Numerical Linear Algebra Methods for Linear Differential-Algebraic Equations

Peter Benner[1], Philip Losse[2], Volker Mehrmann[3], and Matthias Voigt[4]

[1] Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, D-39106 Magdeburg, Germany; E-mail: `benner@mpi-magdeburg.mpg.de`.

[2] Pestalozzistraße 13B, D-10625 Berlin, Germany; E-mail: `philip.losse@gmail.com`. Supported by the DFG Research Center MATHEON in Berlin.

[3] Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni 136, D-10623 Berlin, Germany; E-mail: `mehrmann@math.tu-berlin.de`. Supported by the European Research Council through ERC Advanced Grant ModSimConMP.

[4] Corresponding author, Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni 136, D-10623 Berlin, Germany; E-mail: `mvoigt@math.tu-berlin.de`. Research supported in the framework of MATHEON project *C-SE1: Reduced order modeling for data assimilation* supported by Einstein Foundation Berlin.

Summary. A survey of methods from numerical linear algebra for linear constant coefficient differential-algebraic equations (DAEs) and descriptor control systems is presented. We discuss numerical methods to check the solvability properties of DAEs as well as index reduction and regularization techniques. For descriptor systems we discuss controllability and observability properties and how these can be checked numerically. These methods are based on staircase forms and derivative arrays, obtained by real orthogonal transformations that are discussed in detail. Then we use the reformulated problems in several control applications for differential-algebraic equations ranging from regular and singular linear-quadratic optimal and robust control to dissipativity checking. We discuss these applications and give a systematic overview of the theory and the numerical solution methods. In particular, we show that all these applications can be treated with a common approach that is based on the computation of eigenvalues and deflating subspaces of even matrix pencils. The unified approach allows us to generalize and improve several techniques that are currently in use in systems and control.

Notation

$\mathbb{N}$, $\mathbb{N}_0$	the set of natural numbers, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$;
$\mathbb{R}$, $\mathbb{C}$	the fields of real and complex numbers, resp.;
$\mathbb{C}^-$, $\mathbb{C}^+$	the sets of complex numbers with negative and positive real parts, resp.;
i	the imaginary unit;
$\mathbf{u}$	the roundoff unit;
$\mathbb{R}[s]$, $\mathbb{C}[s]$	the rings of polynomials with real and complex coefficients in the indeterminate s, resp.;
$\mathbb{R}(s)$	the field of real-rational functions in the indeterminate s;
$\mathscr{R}^{m,n}$	the sets of $m \times n$ matrices with entries in a ring $\mathscr{R}$;
A^T, A^H, A^{-1}	transpose, conjugate transpose, and inverse of the matrix A;
$\operatorname{range} A$, $\ker A$	range and kernel of the matrix A, resp.;
$\operatorname{diag}(A_1,\ldots,A_k)$	$:= \begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_k \end{bmatrix}$;
$\Lambda(A)$	the spectrum of $A \in \mathbb{R}^{n,n}$:
$\Lambda(E,A)$	the set of finite eigenvalues of $sE - A \in \mathbb{R}[s]^{m,n}$.

1 Introduction

In modern modeling and simulation software packages such as MODELICA[1] or MATLAB/SIMULINK[2], the mathematical models are generated via a network of standardized submodels. This network approach has become the industrial standard in many physical and engineering domains, see, e.g., [8, 58, 68, 83, 101, 103, 104,

[1] `https://www.modelica.org/`
[2] `http://www.mathworks.com/`

105, 106, 112], and leads to differential-algebraic equations (DAEs), or descriptor systems in the control setting. The models include differential equations that model the dynamical behavior and algebraic equations that model constraints, interface and boundary conditions, or balance equations.

In this survey we study linear constant coefficient DAEs and descriptor systems, which arise from general nonlinear DAEs or descriptor systems by linearizing around a stationary solution [46], or via realization procedures [3, 4]. Linear constant coefficient DAEs take the form

$$E\dot{x}(t) = Ax(t) + f(t), \quad x(0) = x_0, \tag{1}$$

and linear time-invariant descriptor systems have the form

$$E\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0, \tag{2a}$$

$$y(t) = Cx(t) + Du(t), \tag{2b}$$

with matrices $E, A \in \mathbb{R}^{k,n}$, $B \in \mathbb{R}^{k,m}$, $C \in \mathbb{R}^{p,n}$ and $D \in \mathbb{R}^{p,m}$. Here, $x : [0,\infty) \to \mathbb{R}^n$ represents the state, $u : [0,\infty) \to \mathbb{R}^m$ denotes a control input signal, $y : [0,\infty) \to \mathbb{R}^p$ is the output signal, and $f : [0,\infty) \to \mathbb{R}^k$ is a given inhomogeneity.

For a uniform presentation, we combine both DAE and descriptor system in the form

$$E\dot{x}(t) = Ax(t) + Bu(t) + f(t), \quad x(0) = x_0, \tag{3a}$$

$$y(t) = Cx(t) + Du(t), \tag{3b}$$

where in the DAE case the term $Bu(\cdot)$ and the output equation are missing, whereas in the descriptor system case the inhomogeneity $f(\cdot)$ is dropped.

The survey is organized as follows. In Section 2 we briefly discuss the existence and uniqueness of solutions, as well the consistency of initial values. With a given DAE or descriptor system we can carry out numerical simulation, control, and optimization tasks. However, in the automatically generated models many difficulties arise which require a preliminary treatment, a reformulation, or a regularization, see [47]. In the case of linear constant coefficient DAEs or descriptor systems, this preliminary treatment is achieved using techniques from numerical linear algebra. In Section 3 the methods are based on derivative arrays and in Section 4 on staircase forms. These numerically stable methods allow us to check solvability and consistency of initial values for DAEs, as well as controllability and observability properties of descriptor systems.

After discussing the analysis and regularization techniques, we can proceed to more advanced control and optimization applications for descriptor systems. All these applications lead to generalized eigenvalue problems for even matrix pencils. Therefore, in Section 5 we discuss their structured condensed forms as well as the appropriate numerical methods. Afterward we consider the linear-quadratic regulator problem in Section 6 and the $\mathscr{H}_\infty$ optimal control problem in Section 7. In Section 8 we consider the computation of the $\mathscr{L}_\infty$-norm for continuous-time descriptor systems and finally, in Section 9 the dissipativity checking problem. Conclusions and comments on open problems complete the paper.

2 Solvability Theory

We begin our survey with the solvability theory of system (3a). This can done in terms of Kronecker canonical form (KCF) of the matrix pencil $sE-A \in \mathbb{R}[s]^{k,n}$, see, e.g., [43, 60].

Theorem 1 (Kronecker canonical form). *Let $sE-A \in \mathbb{R}[s]^{k,n}$ be given. Then there exist nonsingular matrices $P \in \mathbb{C}^{k,k}$ and $Q \in \mathbb{C}^{n,n}$ such that*

$$P(sE-A)Q = \operatorname{diag}\Big(\mathscr{L}_{\varepsilon_1}(s),\ldots,\mathscr{L}_{\varepsilon_k}(s),\mathscr{L}_{\delta_1}(s)^T,\ldots,\mathscr{L}_{\delta_\ell}(s)^T,\\ \mathscr{N}_{\sigma_1}(s),\ldots,\mathscr{N}_{\sigma_q}(s),\mathscr{J}_{\rho_1}(s),\ldots,\mathscr{J}_{\rho_r}(s)\Big),$$

where

(i) each $\mathscr{L}_{\varepsilon_j}(s)$ is an $\varepsilon_j \times (\varepsilon_j+1)$ right singular block with right minimal index $\varepsilon_j \in \mathbb{N}_0$ and form

$$s\begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix};$$

(ii) each $\mathscr{L}_{\delta_j}(s)^T$ is a $(\delta_j+1) \times \delta_j$ left singular block with left minimal index $\delta_j \in \mathbb{N}_0$ and form

$$s\begin{bmatrix} 0 & & \\ 1 & \ddots & \\ & \ddots & 0 \\ & & 1 \end{bmatrix} - \begin{bmatrix} 1 & & \\ 0 & \ddots & \\ & \ddots & 1 \\ & & 0 \end{bmatrix};$$

(iii) each $\mathscr{N}_{\sigma_j}(s)$ is a $\sigma_j \times \sigma_j$ infinite eigenvalue block with index $\sigma_j \in \mathbb{N}$ and form

$$s\begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix} - \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1 \end{bmatrix};$$

(iv) each $\mathscr{J}_{\rho_j}(s)$ is a $\rho_j \times \rho_j$ Jordan block with index $\rho_j \in \mathbb{N}$ and finite eigenvalue $\lambda_j \in \mathbb{C}$ and form

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

The Kronecker canonical form is unique up to permutation of the blocks, i.e., the kind, size, and number of the blocks are invariants of the pencil $sE-A$.

In the real version of the KCF, the blocks $\mathscr{J}_{\rho_j}(s)$ are in real Jordan form [69] and the transformation matrices P, Q are real. Based on the KCF we have the following definition.

Definition 1.

(i) A matrix pencil $sE - A \in \mathbb{R}[s]^{k,n}$ is called regular, *if $k = n$ and* $\det(\lambda E - A) \neq 0$ *for some $\lambda \in \mathbb{C}$. Otherwise the pencil is called* singular.

(ii) If $sE - A$ is regular, then a complex number λ_0 is a finite eigenvalue *of $sE - A$, if* $\det(\lambda_0 E - A) = 0$. *The finite eigenvalues are associated to the $\mathscr{J}_{\rho_j}(s)$ blocks in the KCF whereas the $\mathscr{N}_{\sigma_j}(s)$ blocks are said to be corresponding the* infinite eigenvalues *of the pencil $sE - A$.*

(iii) If $sE - A$ is a singular pencil, then its eigenvalues are the eigenvalues of the regular blocks in its Kronecker canonical form, i.e., the union of the eigenvalues of the $\mathscr{N}_{\sigma_j}(s)$ and $\mathscr{J}_{\rho_j}(s)$ blocks in Theorem 1.

(iv) The Kronecker index *of a regular matrix pencil $sE - A$ is the size of the largest block $\mathscr{N}_{\sigma_j}(s)$ in Theorem 1. It is denoted by* $\nu = \operatorname{ind}(E, A)$.

In the DAE case ($Bu \equiv 0$), it is clear from the KCF that for an arbitrary inhomogeneity $f(\cdot)$ and for arbitrary consistent initial conditions, to have a chance for a unique solution of (1), the pencil $sE - A$ has to be regular [44]. Nevertheless, if the pencil is singular, then for special $f(\cdot)$ and special initial conditions, a solution may exist and it even may be unique. Characterizations of existence and uniqueness of solutions can also be analyzed by different condensed forms, for instance the quasi-Kronecker form. see [25, 26]. For descriptor control systems (2a) the regularity of $sE - A$ is good to have, but not necessary.

In the regular case, the KCF specializes to the Weierstraß canonical form (WCF), see, e.g., [40, 60].

Theorem 2 (Weierstraß canonical form (WCF)). *If $sE - A \in \mathbb{R}[s]^{n,n}$ is a regular pencil, then there exist nonsingular matrices $X = \begin{bmatrix} X_\mathrm{f} \; X_\infty \end{bmatrix} \in \mathbb{C}^{n,n}$ and $Y = \begin{bmatrix} Y_\mathrm{f} \; Y_\infty \end{bmatrix} \in \mathbb{C}^{n,n}$ for which*

$$Y^H(sE - A)X = \begin{bmatrix} Y_\mathrm{f}^H \\ Y_\infty^H \end{bmatrix} (sE - A) \begin{bmatrix} X_\mathrm{f} \; X_\infty \end{bmatrix} = s \begin{bmatrix} I_r & 0 \\ 0 & N \end{bmatrix} - \begin{bmatrix} J & 0 \\ 0 & I_{n-r} \end{bmatrix}, \tag{4}$$

where $sI_r - J \in \mathbb{C}[s]^{r,r}$ with $J \in \mathbb{C}^{r,r}$ in Jordan canonical form contains the finite eigenvalues of $sE - A$, whereas the pencil $sN - I_{n-r} \in \mathbb{R}[s]^{n-r,n-r}$ with a nilpotent $N \in \mathbb{R}^{r,r}$ in Jordan canonical form corresponds to the infinite eigenvalues of $sE - A$.

Again there exists a real version of the WCF where J is in real Jordan canonical form and the transformation matrices X and Y are real. Since we prefer real-valued solutions we assume in the following considerations that the pencil $sE - A$ is transformed to real WCF and hence that X, Y as in Theorem 2 are real. With the notation of (4), classical continuously differentiable solutions of (3a) attain the form

$$x(t) = X_\mathrm{f} x_1(t) + X_\infty x_2(t),$$

where $x_1(\cdot), x_2(\cdot)$ are solutions of

$$\dot{x}_1(t) = Jx_1(t) + Y_{\mathrm{f}}^T(Bu(t) + f(t))$$
$$N\dot{x}_2(t) = x_2(t) + Y_{\infty}^T(Bu(t) + f(t)),$$

respectively. With $\nu = \mathrm{ind}(E,A)$, there are the explicit solution formulas

$$\begin{aligned} x_1(t) &= e^{Jt}x_1(0) + \int_0^t e^{J(t-\tau)}Y_{\mathrm{f}}^T(Bu(\tau) + f(\tau))\,\mathrm{d}\tau, \\ x_2(t) &= -\sum_{i=0}^{\nu-1} \frac{\mathrm{d}^i}{\mathrm{d}t^i}\left(N^i Y_{\infty}^T(Bu(t) + f(t))\right). \end{aligned} \tag{5}$$

In the descriptor system case this shows that the input functions must belong to some suitable function space $\mathscr{U}_{\mathrm{ad}}$. In particular, they must be sufficiently smooth.

Equation (5) also shows that the possible values of the initial condition x_0 are restricted. The initial state must be an element of the set of *consistent* initial conditions

$$\mathscr{S} := \left\{ X_{\mathrm{f}}x_{0,1} + X_{\infty}x_{0,2} : x_{0,1} \in \mathbb{R}^r, \right. \\ \left. x_{0,2} = -\sum_{i=0}^{\nu-1} \frac{\mathrm{d}^i}{\mathrm{d}t^i}\left(N^i Y_{\infty}^T(Bu(0) + f(0))\right), u(\cdot) \in \mathscr{U}_{\mathrm{ad}} \right\}.$$

To ensure a smooth response for every continuous input $u(\cdot)$ and every consistent initial value, it is necessary for the system to be regular and have index less than or equal to one.

The presented existence and uniqueness results are useful from a theoretical point of view, but it is well known that arbitrarily small perturbations can radically change the kind and number of the Kronecker blocks and thus it is problematic to compute the KCF or WCF with a numerical algorithm in finite precision arithmetic [109].

A better way to obtain the full information about the characteristic invariants in the WCF and KCF are *staircase algorithms*, which use a sequence of rank decisions, orthogonal matrix multiplications, and small perturbations to transform a pencil into a generalized upper triangular (GUPTRI) form [52, 53, 54, 113], see Section 4. These staircase forms can be used to check solvability conditions and consistency of initial conditions. However, if the system violates the above mentioned consistency and solvability conditions, or is close to such a system (in the sense that there exist small perturbations of the data that lead to a system that violates these conditions), then it is necessary to remodel or regularize the system such that further simulation and control methods are applicable. Again this should be done via numerically stable methods and this is the topic of the next section.

3 Regularization and Derivative Arrays

If not all the information about the characteristic quantities in the KCF is needed, then a very good alternative to the staircase form is to use a derivative array, see [45, 75]. This leads to a numerically stable method that allows us to check solvability and consistency of initial conditions [51]. Furthermore, if some of the conditions do not hold, then this approach can be used to obtain a regularization of the system. For general nonlinear DAEs and descriptor systems this general procedure has been presented in [47]. Here we briefly summarize this regularization procedure for the linear constant coefficient case.

In the following we assume that B and C^T have full column rank, otherwise we can just remove the kernels, by considering fewer inputs and outputs, respectively. This can be achieved by performing a singular value decomposition (SVD) or QR decomposition with column pivoting.

One first writes (3a) in behavior form, by combining input and state to a joint vector $z(\cdot) = \left[x(\cdot)^T \; u(\cdot)^T\right]^T$ as

$$\mathscr{E}\dot{z}(t) = \mathscr{A}z(t) + f(t) \tag{6}$$

with $\mathscr{E} = \left[E \; 0\right]$, $\mathscr{A} = \left[A \; B\right]$ partitioned analogously. Then for given $\mu \in \mathbb{N}$, one forms an enlarged DAE, namely

$$\mathscr{M}_\mu \dot{z}_\mu(t) = \mathscr{N}_\mu z_\mu(t) + \phi_\mu(t),$$

where

$$\mathscr{M}_\mu = \begin{bmatrix} \mathscr{E} & & & \\ -\mathscr{A} & \mathscr{E} & & \\ & \ddots & \ddots & \\ & & -\mathscr{A} & \mathscr{E} \end{bmatrix} \in \mathbb{R}^{(\mu+1)k,(\mu+1)(n+m)},$$

$$\mathscr{N}_\mu = \begin{bmatrix} \mathscr{A} & 0 & \dots & 0 \\ 0 & 0 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & 0 \end{bmatrix} \in \mathbb{R}^{(\mu+1)k,(\mu+1)(n+m)},$$

$$z_\mu(\cdot) = \begin{bmatrix} z(\cdot) \\ \dot{z}(\cdot) \\ \vdots \\ z^{(\mu)}(\cdot) \end{bmatrix}, \quad \phi_\mu(\cdot) = \begin{bmatrix} f(\cdot) \\ \dot{f}(\cdot) \\ \vdots \\ f^{(\mu)}(\cdot) \end{bmatrix}.$$

With the above notation, the pair $(\mathscr{M}_\mu, \mathscr{N}_\mu)$ is called derivative array [45]. One obtains the following theorem, see [74, 75, 78].

Theorem 3. *Let the system* (6) *be given. Then there exist integers μ, d, a, and v such that $(\mathscr{M}_\mu, \mathscr{N}_\mu)$ has the following properties:*

(i) $\operatorname{corank}\mathscr{M}_{\mu+1} - \operatorname{corank}\mathscr{M}_{\mu} = v$.

(ii) $\operatorname{rank}\mathscr{M}_{\mu} = (\mu+1)k - a - v$, *i.e., there exists a matrix* $Z \in \mathbb{R}^{(\mu+1)k,a+v}$ *with orthonormal columns and maximal rank, satisfying* $Z^T \mathscr{M}_{\mu} = 0$.

(iii) $\operatorname{rank} Z^T \mathscr{N}_{\mu} \begin{bmatrix} I_{n+m} \; 0 \; \ldots \; 0 \end{bmatrix}^T = a$, *i.e., Z can be partitioned as* $Z = \begin{bmatrix} Z_2 \; Z_3 \end{bmatrix}$, *with* $Z_2 \in \mathbb{R}^{(\mu+1)k,a}$ *and* $Z_3 \in \mathbb{R}^{(\mu+1)k,v}$ *such that* $\widehat{A}_2 := Z_2^T \mathscr{N}_{\mu} \begin{bmatrix} I_{n+m} \; 0 \; \ldots \; 0 \end{bmatrix}^T$ *has full row rank a and* $Z_3^T \mathscr{N}_{\mu} \begin{bmatrix} I_{n+m} \; 0 \; \ldots \; 0 \end{bmatrix}^T = 0$. *Furthermore, there exists a matrix* T_2 *with orthonormal columns of maximal rank satisfying* $\widehat{A}_2 T_2 = 0$.

(iv) $\operatorname{rank}\mathscr{E} T_2 = d = k - a - v$, *i.e., there exists* $Z_1 \in \mathbb{R}^{k,d}$ *with orthonormal columns and maximal rank satisfying* $\widehat{E}_1 := Z_1^T \mathscr{E}$ *with* $\operatorname{rank}\widehat{E}_1 = d$.

Furthermore, system (6) *has the same solution set as the* strangeness-free *system*

$$\begin{bmatrix} \widehat{E}_1 \\ 0 \\ 0 \end{bmatrix} \dot{z}(t) = \begin{bmatrix} \widehat{A}_1 \\ \widehat{A}_2 \\ 0 \end{bmatrix} z(t) + \begin{bmatrix} \widehat{f}_1(t) \\ \widehat{f}_2(t) \\ \widehat{f}_3(t) \end{bmatrix}, \tag{7}$$

where $\widehat{E}_1 = Z_1^T \mathscr{E}$, $\widehat{A}_1 = Z_1^T \mathscr{A}$, $\widehat{f}_1(\cdot) = Z_1^T f(\cdot)$, *and* $\widehat{f}_i(\cdot) = Z_i^T \phi_{\mu}(\cdot)$, $i = 2, 3$.

The number μ is called the *strangeness-index* of the DAE. It is equal to the size of the largest block of types $\mathscr{L}_{\varepsilon}(s)$ or $\mathscr{N}_{\sigma}(s)$ and is equal to $\nu - 1$ with $\nu = \operatorname{ind}(E,A)$ if the pencil is regular with $\nu > 0$, see [75, 78]. It satisfies $\mu = 0$ if the system is regular and of index at most one. If μ is known, then the coefficients of the differential-algebraic system (7) can be computed by using three nullspace computations, which can be implemented via singular value decompositions or QR decompositions with column pivoting. If μ is not known, then one proceeds recursively, increasing μ until the form (7) can be numerically safely determined.

System (7) is a *reformulation* of (6) (using the original system and its derivatives), without changing the solution set, since no transformation of the vector $z(\cdot)$ has been made. The constructed submatrices $\widehat{A}_1$ and $\widehat{A}_2$ have been obtained from the block matrix

$$\begin{bmatrix} A & B \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{(\mu+1)k,n+m}$$

by transformations from the left, and this means that no derivatives of $u(\cdot)$ are needed, and hence, there are no additional smoothness requirements for $u(\cdot)$. Furthermore, we immediately obtain again a descriptor system of the form

$$E_1 \dot{x}(t) = A_1 x(t) + B_1 u(t) + \widehat{f}_1(t), \quad x(0) = x_0 \tag{8a}$$

$$0 = A_2 x(t) + B_2 u(t) + \widehat{f}_2(t), \tag{8b}$$

$$0 = \widehat{f}_3(t), \tag{8c}$$

$$y(t) = Cx(t) + Du(t), \tag{8d}$$

where

$$E_1 = \widehat{E}_1 \begin{bmatrix} I_n \\ 0 \end{bmatrix}, \quad A_i = \widehat{A}_i \begin{bmatrix} I_n \\ 0 \end{bmatrix}, \quad B_i = \widehat{A}_i \begin{bmatrix} 0 \\ I_m \end{bmatrix}, \quad i = 1, 2,$$

and $E_1, A_1 \in \mathbb{R}^{d,n}$, while $A_2 \in \mathbb{R}^{a,n}$.

