Balanced Truncation Model Reduction for Large-Scale Systems in Descriptor Form

Volker Mehrmann, Tatjana Stykel

Technical Report 157

DFG Research Center Matheon *Mathematics for key technologies*
http://www.math.tu-berlin.de/DFG-Forschungszentrum/preprints/

Balanced Truncation Model Reduction for Large Scale Systems in Descriptor Form

Volker Mehrmann[1] and Tatjana Stykel[2]

[1] Institut für Mathematik, MA 4-5, Technische Universität Berlin, Straße des 17. Juni 136, D-10623 Berlin, Germany, mehrmann@math.tu-berlin.de. Supported by the DFG Research Center Matheon "Mathematics for key technologies" in Berlin.

[2] Institut für Mathematik, MA 3-3, Technische Universität Berlin, Straße des 17. Juni 136, D-10623 Berlin, Germany, stykel@math.tu-berlin.de. Supported by the DFG Research Center Matheon "Mathematics for key technologies" in Berlin.

Summary. In this paper we give a survey on balanced truncation model order reduction for linear time-invariant continuous-time systems in descriptor form. We first give a brief overview of the basic concepts from linear system theory and then present balanced truncation model reduction methods for descriptor systems and discuss their algorithmic aspects. The efficiency of these methods is demonstrated by numerical experiments.

1 Introduction

We study model order reduction for linear time-invariant continuous-time systems

$$\begin{aligned} E\,\dot{x}(t) &= A\,x(t) + B\,u(t), \quad x(0) = x_0, \\ y(t) &= C\,x(t), \end{aligned} \tag{1}$$

where $E, A \in \mathbb{R}^{n,n}$, $B \in \mathbb{R}^{n,m}$, $C \in \mathbb{R}^{p,n}$, $x(t) \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}^m$ is the control input, $y(t) \in \mathbb{R}^p$ is the output and $x_0 \in \mathbb{R}^n$ is the initial value. The number of state variables n is called the *order* of system (1). If $I = E$, then (1) is a *standard state space system*. Otherwise, (1) is a *descriptor system* or *generalized state space system*. Such systems arise in a variety of applications including multibody dynamics with constrains, electrical circuit simulation and semidiscretization of partial differential equations, see [Ber90, BCP89, Cam80, Dai89, GF99, Sch95].

Modeling of complex physical and technical processes such as fluid flow, very large system integrated (VLSI) chip design or mechanical systems simulation, leads to descriptor systems of very large order n, while the number m of inputs and the number p of outputs are typically small compared to n.

Despite the ever increasing computational speed, simulation, optimization or real time controller design for such large scale systems is difficult because of storage requirements and expensive computations. In this case *model order reduction* plays an important role. It consists in approximating the descriptor system (1) by a reduced-order system

$$\begin{aligned} \widetilde{E}\,\dot{\widetilde{x}}(t) &= \widetilde{A}\,\widetilde{x}(t) + \widetilde{B}\,u(t), \quad \widetilde{x}(0) = \widetilde{x}_0, \\ \widetilde{y}(t) &= \widetilde{C}\,\widetilde{x}(t), \end{aligned} \tag{2}$$

where $\widetilde{E}$, $\widetilde{A} \in \mathbb{R}^{\ell,\ell}$, $\widetilde{B} \in \mathbb{R}^{\ell,m}$, $\widetilde{C} \in \mathbb{R}^{p,\ell}$ and $\ell \ll n$. Note that systems (1) and (2) have the same input $u(t)$. We require the approximate model (2) to preserve properties of the original system (1) like regularity, stability and passivity. It is also desirable for the approximation error to be small. Moreover, the computation of the reduced-order system should be numerically reliable and efficient.

There exist various model reduction approaches for standard state space systems such as balanced truncation [LHPW87, Moo81, SC89, TP84, Var87], moment matching approximation [Bai02, FF95, Fre00, GGV94], singular perturbation approximation [LA89] and optimal Hankel norm approximation [Glo84]. Surveys on standard state space system approximation and model reduction can be found in [Ant04, ASG01, FNG92].

A popular model reduction technique for large scale standard state space systems is *moment matching approximation* considered first in [FF95, GGV94]. This approach consists in projecting the dynamical system onto Krylov subspaces computed by an Arnoldi or Lanczos process. Krylov subspace methods are attractive for large scale sparse systems, since only matrix-vector multiplications are required, and they can easily be generalized for descriptor systems, e.g., [BF01, Fre00, GGV96, Gri97]. Drawbacks of this technique are that stability and passivity are not necessarily preserved in the reduced-order system and that there is no global approximation error bound, see[Bai02, BF01, BSSY99, Bea04, Gug03] for recent contributions on this topic.

Balanced truncation [LHPW87, Moo81, SC89, TP84, Var87] is another well studied model reduction approach for standard state space systems. The method makes use of the two Lyapunov equations

$$A\mathcal{P} + \mathcal{P}A^T = -BB^T, \qquad A^T\mathcal{Q} + \mathcal{Q}A = -C^TC.$$

The solutions $\mathcal{P}$ and $\mathcal{Q}$ of these equations are called the *controllability* and *observability Gramians*, respectively. The balanced truncation method consists in transforming the state space system into a balanced form whose controllability and observability Gramians become diagonal and equal, together with a truncation of those states that are both difficult to reach and to observe [Moo81]. An important property of this method is that the asymptotical stability is preserved in the reduced-order system. Moreover, the existence of a priory error bounds [Enn84, Glo84] allows an adaptive choice of the

state space dimension ℓ of the reduced model depending on how accurate the approximation is needed. A difficulty in balanced truncation model reduction for large scale problems is that two matrix Lyapunov equations have to be solved. However, recent results on low rank approximations to the solutions of Lyapunov equations [ASG03, Gra04, LW02, Pen99, Pen00b] make the balanced truncation model reduction approach attractive for large scale systems, see [Li00, LWW99, Pen99b]. The extension of balanced truncation model reduction to descriptor systems has only recently been considered in [LS00, PS94, Sty04a, Sty04b].

In this paper we briefly review some basic linear system concepts including fundamental solution matrix, transfer function, realizations, controllability and observability Gramians, Hankel operators as well as Hankel singular values that play a key role in balanced truncation. We also present generalizations of balanced truncation model reduction methods for descriptor systems and discuss their numerical aspects.

Throughout the paper we will denote by $\mathbb{R}^{n,m}$ the space of $n \times m$ real matrices. The complex plane is denoted by $\mathbb{C}$, the open left half-plane is denoted by $\mathbb{C}^-$, and $\mathrm{i}\mathbb{R}$ is the imaginary axis. Furthermore, $\mathbb{R}^- = (-\infty, 0)$ and $\mathbb{R}_0^+ = [\,0, \infty\,)$. The matrix A^T stands for the transpose of $A \in \mathbb{R}^{n,m}$ and $A^{-T} = (A^{-1})^T$. We will denote by $\mathrm{rank}(A)$ the rank, by $\mathrm{Im}\,(A)$ the image and by $\mathrm{Ker}\,(A)$ the null space of a matrix A. An identity matrix of order n is denoted by I_n. We will use $\mathbb{L}_2^m(\mathbb{I})$ to denote the Hilbert space of vector-valued functions of dimension m whose elements are quadratically integrable on $\mathbb{I}$, where $\mathbb{I} \subseteq \mathbb{R}$ or $\mathbb{I} = \mathrm{i}\mathbb{R}$.

2 Descriptor systems

In this section we give a brief overview of linear system concepts and discuss the main differences between standard state space systems and systems in descriptor form.

Consider the continuous-time descriptor system (1). Assume that the pencil $\lambda E - A$ is *regular*, i.e., $\det(\lambda E - A) \neq 0$ for some $\lambda \in \mathbb{C}$. In this case $\lambda E - A$ can be reduced to the *Weierstrass canonical form* [SS90]. There exist nonsingular matrices W and T such that

$$E = W \begin{bmatrix} I_{n_f} & 0 \\ 0 & N \end{bmatrix} T \qquad \text{and} \qquad A = W \begin{bmatrix} J & 0 \\ 0 & I_{n_\infty} \end{bmatrix} T, \tag{3}$$

where J and N are matrices in Jordan canonical form and N is nilpotent with index of nilpotency ν. The numbers n_f and n_∞ are the dimensions of the deflating subspaces of $\lambda E - A$ corresponding to the finite and infinite eigenvalues, respectively, and ν is the *index* of the pencil $\lambda E - A$ and also the index of the descriptor system (1). The matrices

$$P_r = T^{-1} \begin{bmatrix} I_{n_f} & 0 \\ 0 & 0 \end{bmatrix} T \qquad \text{and} \qquad P_l = W \begin{bmatrix} I_{n_f} & 0 \\ 0 & 0 \end{bmatrix} W^{-1} \tag{4}$$

are the *spectral projections* onto the right and left deflating subspaces of the pencil $\lambda E - A$ corresponding to the finite eigenvalues.

Using the Weierstrass canonical form (3), we obtain the following Laurent expansion at infinity for the *generalized resolvent*

$$(\lambda E - A)^{-1} = \sum_{k=-\infty}^{\infty} F_k \lambda^{-k-1}, \tag{5}$$

where the coefficients F_k have the form

$$F_k = \begin{cases} T^{-1} \begin{bmatrix} J^k & 0 \\ 0 & 0 \end{bmatrix} W^{-1}, & k = 0, 1, 2 \ldots, \\ T^{-1} \begin{bmatrix} 0 & 0 \\ 0 & -N^{-k-1} \end{bmatrix} W^{-1}, & k = -1, -2, \ldots . \end{cases} \tag{6}$$

Let the matrices

$$W^{-1}B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \qquad \text{and} \qquad CT^{-1} = [\, C_1, \; C_2 \,]$$

be partitioned in blocks conformally to E and A in (3). Under the coordinate transformation $Tx(t) = \big[\, z_1^T(t), \; z_2^T(t) \,\big]^T$, system (1) is decoupled in the *slow* subsystem

$$\dot{z}_1(t) = Jz_1(t) + B_1 u(t), \quad z_1(0) = z_1^0, \tag{7}$$

and the *fast* subsystem

$$N\dot{z}_2(t) = z_2(t) + B_2 u(t), \quad z_2(0) = z_2^0 \tag{8}$$

with $y(t) = C_1 z_1(t) + C_2 z_2(t)$ and $Tx_0 = \big[\, (z_1^0)^T, \; (z_2^0)^T \,\big]^T$.

Equation (7) has a unique solution for any integrable input $u(t)$ and any given initial value $z_1^0 \in \mathbb{R}^{n_f}$, see [Kai80]. This solution has the form

$$z_1(t) = e^{tJ} z_1^0 + \int_0^t e^{(t-\tau)J} B_1 u(\tau)\, d\tau.$$

The unique solution of (8) is given by

$$z_2(t) = -\sum_{k=0}^{\nu-1} N^k B_2 u^{(k)}(t). \tag{9}$$

We see from (9) that for the existence of a classical smooth solution $z(t)$, it is necessary that the input function $u(t)$ is sufficiently smooth and the initial value z_2^0 satisfies

$$z_2^0 = -\sum_{k=0}^{\nu-1} N^k B_2 u^{(k)}(0).$$

Therefore, unlike standard state space systems, the initial value x_0 of the descriptor system (1) has to be *consistent*, i.e., it must satisfy the condition

$$(I - P_r)x_0 = \sum_{k=0}^{\nu-1} F_{-k-1} B u^{(k)}(0),$$

where P_r is the spectral projector as in (4) and the matrices F_k are given in (6).

Thus, if the pencil $\lambda E - A$ is regular, $u(t)$ is ν times continuously differentiable and the initial value x_0 is consistent, then system (1) has a unique, continuously differentiable solution $x(t)$ given by

$$x(t) = \mathcal{F}(t)Ex_0 + \int_0^t \mathcal{F}(t-\tau)Bu(\tau)\,d\tau + \sum_{k=0}^{\nu-1} F_{-k-1} B u^{(k)}(t),$$

where

$$\mathcal{F}(t) = T^{-1} \begin{bmatrix} \mathrm{e}^{tJ} & 0 \\ 0 & 0 \end{bmatrix} W^{-1} \tag{10}$$

is a *fundamental solution matrix* of the descriptor system (1).

