial and Applied Mathematics.

M. H. Ramirez Estay (2019), Modeling and control of irreversible thermodynamic processes and systems described by partial differential equations: A port-Hamiltonian approach. Habilitation thesis, Université Bourgogne Franche-Comté.

J. Ramsebner, R. Haas, A. Ajanovic and M. Wietschel (2021), The sector coupling concept: A critical review, *WIREs Energy and Environment* **10**, e396.

R. Rannacher (2000), Finite element methods for the incompressible Navier–Stokes equations, in *Fundamental Directions in Mathematical Fluid Mechanics* (P. Galdi, J. Heywood and R. Rannacher, eds), Birkhäuser, pp. 191–293.

D. Rapoport (1978), A nonlinear Lanczos algorithm and the stationary Navier–Stokes equation. PhD thesis, Department of Mathematics, Courant Institute, New York University.

R. Rashad, F. Califano, A. van der Schaft and S. Stramigioli (2020), Twenty years of distributed port-Hamiltonian systems: A literature review, *IMA J. Math. Control I.* pp. 1–23.

S. Reich (1990), On a geometrical interpretation of differential-algebraic equations, *Circuits Systems Signal Process.* **9**, 367–382.

T. Reis and T. Stykel (2010*a*), PABTEC: Passivity-preserving balanced truncation for electrical circuits, *IEEE Trans. Circuits and Systems* **29**, 1354–1367.

T. Reis and T. Stykel (2010*b*), Passivity-preserving balanced truncation model reduction of circuit equations, in *Scientific Computing in Electrical Engineering SCEE 2008* (J. Roos and L. Costa, eds), Vol. 14 of Mathematics in Industry, Springer, pp. 483–490.

T. Reis and M. Voigt (2015), The Kalman–Yakubovich–Popov inequality for differential-algebraic systems: Existence of nonpositive solutions, *Systems Control Lett.* **86**, 1–8.

T. Reis, O. Rendel and M. Voigt (2015), The Kalman–Yakubovich–Popov inequality for differential-algebraic systems, *Linear Algebra Appl.* **485**, 153–193.

W. C. Rheinboldt (1984), Differential-algebraic systems as differential equations on manifolds, *Math. Comp.* **43**, 473–482.

A. Romer, J. Berberich, H. Köhler and F. Allgöwer (2019), One-shot verification of dissipativity properties from input–output data, *IEEE Control Systems Lett.* **3**, 709–714.

A. Romer, J. M. Montenbruck and F. Allgöwer (2017), Determining dissipation inequalities from input-output samples, *IFAC-PapersOnLine* **50**, 7789–7794.

Y. Saad (2003), *Iterative Methods for Sparse Linear Systems*, second edition, Society for Industrial and Applied Mathematics.

J. Saak, M. Köhler and P. Benner (2021), M-M.E.S.S.-2.1: The matrix equations sparse solvers library. See also https://www.mpi-magdeburg.mpg.de/projects/mess.

K. Sato and H. Sato (2018), Structure-preserving H^2 optimal model reduction based on the Riemannian trust-region method, *IEEE Trans. Automat. Control* **63**, 505–512.

M. Schaller, F. Philipp, T. Faulwasser, K. Worthmann and B. Maschke (2021), Control of port-Hamiltonian systems with minimal energy supply, *Eur. J. Control* **62**, 33–40.

https://doi.org/10.1017/S0962492922000083 Published online by Cambridge University Press

T. M. Scheuermann, P. Kotyczka and B. Lohmann (2019), On parametric structure preserving model order reduction of linear port-Hamiltonian systems, *Automatisierungstechnik* **67**, 521–525.
J. Schiffer, E. Fridman, R. Ortega and J. Raisch (2016), Stability of a class of delayed port-Hamiltonian systems with application to microgrids with distributed rotational and electronic generation, *Automatica* **74**, 71–79.
M. Schöberl and K. Schlacher (2017), Variational principles for different representations of Lagranian and Hamiltonian systems, in *Dynamics and Control of Advanced Structures and Machines* (H. Irschik, A. Belyaev and M. Krommer, eds), Springer, pp. 65–73.
L. Scholz (2019), Condensed forms for linear port-Hamiltonian descriptor systems, *Electron. J. Linear Algebra* **35**, 65–89.
P. Schulze (2023), Energy-based model reduction of transport-dominated phenomena. PhD thesis, Institut für Mathematik, Technische Universität, Berlin.
P. Schulze and B. Unger (2018), Model reduction for linear systems with low-rank switching, *SIAM J. Control Optim.* **56**, 4365–4384.
P. Schulze, B. Unger, C. Beattie and S. Gugercin (2018), Data-driven structured realization, *Linear Algebra Appl.* **537**, 250–286.
P. Schwerdtner (2021), Port-Hamiltonian system identification from noisy frequency response data. Available at arXiv:2106.11355.
P. Schwerdtner and M. Voigt (2020), SOBMOR: Structured optimization-based model order reduction. Available at arXiv:2011.07567.
P. Schwerdtner and M. Voigt (2021), Adaptive sampling for structure-preserving model order reduction of port-Hamiltonian systems, *IFAC-PapersOnLine* **54**(19), 143–148.
P. Schwerdtner, T. Moser, V. Mehrmann and M. Voigt (2022), Structure-preserving model order reduction for index one port-Hamiltonian descriptor systems. Available at arXiv:2206.01608.
A. Serhani, D. Matignon and G. Haine (2019), A partitioned finite element method for the structure-preserving discretization of damped infinite-dimensional port-Hamiltonian systems with boundary control, in *Geometric Science of Information* (F. Nielsen and F. Barbaresco, eds), Springer, pp. 549–558.
H. Sharma, Z. Wang and B. Kramer (2022), Hamiltonian operator inference: Physics-preserving learning of reduced-order models for canonical Hamiltonian systems, *Phys. D* **431**, 133122.
B. Simeon (2013), *Computational Flexible Multibody Dynamics: A Differential-Algebraic Approach*, Differential-Algebraic Equations Forum, Springer.
I. Sobey, A. Eisenträger, B. Wirth and M. Czosnyka (2012), Simulation of cerebral infusion tests using a poroelastic model, *Int. J. Numer. Anal. Model. Ser. B* **3**, 52–64.
D. C. Sorensen (2005), Passivity preserving model reduction via interpolation of spectral zeros, *Systems Control Lett.* **54**, 347–360.
T. Stykel (2002), Stability and inertia theorems for generalized Lyapunov equations, *Linear Algebra Appl.* **355**, 297–314.
T. Stykel (2006), On some norms for descriptor systems, *IEEE Trans. Automat. Control* **51**, 842–847.
R. Temam (1977), *Navier–Stokes Equations: Theory and Numerical Analysis*, North-Holland.
S. Trenn (2013), Solution concepts for linear DAEs: A survey, in *Surveys in Differential-Algebraic Equations I* (A. Ilchmann and T. Reis, eds), Differential-Algebraic Equations Forum, Springer, pp. 137–172.