The equations in (8c) characterize the solvability of (3a), which is given if $\widehat{f}_3 \equiv 0$. If $\widehat{f}_3 \not\equiv 0$, then the system does not have a classical solution. In this case either the model should be discarded or one can perform a regularization by just setting $\widehat{f}_3 \equiv 0$ and release a warning that the system has been modified. In the latter case these equations can just be removed from the system and one continues with a modified state equation of $d + a$ equations in n state variables

$$E_1 \dot{x}(t) = A_1 x(t) + B_1 u(t) + \widehat{f}_1(t), \quad x(0) = x_0 \tag{9a}$$

$$0 = A_2 x(t) + B_2 u(t) + \widehat{f}_2(t), \tag{9b}$$

$$y(t) = Cx(t) + Du(t), \tag{9c}$$

together with the given initial conditions.

Consistency of initial values can be easily checked, they have to satisfy the equation

$$A_2 x_0 + B_2 u(0) + \widehat{f}_2(0) = 0, \tag{10}$$

and this also restricts the set of admissible inputs $u(\cdot)$. Again if the given initial conditions do not satisfy (10), then a regularization would make them consistent.

In (9a) and (9b) we have $d + a$ equations and n variables in $x(\cdot)$ and m variables in $u(\cdot)$. In order for this system to be uniquely solvable for all sufficiently smooth inputs $u(\cdot)$, and all consistent initial conditions, as a necessary condition we would need that $d + a = n$ [44, 75]. In a reasonable model, this should be the case, but since automatically generated models typically have redundancies, and also there may be modeling errors, a mismatch may happen which can, however, be easily fixed. If $d + a < n$, then for given $u(\cdot)$ we cannot expect a unique solution, so we can just attach $n - d - a$ variables from $x(\cdot)$ to $u(\cdot)$ and if $d + a > n$, then we just attach $d + a - n$ of the input variables in $u(\cdot)$ to $x(\cdot)$. Note that we must also change the output equation by moving appropriate columns from D to C or vice versa. There is some freedom in this renaming of variables, which should be resolved by considering the application, and actually this step is not needed in some of the applications that we discuss below.

As a result of a possible reinterpretation of variables we obtain a remodeled system

$$\begin{aligned} \widetilde{E}_1 \dot{\widetilde{x}}(t) &= \widetilde{A}_1 \widetilde{x}(t) + \widetilde{B}_1 \widetilde{u}(t) + \widetilde{f}_1(t), \quad \widetilde{x}(0) = \widetilde{x}_0, \\ 0 &= \widetilde{A}_2 \widetilde{x}(t) + \widetilde{B}_2 \widetilde{u}(t) + \widetilde{f}_2(t), \\ y(t) &= \widetilde{C} \widetilde{x}(t) + \widetilde{D} \widetilde{u}(t), \end{aligned}$$

where $\begin{bmatrix} \widetilde{E}_1 \\ 0 \end{bmatrix}, \begin{bmatrix} \widetilde{A}_1 \\ \widetilde{A}_2 \end{bmatrix} \in \mathbb{R}^{\widetilde{n},\widetilde{n}}$, $\begin{bmatrix} \widetilde{B}_1 \\ \widetilde{B}_2 \end{bmatrix} \in \mathbb{R}^{\widetilde{n},\widetilde{m}}$, and $\widetilde{n}$, $\widetilde{m}$ are the numbers of state and input variables of the reinterpreted system, respectively.

It is also often useful to remove the feed-through term $\widetilde{D}\widetilde{u}(\cdot)$ in the output equation. This can be achieved by performing a row compression with an orthogonal matrix P such that $P^T\widetilde{D} = \begin{bmatrix} \widetilde{D}_1 \\ 0 \end{bmatrix}$ with $\widetilde{D}_1 \in \mathbb{R}^{p_1,\widetilde{m}}$ having full row rank. By setting (with an accordant partitioning)

$$P^T y(\cdot) = \begin{bmatrix} \widetilde{y}_1(\cdot) \\ \widetilde{y}_2(\cdot) \end{bmatrix}, \quad P^T\widetilde{C} = \begin{bmatrix} \widetilde{C}_1 \\ \widetilde{C}_2 \end{bmatrix},$$

with $\widetilde{C}_1 \in \mathbb{R}^{p_1,\widetilde{n}}$, then we obtain a new system without feed-through term of the form

$$\overline{E}\dot{\overline{x}}(t) = \overline{A}\overline{x}(t) + \overline{B}\overline{u}(t) + \overline{f}(t), \quad \overline{x}(0) = \overline{x}_0, \tag{11a}$$

$$\overline{y}(t) = \overline{C}\overline{x}(t), \tag{11b}$$

with data

$$\overline{E} = \begin{bmatrix} \widetilde{E}_1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{\overline{n},\overline{n}}, \quad \overline{A} = \begin{bmatrix} \widetilde{A}_1 & 0 \\ \widetilde{A}_2 & 0 \\ \widetilde{C}_1 & -I_{p_1} \end{bmatrix} \in \mathbb{R}^{\overline{n},\overline{n}}, \quad \overline{B} = \begin{bmatrix} \widetilde{B}_1 \\ \widetilde{B}_2 \\ \widetilde{D}_1 \end{bmatrix} \in \mathbb{R}^{\overline{n},\overline{m}},$$

$$\overline{C} = \widetilde{C}_2 \in \mathbb{R}^{\overline{p},\overline{n}},$$

$$\overline{x}(\cdot) = \begin{bmatrix} \widetilde{x}(\cdot) \\ \widetilde{y}_1(\cdot) \end{bmatrix}, \quad \overline{y}(\cdot) = \widetilde{y}_2(\cdot), \quad \overline{u}(\cdot) = \widetilde{u}(\cdot), \quad \overline{f}(\cdot) = \begin{bmatrix} \widetilde{f}_1(\cdot) \\ \widetilde{f}_2(\cdot) \\ 0 \end{bmatrix},$$

where $\overline{n} = \widetilde{n} + p_1$, $\overline{m} = \widetilde{m}$, and $\overline{p} = p - p_1$. The resulting system may not be of index at most one as a free system with $\overline{u} \equiv 0$. In this case, see [40, 78], one can construct a linear state feedback $\overline{u}(t) = K\overline{x}(t) + w(t)$, with $K \in \mathbb{R}^{\overline{m},\overline{n}}$ such that in the closed-loop system

$$\overline{E}\dot{\overline{x}}(t) = (\overline{A} + \overline{B}K)\overline{x}(t) + \overline{B}w(t) + \overline{f}(t), \quad \overline{x}(0) = \overline{x}_0, \tag{12a}$$

$$\overline{y}(t) = \overline{C}\overline{x}(t), \tag{12b}$$

the matrix $(\overline{A}_2 + \overline{B}_2K)\overline{S}_\infty$ is nonsingular, where $\overline{S}_\infty$ is a matrix that spans $\ker \overline{E}_1$. This implies that the DAE in (12a) is regular and of index at most one as a free system with $w \equiv 0$. A similar index reduction can also be constructed via output feedback [40, 41, 78], it would however require a change of basis in the state space, and thus a change of the physical meaning of the state variables. See Section 4 for more details. The flowchart given in Figure 1 summarizes the regularization procedure.

Note that several of the steps in the regularization procedure may be void if the system has adequate properties and for some of the applications discussed below also a preliminary regularization may not be necessary. Note furthermore, that in this procedure no changes have been performed in the state variables, except for the possible reinterpretation of variables or the extension of the state space in the

$$\begin{array}{rcl} E\dot{x}(t) &=& Ax(t)+Bu(t)+f(t), \quad x(0)=x_0, \\ y(t) &=& Cx(t)+Du(t) \end{array}$$

$\mu \neq 0$ ↓ index reduction in behavior

$$\begin{array}{rcl} E_1\dot{x}(t) &=& A_1x(t)+B_1u(t)+\widehat{f}_1(t), \quad x(0)=x_0, \\ 0 &=& A_2x(t)+B_2u(t)+\widehat{f}_2(t), \\ 0 &=& \widehat{f}_3(t), \\ y(t) &=& Cx(t)+Du(t) \end{array}$$

↓ remove $0 \equiv \widehat{f}_3$

$0 \neq A_2x_0+B_2u(0)+\widehat{f}_2(0)$ ↓ cond. for consistency

$a+d \neq n$ ↓ reinterpret variables

$$\begin{array}{rcl} \widetilde{E}_1\dot{\widetilde{x}}(t) &=& \widetilde{A}_1\widetilde{x}(t)+\widetilde{B}_1\widetilde{u}(t)+\widetilde{f}_1(t), \quad \widetilde{x}(0)=\widetilde{x}_0, \\ 0 &=& \widetilde{A}_2\widetilde{x}(t)+\widetilde{B}_2\widetilde{u}(t)+\widetilde{f}_2(t), \\ \widetilde{y}(t) &=& \widetilde{C}\widetilde{x}(t)+\widetilde{D}\widetilde{u}(t) \end{array}$$

$\widetilde{D} \neq 0$ ↓ remove feed-through

$$\begin{array}{rcl} \overline{E}_1\dot{\overline{x}}(t) &=& \overline{A}_1\overline{x}(t)+\overline{B}_1\overline{u}(t)+\overline{f}_1(t), \quad \overline{x}(0)=\overline{x}_0, \\ 0 &=& \overline{A}_2\overline{x}(t)+\overline{B}_2\overline{u}(t)+\overline{f}_2(t), \\ \overline{y}(t) &=& \overline{C}\overline{x}(t) \end{array}$$

not index at most one for $u \equiv 0$ ↓ feedback $\overline{u}(t)=K\overline{x}(t)+w(t)$

$$\begin{array}{rcl} \overline{E}_1\dot{\overline{x}}(t) &=& (\overline{A}_1+\overline{B}_1K)\overline{x}(t)+\overline{B}_1w(t)+\overline{f}_1(t), \quad \overline{x}(0)=\overline{x}_0, \\ 0 &=& (\overline{A}_2+\overline{B}_2K)\overline{x}(t)+\overline{B}_2w(t)+\overline{f}_2(t), \\ \overline{y}(t) &=& \overline{C}\overline{x}(t) \end{array}$$

Fig. 1. Regularization procedure

case of feed-through removal. In any case, the original physical meaning of the state variables is still present in the system. This is of great importance and a clear advantage of the derivative array approach compared to the staircase forms that we discuss below.

Example 1. To illustrate the regularization procedure, consider the following example:

$$E = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad f(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \end{bmatrix},$$
$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

The corresponding behavioral system (6) is given by

$$\mathscr{E} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathscr{A} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad f(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \end{bmatrix}$$

The strangeness index of this system is 1. Thus we obtain the derivative array

$$(\mathscr{M}_1, \mathscr{N}_1) = \left(\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \right).$$

With the notation of Theorem 3 we obtain

$$Z_2^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad Z_3^T = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then we obtain

$$\widehat{A}_2 := Z_2^T \mathscr{N}_1 \begin{bmatrix} I_4 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

and consequently we have

$$T_2 = \begin{bmatrix} -1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

We obtain $\mathscr{E}T_2 = 0$ and thus Z_1 is void. Overall, the strangeness-free system (7) reads

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

We see, that in order for the system to be solvable at all, we need $f_3 \equiv 0$. Therefore, assume that $f_3 \equiv 0$. Then the reduced system (9) is given by

$$\begin{aligned} 0 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} f_1(t) + \dot{f}_2(t) \\ f_2(t) \end{bmatrix}, \\ y(t) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t), \end{aligned}$$

from which we can directly read off the condition for consistency by setting $t = 0$. Since $2 = a + d \neq n = 3$, a reinterpretation of variables is necessary. Thus, by setting $u_1(\cdot) := u(\cdot)$ and $u_2(\cdot) := x_3(\cdot)$, we obtain the square system

$$\begin{aligned} 0 &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} + \begin{bmatrix} f_1(t) + \dot{f}_2(t) \\ f_2(t) \end{bmatrix}, \\ y(t) &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}. \end{aligned}$$

Now we remove the feed-through matrix D. Thus the feed-through-free system (11) reads

$$0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ y_1(t) \\ y_2(t) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} + \begin{bmatrix} f_1(t) + \dot{f}_2(t) \\ f_2(t) \\ 0 \\ 0 \end{bmatrix},$$

which is regular and of index one. Note that the output equation has vanished during the last step in the regularization procedure. In fact, it is included in the state equation.

4 Staircase Forms and Properties of Descriptor Systems

In this section we discuss the system theoretic properties of descriptor systems and present the staircase forms that allow us to check these properties. We focus on concepts like *controllability, stabilizability* and the related dual notions of *observability* and *detectability*. For brevity we only introduce these for the case of square systems and systems where the feed-through term D has been removed, so we assume that the system is already in the form (11) as generated by the regularization procedure of Section 3. Also, instead of defining these properties in system theoretic terms, we directly introduce equivalent algebraic characterizations. These are very useful for numerically checking these properties. Note that there are several different concepts of controllability at infinity introduced in [102, 115] and compared in [22, 23, 40, 49, 50]. Furthermore, different observability notions are reviewed in [24].

Proposition 1. *Let* $sE - A \in \mathbb{R}[s]^{n,n}$ *be regular,* $B \in \mathbb{R}^{n,m}$, $C \in \mathbb{R}^{p,n}$*. Furthermore, let* S_∞, T_∞ *be matrices with* $\operatorname{range} S_\infty = \ker E$ *and* $\operatorname{range} T_\infty = \ker E^T$*. Then the triple* (E, A, B) *is called*

(i) behaviorally stabilizable *if* rank $\left[\lambda E - A \ B\right] = n$ *for all* $\lambda \in \mathbb{C}^+ \cup \mathrm{i}\mathbb{R}$*;*
(ii) behaviorally controllable *if* rank $\left[\lambda E - A \ B\right] = n$ *for all* $\lambda \in \mathbb{C}$*;*
(iii) impulse controllable *if* rank $\left[E \ AS_\infty \ B\right] = n$*;*
(iv) strongly stabilizable *if it is both behaviorally stabilizable and impulse controllable;*
(v) strongly controllable *if it is both behaviorally controllable and impulse controllable;*
(vi) completely controllable *if it is both behaviorally controllable and* rank $\left[E \ B\right] = n$*.*

In a dual fashion, the triple (E,A,C) *is called*

(vii) behaviorally detectable *if* rank $\left[\lambda E^T - A^T \ C^T\right] = n$ *for all* $\lambda \in \mathbb{C}^+ \cup \mathrm{i}\mathbb{R}$*;*
(viii) behaviorally observable *if* rank $\left[\lambda E^T - A^T \ C^T\right] = n$ *for all* $\lambda \in \mathbb{C}$*;*
(ix) impulse observable *if* rank $\left[E^T \ A^T T_\infty \ C^T\right] = n$*;*
(x) strongly detectable *if it is both behaviorally detectable and impulse observable;*
(xi) strongly observable *if it is both behaviorally observable and impulse observable;*
(xii) completely observable *if it is both behaviorally observable and* rank $\left[E^T \ C^T\right] = n$*.*

To check whether a given descriptor system satisfies these conditions one can use the staircase form of [40, 41, 113], which can be implemented as a sequence of orthogonal transformations to the system [28].

Theorem 4. *[41] If* $E, A \in \mathbb{R}^{n,n}$*,* $B \in \mathbb{R}^{n,m}$*,* $C \in \mathbb{R}^{p,n}$*, then there exist orthogonal matrices* $U, V \in \mathbb{R}^{n,n}$*,* $W \in \mathbb{R}^{m,m}$*,* $Y \in \mathbb{R}^{p,p}$ *such that*

$$U^T E V = \begin{array}{c} \\ t_1 \\ n-t_1 \end{array} \begin{array}{c} \begin{array}{cc} t_1 & n-t_1 \end{array} \\ \begin{bmatrix} \Sigma_E & 0 \\ 0 & 0 \end{bmatrix} \end{array},$$

$$U^T A V = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{array} \begin{array}{c} \begin{array}{cccccc} t_1 & s_2 & t_5 & t_4 & t_3 & s_6 \end{array} \\ \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} \\ A_{21} & A_{22} & A_{23} & A_{24} & 0 & 0 \\ A_{31} & A_{32} & A_{33} & A_{34} & \Sigma_{35} & 0 \\ A_{41} & A_{42} & A_{43} & \Sigma_{44} & 0 & 0 \\ A_{51} & 0 & \Sigma_{53} & 0 & 0 & 0 \\ A_{61} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array},$$

$$U^T B W = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{array} \begin{array}{c} \begin{array}{ccc} k_1 & k_2 & k_3 \end{array} \\ \begin{bmatrix} B_{11} & B_{12} & 0 \\ B_{21} & 0 & 0 \\ B_{31} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}, \tag{13}$$

$$Y^T C V = \begin{array}{c} \\ l_1 \\ l_2 \\ l_3 \end{array} \begin{array}{c} \begin{array}{cccccc} t_1 & s_2 & t_5 & t_4 & t_3 & s_6 \end{array} \\ \begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{21} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}.$$

The matrices $\Sigma_E, \Sigma_{35}, \Sigma_{44}, \Sigma_{53}$ are nonsingular diagonal matrices, B_{12} has full column rank, C_{21} has full row rank and the matrices

$$\begin{bmatrix} B_{21} \\ B_{31} \end{bmatrix} \in \mathbb{R}^{k_1,k_1}, \quad \begin{bmatrix} C_{12} \; C_{13} \end{bmatrix} \in \mathbb{R}^{l_1,l_1},$$

with $k_1 = t_2 + t_3$ and $l_1 = s_2 + t_5$ are nonsingular.

For numerical examples for the above decomposition we refer to the appendix of [40]. Impulse controllability and observability and some further properties can be checked via the following corollary.

Corollary 1. *[41] Let E, A, B, C with regular $sE - A$ be given as in the condensed form* (13). *Then the following statements are satisfied.*

(i) The triple (E,A,B) is impulse controllable if and only if $t_6 = 0$, i.e., the last block row of A is void.
(ii) The triple (E,A,C) is impulse observable if and only if $s_6 = 0$, i.e., the last block column of A is void.

(iii) The condition rank $\begin{bmatrix} E & B \end{bmatrix} = n$ *is satisfied if and only if* $t_4 = t_5 = t_6 = 0$.
(iv) The condition rank $\begin{bmatrix} E^T & C^T \end{bmatrix} = n$ *is satisfied if and only if* $t_4 = t_3 = s_6 = 0$.
(v) The triple (E, A, B) *is completely controllable if and only if* $t_4 = t_5 = t_6 = 0$ *and the system is behaviorally controllable.*
(vi) The triple (E, A, C) *is completely observable if and only if* $t_4 = t_3 = s_6 = 0$ *and the system is behaviorally observable.*

If properly implemented, see [52, 53, 54, 113], these techniques determine the characteristic invariants of a *least generic* system within a tolerated perturbation, see [56, 57]. In this way the staircase form (13) presents an alternative way to check some of the properties compared to the derivative array as in Section 3. But the computation of the staircase form is much more subtle numerically, since the consecutive rank decisions of the transformed matrices have to be made in a proper way, see [113]. In contrast to the derivative array approach, two-sided transformations are used, i.e., also changes of basis in the state space. This allows us to check observability and controllability conditions simultaneously, but at the cost of changing the physical meaning of the state variables. This is clearly no problem when the data is produced from a realization or model reduction process [3, 4], where the state variables have no direct physical interpretation, but this may be a problem when the model directly arises from a physical model. In this case it is suggested to first perform the regularization procedure of Section 3 and then perform the staircase algorithm to check the properties.

If only the first step of the regularization via derivative arrays has been performed and the system (9) has been filtered out of the derivative array and $d + a = n$, i.e., no more reinterpretation of variables is necessary, then the system is already impulse controllable. If the system is not impulse observable, then this is critical because impulse observability cannot be achieved by removing equations and variables. In this case, impulses in the solution (that appear, e.g., for inconsistent initial values) cannot be observed and this is an indication of a problem in the modeling, see [42]. In some of the applications that we discuss below, the solvability depends on these properties and an alternative model should be created to ensure that they are satisfied.

If the system is impulse controllable and impulse observable, then the other properties, i.e., behavioral controllability or stabilizability and behavioral observability or detectability can be checked via the following *controllability/observability decompositions*, see [50, 113, 114]. Let Q_c, Z_c be real orthogonal matrices, such that

$$\begin{aligned} Q_c^T E Z_c &= \begin{bmatrix} E_c & * \\ 0 & E_{nc} \end{bmatrix}, \; Q_c^T A Z_c = \begin{bmatrix} A_c & * \\ 0 & A_{nc} \end{bmatrix}, \\ Q_c^T B &= \begin{bmatrix} B_c \\ 0 \end{bmatrix}, \qquad C Z_c = \begin{bmatrix} C_c & C_{nc} \end{bmatrix}, \end{aligned} \tag{14}$$

where the subsystem given by the matrices E_c, A_c, B_c, C_c contains the *controllable subsystem* of the original system, i.e., the triple (E_c, A_c, B_c) is behaviorally controllable. If the subpencil $sE_{nc} - A_{nc}$ corresponding to the uncontrollable part of the system has no finite eigenvalues with nonnegative real part, then the system is behaviorally stabilizable, otherwise it is not.