If the initial condition x_0 is inconsistent or the input $u(t)$ is not sufficiently smooth, then the solution of the descriptor system (1) may have impulsive modes [Cob84, Dai89].

2.1 The transfer function

Consider the *Laplace transform* of a function $f(t)$, $t \in \mathbb{R}$, given by

$$\mathbf{f}(s) = \mathfrak{L}[f(t)] = \int_0^\infty e^{-st} f(t)\,dt, \tag{11}$$

where s is a complex variable called *frequency*. A discussion of the convergence region of the integral (11) in the complex plane and properties of the Laplace transform may be found in [Doe71]. Applying the Laplace transform to (1) and taking into account that $\mathfrak{L}[\dot{x}(t)] = s\mathbf{x}(s) - x(0)$, we have

$$\mathbf{y}(s) = C(sE - A)^{-1}B\mathbf{u}(s) + C(sE - A)^{-1}Ex(0), \tag{12}$$

where $\mathbf{x}(s)$, $\mathbf{u}(s)$ and $\mathbf{y}(s)$ are the Laplace transforms of $x(t)$, $u(t)$ and $y(t)$, respectively. The rational matrix-valued function

$$\mathbf{G}(s) = C(sE - A)^{-1}B$$

is called the *transfer function* of the continuous-time descriptor system (1). Equation (12) shows that if $Ex(0) = 0$, then $\mathbf{G}(s)$ gives the relation between the Laplace transforms of the input $u(t)$ and the output $y(t)$. In other words, $\mathbf{G}(s)$ describes the input-output behavior of (1) in the frequency domain.

A *frequency response* of the descriptor system (1) is given by $\mathbf{G}(\mathrm{i}\omega)$, i.e., the values of the transfer function on the imaginary axis. For an input function $u(t) = \mathrm{e}^{\mathrm{i}\omega t}u_0$ with $\omega \in \mathbb{R}$ and $u_0 \in \mathbb{R}^m$, we get from (1) that $y(t) = \mathbf{G}(\mathrm{i}\omega)\mathrm{e}^{\mathrm{i}\omega t}u_0$. Thus, the frequency response $\mathbf{G}(\mathrm{i}\omega)$ gives a transfer relation from the periodic input $u(t) = \mathrm{e}^{\mathrm{i}\omega t}u_0$ into the output $y(t)$.

Definition 2.1. *The transfer function* $\mathbf{G}(s)$ *is called* proper *if* $\lim\limits_{s\to\infty} \mathbf{G}(s) < \infty$, *and* improper *otherwise. If* $\lim\limits_{s\to\infty} \mathbf{G}(s) = 0$, *then* $\mathbf{G}(s)$ *is called* strictly proper.

Using the generalized resolvent equation (5), the transfer function $\mathbf{G}(s)$ can be expanded into a Laurent series at $s = \infty$ as

$$\mathbf{G}(s) = \sum_{k=-\infty}^{\infty} CF_{k-1}Bs^{-k},$$

where $CF_{k-1}B$ are the *Markov parameters* of system (1). Note that $CF_{k-1}B = 0$ for $k \le -\nu$, where ν is the index of the pencil $\lambda E - A$. One can see that the transfer function $\mathbf{G}(s)$ is additively decomposed as $\mathbf{G}(s) = \mathbf{G}_{sp}(s) + \mathbf{P}(s)$, where

$$\mathbf{G}_{sp}(s) = \sum_{k=1}^{\infty} CF_{k-1}Bs^{-k} \qquad \text{and} \qquad \mathbf{P}(s) = \sum_{k=-\nu+1}^{0} CF_{k-1}Bs^{-k} \tag{13}$$

are, respectively, the *strictly proper part* and the *polynomial part* of $\mathbf{G}(s)$. The transfer function $\mathbf{G}(s)$ is strictly proper if and only if $CF_{k-1}B = 0$ for $k \le 0$. Moreover, $\mathbf{G}(s)$ is proper if and only if $CF_{k-1}B = 0$ for $k < 0$. Obviously, if the pencil $\lambda E - A$ is of index at most one, then $\mathbf{G}(s)$ is proper.

Let $\mathbb{H}_\infty$ be a space of all proper rational transfer functions that are analytic and bounded in the closed right half-plane. The $\mathbb{H}_\infty$-*norm* of $\mathbf{G}(s) \in \mathbb{H}_\infty$ is defined via

$$\|\mathbf{G}\|_{\mathbb{H}_\infty} = \sup_{\mathbf{u}\neq 0} \frac{\|\mathbf{G}\mathbf{u}\|_{\mathbb{L}_2^p(\mathrm{i}\mathbb{R})}}{\|\mathbf{u}\|_{\mathbb{L}_2^m(\mathrm{i}\mathbb{R})}} = \sup_{\omega\in\mathbb{R}} \|\mathbf{G}(i\omega)\|_2,$$

where $\|\cdot\|_2$ denotes the spectral matrix norm. By the Parseval identity [Rud87] we have $\|\mathbf{G}\|_{\mathbb{H}_\infty} = \sup\limits_{u\neq 0} \|y\|_{\mathbb{L}_2^p(\mathbb{R})}/\|u\|_{\mathbb{L}_2^m(\mathbb{R})}$, i.e., the $\mathbb{H}_\infty$-norm of $\mathbf{G}(s)$ gives the ratio of the output energy to the input energy of the descriptor system (1).

2.2 Controllability and observability

In contrast to standard state space systems, for descriptor systems, there are several different notions of controllability and observability, see [BBMN99, Cob84, Dai89, YS81] and the references therein. We consider only complete controllability and observability here.

Definition 2.2. *The descriptor system* (1) *is called* completely controllable (C-controllable) *if* $\operatorname{rank}[\,\alpha E - \beta A,\ \ B\,] = n$ *for all* $(\alpha, \beta) \in (\mathbb{C}\times\mathbb{C}) \setminus \{(0,0)\}$.

C-controllability implies that for any given initial state $x_0 \in \mathbb{R}^n$ and final state $x_f \in \mathbb{R}^n$, there exists a control input $u(t)$ that transfers the system from x_0 to x_f in finite time. This notion follows [BBMN99, YS81] and is consistent with the definition of *controllability* given in [Dai89].

Observability is the dual property of controllability.

Definition 2.3. *The descriptor system* (1) *is called* completely observable (C-observable) *if* $\operatorname{rank}[\,\alpha E^T - \beta A^T,\ C^T\,] = n$ *for all* $(\alpha, \beta) \in (\mathbb{C}\times\mathbb{C})\backslash\{(0,0)\}$.

C-observability implies that if the output is zero for all solutions of the descriptor system (1) with a zero input, then this system has only the trivial solution.

The following theorem gives equivalent conditions for system (1) to be C-controllable and C-observable.

Theorem 2.4. [YS81] *Consider a descriptor system* (1)*, where* $\lambda E - A$ *is regular.*

1. *System* (1) *is C-controllable if and only if* $\operatorname{rank}[\,\lambda E - A,\ B\,] = n$ *for all finite* $\lambda \in \mathbb{C}$ *and* $\operatorname{rank}[\,E,\ B\,] = n$.
2. *System* (1) *is C-observable if and only if* $\operatorname{rank}[\,\lambda E^T - A^T,\ C^T\,] = n$ *for all finite* $\lambda \in \mathbb{C}$ *and* $\operatorname{rank}[\,E^T,\ C^T\,] = n$.

Other equivalent algebraic and geometric characterizations of controllability and observability for descriptor systems can be found in [Cob84, Dai89].

2.3 Stability

In this subsection we present some results from [Dai89, Sty02c] on stability for the descriptor system (1).

Definition 2.5. *The descriptor system* (1) *is called* asymptotically stable *if* $\lim_{t\to\infty} x(t) = 0$ *for all solutions* $x(t)$ *of the homogeneous system* $E\dot{x}(t) = Ax(t)$.

The following theorem collects equivalent conditions for system (1) to be asymptotically stable.

Theorem 2.6. [Dai89, Sty02c] *Consider a descriptor system* (1) *with a regular pencil* $\lambda E - A$*. The following statements are equivalent.*

1. *System* (1) *is asymptotically stable.*
2. *All finite eigenvalues of the pencil* $\lambda E - A$ *lie in the open left half-plane.*
3. *The projected generalized continuous-time Lyapunov equation*

$$E^T X A + A^T X E = -P_r^T Q P_r, \qquad X = P_l^T X P_l$$

has a unique Hermitian, positive semidefinite solution X *for every Hermitian, positive definite matrix* Q.

In the sequel, the pencil $\lambda E - A$ will be called *c-stable* if it is regular and all the finite eigenvalues of $\lambda E - A$ have negative real part. Note that the infinite eigenvalues of $\lambda E - A$ do not affect the behavior of the homogeneous system at infinity.

2.4 Gramians and Hankel singular values

Assume that the pencil $\lambda E - A$ is c-stable. Then the integrals

$$\mathcal{G}_{pc} = \int_0^\infty \mathcal{F}(t) B B^T \mathcal{F}^T(t)\, dt \quad \text{and} \quad \mathcal{G}_{po} = \int_0^\infty \mathcal{F}^T(t) C^T C \mathcal{F}(t)\, dt$$

exist, where $\mathcal{F}(t)$ is as in (10). The matrix $\mathcal{G}_{pc}$ is called the *proper controllability Gramian* and the matrix $\mathcal{G}_{po}$ is called the *proper observability Gramian* of the continuous-time descriptor system (1), see [Ben97, Sty02a]. The *improper controllability Gramian* and the *improper observability Gramian* of system (1) are defined by

$$\mathcal{G}_{ic} = \sum_{k=-\nu}^{-1} F_k B B^T F_k^T \quad \text{and} \quad \mathcal{G}_{io} = \sum_{k=-\nu}^{-1} F_k^T C^T C F_k,$$

respectively. Here the matrices F_k are as in (6). If $E = I$, then $\mathcal{G}_{pc}$ and $\mathcal{G}_{po}$ are the usual controllability and observability Gramians for standard state space systems [Glo84]. Using the Parseval identity [Rud87], the Gramians can be rewritten in frequency domain as

$$\begin{aligned}
\mathcal{G}_{pc} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} (\mathrm{i}\omega E - A)^{-1} P_l B B^T P_l^T (-\mathrm{i}\omega E - A)^{-T} dt,\\
\mathcal{G}_{po} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} (-\mathrm{i}\omega E - A)^{-T} P_r^T C^T C P_r (\mathrm{i}\omega E - A)^{-1} dt,\\
\mathcal{G}_{ic} &= \frac{1}{2\pi} \int_{0}^{2\pi} (e^{\mathrm{i}\omega} E - A)^{-1} (I - P_l) B B^T (I - P_l)^T (e^{-\mathrm{i}\omega} E - A)^{-T} dt,\\
\mathcal{G}_{io} &= \frac{1}{2\pi} \int_{0}^{2\pi} (e^{-\mathrm{i}\omega} E - A)^{-T} (I - P_r)^T C^T C (I - P_r) (e^{\mathrm{i}\omega} E - A)^{-1} dt.
\end{aligned}$$

It has been proven in [Sty02a] that the proper controllability and observability Gramians are the unique symmetric, positive semidefinite solutions of the *projected generalized continuous-time Lyapunov equations*

$$E\,\mathcal{G}_{pc} A^T + A\,\mathcal{G}_{pc} E^T = -P_l B B^T P_l^T, \qquad \mathcal{G}_{pc} = P_r \mathcal{G}_{pc} P_r^T, \tag{14}$$

$$E^T \mathcal{G}_{po} A + A^T \mathcal{G}_{po} E = -P_r^T C^T C P_r, \qquad \mathcal{G}_{po} = P_l^T \mathcal{G}_{po} P_l. \tag{15}$$

Furthermore, the improper controllability and observability Gramians are the unique symmetric, positive semidefinite solutions of the *projected generalized discrete-time Lyapunov equations*

$$A\,\mathcal{G}_{ic} A^T - E\,\mathcal{G}_{ic} E^T = (I - P_l) B B^T (I - P_l)^T, \quad P_r \mathcal{G}_{ic} P_r^T = 0, \tag{16}$$

$$A^T \mathcal{G}_{io} A - E^T \mathcal{G}_{io} E = (I - P_r)^T C^T C (I - P_r), \quad P_l^T \mathcal{G}_{io} P_l = 0. \tag{17}$$

Similarly to standard state space systems [Glo84], the controllability and observability Gramians can be used to define Hankel singular values for the

descriptor system (1) that are of great importance in model reduction via balanced truncation.