S. Trenn and B. Unger (2019), Delay regularity of differential-algebraic equations, in *58th IEEE Conference on Decision and Control (CDC 2019)*, pp. 989–994.
B. Tully and Y. Ventikos (2011), Cerebral water transport using multiple-network poroelastic theory: Application to normal pressure hydrocephalus, *J. Fluid Mech.* **667**, 188–215.
B. Unger (2020), Well-posedness and realization theory for delay differential-algebraic equations. PhD thesis, Technische Universität Berlin.
B. Unger and S. Gugercin (2019), Kolmogorov n-widths for linear dynamical systems, *Adv. Comput. Math.* **45**, 2273–2286.
A. van der Schaft (2000), *L2-gain and Passivity Techniques in Nonlinear Control*, Springer.
A. van der Schaft (2013), Port-Hamiltonian differential-algebraic systems, in *Surveys in Differential-Algebraic Equations I* (A. Ilchmann and T. Reis, eds), Differential-Algebraic Equations Forum, Springer, pp. 173–226.
A. van der Schaft and D. Jeltsema (2014), Port-Hamiltonian systems theory: An introductory overview, *Found. Trends Systems Control* **1**, 173–378.
A. van der Schaft and B. Maschke (2002), Hamiltonian formulation of distributed parameter systems with boundary energy flow, *J. Geom. Phys.* **42**, 166–174.
A. van der Schaft and B. Maschke (2018), Generalized port-Hamiltonian DAE systems, *Systems Control Lett.* **121**, 31–37.
A. van der Schaft and B. Maschke (2020), Dirac and Lagrange algebraic constraints in nonlinear port-Hamiltonian systems, *Vietnam J. Math.* **48**, 929–939.
H. J. van Waarde, M. K. Camlibel, P. Rapisarda and H. L. Trentelman (2022), Data-driven dissipativity analysis: Application of the matrix S-lemma, *IEEE Control Syst. Mag.* **42**, 140–149.
J. A. Villegas (2007), A port-Hamiltonian approach to distributed parameter systems. PhD thesis, University of Twente.
O. Widlund (1978), A Lanczos method for a class of nonsymmetric systems of linear equations, *SIAM J. Numer. Anal.* **15**, 801–812.
J. C. Willems (1971), Least squares stationary optimal control and the algebraic Riccati equation, *IEEE Trans. Automat. Control* **16**, 621–634.
T. Wolf, B. Lohmann, R. Eid and P. Kotyczka (2010), Passivity and structure preserving order reduction of linear port-Hamiltonian systems using Krylov subspaces, *Eur. J. Control* **16**, 401–406.
Y. Wu, B. Hamroun, Y. Le Gorrec and B. Maschke (2014), Structure preserving reduction of port Hamiltonian system using a modified LQG method, in *Proceedings of the 33rd Chinese Control Conference*, pp. 3528–3533.
Y. Wu, B. Hamroun, Y. Le Gorrec and B. Maschke (2018), Reduced order LQG control design for port Hamiltonian systems, *Automatica* **95**, 86–92.
H. Yoshimura and J. E. Marsden (2006*a*), Dirac structures in Lagrangian mechanics, I: Implicit Lagranian systems, *J. Geom. Phys.* **57**, 133–156.
H. Yoshimura and J. E. Marsden (2006*b*), Dirac structures in Lagrangian mechanics, II: Variational structures, *J. Geom. Phys.* **57**, 209–250.
O. C. Zienkiewicz and R. L. Taylor (2005), *The Finite Element Method for Solid and Structural Mechanics*, Elsevier.
M. D. Zoback (2010), *Reservoir Geomechanics*, Cambridge University Press.