Similarly, one can determine an *observability decomposition*

$$\begin{aligned} Q_{\rm o}^T E Z_{\rm o} &= \begin{bmatrix} E_{\rm o} & 0 \\ * & E_{\rm no} \end{bmatrix},\ Q_{\rm o}^T A Z_{\rm o} = \begin{bmatrix} A_{\rm o} & 0 \\ * & A_{\rm no} \end{bmatrix}, \\ Q_{\rm o}^T B &= \begin{bmatrix} B_{\rm o} \\ B_{\rm no} \end{bmatrix}, \qquad C Z_{\rm o} = \begin{bmatrix} C_{\rm o} & 0 \end{bmatrix}, \end{aligned} \tag{15}$$

where $Q_{\rm o}, Z_{\rm o}$ are orthogonal matrices and the subsystem given by the matrices $E_{\rm o}, A_{\rm o}, B_{\rm o}, C_{\rm o}$ contains the *observable subsystem* of the original system, i.e., the system $(E_{\rm o}, A_{\rm o}, C_{\rm o})$ is behaviorally observable. If the subpencil $sE_{\rm no} - A_{\rm no}$ corresponding to the unobservable part of the system has no finite eigenvalues with nonnegative real part, then the system is behaviorally detectable, otherwise it is not. Methods for the computation of these decompositions are described in [114] and implemented as `TG01HD`, `TG01ID` in the SLICOT library[3].

For some applications, in particular those where the influence of the inputs to the outputs is crucial, it is not suitable to analyze the descriptor system in the time domain, i.e., in the form (2). Instead, one turns to the frequency domain. For this, assume that the system is square and that the pencil $sE - A$ is regular. Then we can apply the Laplace transformation to the functions $x(\cdot)$, $u(\cdot)$, and $y(\cdot)$ and under the assumption that $Ex(0) = 0$ we obtain the *transfer function*

$$G(s) := C(sE - A)^{-1}B + D \in \mathbb{R}(s)^{p,m}, \tag{16}$$

that directly maps the Laplace transformed inputs to the Laplace transformed outputs [50]. These transfer functions are typically associated to certain function spaces. Consider the normed spaces $\mathscr{RH}_\infty^{p,m}$ and $\mathscr{RL}_\infty^{p,m}$ of all *real-rational* $\mathbb{C}^{p,m}$-valued functions that are analytic and bounded in the open right half-plane $\mathbb{C}^+$; and bounded on the imaginary axis $\mathrm{i}\mathbb{R}$, respectively. Obviously, the inclusion $\mathscr{RH}_\infty^{p,m} \subset \mathscr{RL}_\infty^{p,m}$ holds. For $G \in \mathscr{RL}_\infty^{p,m}$, the $\mathscr{L}_\infty$-norm is given by

$$\|G\|_{\mathscr{L}_\infty} = \sup_{\omega \in \mathbb{R}} \sigma_{\max}(G(\mathrm{i}\omega)),$$

where $\sigma_{\max}(\cdot)$ denotes the maximum singular value. For $G \in \mathscr{RH}_\infty^{p,m}$, the $\mathscr{L}_\infty$-norm is equal to the $\mathscr{H}_\infty$-norm. These norms play an important role in many applications, in particular as robustness measures in robust control. Details on this will be pointed out in Sections 7 and 8.

5 Even Matrix Pencils

After briefly introducing the basic concepts, some of the system theoretic properties and numerical methods to check these properties, we now turn to several important

[3] `http://slicot.org/`

applications in control theory. As we will see later in the forthcoming sections, these are based on generalized eigenvalue problems for *even* matrix pencils. A matrix pencil $sN - M \in \mathbb{R}[s]^{n,n}$ is called even, if $N^T = -N$ and $M^T = M$. Besides the applications presented in this paper, even matrix pencils also play a role in linearized models that occur in the vibration analysis of buildings, machines, and vehicles [27, 61, 82, 87, 93, 111].

If the dimension of an even matrix pencil is even, i.e., $n = 2m$, then it is closely related to so-called *skew-Hamiltonian/Hamiltonian* matrix pencils [12, 90, 91, 94]. A matrix pencil $sS - H \in \mathbb{R}[s]^{2m,2m}$ is called *skew-Hamiltonian/Hamiltonian* if $\mathscr{J}_m(sS - H)$ is even, where

$$\mathscr{J}_m = \begin{bmatrix} 0 & I_m \\ -I_m & 0 \end{bmatrix},$$

that means S is skew-Hamiltonian (i.e., $(\mathscr{J}_m S)^T = -\mathscr{J}_m S$) and H is Hamiltonian (i.e., $(\mathscr{J}_m H)^T = \mathscr{J}_m H$).

Since even pencils are so closely related to skew-Hamiltonian/Hamiltonian pencils, it is easy to show that they exhibit the Hamiltonian spectral symmetry, i.e., if λ is a finite eigenvalue of $sN - M$, then $-\overline{\lambda}$ is an eigenvalue as well. This means that nonreal and nonimaginary finite eigenvalues of an even pencil appear in quadruples $(\lambda, -\lambda, \overline{\lambda}, -\overline{\lambda})$ while for purely real or purely imaginary eigenvalues they form pairs $(\lambda, -\lambda)$, $(\lambda, \overline{\lambda})$ on the real or imaginary axis, respectively. The only exceptions are the eigenvalues 0 and ∞. Furthermore, it is also well-known that even pencils possess a structured Kronecker canonical form [110] as well as a corresponding staircase form under orthogonal congruence transformations [37, 43]. We briefly recall these forms within the next subsection. A structured Smith form is available as well [88]. The staircase form allows us to filter out a regular even pencil which has Kronecker blocks at ∞ of size at most one for which we can apply structure-preserving methods for skew-Hamiltonian/Hamiltonian eigenvalue problems. These are discussed in Subsection 5.2.

5.1 Structured Condensed Forms

Even pencils have a special Kronecker canonical form under congruence transformations which preserve the even structure, see [110]. This canonical form is described in the following theorem. Besides the usual invariants occurring in the Kronecker canonical form, the even Kronecker form has further invariants associated to each purely imaginary eigenvalue, called *sign-characteristics*. These arise due to the fact that congruence transformations preserve inertia.

Theorem 5. *If $sN - M \in \mathbb{R}[s]^{n,n}$ with $N = -N^T$ and $M = M^T$, then there exists a nonsingular matrix $X \in \mathbb{R}^{n,n}$ such that*

$$X^T(sN - M)X = \operatorname{diag}(\mathscr{K}_{\mathscr{S}}(s), \mathscr{K}_{\mathscr{I}}(s), \mathscr{K}_{\mathscr{Z}}(s), \mathscr{K}_{\mathscr{F}}(s)),$$

where

$$\begin{aligned}
\mathscr{K}_{\mathscr{S}}(s) &= \operatorname{diag}\left(\mathscr{S}_{\xi_1}(s),\dots,\mathscr{S}_{\xi_k}(s)\right),\\
\mathscr{K}_{\mathscr{I}}(s) &= \operatorname{diag}\left(\mathscr{I}_{2\varepsilon_1+1}(s),\dots,\mathscr{I}_{2\varepsilon_l+1}(s),\mathscr{I}_{2\delta_1}(s),\dots,\mathscr{I}_{2\delta_m}(s)\right),\\
\mathscr{K}_{\mathscr{Z}}(s) &= \operatorname{diag}\left(\mathscr{Z}_{2\sigma_1+1}(s),\dots,\mathscr{Z}_{2\sigma_r+1}(s),\mathscr{Z}_{2\rho_1}(s),\dots,\mathscr{Z}_{2\rho_t}(s)\right),\\
\mathscr{K}_{\mathscr{F}}(s) &= \operatorname{diag}\left(\mathscr{R}_{\phi_1}(s),\dots,\mathscr{R}_{\phi_u}(s),\mathscr{C}_{\psi_1}(s),\dots,\mathscr{C}_{\psi_v}(s)\right)
\end{aligned}$$

and the blocks have the following properties:

(i) each $\mathscr{S}_{\xi_j}(s)$ is a $(2\xi_j+1)\times(2\xi_j+1)$ block ($\xi_j\in\mathbb{N}_0$) that combines a right singular block and a left singular block, both of minimal index ξ_j. It has the form

$$s\left[\begin{array}{ccc|ccc} & & & & 1 & 0\\ & & & & \iddots & \iddots\\ & & & 1 & 0 & \\ \hline & & -1 & & & \\ & \iddots & 0 & & & \\ -1 & \iddots & & & & \\ 0 & & & & & \end{array}\right] - \left[\begin{array}{ccc|ccc} & & & & 0 & 1\\ & & & & \iddots & \iddots\\ & & & 0 & 1 & \\ \hline & & 0 & & & \\ & \iddots & 1 & & & \\ 0 & \iddots & & & & \\ 1 & & & & & \end{array}\right];$$

(ii) each $\mathscr{I}_{2\varepsilon_j+1}(s)$ is a $(2\varepsilon_j+1)\times(2\varepsilon_j+1)$ block ($\varepsilon_j\in\mathbb{N}_0$) that contains a single block corresponding to the eigenvalue $\lambda=\infty$ of size $2\varepsilon_j+1$. It has the form

$$s\left[\begin{array}{ccc|ccc} & & & & 1 & 0\\ & & & & \iddots & \iddots\\ & & & 1 & 0 & \\ \hline & & -1 & 0 & & \\ & \iddots & 0 & & & \\ -1 & \iddots & & & & \\ 0 & & & & & \end{array}\right] - \left[\begin{array}{ccc|ccc} & & & & 0 & 1\\ & & & & \iddots & \iddots\\ & & & 0 & 1 & \\ \hline & & 0 & \gamma & & \\ & \iddots & 1 & & & \\ 0 & \iddots & & & & \\ 1 & & & & & \end{array}\right],$$

where $\gamma\in\{1,-1\}$ is the sign-characteristic of the block;

(iii) each $\mathscr{I}_{2\delta_j}(s)$ is a $4\delta_j\times 4\delta_j$ block ($\delta_j\in\mathbb{N}$) that combines two $2\delta_j\times 2\delta_j$ blocks associated to $\lambda=\infty$. It has the form

$$s\left[\begin{array}{ccc|ccc} & & & & 1 & 0\\ & & & & \iddots & \iddots\\ & & & 1 & \iddots & \\ & & & 0 & & \\ \hline & -1 & 0 & & & \\ \iddots & \iddots & & & & \\ -1 & \iddots & & & & \\ 0 & & & & & \end{array}\right] - \left[\begin{array}{ccc|ccc} & & & & & 1\\ & & & & \iddots & \\ & & & 1 & & \\ \hline & & 1 & & & \\ & \iddots & & & & \\ 1 & & & & & \end{array}\right];$$

(iv) each $\mathscr{Z}_{2\sigma_j+1}(s)$ is a $(4\sigma_j+2)\times(4\sigma_j+2)$ block ($\sigma_j\in\mathbb{N}_0$) that combines two $(2\sigma_j+1)\times(2\sigma_j+1)$ Jordan blocks corresponding to the eigenvalue $\lambda=0$. It has the form

$$s\left[\begin{array}{ccc|ccc} & & & & & 1 \\ & & & & \iddots & \\ & & & 1 & & \\ \hline & & -1 & & & \\ & \iddots & & & & \\ -1 & & & & & \end{array}\right] - \left[\begin{array}{cccc|cccc} & & & & & & 1 & 0 \\ & & & & & \iddots & \iddots & \\ & & & & 1 & \iddots & & \\ & & & & 0 & & & \\ \hline & & 1 & 0 & & & & \\ & \iddots & \iddots & & & & & \\ 1 & \iddots & & & & & & \\ 0 & & & & & & & \end{array}\right];$$

(v) each $\mathscr{Z}_{2\rho_j}(s)$ *is a* $2\rho_j \times 2\rho_j$ *block (*$\rho_j \in \mathbb{N}$*) that contains a single Jordan block corresponding to the eigenvalue* $\lambda = 0$*. It has the form*

$$s\left[\begin{array}{ccc|ccc} & & & & & 1 \\ & & & & \iddots & \\ & & & 1 & & \\ \hline & & -1 & & & \\ & \iddots & & & & \\ -1 & & & & & \end{array}\right] - \left[\begin{array}{cccc|cccc} & & & & & & 1 & 0 \\ & & & & & \iddots & \iddots & \\ & & & & 1 & \iddots & & \\ & & & \gamma & 0 & & & \\ \hline & & 1 & 0 & & & & \\ & \iddots & \iddots & & & & & \\ 1 & \iddots & & & & & & \\ 0 & & & & & & & \end{array}\right],$$

where $\gamma \in \{1, -1\}$ *is the sign-characteristic of this block;*

(vi) each $\mathscr{R}_{\phi_j}(s)$ *is a* $2\phi_j \times 2\phi_j$ *block (*$\phi_j \in \mathbb{N}$*) that combines two* $\phi_j \times \phi_j$ *Jordan blocks corresponding to nonzero real eigenvalues* a_j *and* $-a_j$*. It has the form*

$$s\left[\begin{array}{ccc|ccc} & & & & & 1 \\ & & & & \iddots & \\ & & & 1 & & \\ \hline & & -1 & & & \\ & \iddots & & & & \\ -1 & & & & & \end{array}\right] - \left[\begin{array}{cccc|cccc} & & & & & & 1 & a_j \\ & & & & & \iddots & \iddots & \\ & & & & 1 & \iddots & & \\ & & & & a_j & & & \\ \hline & & 1 & a_j & & & & \\ & \iddots & \iddots & & & & & \\ 1 & \iddots & & & & & & \\ a_j & & & & & & & \end{array}\right];$$

(vii) the entries $\mathscr{C}_{\psi_j}(s)$ *take two slightly different forms:*

(a) one possibility is that $\mathscr{C}_{\psi_j}(s)$ *is a* $2\psi_j \times 2\psi_j$ *block (*$\psi_j \in \mathbb{N}$*) combining two* $\psi_j \times \psi_j$ *Jordan blocks corresponding to purely imaginary eigenvalues* $\mathrm{i}b_j, -\mathrm{i}b_j$ *(*$b_j > 0$*). In this case it has the form*

$$s\left[\begin{array}{ccc|ccc} & & & & & 1 \\ & & & & \iddots & \\ & & & 1 & & \\ \hline & & -1 & & & \\ & \iddots & & & & \\ -1 & & & & & \end{array}\right] - \gamma \left[\begin{array}{cccc|cccc} & & 1 & b_j & & & & \\ & \iddots & \iddots & & & & & \\ 1 & \iddots & & & & & & \\ b_j & & & & & & & \\ \hline & & & & & & 1 & b_j \\ & & & & & \iddots & \iddots & \\ & & & & 1 & \iddots & & \\ & & & & b_j & & & \end{array}\right],$$

where $\gamma \in \{1, -1\}$ *is the sign-characteristic;*

(b) the other possibility is that $\mathscr{C}_{\psi_j}(s)$ is a $4\psi_j \times 4\psi_j$ block ($\psi_j \in \mathbb{N}$) combining $\psi_j \times \psi_j$ Jordan blocks for each of the complex eigenvalues $a_j + \mathrm{i}b_j, a_j - \mathrm{i}b_j, -a_j + \mathrm{i}b_j, -a_j - \mathrm{i}b_j$ (with $a_j \neq 0$ and $b_j \neq 0$). In this case it has the form

$$
s\left[\begin{array}{ccc|ccc}
 & & & & & \Omega \\
 & & & & \iddots & \\
 & & & \Omega & & \\
\hline
 & & -\Omega & & & \\
 & \iddots & & & & \\
-\Omega & & & & &
\end{array}\right]
-
\left[\begin{array}{cccc|cccc}
 & & & & & & \Omega & \Lambda_j \\
 & & & & & \iddots & \iddots & \\
 & & & & \Omega & \iddots & & \\
 & & & & \Lambda_j & & & \\
\hline
 & & \Omega & \Lambda_j & & & & \\
 & \iddots & \iddots & & & & & \\
\Omega & \iddots & & & & & & \\
\Lambda_j & & & & & & &
\end{array}\right]
$$

with $\Omega = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ *and* $\Lambda_j = \begin{bmatrix} -b_j & a_j \\ a_j & b_j \end{bmatrix}$.

This structured Kronecker canonical form is unique up to permutation of the blocks, i.e., the kind, size, and number of the blocks as well as the sign-characteristics are invariants of the pencil $sN - M$ under congruence transformations.

An even pencil is called *regular* if and only if no blocks of type (i) occur in the even Kronecker form. The *(Kronecker) index* of the pencil is the size of the largest block of type (ii) and (iii) in the even Kronecker form, thus a regular pencil is of *index at most one* if and only if there are no blocks of type (iii) and the blocks of type (ii) are of size at most one. In some of the applications discussed below, it will be necessary to detect whether an even matrix pencil is regular and of index at most one and whether there exist *finite eigenvalues* with real part 0. In other applications the computation of the stable deflating subspace, i.e., the subspace spanned by the eigenvectors and generalized eigenvectors, associated to all eigenvalues in the open left-half plane is the goal. The structured Kronecker form reveals this information but usually it cannot be computed numerically, because arbitrary small perturbations may change the structural information and since the transformation matrices may be unbounded.

A computationally attractive alternative is the staircase form under orthogonal transformations. It allows us to check regularity and to determine the index within the usual limitations of rank computations in finite precision arithmetic, see [43] for a detailed discussion of the difficulties. This is an essential preparation for the computation of the eigenvalues and deflating subspaces.

Theorem 6. *[43] For every even pencil $sN - M \in \mathbb{R}[s]^{n,n}$, there exists a real orthogonal matrix $U \in \mathbb{R}^{n,n}$ such that*

$$
U^T N U =
\begin{array}{c}
s_1\\ \vdots\\ \vdots\\ s_w\\ l\\ q_w\\ \vdots\\ q_2\\ q_1
\end{array}
\left[\begin{array}{cccc|c|cccc}
N_{1,1} & \cdots & \cdots & N_{1,w} & N_{1,w+1} & N_{1,w+2} & \cdots & N_{1,2w} & 0\\
\vdots & \ddots & & \vdots & \vdots & \vdots & \iddots & \iddots & \\
\vdots & & \ddots & \vdots & \vdots & N_{w-1,w+2} & \iddots & & \\
-N_{1,w}^T & \cdots & \cdots & N_{w,w} & N_{w,w+1} & 0 & & & \\
\hline
-N_{1,w+1}^T & \cdots & \cdots & -N_{w,w+1}^T & N_{w+1,w+1} & & & & \\
\hline
-N_{1,w+2}^T & \cdots & -N_{w-1,w+2}^T & 0 & & & & & \\
\vdots & \iddots & \iddots & & & & & & \\
-N_{1,2w}^T & \iddots & & & & & & & \\
0 & & & & & & & &
\end{array}\right]
\tag{17}
$$

$$
U^T M U =
\begin{array}{c}
s_1\\ \vdots\\ \vdots\\ s_w\\ l\\ q_w\\ \vdots\\ \vdots\\ q_1
\end{array}
\left[\begin{array}{cccc|c|cccc}
M_{1,1} & \cdots & \cdots & M_{1,w} & M_{1,w+1} & M_{1,w+2} & \cdots & \cdots & M_{1,2w+1}\\
\vdots & \ddots & & \vdots & \vdots & \vdots & \iddots & & \\
\vdots & & \ddots & \vdots & \vdots & \vdots & \iddots & & \\
M_{1,w}^T & \cdots & \cdots & M_{w,w} & M_{w,w+1} & M_{w,w+2} & & & \\
\hline
M_{1,w+1}^T & \cdots & \cdots & M_{w,w+1}^T & M_{w+1,w+1} & & & & \\
\hline
M_{1,w+2}^T & \cdots & \cdots & M_{w,w+2}^T & & & & & \\
\vdots & \iddots & & & & & & & \\
\vdots & \iddots & & & & & & & \\
M_{1,2w+1}^T & & & & & & & &
\end{array}\right],
$$

where $q_1 \geq s_1 \geq q_2 \geq s_2 \geq \ldots \geq q_w \geq s_w$, $l = r_{w+1} + a_{w+1}$, *and for* $i = 1, \ldots, w$, *we have* $N_{i,i} = -N_{i,i}^T$, $M_{i,i} = M_{i,i}^T$. *Furthermore,*

$$
\begin{aligned}
& N_{j,2w+1-j} \in \mathbb{R}^{s_j, q_{j+1}}, \quad 1 \leq j \leq w-1,\\
& N_{w+1,w+1} = \begin{bmatrix} \Delta & 0\\ 0 & 0 \end{bmatrix}, \quad \Delta = -\Delta^T \in \mathbb{R}^{r_{w+1}, r_{w+1}},\\
& M_{j,2w+2-j} = \begin{bmatrix} \Gamma_j & 0 \end{bmatrix} \in \mathbb{R}^{s_j, q_j}, \quad \Gamma_j \in \mathbb{R}^{s_j, s_j}, \quad 1 \leq j \leq w,\\
& M_{w+1,w+1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Sigma_{11} \in \mathbb{R}^{r_{w+1}, r_{w+1}}, \quad \Sigma_{22} \in \mathbb{R}^{a_{w+1}, a_{w+1}},\\
& M_{w+1,w+1} = M_{w+1,w+1}^T,
\end{aligned}
$$

and the blocks Σ_{22} *and* Δ *and* Γ_j, $j = 1, \ldots, w$ *(if they occur) are nonsingular.*

Production code implementations for the computation of these and other related structured staircase forms via a sequence of singular value decompositions have been presented in [37]. Since the staircase form uses congruence transformations, all the invariants of the even Kronecker canonical form are preserved, as discussed in the following corollary.

Corollary 2. *[43] Consider an even pencil and its staircase form* (17).

(i) The pencil is regular if and only if $s_i = q_i$ *for* $i = 1, \ldots, w$.
(ii) The pencil is regular and of index at most one if and only if $w = 0$.
(iii) The block $(N_{w+1,w+1}, M_{w+1,w+1})$ *contains the regular part associated to finite eigenvalues and blocks associated to the infinite eigenvalues of index at most one.*
(iv) The finite eigenvalues of the pencil are the eigenvalues of

$$s\Delta - \left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

(v) For every purely imaginary eigenvalue $\lambda_0 \in \mathrm{i}\mathbb{R}$*, satisfying*

$$\left(\lambda_0\Delta - \left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)\right)x_0 = 0$$

for $x_0 \in \mathbb{C}^{r_{w+1}} \setminus \{0\}$*, the sign-characteristic of* λ_0 *is given by the sign of the real number* $\mathrm{i}x_0^H \Delta x_0$.

Thus, once the staircase form has been computed, for the computation of eigenvalues and invariant subspaces one can restrict the methods to the middle regular index one block of the staircase form. We recall the appropriate methods in the next subsection.