Consider the matrices $\mathcal{G}_{pc}E^T\mathcal{G}_{po}E$ and $\mathcal{G}_{ic}A^T\mathcal{G}_{io}A$. These matrices play the same role for descriptor systems as the product of the controllability and observability Gramians for standard state space systems [Glo84, ZDG96]. It has been shown in [Sty04b] that all the eigenvalues of $\mathcal{G}_{pc}E^T\mathcal{G}_{po}E$ and $\mathcal{G}_{ic}A^T\mathcal{G}_{io}A$ are real and non-negative. The square roots of the largest n_f eigenvalues of the matrix $\mathcal{G}_{pc}E^T\mathcal{G}_{po}E$, denoted by ς_j, are called the *proper Hankel singular values* of the continuous-time descriptor system (1). The square roots of the largest n_∞ eigenvalues of the matrix $\mathcal{G}_{ic}A^T\mathcal{G}_{io}A$, denoted by θ_j, are called the *improper Hankel singular values* of system (1). Recall that n_f and n_∞ are the dimensions of the deflating subspaces of the pencil $\lambda E - A$ corresponding to the finite and infinite eigenvalues, respectively.

We will assume that the proper and improper Hankel singular values are ordered decreasingly, i.e., $\varsigma_1 \geq \varsigma_2 \geq \ldots \geq \varsigma_{n_f} \geq 0$ and $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_{n_\infty} \geq 0$. For $E = I$, the proper Hankel singular values are the classical Hankel singular values of standard state space systems [Glo84, Moo81].

Since the proper and improper controllability and observability Gramians are symmetric and positive semidefinite, there exist *Cholesky factorizations*

$$\begin{aligned} \mathcal{G}_{pc} &= R_p R_p^T, \qquad & \mathcal{G}_{po} &= L_p L_p^T, \\ \mathcal{G}_{ic} &= R_i R_i^T, \qquad & \mathcal{G}_{io} &= L_i L_i^T, \end{aligned} \tag{18}$$

where the matrices R_p, L_p, R_i, $L_i \in \mathbb{R}^{n,n}$ are Cholesky factors [GV96] of the Gramians. In this case the proper Hankel singular values of system (1) can be computed as the n_f largest singular values of the matrix $L_p^T E R_p$, and the improper Hankel singular values of (1) are the n_∞ largest singular values of the matrix $L_i^T A R_i$, see [Sty04b].

For the descriptor system (1), we consider a *proper Hankel operator* $\mathcal{H}_p$ that transforms the past inputs $u_-(t)$ ($u_-(t) = 0$ for $t \geq 0$) into the present and future outputs $y_+(t)$ ($y_+(t) = 0$ for $t < 0$) through the state $x(0) \in \mathrm{Im}\,(P_r)$, see [Sty03]. This operator is defined via

$$y_+(t) = (\mathcal{H}_p u_-)(t) = \int_{-\infty}^{0} G_{sp}(t-\tau)u_-(\tau)\,d\tau, \qquad t \geq 0, \tag{19}$$

where $G_{sp}(t) = C\mathcal{F}(t)B$, $t \geq 0$. If the pencil $\lambda E - A$ is c-stable, then $\mathcal{H}_p$ acts from $\mathbb{L}_2^m(\mathbb{R}^-)$ into $\mathbb{L}_2^p(\mathbb{R}_0^+)$. In this case one can show that $\mathcal{H}_p$ is a Hilbert-Schmidt operator and its non-zero singular values coincide with the non-zero proper Hankel singular values of system (1).

Unfortunately, we do not know a physically meaningful improper Hankel operator. We can only show that the non-zero improper Hankel singular values of system (1) are the non-zero singular values of the improper Hankel matrix

$$\mathcal{H}_i = \begin{bmatrix} CF_{-1}B & CF_{-2}B & \cdots & CF_{-\nu}B \\ CF_{-2}B & & \iddots & 0 \\ \vdots & \iddots & \iddots & \vdots \\ CF_{-\nu}B & 0 & \cdots & 0 \end{bmatrix}$$

with the Markov parameters $CF_{k-1}B$, see [Sty03].

2.5 Realizations

For any rational matrix-valued function $\mathbf{G}(s)$, there exist matrices E, A, B and C such that $\mathbf{G}(s) = C(sE - A)^{-1}B$, see [Dai89]. A descriptor system (1) with these matrices is called a *realization* of $\mathbf{G}(s)$. We will also denote a realization of $\mathbf{G}(s)$ by $\mathbf{G} = [\, E,\, A,\, B,\, C\,]$ or by

$$\mathbf{G} = \left[\begin{array}{c|c} sE - A & B \\ \hline C & \end{array}\right].$$

Note that the realization of $\mathbf{G}(s)$ is, in general, not unique [Dai89]. Among different realizations of $\mathbf{G}(s)$ we are interested only in particular realizations that are useful for reduced-order modeling.

Definition 2.7. *A realization* $[E,\, A,\, B,\, C]$ *of the transfer function* $\mathbf{G}(s)$ *is called* minimal *if the dimension of the matrices* E *and* A *is as small as possible.*

The following theorem gives necessary and sufficient conditions for a realization of $\mathbf{G}(s)$ to be minimal.

Theorem 2.8. [Dai89, Sty04b] *Consider a descriptor system* (1)*, where the pencil* $\lambda E - A$ *is c-stable. The following statements are equivalent:*

1. *The realization* $[\, E,\, A,\, B,\, C\,]$ *is minimal.*
2. *The descriptor system* (1) *is C-controllable and C-observable.*
3. *The rank conditions* $\operatorname{rank}(\mathcal{G}_{pc}) = \operatorname{rank}(\mathcal{G}_{po}) = \operatorname{rank}(\mathcal{G}_{pc}E^T\mathcal{G}_{po}E) = n_f$ *and* $\operatorname{rank}(\mathcal{G}_{ic}) = \operatorname{rank}(\mathcal{G}_{io}) = \operatorname{rank}(\mathcal{G}_{ic}A^T\mathcal{G}_{io}A) = n_\infty$ *hold.*
4. *The proper and improper Hankel singular values of* (1) *are positive.*
5. *The rank conditions* $\operatorname{rank}(\mathcal{H}_p) = n_f$ *and* $\operatorname{rank}(\mathcal{H}_i) = n_\infty$ *hold.*

Remark 2.9. So far we have considered only descriptor systems without a feedthrough term, i.e., $D = 0$ in the output equation $y(t) = Cx(t) + Du(t)$. However, if we allow for the matrix D to be non-zero, then the condition for the realization of the transfer function $\mathbf{G}(s) = C(sE - A)^{-1}B + D$ to be minimal should be reformulated as follows: the realization $[\, E,\, A,\, B,\, C,\, D\,]$ is minimal if and only if the descriptor system is C-controllable and C-observable, and $A\operatorname{Ker}(E) \subseteq \operatorname{Im}(E)$, see [Sok03, VLK81]. The latter condition implies that the nilpotent matrix N in the Weierstrass canonical form (3) does not have any 1×1 Jordan blocks.

Definition 2.10. *A realization* $[\,E,\,A,\,B,\,C\,]$ *of the transfer function* $\mathbf{G}(s)$ *is called* balanced *if*

$$\mathcal{G}_{pc} = \mathcal{G}_{po} = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \qquad \textit{and} \qquad \mathcal{G}_{ic} = \mathcal{G}_{io} = \begin{bmatrix} 0 & 0 \\ 0 & \Theta \end{bmatrix},$$

where $\Sigma = \operatorname{diag}(\varsigma_1, \ldots, \varsigma_{n_f})$ *and* $\Theta = \operatorname{diag}(\theta_1, \ldots, \theta_{n_\infty})$.

For a minimal realization $[\,E, A, B, C\,]$ with a c-stable pencil $\lambda E - A$, it is possible to find nonsingular transformation matrices W_b and T_b such that the transformed realization $[\,W_b^T E T_b,\, W_b^T A T_b,\, W_b^T B,\, C T_b\,]$ is balanced, see [Sty04a]. These matrices are given by

$$\begin{aligned} W_b &= \begin{bmatrix} L_p U_p \Sigma^{-1/2}, & L_i U_i \Theta^{-1/2} \end{bmatrix}, \\ T_b &= \begin{bmatrix} R_p V_p \Sigma^{-1/2}, & R_i V_i \Theta^{-1/2} \end{bmatrix}. \end{aligned} \tag{20}$$

Observe, however, as for standard state space systems [Glo84, Moo81], the balancing transformation for descriptor systems is not unique. It should also be noted that for the matrices W_b and T_b as in (20), we have

$$E_b = W_b^T E T_b = \begin{bmatrix} I_{n_f} & 0 \\ 0 & E_2 \end{bmatrix}, \qquad A_b = W_b^T A T_b = \begin{bmatrix} A_1 & 0 \\ 0 & I_{n_\infty} \end{bmatrix}, \tag{21}$$

where the matrix $E_2 = \Theta^{-1/2} U_i^T L_i^T E R_i V_i \Theta^{-1/2}$ is nilpotent and the matrix $A_1 = \Sigma^{-1/2} U_p^T L_p^T A R_p V_p \Sigma^{-1/2}$ is nonsingular. Thus, the pencil $\lambda E_b - A_b$ of a balanced descriptor system is in a form that resembles the Weierstrass canonical form.

3 Balanced truncation

In this section we present a generalization of balanced truncation model reduction to descriptor systems.

Note that computing the balanced realization may be an ill-conditioned problem if the descriptor system (1) has small proper or improper Hankel singular values. Moreover, if system (1) is not minimal, then it has states that are uncontrollable or/and unobservable. These states correspond to the zero proper and improper Hankel singular values and can be truncated without changing the input-output relation in the system. Note that the number of non-zero improper Hankel singular values of (1) is equal to $\operatorname{rank}(\mathcal{G}_{ic} A^T \mathcal{G}_{io} A)$ which can in turn be bounded by

$$\operatorname{rank}(\mathcal{G}_{ic} A^T \mathcal{G}_{io} A) \leq \min(\nu m, \nu p, n_\infty),$$

where ν is the index of the pencil $\lambda E - A$, m is the number of inputs, p is the number of outputs and n_∞ is the dimension of the deflating subspace of

$\lambda E - A$ corresponding to the infinite eigenvalues. This estimate shows that if the number of inputs or outputs multiplied by the index ν is much smaller than the dimension n_∞, then the order of system (1) can be reduced significantly by the truncation of the states corresponding to the zero improper Hankel singular values.

Furthermore, we have the following theorem that gives an energy interpretation of the proper controllability and observability Gramians.

Theorem 3.1. [Sty04b] *Consider a descriptor system* (1) *that is asymptotically stable and C-controllable. Let* $\mathcal{G}_{pc}$ *and* $\mathcal{G}_{po}$ *be the proper controllability and observability Gramians of* (1). *If* $x_0 \in \operatorname{Im}(P_r)$ *and* $u(t) = 0$ *for* $t \geq 0$, *then*

$$\mathbf{E}_y := \int_0^\infty y^T(t)y(t)\,dt = x_0^T E^T \mathcal{G}_{po} E x_0.$$

Moreover, for $u_{\min}(t) = B^T \mathcal{F}^T(-t)\mathcal{G}_{pc}^- x_0$, *we have*

$$\mathbf{E}_{u_{\min}} := \min_{u \in \mathbb{L}_2^m(\mathbb{R}^-)} \int_{-\infty}^0 u^T(t)u(t)\,dt = x_0^T \mathcal{G}_{pc}^- x_0,$$

where the matrix $\mathcal{G}_{pc}^-$ *is a solution of the three matrix equations*

$$\mathcal{G}_{pc}\mathcal{G}_{pc}^-\mathcal{G}_{pc} = \mathcal{G}_{pc}, \qquad \mathcal{G}_{pc}^-\mathcal{G}_{pc}\mathcal{G}_{pc}^- = \mathcal{G}_{pc}^-, \qquad (\mathcal{G}_{pc}^-)^T = \mathcal{G}_{pc}^-.$$

Theorem 3.1 implies that a large past input energy $\mathbf{E}_u = \|u\|^2_{\mathbb{L}_2^m(\mathbb{R}^-)}$ is required to reach the state $x(0) = P_r x_0$ which lies in an invariant subspace of the proper controllability Gramian $\mathcal{G}_{pc}$ corresponding to its small non-zero eigenvalues from the state $x(-\infty) = 0$. Moreover, if x_0 is contained in an invariant subspace of the matrix $E^T\mathcal{G}_{po}E$ corresponding to its small non-zero eigenvalues, then the initial state $x(0) = x_0$ has a small effect on the future output energy $\mathbf{E}_y = \|y\|^2_{\mathbb{L}_2^p(\mathbb{R}_0^+)}$. For the balanced system, we have

$$\mathcal{G}_{pc} = E^T\mathcal{G}_{po}E = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$$

In this case the states related to the small proper Hankel singular values are difficult to reach and to observe at the same time. The truncation of these states essentially does not change system properties.

Unfortunately, this does not hold for the improper Hankel singular values. If we truncate the states that correspond to the small non-zero improper Hankel singular values, then the pencil of the reduced-order system may get finite eigenvalues in the closed right half-plane, see [LS00]. In this case the approximation may be inaccurate.

Remark 3.2. The equations associated to the improper Hankel singular values describe constraints of the system, i.e., they define a manifold in which the solution dynamics takes place. For this reason, a truncation of these equations corresponds to ignoring constrains and, hence, physically meaningless results may be expected.

Note that to perform order reduction we do not need to transform the descriptor system into a balanced form explicitly. It is sufficient to determine the subspaces associated with dominant proper and non-zero improper Hankel singular values and project the descriptor system on these subspaces. To compute a reduced-order system we can use the following algorithm which is a generalization of the *square root balanced truncation method* [LHPW87, TP84] to the descriptor system (1).

Algorithm 3.1. Generalized Square Root (GSR) method.

INPUT: A realization $\mathbf{G} = [\, E,\, A,\, B,\, C\,]$ such that $\lambda E - A$ is c-stable.
OUTPUT: A reduced-order system $\widetilde{\mathbf{G}} = [\, \widetilde{E},\, \widetilde{A},\, \widetilde{B},\, \widetilde{C}\,]$.

1. Compute the Cholesky factors R_p and L_p of the proper Gramians $\mathcal{G}_{pc} = R_p R_p^T$ and $\mathcal{G}_{po} = L_p L_p^T$ that satisfy (14) and (15), respectively.
2. Compute the Cholesky factors R_i and L_i of the improper Gramians $\mathcal{G}_{ic} = R_i R_i^T$ and $\mathcal{G}_{io} = L_i L_i^T$ that satisfy (16) and (17), respectively.
3. Compute the skinny singular value decomposition

$$L_p^T E R_p = [\, U_1,\, U_2\,] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} [\, V_1,\, V_2\,]^T, \tag{22}$$

   where the matrices $[\, U_1,\, U_2\,]$ and $[\, V_1,\, V_2\,]$ have orthonormal columns, $\Sigma_1 = \mathrm{diag}(\varsigma_1, \ldots, \varsigma_{\ell_f})$, $\Sigma_2 = \mathrm{diag}(\varsigma_{\ell_f+1}, \ldots, \varsigma_{r_p})$ with $r_p = \mathrm{rank}(L_p^T E R_p)$.
4. Compute the skinny singular value decomposition

$$L_i^T A R_i = U_3 \Theta_3 V_3^T, \tag{23}$$

   where U_3 and V_3 have orthonormal columns, $\Theta_3 = \mathrm{diag}(\theta_1, \ldots, \theta_{\ell_\infty})$ with $\ell_\infty = \mathrm{rank}(L_i^T A R_i)$.
5. Compute the projection matrices

$$W_\ell = [\, L_p U_1 \Sigma_1^{-1/2},\;\; L_i U_3 \Theta_3^{-1/2}\,], \quad T_\ell = [\, R_p V_1 \Sigma_1^{-1/2},\;\; R_i V_3 \Theta_3^{-1/2}\,].$$

6. Compute the reduced-order system

$$[\, \widetilde{E},\; \widetilde{A},\; \widetilde{B},\; \widetilde{C}\,] = [\, W_\ell^T E T_\ell,\; W_\ell^T A T_\ell,\; W_\ell^T B,\; C T_\ell\,].$$

This method has to be used with care, since if the original system (1) is highly unbalanced or if the angle between the deflating subspaces of the pencil $\lambda E - A$ corresponding to the finite and infinite eigenvalues is small, then the projection matrices W_ℓ and T_ℓ will be ill-conditioned. To avoid accuracy loss in the reduced-order model, a *square root balancing free method* has been proposed in [Var87] for standard state space systems. This method can be generalized to descriptor systems as follows.

Algorithm 3.2. Generalized Square Root Balancing Free (GSRBF) method.

INPUT: A realization $\mathbf{G} = [\,E,\, A,\, B,\, C\,]$ such that $\lambda E - A$ is c-stable.
OUTPUT: A reduced-order system $\widehat{\mathbf{G}} = [\,\widehat{E},\, \widehat{A},\, \widehat{B},\, \widehat{C}\,]$.

1. Compute the Cholesky factors R_p and L_p of the proper Gramians $\mathcal{G}_{pc} = R_p R_p^T$ and $\mathcal{G}_{po} = L_p L_p^T$ that satisfy (14) and (15), respectively.
2. Compute the Cholesky factors R_i and L_i of the improper Gramians $\mathcal{G}_{ic} = R_i R_i^T$ and $\mathcal{G}_{io} = L_i L_i^T$ that satisfy (16) and (17), respectively.
3. Compute the skinny singular value decompositions (22).
4. Compute the skinny singular value decomposition (23).
5. Compute the skinny QR decompositions
$$[\,R_p V_1,\, R_i V_3\,] = Q_R R_0, \qquad [\,L_p U_1,\, L_i U_3\,] = Q_L L_0,$$
where $Q_R,\, Q_L \in \mathbb{R}^{n,\ell}$ have orthonormal columns and $R_0,\, L_0 \in \mathbb{R}^{\ell,\ell}$ are nonsingular.
6. Compute the reduced-order system
$$[\,\widehat{E},\, \widehat{A},\, \widehat{B},\, \widehat{C}\,] = [\,Q_L^T E Q_R,\, Q_L^T A Q_R,\, Q_L^T B,\, C Q_R\,].$$

The GSR and GSRBF methods are formally equivalent in the sense that in exact arithmetic they return reduced systems with the same transfer function. However, since the projection matrices Q_L and Q_R computed by the GSRBF method have orthonormal columns, they may be significantly less sensitive to perturbations than the projection matrices W_ℓ and T_ℓ computed by the GSR method. Observe that the realization $[\,\widehat{E},\, \widehat{A},\, \widehat{B},\, \widehat{C}\,]$ is, in general, not balanced and the pencil $\lambda\widehat{E} - \widehat{A}$ is not in the block diagonal form (21).

3.1 Stability and approximation error

Computing the reduced-order descriptor system via balanced truncation can be interpreted as follows. At first we transform the asymptotically stable descriptor system (1) to the block diagonal form

$$\left[\begin{array}{c|c} \check{W}(sE - A)\check{T} & \check{W}B \\ \hline C\check{T} & \end{array}\right] = \left[\begin{array}{cc|c} sE_f - A_f & 0 & B_f \\ 0 & sE_\infty - A_\infty & B_\infty \\ \hline C_f & C_\infty & \end{array}\right],$$

where $\check{W}$ and $\check{T}$ are nonsingular, the pencil $\lambda E_f - A_f$ has only the finite eigenvalues that are the finite eigenvalues of $\lambda E - A$, and all the eigenvalues of $\lambda E_\infty - A_\infty$ are infinite. Then we reduce the order of the subsystems $[\,E_f,\, A_f,\, B_f,\, C_f\,]$ and $[\,E_\infty,\, A_\infty,\, B_\infty,\, C_\infty\,]$ separately. Clearly, the reduced-order system (2) is asymptotically stable and minimal.

The described decoupling of system matrices is equivalent to the additive decomposition of the transfer function as $\mathbf{G}(s) = \mathbf{G}_{sp}(s) + \mathbf{P}(s)$, where

$$\mathbf{G}_{sp}(s) = C_f(sE_f - A_f)^{-1}B_f \quad \text{and} \quad \mathbf{P}(s) = C_\infty(sE_\infty - A_\infty)^{-1}B_\infty$$

are the strictly proper part and the polynomial part of $\mathbf{G}(s)$. The reduced-order system (2) has the transfer function $\widetilde{\mathbf{G}}(s) = \widetilde{\mathbf{G}}_{sp}(s) + \widetilde{\mathbf{P}}(s)$, where

$$\widetilde{\mathbf{G}}_{sp}(s) = \widetilde{C}_f(s\widetilde{E}_f - \widetilde{A}_f)^{-1}\widetilde{B}_f \quad \text{and} \quad \widetilde{\mathbf{P}}(s) = \widetilde{C}_\infty(s\widetilde{E}_\infty - \widetilde{A}_\infty)^{-1}\widetilde{B}_\infty$$

are the transfer functions of the reduced-order subsystems. For the subsystem $\mathbf{G}_{sp} = [\, E_f,\, A_f,\, B_f,\, C_f \,]$ with nonsingular E_f, we have the following upper bound on the $\mathbb{H}_\infty$-norm of the absolute error

$$\|\mathbf{G}_{sp} - \widetilde{\mathbf{G}}_{sp}\|_{\mathbb{H}_\infty} = \sup_{\omega\in\mathbb{R}} \|\mathbf{G}_{sp}(i\omega) - \widetilde{\mathbf{G}}_{sp}(\mathrm{i}\omega)\|_2 \le 2(\varsigma_{\ell_f+1} + \ldots + \varsigma_{n_f})$$

that can be derived similarly as in [Enn84, Glo84] for the standard state space case.

Reducing the order of the subsystem $\mathbf{P} = [\, E_\infty,\, A_\infty,\, B_\infty,\, C_\infty \,]$ is equivalent to the balanced model reduction of the discrete-time system

$$\begin{aligned} A_\infty\xi_{k+1} &= E_\infty\xi_k + B_\infty\eta_k, \\ w_k &= C_\infty\xi_k \end{aligned}$$

with a nonsingular matrix A_∞. The Hankel singular values of this system are just the improper Hankel singular values of (1). Since we truncate only the states corresponding to the zero improper Hankel singular values, the equality $\mathbf{P}(s) = \widetilde{\mathbf{P}}(s)$ holds and the index of the reduced-order system is equal to $\deg(\mathbf{P})+1$, where $\deg(\mathbf{P})$ denotes the degree of the polynomial $\mathbf{P}(s)$, or, equivalently, the multiplicity of the pole at infinity of the transfer function $\mathbf{G}(s)$. In this case the error system $\mathbf{G}(s) - \widetilde{\mathbf{G}}(s) = \mathbf{G}_{sp}(s) - \widetilde{\mathbf{G}}_{sp}(s)$ is strictly proper, and we have the following $\mathbb{H}_\infty$-norm error bound

$$\|\mathbf{G}(s) - \widetilde{\mathbf{G}}(s)\|_{\mathbb{H}_\infty} \le 2(\varsigma_{\ell_f+1} + \ldots + \varsigma_{n_f}).$$

Existence of this error bound is an important property of the balanced truncation model reduction approach for descriptor systems. It makes this approach preferable compared, for instance, to moment matching techniques as in [FF95, Fre00, GGV96, Gri97].

3.2 Numerical aspects

To reduce the order of the descriptor system (1) we have to compute the Cholesky factors of the proper and improper controllability and observability Gramians that satisfy the projected generalized Lyapunov equations (14), (15), (16) and (17). These factors can be determined using the *generalized Schur-Hammarling method* [Sty02a, Sty02b] without computing the solutions of Lyapunov equations explicitly. Combining this method with the GSR method we obtain the following algorithm for computing the reduced-order descriptor system (2).