5.2 Computing Eigenvalues and Deflating Subspaces of Regular Index One Even Pencils

For the computation of eigenvalues, eigenvectors, and deflating subspaces associated to finite eigenvalues of even pencils, we need eigenvalue methods for regular even pencils of index at most one that can be applied to the middle block in the staircase form (17)

$$sN_{w+1,w+1} - M_{w+1,w+1} = s\begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}. \tag{18}$$

In the special case that this even pencil has no infinite eigenvalues, i.e., if the second block row and column are not occurring, and hence $a_{w+1} = 0$, then we have a pencil $s\Delta - \Sigma_{11}$, where Δ is nonsingular (and thus of even dimension). In this case one can perform a Cholesky-like decomposition, see [9, 39] of the form $\Delta = \mathscr{U}^T \mathscr{J}_{r_{w+1}/2} \mathscr{U}$ with an upper-triangular matrix $\mathscr{U}$. If the factorization is well-conditioned and if $\mathscr{U}$ is well-conditioned with respect to inversion, then one can turn this even eigenvalue problem into an eigenvalue problem for the Hamiltonian matrix $\mathscr{H} = \mathscr{J}_{r_{w+1}/2}^T \mathscr{U}^{-T} \Sigma_{11} \mathscr{U}^{-1}$ and apply the structure-preserving methods for Hamiltonian eigenvalue problems [48, 92]. If, however, the computation and inversion of $\mathscr{U}$ is ill-conditioned or if the pencil $sN_{w+1,w+1} - M_{w+1,w+1}$ has infinite eigenvalues, then it is better to proceed with the pencil formulation.

Recently, in [95], a new structure-preserving method to deflate the infinite eigenvalues via an orthogonal congruence transformation has been derived for the pencil case. Consider the even pencil $sN_{w+1,w+1} - M_{w+1,w+1}$ as in (18). This procedure works by using a rank-revealing QR-decomposition or a singular value decomposition to determine an orthogonal matrix V_{w+1} such that

$$\begin{bmatrix} \Sigma_{21} & \Sigma_{22} \end{bmatrix} V_{w+1} = \begin{bmatrix} 0 & \widehat{\Sigma}_{22} \end{bmatrix},$$

with nonsingular $\widehat{\Sigma}_{22}$. By forming

$$V_{w+1}^T \left(sM_{w+1,w+1} - N_{w+1,w+1} \right) V_{w+1} = s \begin{bmatrix} \widetilde{\Delta}_{11} & \widetilde{\Delta}_{12} \\ -\widetilde{\Delta}_{12}^T & \widetilde{\Delta}_{22} \end{bmatrix} - \begin{bmatrix} \widetilde{\Sigma}_{11} & \widetilde{\Sigma}_{12} \\ \widetilde{\Sigma}_{12}^T & \widetilde{\Sigma}_{22} \end{bmatrix},$$

partitioned accordingly, it has been shown in [95] that the eigenvalues of the even pencil $s\widetilde{\Delta}_{11} - \widetilde{\Sigma}_{11}$ are exactly the finite eigenvalues of $sN_{w+1,w+1} - M_{w+1,w+1}$ and also the eigenvectors and invariant subspaces can be easily recovered.

The detailed error analysis of this procedure in [95] analyzes when this deflation procedure is reliable and when it is more reasonable to proceed with the index one pencil formulation. In the following we assume that this decision has been made. and that we either proceed with an even pencil with only finite eigenvalues, which means that the dimension is even or with an index one even pencil. Since for skew-Hamiltonian/Hamiltonian pencils eigenvalue methods are well established and have been professionally implemented [12, 14, 20, 21, 59, 85, 91, 94], we just adapt these for the even pencil case. However, we suggest that in the long run these methods should be implemented to directly work for the even case, since it may happen that the middle block $sN_{w+1,w+1} - M_{w+1,w+1}$ in the even staircase form (i.e., the regular index one part) is of odd dimension. To apply the methods for skew-Hamiltonian/Hamiltonian pencils to this middle block in the odd-dimensional case, we consider an embedded $2k \times 2k$ pencil

$$sS - H = \mathscr{J}_k \left(s \begin{bmatrix} N_{w+1,w+1} & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} M_{w+1,w+1} & 0 \\ 0 & 1 \end{bmatrix} \right)$$

which has an additional eigenvalue ∞, right eigenvector e_{2k} (the $2k$-th unit vector) and left eigenvector $\mathscr{J}_k^T e_{2k}$, which are orthogonal to all the other eigenvectors. So in the following, whenever an eigenvalue method for regular even pencils of index at most one is needed, then we can perform this embedding and employ a solver for the skew-Hamiltonian/Hamiltonian pencil $sS - H \in \mathbb{R}[s]^{2k,2k}$.

For the computation of the eigenvalues and deflating subspaces of skew-Hamiltonian/Hamiltonian pencils we make use of $\mathscr{J}_k$-congruence transformations of the form

$$s\widetilde{S} - \widetilde{H} := \mathscr{J}_k \mathscr{Q}^T \mathscr{J}_k^T (sS - H) \mathscr{Q}$$

with nonsingular matrices $\mathscr{Q}$, which preserve the skew-Hamiltonian/Hamiltonian structure. In general we would hope that we can compute an *orthogonal* matrix $\mathscr{Q}$ such that

$$\mathscr{J}_k \mathscr{Q}^T \mathscr{J}_k^T (sS - H) \mathscr{Q} = s \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{11}^T \end{bmatrix} - \begin{bmatrix} H_{11} & H_{12} \\ 0 & -H_{11}^T \end{bmatrix}$$

is in skew-Hamiltonian/Hamiltonian Schur form, i.e., the subpencil $sS_{11} - H_{11}$ is in generalized Schur form [62]. Unfortunately, not every skew-Hamiltonian/Hamiltonian pencil has this structured Schur form, since certain simple purely imaginary eigenvalues, or multiple purely imaginary eigenvalues with even algebraic multiplicity, but uniform sign-characteristic, cannot be represented in this structure. An embedding into a pencil of the double size solves this issue as follows.

We introduce the orthogonal matrices

$$\mathscr{Y} = \frac{\sqrt{2}}{2}\begin{bmatrix} I_{2k} & I_{2k} \\ -I_{2k} & I_{2k} \end{bmatrix}, \quad \mathscr{P} = \begin{bmatrix} I_k & 0 & 0 & 0 \\ 0 & 0 & I_k & 0 \\ 0 & I_k & 0 & 0 \\ 0 & 0 & 0 & I_k \end{bmatrix}, \quad \mathscr{X} = \mathscr{Y}\mathscr{P},$$

and define the matrix pencil

$$s\mathscr{B}_S - \mathscr{B}_H := \mathscr{X}^T \left(s\begin{bmatrix} S & 0 \\ 0 & S \end{bmatrix} - \begin{bmatrix} H & 0 \\ 0 & -H \end{bmatrix} \right) \mathscr{X} \in \mathbb{R}[s]^{4k,4k},$$

which is still regular and of index at most one.

It can be easily observed, that $s\mathscr{B}_S - \mathscr{B}_H$ is again real skew-Hamiltonian/Hamiltonian with the same eigenvalues (now with double algebraic, geometric and partial multiplicities, but with appropriate mixed sign-characteristic) as the pencil $sS - H$. To compute the eigenvalues of $s\mathscr{B}_S - \mathscr{B}_H$ one uses the generalized symplectic URV decomposition of $sS - H$, see [16, 17], i.e., there exist orthogonal matrices $Q_1, Q_2 \in \mathbb{R}^{4k,4k}$ such that

$$\begin{aligned} Q_1^T S \mathscr{J}_k Q_1 \mathscr{J}_k^T &= \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{11}^T \end{bmatrix}, \\ \mathscr{J}_k Q_2^T \mathscr{J}_k^T S Q_2 &= \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{11}^T \end{bmatrix}, \\ Q_1^T H Q_2 &= \begin{bmatrix} H_{11} & H_{12} \\ 0 & H_{22} \end{bmatrix}, \end{aligned} \tag{19}$$

where S_{12} and T_{12} are skew-symmetric and the formal matrix product $S_{11}^{-1} H_{11} T_{11}^{-1} H_{22}^T$ is in periodic Schur form [31, 67, 71].

Applying this result to to the specially structured pencil $s\mathscr{B}_S - \mathscr{B}_H$, we can compute an orthogonal matrix $\mathscr{Q}$ such that

$$\mathscr{J}_{2k}\mathscr{Q}^T \mathscr{J}_{2k}^T (s\mathscr{B}_S - \mathscr{B}_H)\mathscr{Q} = s\left[\begin{array}{cc|cc} S_{11} & 0 & S_{12} & 0 \\ 0 & T_{11} & 0 & T_{12} \\ \hline 0 & 0 & S_{11}^T & 0 \\ 0 & 0 & 0 & T_{11}^T \end{array}\right] - \left[\begin{array}{cc|cc} 0 & H_{11} & 0 & H_{12} \\ -H_{22}^T & 0 & H_{12}^T & 0 \\ \hline 0 & 0 & 0 & H_{22} \\ 0 & 0 & -H_{11}^T & 0 \end{array}\right]$$

with $\mathscr{Q} = \mathscr{P}^T \begin{bmatrix} \mathscr{J}_k Q_1 \mathscr{J}_k^T & 0 \\ 0 & Q_2 \end{bmatrix} \mathscr{P}$.

Note, that for these computations we never explicitly construct the embedded pencils. It is sufficient to compute the necessary parts of the matrices in (19).

The eigenvalues of $sS-H$ can then be computed as $\pm \mathrm{i}\sqrt{\lambda_j}$ where the λ_j, $j = 1,\ldots,k$, are the eigenvalues of the *formal matrix product* $S_{11}^{-1}H_{11}T_{11}^{-1}H_{22}^T$ which can be determined by evaluating the entries on the 1×1 and 2×2 diagonal blocks of the matrices only. In particular, the finite, purely imaginary eigenvalues correspond to the 1×1 diagonal blocks of this matrix product. Provided that the pairwise distance of the simple, finite, purely imaginary eigenvalues with mixed sign-characteristics is sufficiently large, they can be computed in a robust way without any error in the real part. This property of the algorithm plays an essential role for many of the applications that we will consider in subsequent sections.

If also the deflating subspaces associated to certain eigenvalues are desired, then one computes the real skew-Hamiltonian/Hamiltonian Schur form of the embedded pencil where the eigenvalues are reordered in such a way such that the desired ones appear in the leading principal subpencil. By determining also the sign-characteristics of the purely imaginary eigenvalues, one can (at least in exact arithmetics) check whether a Hamiltonian Schur form exists. It should be noted that if the problem has computed eigenvalues very close to the imaginary axis (within a strip of width $\sqrt{\mathbf{u}}$), then these may be the result of a perturbation of size $\mathbf{u}$ of a double purely imaginary eigenvalue with mixed sign-characteristic. This does not prevent the existence of a Hamiltonian Schur form, however, in the neighborhood of this problem there is then a problem with two simple purely imaginary eigenvalues of mixed sign-characteristic, but with no Hamiltonian Schur form, see [1].

The structure-preserving Algorithm 1 was introduced in [11] and has been updated and improved in [85]. It is available as the SLICOT subroutine `MB04BD`. While the classic unstructured QZ algorithm applied to the $2k \times 2k$ pencil would require $528k^3$ flops or $240k^3$ flops for the eigenvalues [62], this algorithm needs roughly 60% of this, see [11]. Note that there are many more structure-exploiting algorithms for Hamiltonian and even eigenvalues problems in the dense but also in the sparse setting, see, e.g., [14, 15, 48, 72, 84, 91, 92, 107].

In later sections, when discussing applications for even pencils, we will always use the algorithm presented here, since the preservation of the spectral symmetry is essential for the robustness of the methods. For illustration, Figure 2 from [21] plots the computed eigenvalues of a skew-Hamiltonian/Hamiltonian pencil that results from the stability analysis of a linearized gyroscopic system. A necessary condition for stability is that all eigenvalues are on the imaginary axis. The figure shows that the structure-preserving algorithm captures this behavior whereas the standard QZ algorithm fails to do so and therefore, does not allow us to make any statement about stability.

6 Linear-Quadratic Optimal Control

In this section we consider the linear quadratic optimal control problem of minimizing

$$\mathscr{J}(x(\cdot),u(\cdot)) = \frac{1}{2}\int_0^\infty \left(x(t)^T Q x(t) + 2x(t)^T S u(t) + u(t)^T R u(t) \right) \mathrm{d}t \qquad (20)$$

Algorithm 1 Computation of stable eigenvalues and associated stable deflating subspaces of a real skew-Hamiltonian/Hamiltonian pencil

Input: A regular real skew-Hamiltonian/Hamiltonian pencil $sS - H \in \mathbb{R}[s]^{2k,2k}$ of index at most one.

Output: The eigenvalues of $sS - H$ and a matrix P_V^- whose columns form an orthogonal basis of the r-dimensional deflating subspace associated to the eigenvalues in the open left half plane.

1: Compute the generalized symplectic URV decomposition [85, Algorithm 2] of the pencil $sS - H$ and determine orthogonal matrices Q_1, Q_2 such that

$$Q_1^T S \mathscr{J}_k Q_1 \mathscr{J}_k^T = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{11}^T \end{bmatrix},$$
$$\mathscr{J}_k Q_2^T \mathscr{J}_k^T S Q_2 = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{11}^T \end{bmatrix},$$
$$Q_1^T H Q_2 = \begin{bmatrix} H_{11} & H_{12} \\ 0 & H_{22} \end{bmatrix},$$

where the generalized matrix product $S_{11}^{-1} H_{11} T_{11}^{-1} H_{22}^T$ is in periodic Schur form.

2: Apply [85, Algorithm 3] to determine orthogonal matrices Q_3 and Q_4 such that

$$s\mathscr{S}_{11} - \mathscr{H}_{11} := Q_4^T \left(s \begin{bmatrix} S_{11} & 0 \\ 0 & T_{11} \end{bmatrix} - \begin{bmatrix} 0 & H_{11} \\ -H_{22}^T & 0 \end{bmatrix} \right) Q_3$$

is in generalized Schur form. Update

$$\mathscr{S}_{12} := Q_4^T \begin{bmatrix} S_{12} & 0 \\ 0 & T_{12} \end{bmatrix} Q_4, \quad \mathscr{H}_{12} := Q_4^T \begin{bmatrix} 0 & H_{12} \\ H_{12}^T & 0 \end{bmatrix} Q_4$$

and set

$$s\mathscr{B}_S - \mathscr{B}_H := s \begin{bmatrix} \mathscr{S}_{11} & \mathscr{S}_{12} \\ 0 & \mathscr{S}_{11}^T \end{bmatrix} - \begin{bmatrix} \mathscr{H}_{11} & \mathscr{H}_{12} \\ 0 & -\mathscr{H}_{11}^T \end{bmatrix}.$$

3: Apply the eigenvalue reordering method [85, Algorithm 4] to the pencil $s\mathscr{B}_S - \mathscr{B}_H$ to determine an orthogonal matrix $\widehat{Q}$ such that

$$s\widetilde{\mathscr{B}}_S - \widetilde{\mathscr{B}}_H := \mathscr{J}_{2k} \widehat{Q}^T \mathscr{J}_{2k}^T \left(s\mathscr{B}_S - \mathscr{B}_H \right) \widehat{Q}$$

is still in structured Schur form, but the eigenvalues with negative real part of $s\widetilde{\mathscr{B}}_S - \widetilde{\mathscr{B}}_H$ are contained in the leading $2r \times 2r$ principal subpencil of $s\mathscr{S}_{11} - \mathscr{H}_{11}$.

4: Set

$$V = \begin{bmatrix} I_{2k} & 0 \end{bmatrix} \left(\mathscr{Y} \begin{bmatrix} \mathscr{J}_k Q_1 \mathscr{J}_k^T & 0 \\ 0 & Q_2 \end{bmatrix} \mathscr{P} \begin{bmatrix} Q_3 & 0 \\ 0 & Q_4 \end{bmatrix} \widehat{Q} \right) \begin{bmatrix} I_{2r} \\ 0 \end{bmatrix}$$

and compute P_V^-, an orthonormal basis of range V, using any numerically stable orthogonalization scheme.

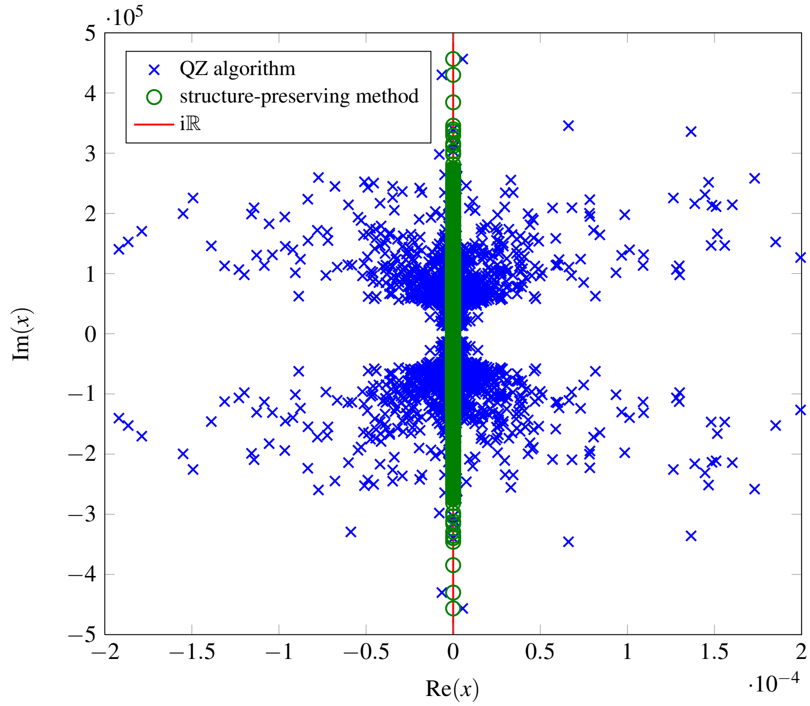

Fig. 2. Computed eigenvalues from a skew-Hamiltonian/Hamiltonian pencil with only purely imaginary eigenvalues resulting from a linearized gyroscopic system

with $Q = Q^T \in \mathbb{R}^{n,n}$, $S \in \mathbb{R}^{n,m}$, and $R = R^T \in \mathbb{R}^{m,m}$ subject to the linear descriptor system of the form (2a) with initial value $x(0) = x_0$ and the stabilization condition $\lim_{t\to\infty} x(t) = 0$. If an output equation (2b) is also given, then the cost functional is usually given as $\widetilde{\mathscr{J}}(y(\cdot),u(\cdot))$ which can then easily be transformed to the form given in (20) by inserting the output equation. This yields

$$\widetilde{\mathscr{J}}(x(\cdot),u(\cdot)) = \frac{1}{2}\int_0^\infty \left(x(t)^T \widetilde{Q} x(t) + 2x(t)^T \widetilde{S} u(t) + u(t)^T \widetilde{R} u(t) \right) \mathrm{d}t$$

with

$$\widetilde{Q} := C^T Q C, \quad \widetilde{S} := C^T Q D + C^T S, \quad \widetilde{R} := D^T Q D + D^T S + S^T D + R. \tag{21}$$

Optimal control problems for equations of this form arise in mechanical multibody systems [64, 65, 108], electrical circuits [63] and many other applications like the linearization of general nonlinear systems along stationary trajectories [46].

In order for an optimal control to exist, for the initial value one needs the condition $E^+ E x_0 = x_0$ with the Moore-Penrose inverse E^+ of E, see [76]. Further additional assumptions are needed to ensure that (20) is bounded from below. A quite common assumption in the literature is

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \geq 0.$$

In [117] it has been further shown that for square systems, strong stabilizability and the feasibility of a certain linear matrix inequality are a sufficient condition for the boundedness of (20) from below, even in the case of an indefinite cost functional.

To solve this problem in the most general situation, we replace the DAE constraint by the *strangeness-free formulation*

$$\widehat{E}\dot{x}(t) = \widehat{A}x(t) + \widehat{B}u(t), \tag{22}$$

where

$$\widehat{E} = \begin{bmatrix} \widehat{E}_1 \\ 0 \end{bmatrix}, \quad \widehat{A} = \begin{bmatrix} \widehat{A}_1 \\ \widehat{A}_2 \end{bmatrix}, \quad \widehat{B} = \begin{bmatrix} \widehat{B}_1 \\ \widehat{B}_2 \end{bmatrix},$$

with the additional property that the matrix

$$\begin{bmatrix} \widehat{E}_1 & 0 \\ \widehat{A}_2 & \widehat{B}_2 \end{bmatrix}$$

has full row rank, see also Section 3. The necessary optimality system is then given by

$$\begin{bmatrix} 0 & \widehat{E} & 0 \\ -\widehat{E}^T & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} \widehat{\lambda}(t) \\ x(t) \\ u(t) \end{bmatrix} = \begin{bmatrix} 0 & \widehat{A} & \widehat{B} \\ \widehat{A}^T & Q & S \\ \widehat{B}^T & S^T & R \end{bmatrix} \begin{bmatrix} \widehat{\lambda}(t) \\ x(t) \\ u(t) \end{bmatrix}, \tag{23}$$

with boundary conditions $x(0) = x_0$, and $\lim_{t\to\infty} \widehat{E}^T\widehat{\lambda}(t) = 0$. Solving this system will give the optimal input $u(\cdot)$, state $x(\cdot)$, and the Lagrange multiplier $\widehat{\lambda}(\cdot)$.

Instead of first computing a strangeness-free formulation and forming the optimality system (23), we can instead directly form and solve the *formal optimality system* [7, 47, 76, 77, 81] given by

$$\begin{bmatrix} 0 & E & 0 \\ -E^T & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} \lambda(t) \\ x(t) \\ u(t) \end{bmatrix} = \begin{bmatrix} 0 & A & B \\ A^T & Q & S \\ B^T & S^T & R \end{bmatrix} \begin{bmatrix} \lambda(t) \\ x(t) \\ u(t) \end{bmatrix}, \tag{24}$$

with boundary conditions $x(0) = x_0$, and $\lim_{t\to\infty} E^T\lambda(t) = 0$. One has the following relation between the true and the formal optimality system which we cite here for constant coefficient systems, for the general case of variable coefficient systems see [77].

Theorem 7. *Suppose that the formal necessary optimality system* (24) *has a solution* $\left[\lambda(\cdot)^T \; x(\cdot)^T \; u(\cdot)^T\right]^T$. *Then there exists a function* $\widehat{\lambda}(\cdot)$ *replacing the function* $\lambda(\cdot)$ *such that* $\left[\widehat{\lambda}(\cdot)^T \; x(\cdot)^T \; u(\cdot)^T\right]^T$ *solves the necessary optimality conditions* (23).

Theorem 7 shows that it is enough to solve the boundary value problem (24) in the original data, provided it is solvable. Since this is a homogeneous differential-algebraic system, the solvability of the boundary value problem depends on the consistency of the boundary conditions and the solvability of the linear system that relates initial and terminal conditions, see [5, 79, 80]. Since the boundary value problem is of the form

$$N\dot{z}(t) = Mz(t), \quad P_1 z(0) = P_1 z_0, \quad \lim_{t\to\infty} P_2 z(t) = 0,$$

with $z(\cdot) = \left[\lambda(\cdot)^T \; x(\cdot)^T \; u(\cdot)^T\right]^T$, and some matrices P_1, and P_2, the simplest way to perform these computations is to apply the congruence transformation to even staircase form

$$U^T NU\dot{\widetilde{z}}(t) = U^T MU\widetilde{z}(t), \quad P_1 U\widetilde{z}(0) = P_1 U\widetilde{z}_0, \quad \lim_{t\to\infty} P_2 U\widetilde{z}(t) = 0,$$

with $\widetilde{z}(\cdot) = U^T z(\cdot)$, and $\widetilde{z}_0 = U^T z_0$.

This allows us to check the unique solvability by checking the regularity as in Corollary 2 and the consistency of the boundary conditions, see [43] for details. By partitioning $\widetilde{z}(\cdot) = \left[\widetilde{z}_1(\cdot)^T, \ldots, \widetilde{z}_{2w+1}(\cdot)^T\right]^T$ analogous to (17), the last w blocks yield the consistency conditions $\widetilde{z}_1 \equiv 0, \ \ldots, \ \widetilde{z}_w \equiv 0$. The middle block system can be expressed as

$$N_{w+1,w+1}\dot{\widetilde{z}}_{w+1}(t) = M_{w+1,w+1}\widetilde{z}_{w+1}(t), \tag{25}$$

with appropriately transformed boundary conditions. This system is regular and has index at most one. If we make use of the semi-explicit form (18) and split

$$\widetilde{z}_{w+1}(\cdot) = \begin{bmatrix} \eta(\cdot) \\ \zeta(\cdot) \end{bmatrix},$$

then we obtain

$$\begin{bmatrix} \Delta & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\eta}(t) \\ \dot{\zeta}(t) \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \eta(t) \\ \zeta(t) \end{bmatrix}.$$

It follows that $\zeta(\cdot) = -\Sigma_{22}^{-1}\Sigma_{21}\eta(\cdot)$, which gives further consistency conditions on $\widetilde{z}_{w+1}(\cdot)$ and

$$\Delta\dot{\eta}(t) = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\eta(t).$$

Then we can perform a factorization $\Delta = \mathscr{U}^T \mathscr{J}_{r_{w+1}/2}\mathscr{U}$ with nonsingular upper triangular matrix $\mathscr{U}$ [39]. If the factorization is well-conditioned and the factor $\mathscr{U}$ is well-conditioned with respect to inversion, then we set $\mathscr{H} := \mathscr{J}_{r_{w+1}/2}^T \mathscr{U}^{-T}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\mathscr{U}^{-1}$ and $\xi(t) := \mathscr{U}\eta(t)$ to obtain the Hamiltonian boundary value problem

$$\dot{\xi}(t) = \mathscr{H}\xi(t). \tag{26}$$

with appropriate boundary conditions $\Pi_1\xi(0) = \Pi_1\xi_0$, and $\lim_{t\to\infty}\Pi_2\xi(t) = 0$. This system has the general solution $\xi(t) = \exp(\mathscr{H}t)\xi_0$ and therefore,

$$\widetilde{z}_{w+1}(t) = \begin{bmatrix} \eta(t) \\ -\Sigma_{22}^{-1}\Sigma_{21}\eta(t) \end{bmatrix} = \begin{bmatrix} \mathscr{U}^{-1}\exp(\mathscr{H}t)\,\xi_0 \\ -\Sigma_{22}^{-1}\Sigma_{21}\mathscr{U}^{-1}\exp(\mathscr{H}t)\,\xi_0 \end{bmatrix}. \tag{27}$$

It is important to note that one does not have to compute the exponential function in (27) explicitly.

With a decomposition of the boundary value problem as

$$\begin{bmatrix} \dot{\xi}_1(t) \\ \dot{\xi}_2(t) \end{bmatrix} = \begin{bmatrix} \mathscr{H}_{11} & \mathscr{H}_{12} \\ \mathscr{H}_{21} & -\mathscr{H}_{11}^T \end{bmatrix} \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \end{bmatrix}, \quad \mathscr{H}_{12} = \mathscr{H}_{12}^T, \quad \mathscr{H}_{21} = \mathscr{H}_{21}^T,$$

one rather computes the stable invariant subspace spanned by the matrix $\mathscr{V} = \begin{bmatrix} \mathscr{V}_1 \\ \mathscr{V}_2 \end{bmatrix} \in \mathbb{R}^{r_{w+1}, r_{w+1}/2}$ which satisfies

$$\begin{bmatrix} \mathscr{H}_{11} & \mathscr{H}_{12} \\ \mathscr{H}_{21} & -\mathscr{H}_{11}^T \end{bmatrix} \begin{bmatrix} \mathscr{V}_1 \\ \mathscr{V}_2 \end{bmatrix} = \begin{bmatrix} \mathscr{V}_1 \\ \mathscr{V}_2 \end{bmatrix} \widetilde{\mathscr{H}}$$

with $\Lambda\left(\widetilde{\mathscr{H}}\right) \subset \mathbb{C}^-$. The appropriate structure-preserving methods for this computation are outlined in [16] and implemented as the routine `MB03ZD` in SLICOT.

By defining $Y := -\mathscr{V}_2\mathscr{V}_1^{-1}$, $\widetilde{\xi}_1(\cdot) := \xi_1(\cdot)$, and $\widetilde{\xi}_2(\cdot) := Y\xi_1(\cdot) + \xi_2(\cdot)$ we get

$$\begin{aligned} \begin{bmatrix} \dot{\widetilde{\xi}}_1(t) \\ \dot{\widetilde{\xi}}_2(t) \end{bmatrix} &= \begin{bmatrix} I_{r_{w+1}/2} & 0 \\ Y & I_{r_{w+1}/2} \end{bmatrix} \begin{bmatrix} \mathscr{H}_{11} & \mathscr{H}_{12} \\ \mathscr{H}_{21} & -\mathscr{H}_{11}^T \end{bmatrix} \begin{bmatrix} I_{r_{w+1}/2} & 0 \\ -Y & I_{r_{w+1}/2} \end{bmatrix} \begin{bmatrix} \widetilde{\xi}_1(t) \\ \widetilde{\xi}_2(t) \end{bmatrix} \\ &= \begin{bmatrix} \widetilde{\mathscr{H}}_{11} & \widetilde{\mathscr{H}}_{12} \\ 0 & -\widetilde{\mathscr{H}}_{11}^T \end{bmatrix} \begin{bmatrix} \widetilde{\xi}_1(t) \\ \widetilde{\xi}_2(t) \end{bmatrix}, \end{aligned} \tag{28}$$

with $\Lambda\left(\widetilde{\mathscr{H}}_{11}\right) \subset \mathbb{C}^-$ and the boundary conditions $\widetilde{\xi}_1(0) = \widetilde{\xi}_{1,0}$, $\lim_{t\to\infty}\widetilde{\xi}_2(t) = 0$. Since $-\widetilde{\mathscr{H}}_{11}^T$ is an unstable matrix, we obtain $\widetilde{\xi}_2(\cdot) \equiv 0$ by backwards integration. This results in

$$\dot{\widetilde{\xi}}_1(t) = \widetilde{\mathscr{H}}_{11}\widetilde{\xi}_1(t), \quad \widetilde{\xi}_1(0) = \widetilde{\xi}_{1,0},$$

which can be efficiently solved by a further transformation of $\widetilde{\mathscr{H}}_{11}$ to Schur form. From that we can easily reconstruct $\widetilde{z}_{w+1}(\cdot)$ as in (27).

This can be further used to determine $\widetilde{z}_{w+2}(\cdot), \ldots, \widetilde{z}_{2w+1}(\cdot)$ in terms of $\widetilde{z}_{w+1}(\cdot)$, and the consistency conditions $\widetilde{z}_1 \equiv 0, \ldots, \widetilde{z}_w \equiv 0$ via a backward substitution process applied to the first w block rows of (17). This recursive process leads to

$$M_{w-j+1,w+j+1}\widetilde{z}_{w+j+1}(t) = \left(\sum_{i=w+1}^{w+j} N_{w-j+1,i}\dot{\widetilde{z}}_i(t) - \sum_{i=w+1}^{w+j} M_{w-j+1,i}\widetilde{z}_i(t) \right), \tag{29}$$

which requires w differentiations to be carried out, see [43]. The complete procedure is graphically displayed in Figure 3. Note that if $sN - M$ is regular, then $M_{w+j+1} = \Gamma_{w-j+1}$ is nonsingular and $\widetilde{z}(\cdot)$ is uniquely determined. Otherwise, some of the variables remain undetermined and thus the optimal solution trajectory might not be unique.

We illustrate the whole procedure using the following example.

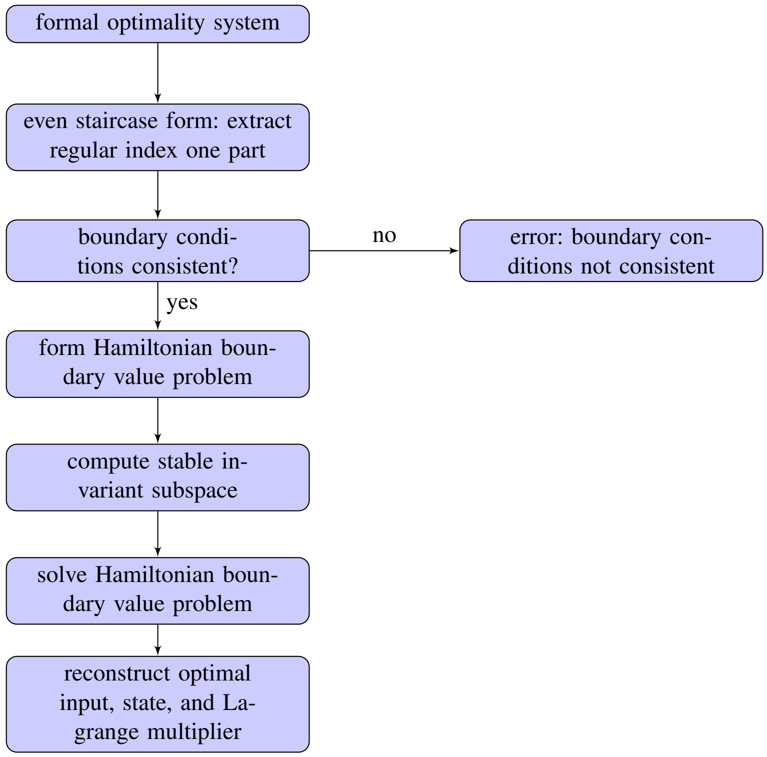

Fig. 3. Algorithm flowchart for solving linear-quadratic optimal control problems

Example 2. Consider the linear-quadratic optimal control problem of minimizing

$$\mathscr{J}(x(\cdot),u(\cdot)) = \int_0^\infty \left(x_1(t)^2 + x_2(t)^2 + u(t)^2\right) \mathrm{d}t$$

subject to

$$\begin{bmatrix} 1&0&0&0&0\\0&1&0&0&0\\0&0&0&1&0\\0&0&0&0&0\\0&0&0&0&0 \end{bmatrix} \begin{bmatrix} \dot{x}_1(t)\\ \dot{x}_2(t)\\ \dot{x}_3(t)\\ \dot{x}_4(t)\\ \dot{x}_5(t) \end{bmatrix} = \begin{bmatrix} 1&1&0&0&0\\-1&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&0 \end{bmatrix} \begin{bmatrix} x_1(t)\\ x_2(t)\\ x_3(t)\\ x_4(t)\\ x_5(t) \end{bmatrix} + \begin{bmatrix} 1\\1\\-1\\0\\0 \end{bmatrix} u(t),$$

$$\begin{bmatrix} x_{1,0}(t)\\ x_{2,0}(t)\\ x_{3,0}(t)\\ x_{4,0}(t)\\ x_{5,0}(t) \end{bmatrix} = \begin{bmatrix} 1\\-1\\0\\0\\0 \end{bmatrix}.$$

with

$$Q = \begin{bmatrix} 1&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \end{bmatrix}, \quad S = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad R = 1.$$

The formal optimality system is given by (24) with the boundary condition for the Lagrange multiplier given by $\lim_{t\to\infty} \lambda_1(t) = \lim_{t\to\infty} \lambda_2(t) = \lim_{t\to\infty} \lambda_3(t) = 0$. With the notation of Theorem 6, a reduction to even staircase form yields the structural characteristics

$$w = 1, \quad s_1 = 2, \quad q_1 = 4.$$

In particular, since $q_1 - s_1 = 2 \neq 0$, the formal optimality system is singular. Thus the transformed formal optimality system attains the form

$$\begin{bmatrix} N_{1,1} & N_{1,2} & 0 \\ -N_{1,2}^T & N_{2,2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\tilde{z}}_1(t) \\ \dot{\tilde{z}}_2(t) \\ \dot{\tilde{z}}_3(t) \end{bmatrix} = \begin{bmatrix} M_{1,1} & M_{1,2} & M_{1,3} \\ M_{1,2}^T & M_{2,2} & 0 \\ M_{1,3}^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{z}_1(t) \\ \tilde{z}_2(t) \\ \tilde{z}_3(t) \end{bmatrix}. \tag{30}$$

The regular index one part is given by

$$\begin{bmatrix} 0&-1&0&0&0 \\ 1&0&0&0&0 \\ 0&0&0&-1&0 \\ 0&0&1&0&0 \\ 0&0&0&0&0 \end{bmatrix} \begin{bmatrix} \dot{z}_{2,1}(t) \\ \dot{z}_{2,2}(t) \\ \dot{z}_{2,3}(t) \\ \dot{z}_{2,4}(t) \\ \dot{z}_{2,5}(t) \end{bmatrix} = \begin{bmatrix} 1&1&0&-1&0 \\ 1&0&1&0&-1 \\ 0&1&1&1&0 \\ -1&0&1&0&-1 \\ 0&-1&0&-1&1 \end{bmatrix} \begin{bmatrix} z_{2,1}(t) \\ z_{2,2}(t) \\ z_{2,3}(t) \\ z_{2,4}(t) \\ z_{2,5}(t) \end{bmatrix},$$

$$\begin{bmatrix} z_{2,1}(t) \\ z_{2,2}(t) \\ z_{2,3}(t) \\ z_{2,4}(t) \\ z_{2,5}(t) \end{bmatrix} = \begin{bmatrix} -x_1(t) \\ -\lambda_1(t) \\ -x_2(t) \\ -\lambda_2(t) \\ u(t) \end{bmatrix}.$$

A further reduction to a Hamiltonian boundary value problem yields

$$\begin{bmatrix} \dot{\xi}_{1,1}(t) \\ \dot{\xi}_{1,2}(t) \\ \dot{\xi}_{2,1}(t) \\ \dot{\xi}_{2,2}(t) \end{bmatrix} = \begin{bmatrix} 1&1&-1&-1 \\ -1&1&-1&-1 \\ -1&0&-1&1 \\ 0&-1&-1&-1 \end{bmatrix} \begin{bmatrix} \xi_{1,1}(t) \\ \xi_{1,2}(t) \\ \xi_{2,1}(t) \\ \xi_{2,2}(t) \end{bmatrix},$$

$$\begin{bmatrix} \xi_{1,1}(0) \\ \xi_{1,2}(0) \end{bmatrix} = \begin{bmatrix} -x_1(0) \\ -x_2(0) \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \lim_{t\to\infty} \begin{bmatrix} \xi_{2,1}(t) \\ \xi_{2,2}(t) \end{bmatrix} = \lim_{t\to\infty} \begin{bmatrix} -\lambda_1(t) \\ -\lambda_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The stable invariant subspace of the Hamiltonian matrix is spanned by

$$\mathscr{V} = \begin{bmatrix} -0.2041 & -0.1727 \\ -0.2317 & -0.4438 \\ -0.9511 & 0.1548 \\ -0.0106 & -0.8656 \end{bmatrix},$$

and thus the Hamiltonian boundary value problem (28) reduces to

$$\begin{bmatrix} \dot{\widetilde{\xi}}_{1,1}(t) \\ \dot{\widetilde{\xi}}_{1,2}(t) \\ \dot{\widetilde{\xi}}_{2,1}(t) \\ \dot{\widetilde{\xi}}_{2,2}(t) \end{bmatrix} = \begin{bmatrix} -4.1813 & 1.4142 & -1.0000 & -1.0000 \\ -6.1813 & 1.4142 & -1.0000 & -1.0000 \\ 0 & 0 & 4.1813 & 6.1813 \\ 0 & 0 & -1.4142 & -1.4142 \end{bmatrix} \begin{bmatrix} \widetilde{\xi}_{1,1}(t) \\ \widetilde{\xi}_{1,2}(t) \\ \widetilde{\xi}_{2,1}(t) \\ \widetilde{\xi}_{2,2}(t) \end{bmatrix},$$

$$\begin{bmatrix} \widetilde{\xi}_{1,1}(0) \\ \widetilde{\xi}_{1,2}(0) \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \lim_{t\to\infty} \begin{bmatrix} \widetilde{\xi}_{2,1}(t) \\ \widetilde{\xi}_{2,2}(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which can now be solved by a numerical integrator. One first integrates the last two equations backward in time and stores the the trajectories either in discrete time points or in a collocation representation. Then the first two equations are integrated forward in time, making use of the already computed variables which then act as inhomogeneities. If the integrator needs this inhomogeneity in points different from the stored points, either these have to be recomputed or obtained by interpolation. It remains to determine $\widetilde{z}_1(\cdot)$ and $\widetilde{z}_3(\cdot)$ in (30). For this we have

$$\widetilde{z}_1(t) = \begin{bmatrix} \lambda_3(t) \\ -x_4(t) \end{bmatrix} = 0,$$

which is consistent to the boundary conditions. For our example, (29) further yields

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_{3,1}(t) \\ z_{3,2}(t) \\ z_{3,3}(t) \\ z_{3,4}(t) \end{bmatrix} = - \begin{bmatrix} 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_{2,1}(t) \\ z_{2,2}(t) \\ z_{2,3}(t) \\ z_{2,4}(t) \\ z_{2,5}(t) \end{bmatrix}, \tag{31}$$

i.e., we have $x_3(t) = z_{3,1}(t) = z_{2,5}(t) = u(t)$, and $-\lambda_2(t) = z_{3,2}(t) = 0$ which are both in agreement to our state equation and the boundary conditions. Note that $z_{3,3}(t) = \lambda_5(t)$ and $z_{3,4}(t) = x_5(t)$ remain undetermined and thus the optimal solution is not unique. This observation is also in agreement to the state equation, since there $x_5(\cdot)$ is already free. To deal with this situation one either removes the last equation and the variable $x_5(\cdot)$ already in the original state equation or one chooses another cost functional, see [73] for a discussion.

Remark 1. A similar decoupling procedure can also be constructed in the finite-time horizon problem by decoupling the forward and backward integration via the solution of a Riccati differential equation or by using other boundary value methods [5].

Remark 2. In [97, 117] the linear-quadratic optimal control problem has been solved by directly using the deflating subspaces of an even matrix pencil that is related to the boundary value problem (24), in particular in the context of singular control problems.

7 $\mathscr{H}_\infty$ Optimal Control

Our second application is the $\mathscr{H}_\infty$ optimal control problem which is one of the major tasks in robust control. We consider descriptor systems of the form

$$\mathbf{P}: \begin{cases} E\dot{x}(t) = \ Ax(t) + \ B_1v(t) + \ B_2u(t), \quad x(0) = x_0, \\ \ z(t) = C_1x(t) + D_{11}v(t) + D_{12}u(t), \\ \ y(t) = C_2x(t) + D_{21}v(t) + D_{22}u(t), \end{cases} \tag{32}$$

where $E, A \in \mathbb{R}^{n,n}$, $B_i \in \mathbb{R}^{n,m_i}$, $C_i \in \mathbb{R}^{p_i,n}$, and $D_{ij} \in \mathbb{R}^{p_i,m_j}$ for $i, j = 1,2$. In this system, $x : [0,\infty) \to \mathbb{R}^n$ is the state, $u : [0,\infty) \to \mathbb{R}^{m_2}$ is the control input, and $v : [0,\infty) \to \mathbb{R}^{m_1}$ is an exogenous input that may include noise, linearization errors and unmodeled dynamics. The function $y : [0,\infty) \to \mathbb{R}^{p_2}$ contains measured outputs, while $z : [0,\infty) \to \mathbb{R}^{p_1}$ is a regulated output or an estimation error.

The $\mathscr{H}_\infty$ optimal control problem is typically formulated in the frequency domain. Its goal is to stabilize the system, while minimizing the $\mathscr{H}_\infty$-norm of the closed-loop transfer function $T_{zv}(\cdot)$ mapping noise or disturbance to error signals [122]. The value of $\|T_{zv}\|_{\mathscr{H}_\infty}$ is used as a measure for the worst-case influence of the disturbances v on the output z. A more rigorous formulation is given in the following definition [86].