Algorithm 3.3. Generalized Schur-Hammarling square root method.

INPUT: A realization $\mathbf{G} = [\,E,\ A,\ B,\ C\,]$ such that $\lambda E - A$ is c-stable.
OUTPUT: A reduced-order realization $\widetilde{\mathbf{G}} = [\,\widetilde{E},\ \widetilde{A},\ \widetilde{B},\ \widetilde{C}\,]$.

1. Compute the generalized Schur form

$$E = V \begin{bmatrix} E_f & E_u \\ 0 & E_\infty \end{bmatrix} U^T \qquad \text{and} \qquad A = V \begin{bmatrix} A_f & A_u \\ 0 & A_\infty \end{bmatrix} U^T, \tag{24}$$

   where U and V are orthogonal, E_f is upper triangular nonsingular, E_∞ is upper triangular nilpotent, A_f is upper quasi-triangular and A_∞ is upper triangular nonsingular.

2. Compute the matrices $V^T B = \begin{bmatrix} B_u \\ B_\infty \end{bmatrix}$ and $CU = [\,C_f,\ \ C_u\,]$.

3. Solve the system of generalized Sylvester equations

$$\begin{aligned} E_f Y - Z E_\infty &= -E_u, \\ A_f Y - Z A_\infty &= -A_u. \end{aligned} \tag{25}$$

4. Compute the Cholesky factors R_f, L_f, R_∞ and L_∞ of the solutions $X_{pc} = R_f R_f^T$, $X_{po} = L_f L_f^T$, $X_{ic} = R_\infty R_\infty^T$ and $X_{io} = L_\infty L_\infty^T$ of the generalized Lyapunov equations

$$E_f X_{pc} A_f^T + A_f X_{pc} E_f^T = -(B_u - Z B_\infty)(B_u - Z B_\infty)^T, \tag{26}$$
$$E_f^T X_{po} A_f + A_f^T X_{po} E_f = -C_f^T C_f, \tag{27}$$
$$A_\infty X_{ic} A_\infty^T - E_\infty X_{ic} E_\infty^T = B_\infty B_\infty^T, \tag{28}$$
$$A_\infty^T X_{io} A_\infty - E_\infty^T X_{io} E_\infty = (C_f Y + C_u)^T (C_f Y + C_u). \tag{29}$$

5. Compute the skinny singular value decompositions

$$L_f^T E_f R_f = [\,U_1,\ U_2\,] \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} [\,V_1,\ V_2\,]^T, \qquad L_\infty^T A_\infty R_\infty = U_3 \Theta_3 V_3^T,$$

   where $[\,U_1,\ \ U_2\,]$, $[\,V_1,\ \ V_2\,]$, U_3 and V_3 have orthonormal columns, $\Sigma_1 = \operatorname{diag}(\varsigma_1, \ldots, \varsigma_{\ell_f})$, $\Sigma_2 = \operatorname{diag}(\varsigma_{\ell_f+1}, \ldots, \varsigma_r)$, $\Theta_3 = \operatorname{diag}(\theta_1, \ldots, \theta_{\ell_\infty})$ with $r = \operatorname{rank}(L_f^T E_f R_f)$ and $\ell_\infty = \operatorname{rank}(L_\infty^T A_\infty R_\infty)$.

6. Compute $W_f = L_f U_1 \Sigma_1^{-1/2}$, $W_\infty = L_\infty U_3 \Theta_3^{-1/2}$, $T_f = R_f V_1 \Sigma_1^{-1/2}$ and $T_\infty = R_\infty V_3 \Theta_3^{-1/2}$.

7. Compute the reduced-order system $[\,\widetilde{E},\ \widetilde{A},\ \widetilde{B},\ \widetilde{C}\,]$ with

$$\widetilde{E} = \begin{bmatrix} I_{\ell_f} & 0 \\ 0 & W_\infty^T E_\infty T_\infty \end{bmatrix}, \qquad \widetilde{A} = \begin{bmatrix} W_f^T A_f T_f & 0 \\ 0 & I_{\ell_\infty} \end{bmatrix},$$
$$\widetilde{B} = \begin{bmatrix} W_f^T (B_u - Z B_\infty) \\ W_\infty^T B_\infty \end{bmatrix}, \qquad \widetilde{C} = [\,C_f T_f, \quad (C_f Y + C_u) T_\infty\,].$$

To compute the generalized Schur form (24) we can use the QZ algorithm [GV96, Wat00] the GUPTRI algorithm [DK93a, DK93b], or algorithms proposed in [BV88, Var98]. To solve the generalized Sylvester equation (25) one can use the generalized Schur method [KW89] or its recursive blocked modification [JK02] that is more suitable for large problems. The upper triangular Cholesky factors R_f, L_f, R_∞ and L_∞ of the solutions of the generalized Lyapunov equations (26)-(29) can be determined without computing the solutions themselves using the generalized Hammarling method [Ham82, Pen98]. Furthermore, the singular value decomposition of $L_f^T E_f R_f$ and $L_\infty^T A_\infty R_\infty$, where all three factors are upper triangular, can be computed without forming these products explicitly, see [BELV91, Drm00, GSV00] and references therein.

Algorithm 3.3 and its balancing-free version have been implemented as a MATLAB-based function `gbta` in Descriptor Systems Toolbox[1] [Var00].

Since the generalized Schur-Hammarling method is based on computing the generalized Schur form (24), it costs $O(n^3)$ flops and has the memory complexity $O(n^2)$. Thus, this method can be used for problems of small and medium size. Unfortunately, it does not take into account the sparsity or any structure of the system and is not attractive for parallelization. Recently, iterative methods related to the ADI method and the Smith method have been proposed to compute low rank approximations of the solutions of standard large scale sparse Lyapunov equations [Li00, LW02, Pen99]. It was observed that the eigenvalues of the symmetric solutions of Lyapunov equations with low rank right-hand side generally decay very rapidly, and such solutions may be well approximated by low rank matrices, see [ASZ02, Pen00a, SZ02]. A similar result holds for projected generalized Lyapunov equations. Consider, for example, the projected GCALE (14). If it is possible to find a matrix X with a small number of columns such that XX^T is an approximate solution of (14), then X is referred to as the *low rank Cholesky factor* of the solution $\mathcal{G}_{pc}$ of the projected GCALE (14). It can be computed by the following algorithm that is a generalization of low rank ADI method for standard Lyapunov equation as suggested in [Li00, LW02, Pen99].

Algorithm 3.4. Generalized low rank ADI method.

INPUT: Matrices $E, A \in \mathbb{R}^{n,n}$, $Q = P_l B \in \mathbb{R}^{n,m}$, shift parameters $\tau_1, \ldots, \tau_q \in \mathbb{C}^-$.
OUTPUT: A low rank Cholesky factor X_k of the Gramian $\mathcal{G}_{pc} \approx X_k X_k^T$.

1. $X^{(1)} = \sqrt{-2\mathrm{Re}(\tau_1)}\,(E + \tau_1 A)^{-1} Q, \quad X_1 = X^{(1)}$,
2. FOR $k = 2, 3, \ldots$

   a. $X^{(k)} = \sqrt{\dfrac{\mathrm{Re}(\tau_k)}{\mathrm{Re}(\tau_{k-1})}}\left(I - (\overline{\tau}_{k-1} + \tau_k)(E + \tau_k A)^{-1} A\right) X^{(k-1)}$,

   b. $X_k = [\, X_{k-1}, \;\; X^{(k)} \,]$.

   END FOR

---

[1] `http://www.robotic.dlr.de/control/num/desctool.html`

If all the finite eigenvalues of the pencil $\lambda E - A$ lie in the open left half-plane, then X_k converges to the solution of the projected GCALE (14). The rate of convergence depends strongly on the choice of the shift parameters $\tau_1, \ldots, \tau_q$. The optimal shift parameters satisfy the generalized ADI minimax problem

$$\{\tau_1, \ldots, \tau_q\} = \underset{\{\tau_1,\ldots,\tau_q\}\in\mathbb{C}^-}{\arg\min} \max_{t\in \mathrm{Sp}_f(E,A)} \frac{|(1-\overline{\tau}_1 t)\cdot\ldots\cdot(1-\overline{\tau}_q t)|}{|(1+\tau_1 t)\cdot\ldots\cdot(1+\tau_q t)|},$$

where $\mathrm{Sp}_f(E, A)$ denotes the finite spectrum of the pencil $\lambda E - A$, see [Sty04c]. The computation of the optimal shift parameters is a difficult problem, since the finite eigenvalues of the pencil $\lambda E - A$ (in particular, if it is large and sparse) are in general unknown and expensive to compute. Instead, sub-optimal ADI shift parameters $\tau_1, \ldots, \tau_q$ can be determined by a heuristic procedure as in [Pen99, Algorithm 5.1] from a set of largest and smallest (in modulus) approximate finite eigenvalues of $\lambda E - A$ that may be computed by an Arnoldi process.

As a stopping criterion one can use the condition $\|X^{(k)}\|/\|X_k\| \leq tol$ with some matrix norm $\|\cdot\|$ and a user-defined tolerance tol. The iteration can also be stopped as soon as a *normalized residual norm*

$$\eta(E, A, P_l B; X_k) = \frac{\|EX_kX_k^TA^T + AX_kX_k^TE^T + P_lBB^TP_l^T\|}{\|P_lBB^TP_l^T\|}$$

satisfies $\eta(E, A, P_l B; X_k) \leq tol$ or a stagnation of $\eta(E, A, P_l B; X_k)$ is observed, see [Pen99] for an efficient computation of the Frobenius norm based normalized residuals. Note that if the low rank ADI method needs more iterations than the number of available ADI shift parameters, then we reuse these parameters in a cyclic manner.

It should also be noted that the matrices $(E + \tau_k A)^{-1}$ in Algorithm 3.4 do not have to be computed explicitly. Instead, we solve linear systems $(E + \tau_k A)x = P_l b$ either by computing (sparse) LU factorizations and forward/backward substitutions or by using iterative Krylov subspace methods [Saa96]. In the latter case the generalized low rank ADI method has the memory complexity $O(k_{ADI}mn)$ and costs $O(k_{\mathrm{ls}}k_{ADI}mn)$ flops, where k_{ls} is the number of linear solver iterations and k_{ADI} is the number of ADI iterations. This method becomes efficient for large scale sparse Lyapunov equations only if $k_{\mathrm{ls}}k_{ADI}m$ is much smaller than n. Note that if the matrices E and A have a particular structure for which the hierarchical matrix arithmetic can be used, then also the methods proposed in [Hac00, HGB02] can be applied to compute the inverse of $E + \tau_k A$.

A major difficulty in the numerical solution of the projected Lyapunov equations by the low rank ADI method is that we need to compute the spectral projections P_l and P_r onto the left and right deflating subspaces of the pencil $\lambda E - A$ corresponding to the finite eigenvalues. This is in general very difficult,

but in many applications, such as control of fluid flow, electrical circuits simulation and constrained multibody systems, the matrices E and A have some special block structure. This structure can be used to construct the projections P_l and P_r explicitly and cheaply, see [ET00, Mar96, Sch95, Sty04a].

3.3 Remarks

We close this section with some concluding remarks.

Remark 3.3. The GSR and the GSRBF methods can also be used to reduce the order of unstable descriptor systems. To do this we first compute the additive decomposition [KV92] of the transfer function $\mathbf{G}(s) = \mathbf{G}_-(s) + \mathbf{G}_+(s)$, where $\mathbf{G}_-(s) = C_-(sE_- - A_-)^{-1}B_-$ and $\mathbf{G}_+(s) = C_+(sE_+ - A_+)^{-1}B_+$. Here the matrix pencil $\lambda E_- - A_-$ is c-stable and all the eigenvalues of the pencil $\lambda E_+ - A_+$ are finite and have non-negative real part. Then we determine the reduced-order system $\widetilde{\mathbf{G}}_-(s) = \widetilde{C}_-(s\widetilde{E}_- - \widetilde{A}_-)^{-1}\widetilde{B}_-$ by applying the balanced truncation model reduction method to the subsystem $\mathbf{G}_- = [\, E_-,\, A_-,\, B_-,\, C_-\,]$. Finally, the reduced-order approximation of $\mathbf{G}(s)$ is given by $\widetilde{\mathbf{G}}(s) = \widetilde{\mathbf{G}}_-(s) + \mathbf{G}_+(s)$, where $\mathbf{G}_+(s)$ is included unmodified.