Definition 2 (The optimal $\mathscr{H}_\infty$ control problem). *For the descriptor system* (32)*, determine a controller (dynamic compensator)*

$$\mathbf{K}: \begin{cases} \widehat{E}\dot{\widehat{x}}(t) = \widehat{A}\widehat{x}(t) + \widehat{B}y(t), \\ \ u(t) = \widehat{C}\widehat{x}(t) + \widehat{D}y(t), \end{cases} \tag{33}$$

with $\widehat{E}, \widehat{A} \in \mathbb{R}^{N,N}$, $\widehat{B} \in \mathbb{R}^{N,p_2}$, $\widehat{C} \in \mathbb{R}^{m_2,N}$, $\widehat{D} \in \mathbb{R}^{m_2,p_2}$, *and transfer function* $K(s) = \widehat{C}(s\widehat{E} - \widehat{A})^{-1}\widehat{B} + \widehat{D}$ *such that the closed-loop system resulting from the combination of (32) and (33), given by*

$$\begin{aligned} E\dot{x}(t) &= \left(A + B_2\widehat{D}Z_1C_2\right)x(t) + B_2Z_2\widehat{C}\widehat{x}(t) + \left(B_1 + B_2\widehat{D}Z_1D_{21}\right)v(t), \\ \widehat{E}\dot{\widehat{x}}(t) &= \widehat{B}Z_1C_2x(t) + \left(\widehat{A} + \widehat{B}Z_1D_{22}\widehat{C}\right)\widehat{x}(t) + \widehat{B}Z_1D_{21}v(t), \\ z(t) &= \left(C_1 + D_{12}Z_2\widehat{D}C_2\right)x(t) + D_{12}Z_2\widehat{C}\widehat{x}(t) + \left(D_{11} + D_{12}\widehat{D}Z_1D_{21}\right)v(t) \end{aligned} \tag{34}$$

with $Z_1 = \left(I_{p_2} - D_{22}\widehat{D}\right)^{-1}$ *and* $Z_2 = \left(I_{m_2} - \widehat{D}D_{22}\right)^{-1}$*, has the following properties:*

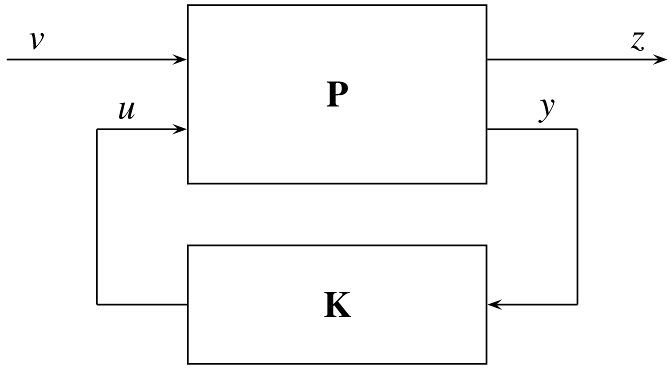

Fig. 4. Interconnection of a system **P** with a controller **K**

(i) System (34) is internally stable*, i.e., the solution* $\begin{bmatrix} x(\cdot) \\ \widehat{x}(\cdot) \end{bmatrix}$ *of the system with* $v \equiv 0$ *is* asymptotically stable*, in other words* $\lim_{t\to\infty} \begin{bmatrix} x(t) \\ \widehat{x}(t) \end{bmatrix} = 0$.

(ii) The closed-loop transfer function $T_{zv}(\cdot)$ *from* v *to* z *satisfies* $T_{zv} \in \mathscr{RH}_\infty^{p_1,m_1}$ *and is minimized in the* $\mathscr{H}_\infty$*-norm.*

Such an interconnection of a system with a controller is depicted in Figure 4. Solving the optimal $\mathscr{H}_\infty$ control problem by trying to directly minimize the $\mathscr{H}_\infty$-norm of $T_{zv}(\cdot)$ over the complicated set of internally stabilizing controllers proves difficult or impossible by conventional optimization methods, since it is often unclear if a minimizing controller exists [122] and if one exists, it is typically not unique, there even exist infinitely many. So usually one studies two closely related optimization problems, the *modified optimal* $\mathscr{H}_\infty$ *control problem* and the *suboptimal* $\mathscr{H}_\infty$ *control problem* [13, 122].

Definition 3 (The modified optimal $\mathscr{H}_\infty$ control problem). *For the descriptor system* (32)*, let* Γ *be the set of positive real numbers* γ *for which there exists an internally stabilizing dynamic controller of the form* (33) *so that the transfer function* $T_{zv}(\cdot)$ *of the closed-loop system* (34) *satisfies* $T_{zv} \in \mathscr{RH}_\infty^{p_1,m_1}$ *with* $\|T_{zv}\|_{\mathscr{H}_\infty} < \gamma$*. Determine* $\gamma_{\mathrm{mo}} = \inf \Gamma$*. If no internally stabilizing dynamic controller exists, we set* $\Gamma = \emptyset$ *and* $\gamma_{\mathrm{mo}} = \infty$*.*

This problem is usually solved by an iterative process, which is often called the *γ-iteration*.

Definition 4 (The suboptimal $\mathscr{H}_\infty$ control problem). *For the descriptor system* (32) *and* $\gamma \in \Gamma$ *with* $\gamma > \gamma_{\mathrm{mo}}$*, determine an internally stabilizing dynamic controller of the form (33) such that the closed-loop transfer function satisfies* $T_{zv} \in \mathscr{RH}_\infty^{p_1,m_1}$ *with* $\|T_{zv}\|_{\mathscr{H}_\infty} < \gamma$*. We call such a controller* γ-suboptimal controller *or simply* suboptimal controller.

To obtain an existence and uniqueness result we make the following assumptions:

A1) The triple (E,A,B_2) is strongly stabilizable and the triple (E,A,C_2) is strongly detectable.

A2) rank $\begin{bmatrix} A-\mathrm{i}\omega E & B_2 \\ C_1 & D_{12} \end{bmatrix} = n+m_2$ for all $\omega \in \mathbb{R}$.

A3) rank $\begin{bmatrix} A-\mathrm{i}\omega E & B_1 \\ C_2 & D_{21} \end{bmatrix} = n+p_2$ for all $\omega \in \mathbb{R}$.

A4) With matrices $S_\infty, T_\infty \in \mathbb{R}^{n,n-r}$ satisfying $\operatorname{range} S_\infty = \ker E$, $\operatorname{range} T_\infty = \ker E^T$ and $r := \operatorname{rank} E$ we have

$$\operatorname{rank} \begin{bmatrix} T_\infty^T A S_\infty & T_\infty^T B_2 \\ C_1 S_\infty & D_{12} \end{bmatrix} = n+m_2-r,$$
$$\operatorname{rank} \begin{bmatrix} T_\infty^T A S_\infty & T_\infty^T B_1 \\ C_2 S_\infty & D_{21} \end{bmatrix} = n+p_2-r.$$

In Assumption **A1)**, the conditions of impulse controllability and impulse observability are necessary to avoid impulsive solutions which cannot be controlled or observed. To check these conditions one can use the condensed forms of Theorem 4 with the characterization of Corollary 1. The property that the system is behavioral stabilizable and behavioral detectable is necessary for the existence of an internally stabilizing controller. To verify these conditions we use the decompositions (14) and (15) which can be computed via the codes `TG01HD`, `TG01ID` in the SLICOT library. These routines can also be used to check **A2)** and **A3)**.

To verify that assumption **A4)** is satisfied, we check that the ranks of the extended matrices fulfill

$$\operatorname{rank} \begin{bmatrix} 0 & E & 0 \\ E^T & A & B_2 \\ 0 & C_1 & D_{12} \end{bmatrix} = n+m_2+r,$$

and

$$\operatorname{rank} \begin{bmatrix} 0 & E & 0 \\ E^T & A & B_1 \\ 0 & C_2 & D_{21} \end{bmatrix} = n+p_2+r.$$

This check is performed by applying a rank-revealing QR (RRQR) decomposition [62]. The corresponding routine `DGEQP3` is available in LAPACK[4]. For details on the implementation we refer to [29, 30, 55].

Once we have assured that the assumptions **A1) – A4)** hold, we can form the two even matrix pencils

$$sN_H - M_H(\gamma) = \left[\begin{array}{cc|ccc} 0 & -sE^T - A^T & 0 & 0 & -C_1^T \\ sE-A & 0 & -B_1 & -B_2 & 0 \\ \hline 0 & -B_1^T & -\gamma^2 I_{m_1} & 0 & -D_{11}^T \\ 0 & -B_2^T & 0 & 0 & -D_{12}^T \\ -C_1 & 0 & -D_{11} & -D_{12} & -I_{p_1} \end{array}\right], \qquad (35)$$

[4] `http://www.netlib.org/lapack/`

and

$$sN_J - M_J(\gamma) = \left[\begin{array}{cc|ccc} 0 & -sE-A & 0 & 0 & -B_1 \\ sE^T - A^T & 0 & -C_1^T & -C_2^T & 0 \\ \hline 0 & -C_1 & -\gamma^2 I_{p_1} & 0 & -D_{11} \\ 0 & -C_2 & 0 & 0 & -D_{21} \\ -B_1^T & 0 & -D_{11}^T & -D_{21}^T & -I_{m_1} \end{array}\right]. \tag{36}$$

We determine the semi-stable deflating subspaces of both pencils, i.e., the deflating subspaces corresponding to the eigenvalues in the open left complex half-plane and a part of the deflating subspaces associated to the purely imaginary eigenvalues with even algebraic multiplicity and uniform sign-characteristic. Suppose that these subspaces are spanned by the columns of the matrices

$$X_H(\gamma) = \begin{bmatrix} X_{H,1}(\gamma) \\ X_{H,2}(\gamma) \\ X_{H,3}(\gamma) \\ X_{H,4}(\gamma) \\ X_{H,5}(\gamma) \end{bmatrix}, \quad X_J(\gamma) = \begin{bmatrix} X_{J,1}(\gamma) \\ X_{J,2}(\gamma) \\ X_{J,3}(\gamma) \\ X_{J,4}(\gamma) \\ X_{J,5}(\gamma) \end{bmatrix},$$

which are partitioned according to the block structure of the pencils $sN_H - M_H$ and $sN_J - M_J$.

We use the following result to solve the modified optimal $\mathscr{H}_\infty$ control problem.

Theorem 8. *[86] Consider system* (32) *and the even pencils $sN_H - M_H(\gamma)$ and $sN_J - M_J(\gamma)$ as in* (35) *and* (36)*, respectively. Suppose that assumptions* **A1)** – **A4)** *hold.*

Then there exists an internally stabilizing controller such that the transfer function from v to z satisfies $T_{zv} \in \mathscr{RH}_\infty^{p_1,m_1}$ with $\|T_{zv}\|_{\mathscr{H}_\infty} < \gamma$ if and only if γ is such that the following conditions **C1)** – **C4)** *hold.*

C1) *The index of both pencils* (35) *and* (36) *is at most one.*
C2) *There exists a matrix $X_H(\gamma)$ such that*
C2.a) *the space* $\operatorname{range} X_H(\gamma)$ *is a semi-stable deflating subspace of $sN_H - M_H(\gamma)$ and* $\operatorname{range} \begin{bmatrix} EX_{H,1}(\gamma) \\ X_{H,2}(\gamma) \end{bmatrix}$ *is an r-dimensional isotropic subspace of* $\mathbb{R}^{2n}$*;*
C2.b) $\operatorname{rank} EX_{H,1}(\gamma) = r$.
C3) *There exists a matrix $X_J(\gamma)$ such that*
C3.a) *the space* $\operatorname{range} X_J(\gamma)$ *is a semi-stable deflating subspace of $sN_J - M_J(\gamma)$ and* $\operatorname{range} \begin{bmatrix} E^T X_{J,1}(\gamma) \\ X_{J,2}(\gamma) \end{bmatrix}$ *is an r-dimensional isotropic subspace of* $\mathbb{R}^{2n}$*;*
C3.b) $\operatorname{rank} E^T X_{J,1}(\gamma) = r$.
C4) *The matrix*

$$\mathscr{Y}(\gamma) = \begin{bmatrix} -\gamma X_{H,2}^T(\gamma) E X_{H,1}(\gamma) & X_{H,2}^T(\gamma) E X_{J,2}(\gamma) \\ X_{J,2}^T(\gamma) E^T X_{H,2}(\gamma) & -\gamma X_{J,2}^T(\gamma) E^T X_{J,1}(\gamma) \end{bmatrix}$$

is symmetric, positive semi-definite and satisfies $\operatorname{rank}\mathscr{Y}(\gamma) = k_H + k_J$*, where* k_H *and* k_J *are such that for all sufficiently large* $\gamma_{H,1}, \gamma_{H,2}$*, and* $\gamma_{J,1}, \gamma_{J,2}$ *the conditions*

$$\begin{aligned} \operatorname{rank} E^T X_{H,2}(\gamma_{H,1}) &= \operatorname{rank} E^T X_{H,2}(\gamma_{H,2}) = k_H, \\ \operatorname{rank} E X_{J,2}(\gamma_{J,1}) &= \operatorname{rank} E X_{J,2}(\gamma_{J,2}) = k_J \end{aligned}$$

hold.

Furthermore, the set of values γ *satisfying the conditions* **C1) – C4)** *is nonempty.*

To check condition **C4)**, we make use of the LDL^T decomposition, described in [6] and implemented in LAPACK by `DSPTRF` which decomposes a real symmetric matrix A as $A = LDL^T$, where L is a product of permutation and lower triangular matrices, and D is symmetric and block diagonal with 1×1 and 2×2 diagonal blocks.

Using Theorem 8, we can use a bisection type algorithm to determine the sub-optimal value γ_{mo}, see [85].

After completing the bisection process, one has the option to either use the result directly, or to perform a strong validation, by dividing the interval $(0, \gamma_{\mathrm{mo}})$ at a desired number of points and checking the four conditions **C1) – C4)** again at these points. If the conditions **C1) – C4)** are fulfilled for another $\gamma \in (0, \gamma_{\mathrm{mo}})$, we have obviously found a better value for γ_{mo}. We can either use this new value or continue with the γ-iteration to find an even better value. Once a satisfactory γ is found, it remains to compute the controller. The trick that we use to determine the controller is to compute an index-reducing static output feedback $u(t) = Fy(t) + \overline{u}(t)$, whose application leads to a new descriptor system of the form (32) with an index of at most one. It can be shown that the application of the feedback does not change the solution of the modified $\mathscr{H}_\infty$ optimal control problem [85, 86]. The feedback is computed using the condensed form (13) and the techniques presented in [41], which yield $s_2 = t_2$ and

$$F = \begin{bmatrix} F_{11} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m,p}, \quad F_{11} = \begin{bmatrix} B_{21} \\ B_{31} \end{bmatrix}^{-1} (I_{s_2} - A_{22}) \begin{bmatrix} C_{12} & C_{13} \end{bmatrix}^{-1}. \tag{37}$$

Note that due to the construction of the condensed form (13), the matrices

$$\begin{bmatrix} B_{21} \\ B_{31} \end{bmatrix}, \quad \begin{bmatrix} C_{12} & C_{13} \end{bmatrix}$$

can be kept in factored form as a product of an orthogonal and a diagonal matrix. So the computation of F can be carried out by the inversion of two diagonal matrices.

We can use this new descriptor system to compute the controller. The controller formulas themselves and their derivation are rather involved. Therefore, we only refer to the robust controller formulas for the standard system case in [10], and based on that, the controller formulas for the descriptor system case in [85].

Figure 5 presents a flow chart for the solution of the optimal solution. First one checks the four assumptions **A1) – A4)**, using the condensed forms from Theorem

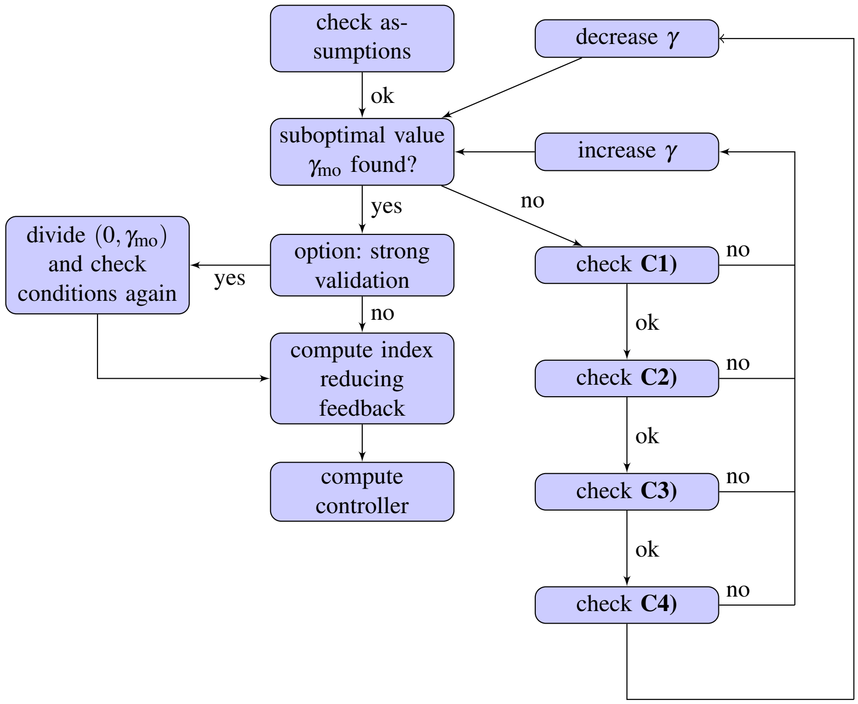

Fig. 5. Algorithm flowchart for solving $\mathscr{H}_\infty$ optimal control problems

4, the decompositions (14) and performing some rank checks. Then one uses a bisection type algorithm to find the optimal value of γ, by checking the four conditions from Theorem 8 in each step by using the staircase form from Theorem 6, the computation of the semi-stable deflating subspaces using Algorithm 1, and the LDL^T decomposition from [6]. Here, the structure-preservation aspect of Algorithm 1 is very important, since using these methods, it *cannot* happen that eigenvalues from the left half-plane move to the right half-plane and vice versa due to round-off errors. Therefore, the computed subspaces are guaranteed to have the correct dimensions. Once the suboptimal value is found, one has the option to use a strong validation by checking the aforementioned four conditions again at a desired number of points. Then it remains to compute an index reducing feedback (37) and to compute the controller formulas given in [10, 85]. For an illustration of the method by numerical examples we refer to [10, 85].

8 $\mathscr{L}_\infty$-Norm Computation

In the previous section we have seen that the $\mathscr{H}_\infty$-norm of a transfer function is an important measure for the robustness of a linear system. This section is devoted to the actual computation of this norm. We will directly present this for the more general case of the $\mathscr{L}_\infty$-norm. Consider a square descriptor system (2) with regular pencil $sE-A$ and transfer function $G(\cdot)$ as in (16).

Before we can turn to the actual norm computation, we have to ensure that $G \in \mathscr{R}\mathscr{L}_\infty^{p,m}$. First, we check whether the transfer function is *proper*, i.e., that $\lim_{\omega\to\infty} \|G(\mathrm{i}\omega)\| < \infty$. For this we make use of the following result of [18, 116] in a modified formulation.

Theorem 9. *Consider a descriptor system* (2a) *given in the condensed form* (13). *Then, $G(\cdot)$ is proper if and only if the subpencil*

$$s\begin{bmatrix}\Sigma_E & 0\\ 0 & 0\end{bmatrix} - \begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}$$

is regular and of index at most one, i.e., if A_{22} is invertible.

Therefore, to check properness, we first reduce the system to the condensed form (13) and subsequently check A_{22} for invertibility, e.g., by employing condition estimators [62].

When we have checked the transfer function for properness, it remains to check whether $G(\cdot)$ has finite, purely imaginary poles. For this, we first determine the controllability and observability decompositions (14) and (15) to extract the controllable and observable subsystem. The finite eigenvalues of the pencil associated to this subsystem are poles of $G(\cdot)$ and we check whether there are eigenvalues that lie in a thin strip around the imaginary axis. The thickness of this strip depends on the multiplicity of the pole which is generally not known. In finite precision, eigenvalues in this region cannot be distinguished from eigenvalues on the imaginary axis. Generically, a pole will be simple and therefore, in the code we choose the thickness as a small multiple of machine precision. After we have ensured that $G \in \mathscr{R}\mathscr{L}_\infty^{p,m}$, we can compute the norm value. For this we make use of the even matrix pencils

$$sN - M(\gamma) = \left[\begin{array}{cc|cc} 0 & sE-A & 0 & -B\\ -sE^T-A^T & 0 & -C^T & 0\\ \hline 0 & -C & \gamma I_p & -D\\ -B^T & 0 & -D^T & \gamma I_m\end{array}\right]. \tag{38}$$

The following theorem connects the singular values of $G(\mathrm{i}\omega)$ with the finite, purely imaginary eigenvalues of $sN-M(\gamma)$, see [18, 19, 116] for details.

Theorem 10. *Assume that $sE-A$ has no purely imaginary eigenvalues, $G \in \mathscr{R}\mathscr{L}_\infty^{p,m}$, $\gamma > 0$ and $\omega_0 \in \mathbb{R}$. Then γ is a singular value of $G(\mathrm{i}\omega_0)$ if and only if $\mathrm{i}\omega_0 N - M(\gamma)$ is singular.*

A direct consequence of Theorem 10 is the following result, see [18, 19].

Theorem 11. *Assume that $sE-A$ has no purely imaginary eigenvalues, $G \in \mathscr{R}\mathscr{L}_\infty^{p,m}$ and suppose that $\gamma > \inf_{\omega\in\mathbb{R}} \sigma_{\max}(G(\mathrm{i}\omega))$. Then $\|G\|_{\mathscr{L}_\infty} \geq \gamma$ if and only if $sN-M(\gamma)$ in* (38) *has finite, purely imaginary eigenvalues.*

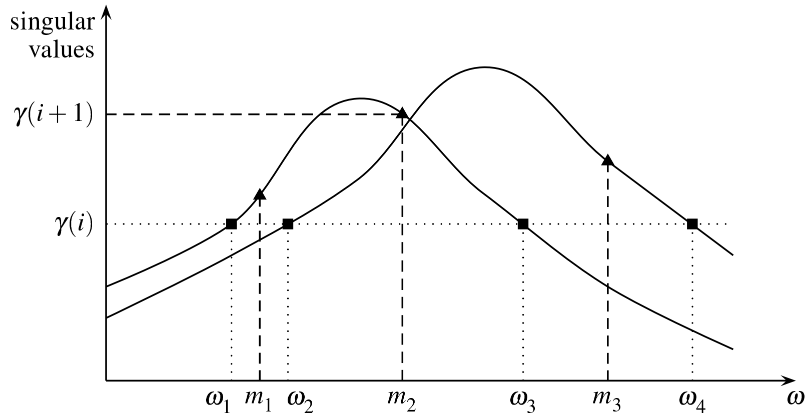

Fig. 6. Graphical interpretation of the algorithm for computing the $\mathscr{L}_\infty$-norm. Here, $\gamma(i)$ and $\gamma(i+1)$ denote the iterates at the i-th and $(i+1)$-st step, respectively.