Remark 3.4. To compute a low order approximation to a large scale descriptor system of index one with dense matrix coefficients E and A we can apply the spectral projection method [BQQ04]. This method is based on the disc and sign functions iterative procedures and can be efficiently implemented on parallel computers.

Remark 3.5. An alternative model reduction approach for descriptor systems is the moment matching approximation which can be formulated as follows. Suppose that $s_0 \in \mathbb{C}$ is not an eigenvalue of the pencil $\lambda E - A$. Then the transfer function $\mathbf{G}(s) = C(sE - A)^{-1}B$ can be expanded into a Laurent series at s_0 as

$$\begin{aligned}\mathbf{G}(s) &= C\left(I - (s - s_0)(s_0E - A)^{-1}E\right)^{-1}(s_0E - A)^{-1}B \\ &= M_0 + M_1(s - s_0) + M_2(s - s_0)^2 + \dots ,\end{aligned}$$

where the matrices $M_j = -C\left((s_0E - A)^{-1}E\right)^j(s_0E - A)^{-1}B$ are called the *moments* of the descriptor system (1) at s_0. The moment matching approximation problem for the descriptor system (1) consists in determining a rational matrix-valued function $\widetilde{\mathbf{G}}(s)$ such that the Laurent series expansion of $\widetilde{\mathbf{G}}(s)$ at s_0 has the form

$$\widetilde{\mathbf{G}}(s) = \widetilde{M}_0 + \widetilde{M}_1(s - s_0) + \widetilde{M}_2(s - s_0)^2 + \dots , \tag{30}$$

where the moments $\widetilde{M}_j$ satisfy the moment matching conditions

$$M_j = \widetilde{M}_j, \qquad j = 0, 1, \dots, k. \tag{31}$$

If $s_0 = \infty$, then $M_j = CF_{j-1}B$ are the Markov parameters of (1) and the corresponding approximation problem is known as *partial realization* [GL83]. Computation of the partial realization for descriptor systems is an open problem. For $s_0 = 0$, the approximation problem (30), (31) reduces to the *Padé approximation* problem [BG96]. Efficient algorithms based on Arnoldi and Lanzcos procedures for solving this problem have been presented in [FF95, GGV94]. For an arbitrary complex number $s_0 \neq 0$, the moment matching approximation is the problem of *rational interpolation* or *shifted Padé approximation* that has been considered in [Bai02, BF01, FF95, Fre00, GGV96]. Apart from a single interpolation point one can construct a reduced-order system with the transfer function $\widetilde{\mathbf{G}}(s)$ that matches $\mathbf{G}(s)$ at multiple points $\{s_0, s_1, \ldots, s_k\}$. Such an approximation is called a *multi-point Padé approximation* or a *rational interpolant* [AA00, BG96]. It can be computed efficiently for descriptor systems by the rational Krylov subspace method [GGV96, Gri97, Rue84].

4 Numerical examples

In this section we give numerical examples to illustrate the effectiveness of the described model reduction methods for descriptor systems. The computations were done on IBM RS 6000 44P Modell 270 with machine precision $\varepsilon = 2.22 \times 10^{-16}$ using MATLAB 6.5. We apply these methods to two different models: a semidiscretized Stokes equation and a constrained damped mass-spring system.

Semidiscretized Stokes equation

Consider the instationary Stokes equation describing the flow of an incompressible fluid

$$\begin{aligned} \frac{\partial v}{\partial t} &= \Delta v - \nabla \rho + f, && (\xi, t) \in \Omega \times (0, t_e), \\ 0 &= \operatorname{div} v, && (\xi, t) \in \Omega \times (0, t_e) \end{aligned} \tag{32}$$

with appropriate initial and boundary conditions. Here $v(\xi, t) \in \mathbb{R}^d$ is the velocity vector ($d = 2$ or 3 is the dimension of the spatial domain), $\rho(\xi, t) \in \mathbb{R}$ is the pressure, $f(\xi, t) \in \mathbb{R}^d$ is the vector of external forces, $\Omega \subset \mathbb{R}^d$ is a bounded open domain and $t_e > 0$ is the endpoint of the time interval. The spatial discretization of the Stokes equation (32) by the finite difference method on a uniform staggered grid leads to a descriptor system

$$\begin{aligned} \dot{\mathbf{v}}_h(t) &= A_{11}\mathbf{v}_h(t) + A_{12}\boldsymbol{\rho}_h(t) + B_1 u(t), \\ 0 &= A_{12}^T \mathbf{v}_h(t) + B_2 u(t), \\ y(t) &= C_1 \mathbf{v}_h(t) + C_2 \boldsymbol{\rho}_h(t), \end{aligned} \tag{33}$$

where $\mathbf{v}_h(t) \in \mathbb{R}^{n_\mathbf{v}}$ and $\boldsymbol{\rho}_h(t) \in \mathbb{R}^{n_\boldsymbol{\rho}}$ are the semidiscretized vectors of velocities and pressures, respectively, see [Ber90]. The matrix $A_{11} \in \mathbb{R}^{n_\mathbf{v},n_\mathbf{v}}$ is the discrete Laplace operator, $-A_{12} \in \mathbb{R}^{n_\mathbf{v},n_\boldsymbol{\rho}}$ and $-A_{12}^T \in \mathbb{R}^{n_\boldsymbol{\rho},n_\mathbf{v}}$ are, respectively, the discrete gradient and divergence operators. Due to the non-uniqueness of the pressure, the matrix A_{12} has a rank defect one. In this case instead of A_{12} we can take a full column rank matrix obtained from A_{12} by discarding the last column. Therefore, in the following we will assume without loss of generality that A_{12} has full column rank. In this case system (33) is of index 2. The matrices $B_1 \in \mathbb{R}^{n_\mathbf{v},m}$, $B_2 \in \mathbb{R}^{n_\boldsymbol{\rho},m}$ and the control input $u(t) \in \mathbb{R}^m$ are resulting from the boundary conditions and external forces, the output $y(t)$ is the vector of interest. The order $n = n_\mathbf{v} + n_\boldsymbol{\rho}$ of system (33) depends on the level of refinement of the discretization and is usually very large, whereas the number m of inputs and the number p of outputs are typically small. Note that the matrix coefficients in (33) given by

$$E = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{and} \qquad A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & 0 \end{bmatrix}$$

are sparse and have a special block structure. Using this structure, the projections P_l and P_r onto the left and right deflating subspaces of the pencil $\lambda E - A$ can be computed as

$$P_l = \begin{bmatrix} \Pi & -\Pi A_{11} A_{12} (A_{12}^T A_{12})^{-1} \\ 0 & 0 \end{bmatrix}, \quad P_r = \begin{bmatrix} \Pi & 0 \\ -(A_{12}^T A_{12})^{-1} A_{12}^T A_{11} \Pi & 0 \end{bmatrix},$$

where $\Pi = I - A_{12}(A_{12}^T A_{12})^{-1} A_{12}^T$ is the orthogonal projection onto $\mathrm{Ker}\,(A_{12}^T)$ along $\mathrm{Im}\,(A_{12})$, see [Sty04a].

The spatial discretization of the Stokes equation (32) on a square domain $\Omega = [0,1] \times [0,1]$ by the finite difference method on a uniform staggered 80×80 grid leads to a problem of order $n = 19520$. The dimensions of the deflating subspaces of the pencil $\lambda E - A$ corresponding to the finite and infinite eigenvalues are $n_f = 6400$ and $n_\infty = 13120$, respectively. In our experiments $B = [\, B_1^T,\ B_2^T\,]^T \in \mathbb{R}^{n,1}$ is chosen at random and we are interested in the first velocity component, i.e., $C = [\,1,\ 0, \ldots, 0\,] \in \mathbb{R}^{1,n}$.

To reduce the order of the semidiscretized Stokes equation (33) we use the GSR and the GSRBF methods, where the exact Cholesky factors R_p and L_p of the proper Gramians are replaced by low rank Cholesky factors R_k and L_k, respectively, such that $\mathcal{G}_{pc} \approx R_k R_k^T$ and $\mathcal{G}_{po} \approx L_k L_k^T$. The matrices R_k and L_k have been computed by the generalized low rank ADI method with 20 shift parameters applied to $(E, A, P_l B)$ and $(E^T, A^T, P_r^T C^T)$, respectively.

In Fig. 1 we present the convergence history for the normalized residuals $\eta(E, A, P_l B; R_k)$ and $\eta(E^T, A^T, P_r^T C^T; L_k)$ versus the iteration step k. Figure 2 shows the approximate dominant proper Hankel singular values $\widetilde{\varsigma}_j$ computed from the singular value decomposition of the matrix $L_{70}^T E R_{39}$ with $R_{39} \in \mathbb{R}^{n,39}$ and $L_{70} \in \mathbb{R}^{n,70}$. Note the Cholesky factors R_i and L_i of the improper Gramians of (33) can be computed in explicit form without solving the generalized Lyapunov equations (16) and (17) numerically,

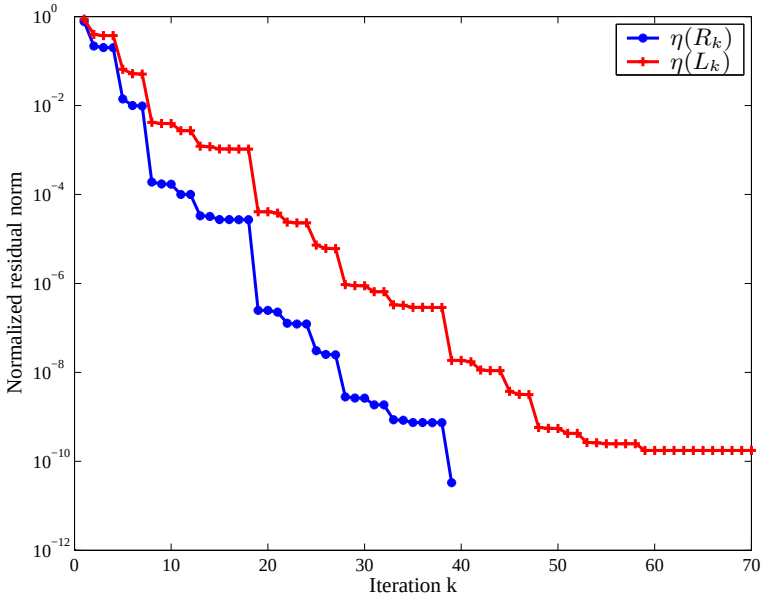

Fig. 1. Convergence history for the normalized residuals $\eta(R_k) = \eta(E, A, P_l B; R_k)$ and $\eta(L_k) = \eta(E^T, A^T, P_r^T C^T; L_k)$ for the semidiscretized Stokes equation.

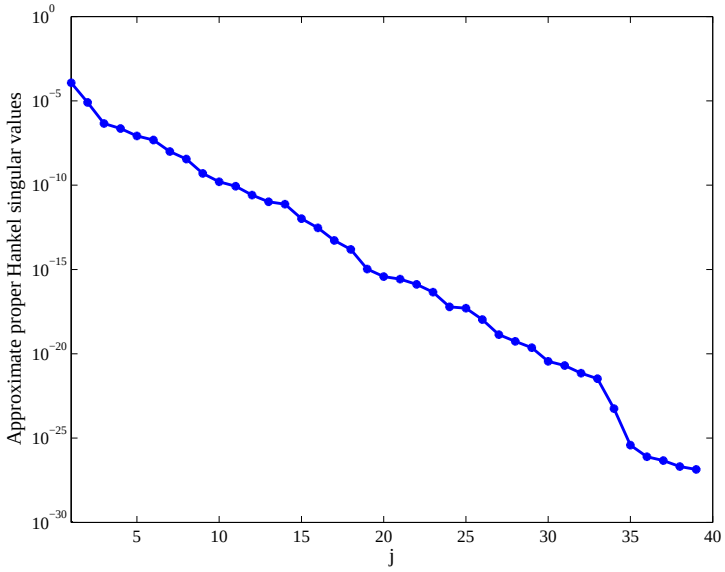

Fig. 2. Approximate proper Hankel singular values for the semidiscretized Stokes equation.