This directly yields an algorithm for the computation of the $\mathscr{L}_\infty$-norm, similarly as in [32, 33, 34]. Given an initial value of γ with $\inf_{\omega\in\mathbb{R}} \sigma_{\max}(G(\mathrm{i}\omega)) < \gamma < \|G\|_{\mathscr{L}_\infty}$, we check if $sN - M(\gamma)$ has purely imaginary eigenvalues. If yes, we denote these eigenvalues with positive imaginary part by $\mathrm{i}\omega_1, \ldots, \mathrm{i}\omega_q$. To obtain the next (larger) value of γ, we determine new test frequencies $m_j = \sqrt{\omega_j \omega_{j+1}}$, $j = 1, \ldots, q-1$. Then, the new value of γ is chosen as

$$\gamma = \max_{1\le j\le q-1} \sigma_{\max}(G(\mathrm{i}m_j)).$$

To check whether a prespecified relative error ε has already been achieved, we would have to check whether the pencil $sN - M(\widehat{\gamma})$ with $\widehat{\gamma} = \gamma(1+2\varepsilon)$ has no purely imaginary eigenvalues. To avoid the additional check in every step, we can directly incorporate this into the algorithm by always working with $\widehat{\gamma}$ instead of γ when determining the eigenvalues of the even pencils.

It can be shown that this algorithm converges globally with a quadratic rate and a guaranteed relative error of ε when assuming exact arithmetics. We refer to [18, 19, 116] for details on the implementation and the algorithm properties. Note again that the decision about the existence of purely imaginary eigenvalues is crucial for a robust execution of this algorithm and does require a structured eigensolver as described in Section 5.2. A graphical interpretation is given in Figure 6.

Note, that when assuming that $G \in \mathscr{R}\mathscr{L}_\infty^{p,m}$, the algorithm runs on the original data without performing any system reductions beforehand. However, $sE - A$ could still have uncontrollable or unobservable eigenvalues on the imaginary axis. If one does not perform the system reductions to extract the behavioral controllable and observable subsystem, then it remains to check whether $sE - A$ has no finite, purely imaginary eigenvalues. The complete procedure is summarized in Figure 7. An illustrative numerical example can be found in [18], whereas in [19] one can find a more detailed analysis of the behavior of the algorithm.

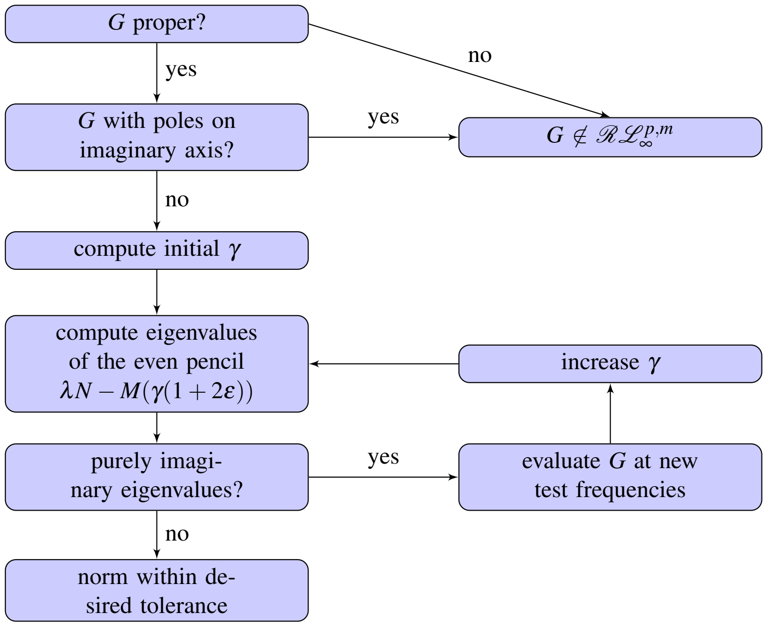

Fig. 7. Flowchart for computing the $\mathscr{L}_\infty$-norm

9 Dissipativity Check

The notion of dissipative systems is one of the most important concepts in systems and control theory, see for instance [118, 119, 120]. It naturally arises in many physical problems, especially when energy considerations are of importance. Roughly speaking, dissipative systems cannot internally generate energy. Equivalently, the system cannot supply more energy to its environment than energy that has been supplied to the system. Typical areas where such systems appear are the modeling of electrical circuits [100] (where, e.g., resistors consume a part of the energy and transform it into heat), or thermodynamic processes (where a part of the energy is transformed into an increase of entropy due to the second law of thermodynamics).

When modeling real-world processes it is often desired or necessary to reflect the dissipative nature of the problem in the model structure. This is important in order to obtain physically meaningful results when performing simulations. This section presents a method to check a certain notion of dissipativity for linear time-invariant descriptor systems of the form (2) based on a spectral characterization for even pencils.

We first introduce a precise mathematical formulation of dissipativity. For this we need the notion of *supply rates* which measure the power supplied to the system at time t. In the following we restrict ourselves to quadratic supply functions of the form

$$s(u(t),y(t)) = \begin{bmatrix} y(t) \\ u(t) \end{bmatrix}^T \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} y(t) \\ u(t) \end{bmatrix}, \tag{39}$$

where $Q = Q^T \in \mathbb{R}^{p,p}$, $S \in \mathbb{R}^{p,m}$, and $R = R^T \in \mathbb{R}^{m,m}$. Then the energy supplied to the system in a time interval $[t_0,t_1]$ is measured by

$$\int_{t_0}^{t_1} s(u(t),y(t))\mathrm{d}t.$$

There are many different notions of dissipativity in the literature. In this survey, we stick to the notion of cyclo-dissipativity which has been introduced in [35, 36] in the context of behavior systems.

Definition 5. *A descriptor system* (2) *is called* cyclo-dissipative *with respect to* $s(\cdot,\cdot)$*, if*

$$\int_0^T s(u(t),y(t))\mathrm{d}t \geq 0$$

for all $T \geq 0$ *and all smooth trajectories* $(u(\cdot),x(\cdot),y(\cdot))$ *solving* (2) *with the boundary conditions* $Ex(0) = Ex(T) = 0$.

Remark 3. Cyclo-dissipativity is only a property of the strongly controllable part of the system. A more general definition of dissipativity would require the existence of a *storage function* $\Theta : \operatorname{im} E \to \mathbb{R}$ with $\Theta(0) = 0$ such that the *dissipation inequality*

$$\Theta(Ex(t_1)) \leq \Theta(Ex(t_0)) + \int_{t_0}^{t_1} s(u(t),y(t))\mathrm{d}t$$

is fulfilled for all $t_0 \leq t_1$ and all smooth solution trajectories $(u(\cdot),x(\cdot),y(\cdot))$ such that the supply rate is locally square-integrable, see [36]. If the system (2) is strongly controllable, then both definitions coincide. However, not every cyclo-dissipative system has to possess a storage function. A counter-example is given in [36].

Remark 4. In the definition of cyclo-dissipativity it is only required that trajectories that start at zero and return to zero in some finite time, do not generate energy. A stronger definition, that would require all trajectories that start at zero not to generate energy, exists as well. Special cases of this stronger notion are passivity and contractivity (see below). Closely related to this is then non-negativity of the storage function (if it exists). Unfortunately, its general treatment is much more involved. However, under the condition that the pencil $sE - A$ is regular, stable, and its Kronecker index is at most one, and Q is negative semidefinite, then this stronger definition coincides with Definition 5, see [38].

In practice, two particular cases for the choice of the supply rate are of great interest. If a descriptor system (2) is dissipative (in the sense of the stronger definition in Remark 4) with respect to the supply rate $s(u(t),y(t)) = u(t)^T y(t)$, i.e., if $k = n$, $p = m$ and

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 0 & I_m \\ I_m & 0 \end{bmatrix},$$

then the system is called *passive*. This situation typically arises in models for RLC circuits [2, 96, 98, 99].

The other special case is that the supply rate is given by $s(u(t),y(t)) = \|u(t)\|_2^2 - \|y(t)\|_2^2$, i.e., $k = n$ and

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} = \begin{bmatrix} -I_p & 0 \\ 0 & I_m \end{bmatrix}.$$

In this case, a dissipative system (in the sense of the stronger definition in Remark 4) is called *contractive*. Usually this structure occurs if (2) is a realization of scattering parameters [89], but similar structures also appear in $\mathscr{H}_\infty$ control, see Sections 7 and 8.

For square systems (with $k = n$), a well-known relation of cyclo-dissipativity defined above between the time and frequency domain is given by the so-called *Popov function*

$$H(\xi,\zeta) := \begin{bmatrix} (\xi E - A)^{-1}B \\ I_m \end{bmatrix}^H \begin{bmatrix} \widetilde{Q} & \widetilde{S} \\ \widetilde{S}^T & \widetilde{R} \end{bmatrix} \begin{bmatrix} (\zeta E - A)^{-1}B \\ I_m \end{bmatrix},$$

with $\widetilde{Q}$, $\widetilde{S}$, and $\widetilde{R}$ as in (21). One has the following theorem of [35, 36].

Theorem 12. *The square descriptor system* (2) *is cyclo-dissipative with respect to* $s(\cdot,\cdot)$ *if and only if* $H(\mathrm{i}\omega,\mathrm{i}\omega) \geq 0$ *for all* $\mathrm{i}\omega \notin \Lambda(E,A)$.

For the cases of passivity and contractivity we get more general relations. These are summarized in the following theorem [2].

Theorem 13. *Consider a square descriptor system of the form* (2) *with* $p = m$.

(i) The system is passive if and only if $G(\cdot)$ *is* positive real*, i.e.,*
- *(a)* $G(\cdot)$ *is analytic in* $\mathbb{C}^+$*; and*
- *(b)* $H(\lambda,\lambda) = G(\lambda) + G(\lambda)^H \geq 0$ *for all* $\lambda \in \mathbb{C}^+$.

(ii) The system is contractive if and only if $G(\cdot)$ *is* bounded real*, i.e.,*
- *(a)* $G(\cdot)$ *is analytic in* $\mathbb{C}^+$*; and*
- *(b)* $H(\lambda,\lambda) = I_m - G(\lambda)^H G(\lambda) \geq 0$ *for all* $\lambda \in \mathbb{C}^+$.

It is very important to note that similar equivalent conditions of Theorem 13 do in general *not* hold for systems that are dissipative in the sense of the stronger definition in Remark 4. A counterexample is given in [121]. There are many algebraic characterizations to check if a given system (2) is cyclo-dissipative. These are mainly based on solvability of certain linear matrix inequalities or matrix equations, see [66]. Instead we make use of the following spectral characterization of even matrix pencils. For this, we need the *sign-sum function* [35, 36, 38] of a Hermitian matrix T which is defined as

$$\eta(T) = \pi_+ + \pi_0 - \pi_-,$$

where π_+, π_0, and π_- are the numbers of positive, zero, and negative eigenvalues of T, respectively. Furthermore, we can define the rank of a polynomial matrix $P(s)$ over the field of real-rational functions (often called normal rank), given by

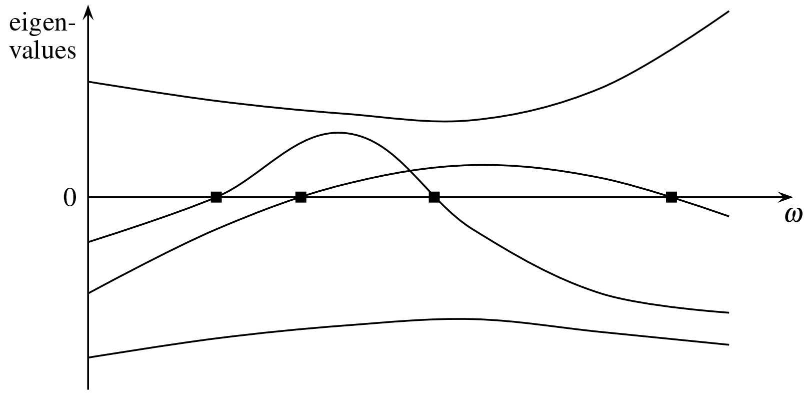

Fig. 8. Spectral plot. Here cyclo-dissipativity is violated, since the sign-sum function changes for varying ω.

$$\operatorname{rank}_{\mathbb{R}(s)}\left(P(s)\right) := \max_{\lambda\in\mathbb{C}} \operatorname{rank}\left(P(\lambda)\right). \tag{40}$$

The maximum in (40) is attained for almost all values of $\lambda \in \mathbb{C}$, there is only a finite set of points, where the rank drops.

Theorem 14. *[36, Theorem 3.11] Consider the system* (2) *with supply rate* (39)*. Let*

$$r := \operatorname{rank}_{\mathbb{R}(s)}\left(\begin{bmatrix} sE - A & -B \end{bmatrix}\right) \tag{41}$$

and define $\ell := k+n+m+2p$. Consider the even pencil

$$\mathscr{N}(s) = sN - M = \begin{bmatrix} 0 & 0 & 0 & sE-A & -B \\ 0 & 0 & I_m & -C & -D \\ 0 & I_m & Q & 0 & S \\ -sE^T - A^T & -C^T & 0 & 0 & 0 \\ -B^T & -D^T & S^T & 0 & R \end{bmatrix} \in \mathbb{R}[s]^{\ell,\ell}. \tag{42}$$

Then the system given by (2) *is cyclo-dissipative if and only if*

$$\eta(\mathscr{N}(\mathrm{i}\omega)) = k+n+m-2r$$

for all $\omega \in \mathbb{R}$ with $\operatorname{rank}\left(\begin{bmatrix} \mathrm{i}\omega E - A & -B \end{bmatrix}\right) = r$.

To better understand this theorem, we present a visualization in terms of the so-called *spectral plot*. This plot is constructed by plotting the ℓ eigenvalues of $\mathscr{N}(\mathrm{i}\omega)$ depending on ω, see Figure 8 for an example.

The general framework for checking cyclo-dissipativity then consists of two steps. First, we check if the assumptions of Theorem 14 are fulfilled. If the normal rank is unknown, then the GUPTRI form [53, 54, 70] is a suitable tool to compute it.

The next step consists in checking the sign-sum condition in Theorem 14. We exploit the fact that $\eta(\mathcal{N}(\mathrm{i}\omega))$ can only change at purely imaginary eigenvalues (of the regular index one part) and remains constant between two subsequent purely imaginary eigenvalues. We construct the pencil (42) and apply the even staircase algorithm from Theorem 6 to get the regular index one part $sN_{w+1,w+1} - M_{w+1,w+1}$. Then we compute its purely imaginary eigenvalues with positive imaginary part, denoted by $\mathrm{i}\omega_1, \ldots, \mathrm{i}\omega_q$, with $\omega_1 < \omega_2 < \ldots < \omega_q$. This is done using Algorithm 1. Next, we set $\omega_0 := 0$ and $\omega_{q+1} := \infty$. For $j = 0, \ldots, q$, we choose points $\alpha_j \in (\omega_j, \omega_{j+1})$ with $\operatorname{rank}\left(\left[\mathrm{i}\alpha_j E - A \;\; -B\right]\right) = r$. Finally, for $j = 0, \ldots, q$ we compute the inertia $(\pi_+^j, \pi_0^j, \pi_-^j)$ of the Hermitian matrix $\mathcal{N}(\mathrm{i}\alpha_j)$ and thus obtain $\eta(\mathcal{N}(\mathrm{i}\alpha_j)) = \pi_+^j + \pi_0^j - \pi_-^j$. Then the system is dissipative if and only if $\eta(\mathcal{N}(\mathrm{i}\alpha_j)) = k + n + m - 2r$ for all j. Figure 9 summarizes the complete procedure in a diagram.

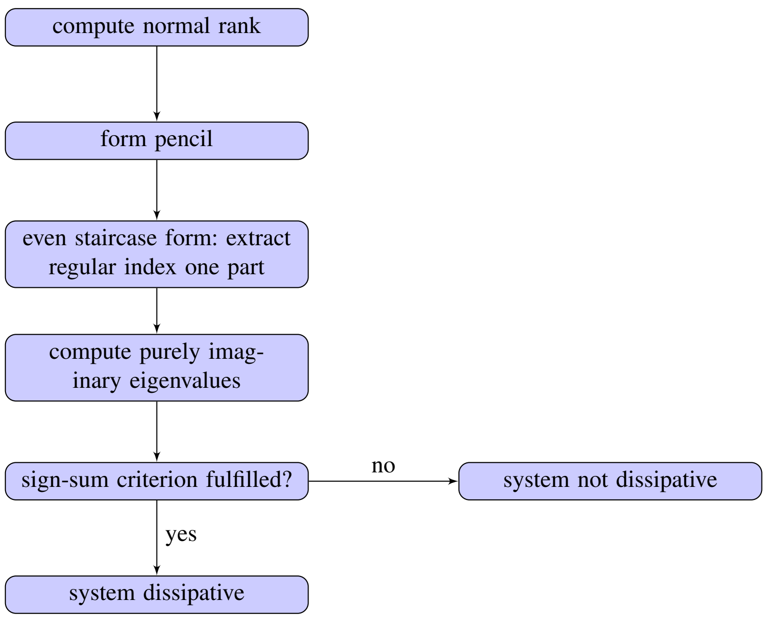

Fig. 9. Algorithm flowchart for dissipativity check

We further illustrate the algorithm by means of the following example.

Example 3. We consider a slightly modified circuit example from [116, Subsec. 1.1.1] resulting from a modified nodal analysis given by the following data:

$$E = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10^4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10^{-3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -10 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 10^{-2} & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$C = B^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \end{bmatrix}, \quad D = 0.$$

Due to the special structure of the matrices, the corresponding descriptor system is passive [96]. In particular, the system is cyclo-dissipative with respect to the supply rate defined by the weighting matrices

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} = \begin{bmatrix} 0 & I_2 \\ I_2 & 0 \end{bmatrix}.$$

We now verify this by checking the spectrum of the even pencil (42). First, we reduce the pencil to even staircase form. Using the notation of Theorem 6 we obtain the following structural information:

$$w = 3, \quad s_1 = 1, \quad s_2 = 1, \quad s_3 = 0, \quad q_1 = 1, \quad q_2 = 1, \quad q_3 = 1.$$

In particular, since $q_3 - s_3 = 1 \neq 0$, the pencil is singular. The regular index one part is given by

$$\begin{aligned} sN_{4,4} - M_{4,4} = -\operatorname{diag}\big(&1.9021, 1.6194, 1.414, 2.1756, 0.6217, \\ &-0.6144, -1.4142, -1.6167, -1.9021, 10.0504, \\ &1.9962, -0.0075, -1.1756, -1.9987, -10.0504\big) \in \mathbb{R}[s]^{15,15}, \end{aligned}$$

which has only (semisimple) infinite eigenvalues. Therefore, it is sufficient to evaluate the sign-sum function at a single point, for instance for $\omega = 0$ we obtain

$$\eta(\mathscr{N}(0)) = \eta(-M) = 2 = k + n + m - 2r.$$

Hence, it is confirmed that the system is cyclo-dissipative.

10 Conclusions

This paper provides a uniform treatment of differential-algebraic equations by methods from numerical linear algebra. First, we have presented the solution theory of such equations as well as regularization procedures. Based on that we have discussed several important applications from control and optimization of DAEs. These are based on the solution of even eigenvalue problems. We have presented several canonical forms of even pencils and discussed their properties. These canonical forms can be employed to numerically treat the presented applications in a uniform framework. The methods discussed here are usable for small-scale problems, i.e., to problems of size up to a few hundred. Here, the main computational bottleneck are the complexity and the storage requirements for solving even and skew-Hamiltonian/Hamiltonian eigenvalue problems.

Thus a big issue is the development of algorithms for large and sparse problems, which are widely unexplored. For instance, it is not clear how to determine *all* desired eigenvalues of a large-scale even pencil, e.g., the purely imaginary ones or how to approximate the complete subspace associated to all eigenvalues in the left half plane by a sparse representation.