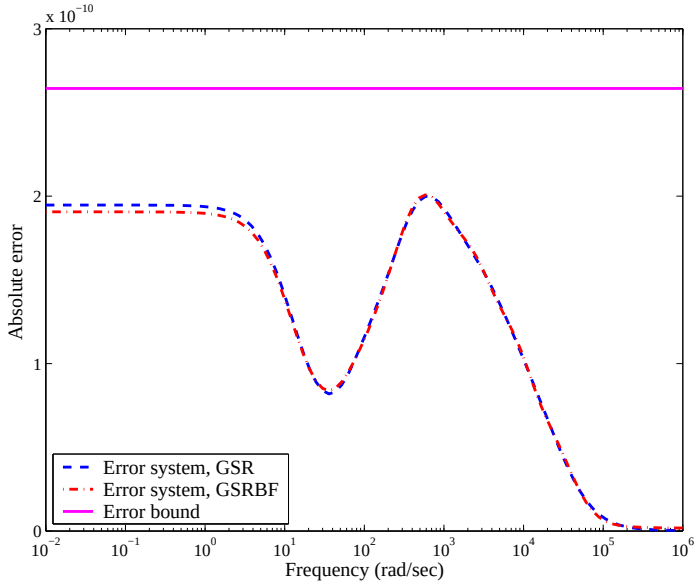

Fig. 3. Absolute error plots and error bound for the semidiscretized Stokes equation.

see [Sty04a]. System (33) has only one non-zero improper Hankel singular value $\theta_1 = 0.0049743$.

We approximate the semidiscretized Stokes equation (33) by two models of order $\ell = 11$ ($\ell_f = 10$, $\ell_\infty = 1$) computed by the approximate GSR and GSRBF methods. The absolute values of the frequency responses of the full order and the reduced-order systems are not presented, since they were impossible to distinguish. In Fig. 3 we display the absolute errors $\|\mathbf{G}(\mathrm{i}\omega) - \widetilde{\mathbf{G}}(\mathrm{i}\omega)\|_2$ and $\|\mathbf{G}(\mathrm{i}\omega) - \widehat{\mathbf{G}}(\mathrm{i}\omega)\|_2$ for a frequency range $\omega \in [\,10^{-2},\ 10^{6}\,]$ as well as the approximate error bound computed as twice the sum of the truncated approximate Hankel singular values $\widetilde{\varsigma}_{11}, \ldots \widetilde{\varsigma}_{39}$. One can see that over the displayed frequency range the absolute errors are smaller than 2×10^{-10} which is much smaller than the discretization error which is of order 10^{-4}.

Constrained damped mass-spring system

Consider the holonomically constrained damped mass-spring system illustrated in Fig. 4. The ith mass of weight m_i is connected to the $(i+1)$st mass by a spring and a damper with constants k_i and d_i, respectively, and also to the ground by a spring and a damper with constants κ_i and δ_i, respectively. Additionally, the first mass is connected to the last one by a rigid bar and and it is influenced by the control $u(t)$. The vibration of this system is

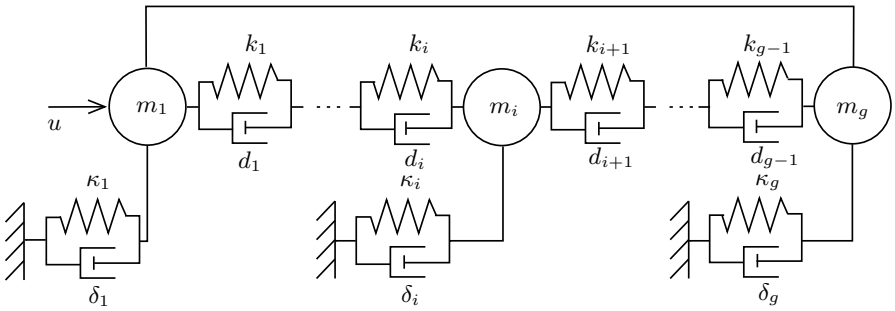

Fig. 4. A damped mass-spring system with a holonomic constraint.

described by a descriptor system

$$\begin{aligned} \dot{\mathbf{p}}(t) &= \mathbf{v}(t), \\ M\dot{\mathbf{v}}(t) &= K\,\mathbf{p}(t) + D\mathbf{v}(t) - G^T\boldsymbol{\lambda}(t) + B_2u(t), \\ 0 &= G\,\mathbf{p}(t), \\ y(t) &= C_1\mathbf{p}(t), \end{aligned} \tag{34}$$

where $\mathbf{p}(t) \in \mathbb{R}^g$ is the position vector, $\mathbf{v}(t) \in \mathbb{R}^g$ is the velocity vector, $\boldsymbol{\lambda}(t) \in \mathbb{R}^2$ is the Lagrange multiplier, $M = \mathrm{diag}(m_1, \ldots, m_g)$ is the mass matrix, D and K are the tridiagonal damping and stiffness matrices, $G = [\,1, 0, \ldots, 0, -1\,] \in \mathbb{R}^{1,g}$ is the constraint matrix, $B_2 = e_1$ and $C_1 = [\,e_1,\ e_2,\ e_{g-1}\,]^T$. Here e_i denotes the ith column of the identity matrix I_g. The descriptor system (34) is of index 3 and the projections P_l and P_r can be computed as

$$P_l = \begin{bmatrix} \Pi_1 & 0 & -\Pi_1 M^{-1} D G_1 \\ -\Pi_1^T D(I - \Pi_1) & \Pi_1^T & -\Pi_1^T (K + D\Pi_1 M^{-1} D) G_1 \\ 0 & 0 & 0 \end{bmatrix},$$

$$P_r = \begin{bmatrix} \Pi_1 & 0 & 0 \\ -\Pi_1 M^{-1} D(I - \Pi_1) & \Pi_1 & 0 \\ G_1^T (K\Pi_1 - D\Pi_1 M^{-1} D(I - \Pi_1)) & G_1^T D \Pi_1 & 0 \end{bmatrix},$$

where $G_1 = M^{-1}G^T(GM^{-1}G^T)^{-1}$ and $\Pi_1 = I - G_1 G$ is a projection onto $\mathrm{Ker}\,(G)$ along $\mathrm{Im}\,(M^{-1}G^T)$, see [Sch95].

In our experiments we take $m_1 = \ldots = m_g = 100$ and

$$\begin{aligned} k_1 = \ldots = k_{g-1} = \kappa_2 = \ldots = \kappa_{g-1} = 2, \quad \kappa_1 = \kappa_g = 4, \\ d_1 = \ldots = d_{g-1} = \delta_2 = \ldots = \delta_{g-1} = 5, \quad \delta_1 = \delta_g = 10. \end{aligned}$$

For $g = 6000$, we obtain the descriptor system of order $n = 12001$ with $m = 1$ input and $p = 3$ outputs. The dimensions of the deflating subspaces of the pencil corresponding to the finite and infinite eigenvalues are $n_f = 11998$ and $n_\infty = 3$, respectively.

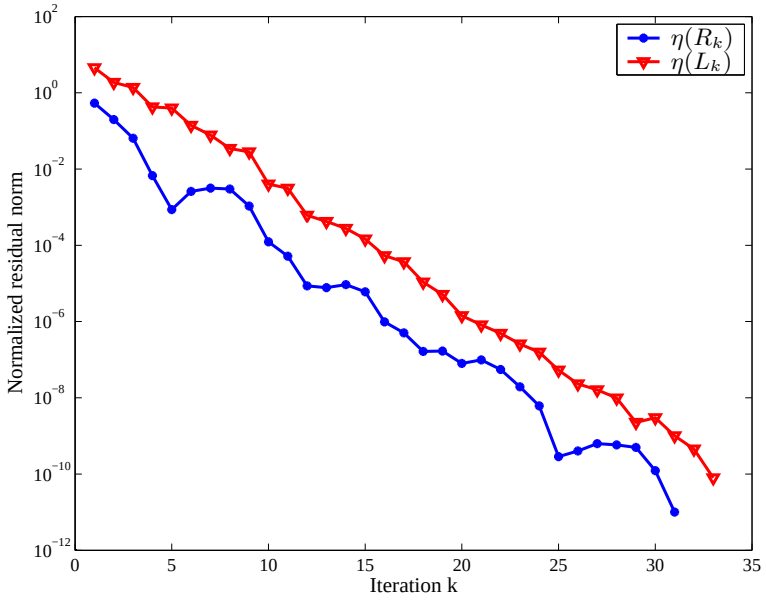

Fig. 5. Convergence history for the normalized residuals $\eta(R_k) = \eta(E, A, P_l B; R_k)$ and $\eta(L_k) = \eta(E^T, A^T, P_r^T C^T; L_k)$ for the damped mass-spring system.

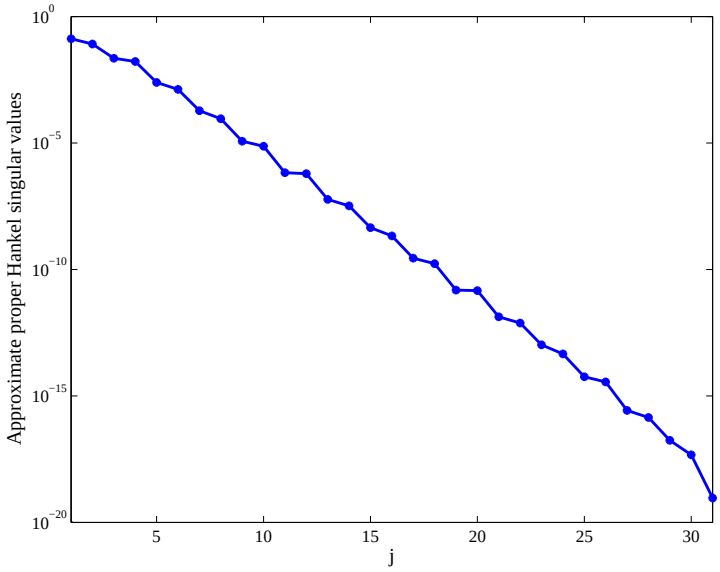

Fig. 6. Approximate proper Hankel singular values for the damped mass-spring system.

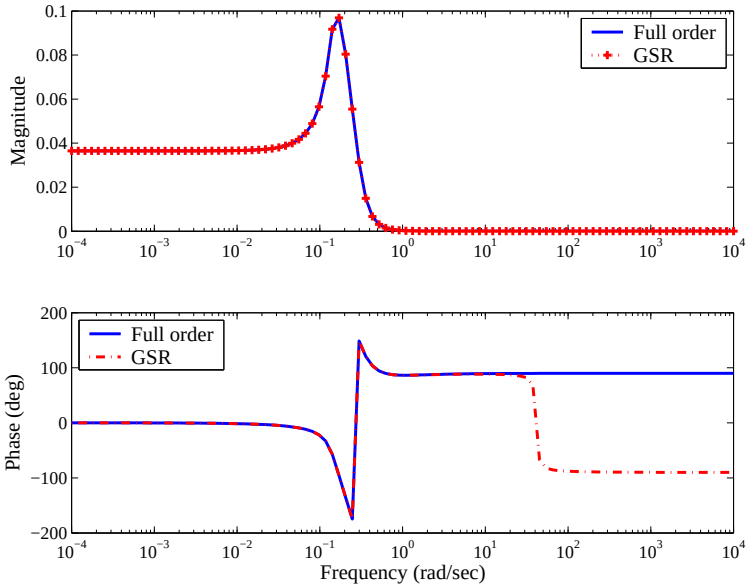

Fig. 7. Magnitude and phase plots of $G_{31}(\mathrm{i}\omega)$ for the damped mass-spring system.