References

1. Alam, R., Bora, S., Karow, M., Mehrmann, V., Moro, J.: Perturbation theory for Hamiltonian matrices and the distance to bounded-realness. SIAM J. Matrix Anal. Appl. **32**(2), 484–514 (2011)
2. Anderson, B.D.O., Vongpanitlerd, S.: Network Analysis and Synthesis. Prentice Hall, Englewood Cliffs, NJ (1973)
3. Antoulas, A.C.: Approximation of Large-Scale Dynamical Systems. Adv. Des. Control. SIAM, Philadelphia, PA (2005)
4. Antoulas, A.C., Mayo, A.J.: A framework for the solution of the generalized realization problem. Linear Algebra Appl. **425**, 634–662 (2007)
5. Ascher, U.M., Mattheij, R., Russell, R.: Numerical Solution of Boundary Value Problems for Ordinary Differential Equations, 2nd edn. SIAM Publications, Philadelphia, PA (1995)
6. Ashcraft, C., Grimes, R.G., Lewis, J.G.: Accurate symmetric indefinite linear equations solvers. SIAM J. Matrix Anal. Appl. **20**(2), 513–561 (1998)
7. Backes, A.: Extremalbedingungen für Optimierungs-Probleme mit Algebro-Differentialgleichungen. Dissertation, Institut für Mathematik, Humboldt-Universität zu Berlin, Berlin (2006)
8. Bals, J., Hofer, G., Pfeiffer, A., Schallert, C.: Virtual iron bird – A multidisciplinary modelling and simulation platform for new aircraft system architectures. In: Deutscher Luft- und Raumfahrtkongress. Friedrichshafen, Germany (2005)
9. Benner, P., Byers, R., Faßbender, H., Mehrmann, V., Watkins, D.: Cholesky-like factorizations of skew-symmetric matrices. Electron. Trans. Numer. Anal. **11**, 85–93 (2000)
10. Benner, P., Byers, R., Losse, P., Mehrmann, V., Xu, H.: Robust formulas for optimal H_∞ controllers. Automatica **47**(12), 2639–2646 (2011)
11. Benner, P., Byers, R., Mehrmann, V., Xu, H.: Numerical computation of deflating subspaces of embedded Hamiltonian pencils. Tech. Rep. SFB393/99-15, Fakultät für Mathematik, TU Chemnitz (1999). Available from `http://www.tu-chemnitz.de/sfb393/sfb99pr.html`
12. Benner, P., Byers, R., Mehrmann, V., Xu, H.: Numerical computation of deflating subspaces of skew-Hamiltonian/Hamiltonian pencils. SIAM J. Matrix Anal. Appl. **24**(1), 165–190 (2002)
13. Benner, P., Byers, R., Mehrmann, V., Xu, H.: A robust numerical method for the γ-iteration in $\mathscr{H}_\infty$ control. Linear Algebra Appl. **425**(2–3), 548–570 (2007)

14. Benner, P., Effenberger, C.: A rational SHIRA method for the Hamiltonian eigenvalue problem. Taiwanese J. Math. **14**(6A), 805–823 (2010)
15. Benner, P., Faßbender, H., Stoll, M.: A Hamiltonian Krylov-Schur-type method based on the symplectic Lanczos process. Linear Algebra Appl **435**(3), 578–600 (2011)
16. Benner, P., Mehrmann, V., Xu, H.: A new method for computing the stable invariant subspace of a real Hamiltonian matrix. J. Comput. Appl. Math. **86**(1), 17–43 (1997)
17. Benner, P., Mehrmann, V., Xu, H.: A numerically stable, structure preserving method for computing the eigenvalues of real Hamiltonian or symplectic pencils. Numer. Math. **78**(3), 329–358 (1998)
18. Benner, P., Sima, V., Voigt, M.: $\mathscr{L}_\infty$-norm computation for continuous-time descriptor systems using structured matrix pencils. IEEE Trans. Automat. Control **57**(1), 233–238 (2012)
19. Benner, P., Sima, V., Voigt, M.: Robust and efficient algorithms for $\mathscr{L}_\infty$-norm computation for descriptor systems. In: Proc. 7th IFAC Symposium on Robust Control Design, pp. 195–200. Aalborg, Denmark (2012). DOI 10.3182/20120620-3-DK-2025.00114
20. Benner, P., Sima, V., Voigt, M.: FORTRAN 77 subroutines for the solution of skew-Hamiltonian/Hamiltonian eigenproblems – Part I: Algorithms and applications. Preprint MPIMD/13-11, Max Planck Institute Magdeburg (2013). Available from `http://www.mpi-magdeburg.mpg.de/preprints/2013/11/`
21. Benner, P., Sima, V., Voigt, M.: FORTRAN 77 subroutines for the solution of skew-Hamiltonian/Hamiltonian eigenproblems – Part II: Implementation and numerical results. Preprint MPIMD/13-12, Max Planck Institute Magdeburg (2013). Available from `http://www.mpi-magdeburg.mpg.de/preprints/2013/12/`
22. Berger, T.: On differential-algebraic control systems. Dissertation, Fakultät für Mathematik und Naturwissenschaften, TU Ilmenau (2013)
23. Berger, T., Reis, T.: Controllability of linear differential-algebraic equations – a survey. In: A. Ilchmann, T. Reis (eds.) Surveys in Differential-Algebraic Equations I, Differential-Algebraic Equations Forum, pp. 1–61. Springer-Verlag, Berlin, Heidelberg (2013)
24. Berger, T., Reis, T., Trenn, S.: Observability of differential-algebraic equations – a survey. Hamburger Beiträge zur Angewandten Mathematik 2015-13, Fachbereich Mathematik, Universität Hamburg (2015). Available from `http://preprint.math.uni-hamburg.de/public/papers/hbam/hbam2015-13.pdf`
25. Berger, T., Trenn, S.: The quasi-Kronecker form for matrix pencils. SIAM J. Matrix Anal. Appl. **33**(2), 336–368 (2012)
26. Berger, T., Trenn, S.: Addition to "The quasi-Kronecker form for matrix pencils". SIAM J. Matrix Anal. Appl. **34**(1), 94–101 (2013)
27. Betcke, T., Higham, N.J., Mehrmann, V., Schröder, C., Tisseur, F.: NLEVP: A collection of nonlinear eigenvalue problems. ACM Trans. Math. Softw. **39**(2), Article 7 (2013)
28. Binder, A., Mehrmann, V., Międlar, A.: A MATLAB toolbox for the regularization of descriptor systems arising from generalized realization procedures (2014). In preparation
29. Bischof, C.H., Quintana-Ortí, G.: Algorithm 782: codes for rank-revealing QR factorizations of dense matrices. ACM Trans. Math. Softw. **24**(2), 254–257 (1998). DOI http://doi.acm.org/10.1145/290200.287638
30. Bischof, C.H., Quintana-Ortí, G.: Computing rank-revealing QR factorizations of dense matrices. ACM Trans. Math. Softw. **24**(2), 226–253 (1998). DOI http://doi.acm.org/10.1145/290200.287637

31. Bojanczyk, A.I., Golub, G.H., Van Dooren, P.: The periodic Schur decomposition. Algorithms and applications. In: F.T. Luk (ed.) Advanced Signal Processing Algorithms, Architectures, and Implementations III, *Proc. SPIE*, vol. 1770, pp. 31–42 (1992)
32. Boyd, S., Balakrishnan, V.: A regularity result for the singular values of a transfer matrix and a quadratically convergent algorithm for computing its L_∞-norm. Systems Control Lett. **15**(1), 1–7 (1990)
33. Boyd, S., Balakrishnan, V., Kabamba, P.: A bisection method for computing the H_∞ norm of a transfer matrix and related problems. Math. Control Signals Systems **2**(3), 207–219 (1989)
34. Bruinsma, N.A., Steinbuch, M.: A fast algorithm to compute the H_∞-norm of a transfer function matrix. Systems Control Lett. **14**(4), 287–293 (1990)
35. Brüll, T.: Checking dissipativity of linear behaviour systems given in kernel representation. Math. Control Signals Systems **23**(1–3), 159–175 (2011)
36. Brüll, T.: Dissipativity of linear quadratic systems. Dissertation, Institut für Mathematik, TU Berlin (2011). Available from `http://opus4.kobv.de/opus4-tuberlin/frontdoor/index/index/docId/2824`
37. Brüll, T., Mehrmann, V.: STCSSP: A FORTRAN 77 routine to compute a structured staircase form for a (skew-)symmetric/(skew-)symmetric matrix pencil. Preprint 31-2007, Institut für Mathematik, TU Berlin (2007)
38. Brüll, T., Schröder, C.: Dissipativity enforcement via perturbation of para-Hermitian pencils. IEEE Trans. Circuits Syst. I. Regul. Pap. **60**(1), 164–177 (2013)
39. Bunch, J.R.: A note on the stable decomposition of skew-symmetric matrices. Math. Comp. **38**(158), 475–479 (1982)
40. Bunse-Gerstner, A., Byers, R., Mehrmann, V., Nichols, N.K.: Feedback design for regularizing descriptor systems. Linear Algebra Appl. **299**, 119–151 (1999)
41. Bunse-Gerstner, A., Mehrmann, V., Nichols, N.K.: Regularization of descriptor systems by output feedback. IEEE Trans. Automat. Control **39**(4), 1742–1748 (1994)
42. Byers, R., Geerts, T., Mehrmann, V.: Descriptor systems without controllability at infinity. SIAM J. Control Optim. **35**(2), 462–479 (1997)
43. Byers, R., Mehrmann, V., Xu, H.: A structured staircase algorithm for skew-symmetric/symmetric pencils. Electron. Trans. Numer. Anal. **26**, 1–13 (2007)
44. Campbell, S.L.: Singular Systems of Differential Equations I. Pitman, San Francisco, CA (1980)
45. Campbell, S.L.: A general form for solvable linear time varying singular systems of differential equations. SIAM J. Math. Anal. **18**(4), 1101–1115 (1987)
46. Campbell, S.L.: Linearization of DAEs along trajectories. Z. Angew. Math. Phys. **46**(1), 70–84 (1995)
47. Campbell, S.L., Kunkel, P., Mehrmann, V.: Regularization of linear and nonlinear descriptor systems. In: L.T. Biegler, S.L. Campbell, V. Mehrmann (eds.) Control and Optimization with Differential-Algebraic Constraints, Adv. Des. Control, chap. 2, pp. 17–36. SIAM, Philadelphia, PA (2012)
48. Chu, D., Liu, X., Mehrmann, V.: A numerical method for computing the Hamiltonian Schur form. Numer. Math. **105**(3), 375–412 (2007)
49. Cobb, J.D.: Controllability, observability and duality in singular systems. IEEE Trans. Automat. Control **AC-29**(12), 1076–1082 (1984)
50. Dai, L.: Singular Control Systems, *Lecture Notes in Control and Inform. Sci.*, vol. 118. Springer-Verlag, Berlin/Heidelberg (1989)
51. Datta, S., Mehrmann, V.: Computation of state reachable points of linear time invariant descriptor systems. Preprint 17/2014, Institut für Mathematik, Technische Universität

Berlin (2014). Submitted, available from http://www.math.tu-berlin.de/preprints/
52. Demmel, J.W., Kågström, B.: Computing stable eigendecompositions of matrix pencils. Linear Algebra Appl. **88**, 139–186 (1987)
53. Demmel, J.W., Kågström, B.: The generalized Schur decomposition of an arbitrary pencil $\lambda A - B$, Part I. ACM Trans. Math. Software **19**, 160–174 (1993)
54. Demmel, J.W., Kågström, B.: The generalized Schur decomposition of an arbitrary pencil $\lambda A - B$, Part II. ACM Trans. Math. Software **19**, 185–201 (1993)
55. Drmač, Z., Bujanović, Z.: On the failure of rank-revealing QR factorization software. ACM Trans. Math. Softw. **35**(2), Article 12 (2008)
56. Edelmann, A., Elmroth, E., Kågström, B.: A geometric approach to perturbation theory of matrices and matrix pencils. Part I: Versal deformations. SIAM J. Matrix Anal. Appl. **18**, 653–692 (1997)
57. Edelmann, A., Elmroth, E., Kågström, B.: A geometric approach to perturbation theory of matrices and matrix pencils. Part II: A stratification-enhanced staircase algorithm. SIAM J. Matrix Anal. Appl. **20**, 667–699 (1999)
58. Eich-Soellner, E., Führer, C.: Numerical Methods in Multibody Dynamics. Teubner, Stuttgart (1998)
59. Fassbender, H.: Symplectic Methods for the Symplectic Eigenproblem. Kluwer Academic/Plenum Publishers, New York (2000)
60. Gantmacher, F.R.: The Theory of Matrices, vol. 2. Chelsea, New York (1959)
61. Gohberg, I., Lancaster, P., Rodman, L.: Matrix Polynomials. Academic Press Inc., New York (1982)
62. Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. Johns Hopkins Univ. Press, Baltimore (1996)
63. Günther, M., Rentrop, P.: Multirate ROW methods and latency of electric circuits. Appl. Numer. Math. **13**(1–3), 83–102 (1993)
64. Hagel, S.: Systematische Modellbildung mechanischer Systeme auf der Basis von Teilkomponenten. In: P.C.M. et al. (ed.) Notes of the Workshop "Identifizierung, Analyse und Entwurfsmethoden für mechanische Mehrkörpersysteme in Deskriptorform", pp. 67–72. Institut für Sicherheitstechnik, BUGH Wuppertal (1994)
65. Hahn, H., Wehage, R.: Dynamic simulation of terrain vehicles. In: W. Schiehlen (ed.) Multibody Systems Handbook, pp. 491–503. Springer-Verlag, Berlin (1990)
66. Hassibi, B., Sayed, A.H., Kailath, T.: Indefinite-Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories. SIAM Studies in Applied and Numerical Mathematics. SIAM, Philadelphia, PA (1999)
67. Hench, J.J., Laub, A.J.: Numerical solution of the discrete-time periodic Riccati equation. IEEE Trans. Automat. Control **39**(6), 1197–1210 (1994)
68. Hiller, M., Hirsch, K.: Multibody system dynamics and mechatronics. Z. Angew. Math. Mech. **86**(2), 87–109 (2006)
69. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (1985)
70. Kågström, B.: RGSVD – An algorithm for computing the Kronecker structure and reducing subspaces of singular $A - zB$ pencils. SIAM J. Sci. Stat. Comput. **7**(1), 185–211 (1986)
71. Kressner, D.: An efficient and reliable implementation of the periodic QZ algorithm. In: Proc. IFAC Workshop on Periodic Control Systems (2001)
72. Kressner, D., Schröder, C., Watkins, D.S.: Implicit QR algorithms for palindromic and even eigenvalue problems. Numer. Algorithms **51**(2), 209–238 (2009)

73. Kunkel, P., Mehrmann, V.: Generalized inverses of differential-algebraic operators. SIAM J. Matrix Anal. Appl. **17**(2), 426–442 (1996)
74. Kunkel, P., Mehrmann, V.: A new class of discretization methods for the solution of linear differential-algebraic equations. SIAM J. Numer. Anal. **5**, 1941–1961 (1996)
75. Kunkel, P., Mehrmann, V.: Differential-Algebraic Equations. Analysis and Numerical Solution. EMS Publishing House, Zürich, Switzerland (2006)
76. Kunkel, P., Mehrmann, V.: Optimal control for unstructured nonlinear differential-algebraic equations of arbitrary index. Math. Control Signals Systems **20**(3), 227–269 (2008)
77. Kunkel, P., Mehrmann, V.: Formal adjoints of linear DAE operators and their role in optimal control. Electron. J. Linear Algebra **22**, 672–693 (2011)
78. Kunkel, P., Mehrmann, V., Rath, W.: Analysis and numerical solution of control problems in descriptor form. Math. Control Signals Systems **14**, 29–61 (2001)
79. Kunkel, P., Mehrmann, V., Stöver, R.: Multiple shooting for unstructured nonlinear differential-algebraic equations of arbitrary index. SIAM J. Numer. Anal. **42**(6), 2277–2297 (2004)
80. Kunkel, P., Mehrmann, V., Stöver, R.: Symmetric collocation for unstructured nonlinear differential-algebraic equations of arbitrary index. Numer. Math. **98**(2), 277–304 (2004)
81. Kurina, G.A., März, R.: On linear-quadratic optimal control problems for time-varying descriptor systems. SIAM J. Control Optim. **42**(6), 2062–2077 (2004)
82. Lancaster, P.: Lambda-matrices and Vibrating Systems, *International Series of Monographs in Pure and Applied Mathematics*, vol. 94. Pergamon Press, Oxford (1966)
83. Landwehr, M., Lefarth, U., Wassmuth, E.: Parameter identification and optimization of nonlinear dynamic systems, exemplified by mechatronic systems. In: Computational Systems Analysis, pp. 257–262. Elsevier, Amsterdam, The Netherlands (1992)
84. Lin, W.W., Xu, S.F.: Convergence analysis of structure-preserving doubling algorithms for Riccati-type matrix equations. SIAM J. Matrix Anal. Appl. **28**(1), 26–39 (2006)
85. Losse, P.: The $\langle_\infty$ optimal control problem for descriptor systems. Dissertation, Fakultät für Mathematik, Technische Universität Chemnitz (2012). Available from `http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-83628`
86. Losse, P., Mehrmann, V., Poppe, L., Reis, T.: The modified optimal $\mathscr{H}_\infty$ control problem for descriptor systems. SIAM J. Control Optim. **47**(6), 2795–2811 (2008)
87. Mackey, D.S., Mackey, N., Mehl, C., Mehrmann, V.: Structured polynomial eigenvalue problems: Good vibrations from good linearizations. SIAM J. Matrix Anal. Appl. **28**(4), 1029–1051 (2006)
88. Mackey, D.S., Mackey, N., Mehl, C., Mehrmann, V.: Jordan structures of alternating matrix polynomials. Linear Algebra Appl. **432**, 867–891 (2010)
89. Mavaddat, R.: Network Scattering Parameters. Adv. Ser. Circuits Systems. World Scientific Publishing Company, River Edge, NJ (1996)
90. Mehl, C.: Condensed forms for skew-Hamiltonian/Hamiltonian pencils. SIAM J. Matrix Anal. Appl. **21**(2), 454–476 (2000)
91. Mehrmann, V., Schröder, C., Simoncini, V.: An implicitly-restarted Krylov method for real symmetric/skew-symmetric eigenproblems. Linear Algebra Appl. **436**(10), 4070–4087 (2012)
92. Mehrmann, V., Schröder, C., Watkins, D.S.: A new block method for computing the Hamiltonian Schur form. Linear Algebra Appl. **431**, 350–368 (2009)
93. Mehrmann, V., Voss, H.: Nonlinear eigenvalue problems: A challenge for modern eigenvalue methods. GAMM-Mitt. **27**, 121–151 (2005)

94. Mehrmann, V., Watkins, D.: Structure-preserving methods for computing eigenpairs of large sparse skew-Hamiltonian/Hamiltonian pencils. SIAM J. Sci. Comput. **22**(6), 1905–1925 (2001)
95. Mehrmann, V., Xu, H.: Structure preserving deflation of infinite eigenvalues in structured pencils. Preprint 1052, DFG Research Center MATHEON, Berlin, Germany (2014). Available from `http://opus4.kobv.de/opus4-matheon`
96. Reis, T.: Circuit synthesis of passive descriptor systems – a modified nodal approach. Int. J. Circuit Theory Appl. **38**(1), 44–68 (2010)
97. Reis, T., Rendel, O., Voigt, M.: The Kalman-Yakubovich-Popov inequality for differential-algebraic systems. Hamburger Beiträge zur Angewandten Mathematik 2014-27, Fachbereich Mathematik, Universität Hamburg (2014). Available from `http://preprint.math.uni-hamburg.de/public/papers/hbam/hbam2014-27.pdf`
98. Reis, T., Stykel, T.: PABTEC: Passivity-preserving balanced truncation for electrical circuits. IEEE Trans. Circuits Syst. I. Regul. Pap. **29**(9), 1354–1367 (2010)
99. Reis, T., Stykel, T.: Passivity-preserving balanced truncation model reduction of circuit equations. In: J. Roos, L. Costa (eds.) Scientific Computing in Electrical Engineering SCEE 2008, *Math. Ind.*, vol. 14, pp. 483–490. Springer-Verlag, Berlin/Heidelberg (2010)
100. Riaza, R.: Differential-Algebraic Systems. Analytical Aspects and Circuit Applications. World Scientific Publishing Co. Pte. Ltd., Singapore (2008)
101. Roberson, R.E., Schwertassek, R.: Dynamics of Multibody Systems. Springer-Verlag, Heidelberg, Germany (1988)
102. Rosenbrock, H.H.: Structural properties of linear dynamical systems. Int. J. Control **20**(2), 191–202 (1974)
103. Scherer, C.: $\mathscr{H}_\infty$-control by state-feedback and fast algorithms for the computation of optimal $\mathscr{H}_\infty$-norms. IEEE Trans. Automat. Control **35**(10), 1090–1099 (1990)
104. Schiehlen, W. (ed.): Advanced Multibody System Dynamics – Simulation and Sofware Tools, *Solid Mech. Appl.*, vol. 20. Kluwer, Dordrecht (1993)
105. Schlacher, K., Kugi, A.: Automatic control of mechatronic systems. Int. J. Appl. Math. Comput. Sci. **11**(1), 131–164 (2001)
106. Schlacher, K., Kugi, A., Scheidl, R.: Tensor analysis based symbolic computation for mechatronic systems. Math. Comput. Simulation **46**(5-6), 517–525 (1998)
107. Schröder, C.: Palindromic and even eigenvalue problems – analysis and numerical methods. Dissertation, Institut für Mathematik, TU Berlin (2008). Available from `http://opus4.kobv.de/opus4-tuberlin/frontdoor/index/index/docId/1770`
108. Simeon, B., Grupp, F., Führer, C., Rentrop, P.: A nonlinear truck model and its treatment as multibody system. J. Comput. Appl. Math. **50**(1–3), 523–532 (1994)
109. Stewart, G.W., Sun, J.G.: Matrix Perturbation Theory. Academic Press, New York (1990)
110. Thompson, R.C.: Pencils of complex and real symmetric and skew matrices. Linear Algebra Appl. **147**, 323–371 (1991)
111. Tisseur, F., Meerbergen, K.: The quadratic eigenvalue problem. SIAM Rev. **43**(2), 235–286 (2001). DOI 10.1137/S0036144500381988. URL `http://dx.doi.org/10.1137/S0036144500381988`
112. Trautenberg, W.: SIMPACK 8.9. Available from `www.simpack.com`
113. Van Dooren, P.: The computation of Kronecker's canonical form of a singular pencil. Linear Algebra Appl. **27**, 103–121 (1979)

114. Varga, A.: Computation of irreducible generalized state-space realizations. Kybernetica **26**(2), 89–106 (1990)
115. Verghese, G.C., Lévy, B.C., Kailath, T.: A generalized state-space for singular systems. IEEE Trans. Automat. Control **AC-26**(4), 811–831 (1981)
116. Voigt, M.: $\mathscr{L}_\infty$-norm computation for descriptor systems. Diploma thesis, Chemnitz University of Technology, Faculty of Mathematics, Germany (2010). Available from `http://nbn-resolving.de/urn:nbn:de:bsz:ch1-201001050`
117. Voigt, M.: On linear-quadratic optimal control and robustness of differential-algebriac systems. Dissertation, Otto-von-Guericke-Universität Magdeburg, Fakultät für Mathematik (2015). Submitted
118. Willems, J.C.: Least squares stationary optimal control and the algebraic Riccati equation. IEEE Trans. Automat. Control **AC-16**(6), 621–634 (1971)
119. Willems, J.C.: Dissipative dynamical systems – Part I: General theory. Arch. Ration. Mech. Anal. **45**, 321–351 (1972)
120. Willems, J.C.: Dissipative dynamical systems – Part II: Linear systems with quadratic supply rates. Arch. Ration. Mech. Anal. **45**, 352–393 (1972)
121. Willems, J.C.: On the existence of a nonpositive solution to the algebraic Riccati equation. IEEE Trans. Automat. Control **AC-19**(10), 592–593 (1974)
122. Zhou, K., Doyle, J.C., Glover, K.: Robust and Optimal Control. Prentice-Hall, Upper Saddle River, NJ (1995)

Index