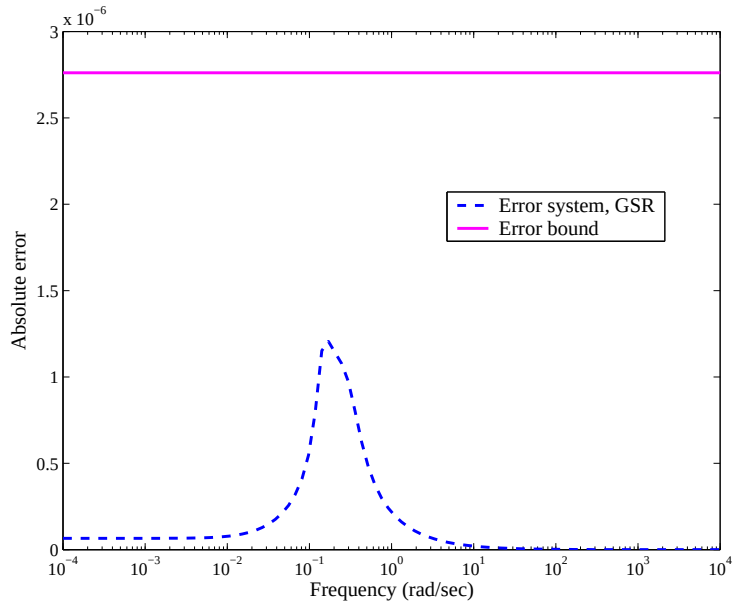

Fig. 8. Absolute error plot and error bound for the damped mass-spring system.

Figure 5 shows the normalized residual norms for the low rank Cholesky factors R_k and L_k of the proper Gramians computed by the generalized ADI method with 20 shift parameters. The approximate dominant proper Hankel singular values presented in Fig. 6 have been determined from the singular value decomposition of the matrix $L_{33}^T E R_{31}$ with $L_{33} \in \mathbb{R}^{n,99}$ and $R_{31} \in \mathbb{R}^{n,31}$. All improper Hankel singular values are zero. This implies that the transfer function $\mathbf{G}(s)$ of (34) is proper. We approximate the descriptor system (34) by a standard state space system of order $\ell = \ell_f = 10$ computed by the approximate GSR method. In Fig. 7 we display the magnitude and phase plots of the $(3,1)$ components of the frequency responses $\mathbf{G}(\mathrm{i}\omega)$ and $\widetilde{\mathbf{G}}(\mathrm{i}\omega)$. Note that there is no visible difference between the magnitude plots for the full order and reduced-order systems. Similar results have been observed for other components of the frequency response. Figure 8 shows the absolute error $\|\mathbf{G}(\mathrm{i}\omega) - \widetilde{\mathbf{G}}(\mathrm{i}\omega)\|_2$ for a frequency rang $\omega \in [\,10^{-4},\ 10^4\,]$ and the approximate error bound computed as twice the sum of the truncated approximate proper Hankel singular values. We see that the reduced-order system approximates the original system satisfactorily.

References

[AA00] Anderson, B.D.O., Antoulas, A.C.: Rational interpolation and state-variable realizations. Linear Algebra Appl., **137-138**, 479–509 (1990)

[Ant04] Antoulas, A.C.: Lectures on the Approximation of Large-Scale Dynamical Systems. SIAM, Philadelphia, PA (2004)

[ASG01] Antoulas, A.C., Sorensen, D.C., Gugercin, S.: A survey of model reduction methods for large-scale systems. In: Olshevsky, V. (ed) Structured Matrices in Mathematics, Computer Science and Engineering, Vol. I. Contemporary Mathematics Series, 280, pages 193–219. American Mathematical Society, (2001)

[ASG03] Antoulas, A.C., Sorensen, D.C., Gugercin, S.: A modified low-rank Smith method for large-scale Lyapunov equations. Numerical Algorithms, **32**, 27–55 (2003)

[ASZ02] Antoulas, A.C., Sorensen, D.C., Zhou, Y.: On the decay rate of the Hankel singular values and related issues. Systems Control Lett., **46**, 323–342 (2002)

[Bai02] Bai, Z.: Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. Appl. Numer. Math., **43**, 9–44 (2002)

[BF01] Bai, Z., Freund, R.W.: A partial Padé-via-Lanczos method for reduced-order modeling. Linear Algebra Appl., **332–334**, 141–166 (2001)

[BSSY99] Bai, Z., Slone, R.D., Smith, W.T., Ye, Q.: Error bound for reduced system model by Padé approximation via the Lanczos process. IEEE Trans. Comput. Aided Design, **18**, 133–141 (1999)

[BG96] Baker Jr., G.A., Graves-Morris P.R.: Padé Approximants. Second edition. Encyclopedia of Mathematics and its Applications, 59. Cambridge University Press, Cambridge (1996)

[Bea04] Beattie Ch.: Potential theory in the analysis of iterative methods. Lecture Notes, Technische Universität Berlin (2004)

[BV88] Beelen, T., Van Dooren, P.: An improved algorithm for the computation of Kronecker's canonical form of a singular pencil. Linear Algebra Appl., **105**, 9–65 (1988)

[Ben97] Bender, D.J.: Lyapunov-like equations and reachability/observability Gramians for descriptor systems. IEEE Trans. Automat. Control, **32**, 343–348 (1987)

[BQQ04] Benner, P., Quintana-Ortí, E.S., Quintana-Ortí, G.: Parallel model reduction of large-scale linear descriptor systems via balanced truncation. In: High Performance Computing for Computational Science. Proceedings of the 6th International Meeting VECPAR'04, (Valencia, Spain, June 28-30, 2004), pages 65–78 (2004)

[Ber90] Bernert, K.: Differenzenverfahren zur Lösung der Navier-Stokes-Gleichungen über orthogonalen Netzen. Wissenschaftliche Schriftenreihe 10/1990, Technische Universität Chemnitz (1990) [German]

[BELV91] Bojanczyk, A.W., Ewerbring, L.M., Luk, F.T., Van Dooren, P.: An accurate product SVD algorithm. Signal Process., **25**, 189–201 (1991)

[BCP89] Brenan, K.E., Campbell, S.L., Petzold, L.R.: The Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations. Elsevier, North-Holland, New York, N.Y. (1989)

[BBMN99] Bunse-Gerstner, A., Byers, R., Mehrmann, V., Nichols, N.K.: Feedback design for regularizing descriptor systems. Linear Algebra Appl., **299**, 119–151 (1999)

[Cam80] Campbell, S.L.: Singular Systems of Differential Equation, I. Pitman, San Francisco (1980)

[Cob84] Cobb, D.: Controllability, observability, and duality in singular systems. IEEE Trans. Automat. Control, **29**, 1076–1082 (1984)

[Dai89] Dai, L.: Singular Control Systems. Lecture Notes in Control and Information Sciences, 118. Springer, Berlin Heidelberg New York (1989)

[DK93a] Demmel, J.W., Kågström, B.: The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: robust software with error bounds and applications. Part I: Theory and algorithms. ACM Trans. Math. Software, **19**, 160–174 (1993)

[DK93b] Demmel, J.W., Kågström, B.: The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: robust software with error bounds and applications. Part II: Software and applications. ACM Trans. Math. Software, **19**, 175–201 (1993)

[Doe71] Doetsch, G.: Guide to the Applications of the Laplace and Z-Transforms. Van Nostrand Reinhold Company, London (1971)

[Drm00] Drmač, Z.: New accurate algorithms for singular value decomposition of matrix triplets. SIAM J. Matrix Anal. Appl., **21**, 1026–1050 (2000)

[Enn84] Enns, D.: Model reduction with balanced realization: an error bound and a frequency weighted generalization. In: Proceedings of the 23rd IEEE Conference on Decision and Control (Las Vegas, 1984), pages 127–132. IEEE, New York (1984)

[ET00] Estévez Schwarz, D., Tischendorf, C.: Structural analysis for electric circuits and consequences for MNA. Int. J. Circ. Theor. Appl., **28**, 131–162 (2000)

[FF95] Feldmann, P., Freund, R.W.: Efficient linear circuit analysis by Padé approximation via the Lanczos process. IEEE Trans. Computer-Aided Design, **14**, 639–649 (1995)

[FNG92] Fortuna, L., Nunnari, G., Gallo, A.: Model Order Reduction Techniques with Applications in Electrical Engineering. Springer, London (1992)

[Fre00] Freund, R.W.: Krylov-subspace methods for reduced-order modeling in circuit simulation. J. Comput. Appl. Math., **123**, 395–421 (2000)

[GGV94] Gallivan, K.,Grimme, E., Van Dooren, P.: Asymptotic waveform evaluation via a Lanczos method. Appl. Math. Lett., **7**, 75–80 (1994)

[GGV96] Gallivan, K., Grimme, E., Van Dooren, P.: A rational Lanczos algorithm for model reduction. Numerical Algorithms, **12**, 33–63 (1996)

[Glo84] Glover, K.: All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-errors bounds. Internat. J. Control, **39**, 1115–1193 (1984)

[GV96] Golub, G.H., Van Loan, C.F.: Matrix Computations. 3rd ed. The Johns Hopkins University Press, Baltimore, London (1996)

[GSV00] Golub, G.H., Sølna, K., Van Dooren, P.: Computing the SVD of a general matrix product/quotient. SIAM J. Matrix Anal. Appl., **22**, 1–19 (2000)

[GL83] Gragg, W.B., Lindquist, A.: On the partial realization problem. Linear Algebra Appl., **50**, 277–319 (1983)

[Gra04] Grasedyck, L.:. Existence of a low rank of $\mathcal{H}$-matrix approximation to the solution of the Sylvester equation. Numer. Linear Algebra Appl., **11**, 371–389 (2004)

[Gri97] Grimme, E.: Krylov projection methods for model reduction. Ph.D. Thesis, Iniversity of Illinois, Urbana-Champaign (1997)

[Gug03] Gugercin, S.: Projection methods for model reduction of large-scale dynamical systems. Ph.D. Thesis, Rice University, Houston, Texas (2003)

[GF99] Günther, M., Feldmann, U.: CAD-based electric-circuit modeling in industry. I. Mathematical structure and index of network equations. Surveys Math. Indust., **8**, 97–129 (1999)

[Hac00] Hackbusch, W.: A sparse matrix arithmetic based on $\mathcal{H}$-matrices. Part I: Introduction to $\mathcal{H}$-matrices. Computing, **62**, 89–108 (2000)

[HGB02] Hackbusch, W., Grasedyck, L., Börm, S.: An introduction to hierarchical matrices. Math. Bohem., **127**, 229–241 (2002)

[Ham82] Hammarling, S.J.: Numerical solution of the stable non-negative definite Lyapunov equation. IMA J. Numer. Anal., **2**, 303–323 (1982)

[JK02] Jonsson, I., Kågström, B.: Recursive blocked algorithms for solving triangular systems – Part I: One-sided and coupled Sylvester-type matrix equations. ACM Trans. Math. Software, **28**, 392–415 (2002)

[KV92] Kågström, B., Van Dooren, P.: A generalized state-space approach for the additive decomposition of a transfer function. J. Numer. Linear Algebra Appl., **1**, 165–181 (1992)

[KW89] Kågström, B., Westin, L.: Generalized Schur methods with condition estimators for solving the generalized Sylvester equation. IEEE Trans. Automat. Control, **34**, 745–751 (1989)

[Kai80] Kailath, T.: Linear Systems. Prentice-Hall Information and System Sciences Series. Prentice Hall, Englewood Cliffs, NJ (1980)

[LHPW87] Laub, A.J., Heath, M.T., Paige, C.C., Ward, R.C.: Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms. IEEE Trans. Automat. Control, **32**, 115–122 (1987)

[Li00] Li, J.-R.: Model reduction of large linear systems via low rank system Gramians. Ph.D. Thesis, Department of Mathematics, Massachusetts Institute of Technology (2000)

[LWW99] Li, J.-R., Wang, F., White, J.: An efficient Lyapunov equation-based approach for generating reduced-order models of interconnect. In: Proceedings of the 36th Design Automation Conference (New Orleans, USA, 1999), pages 1–6. IEEE (1999)

[LW02] Li, J.-R., White, J.: Low rank solution of Lyapunov equations. SIAM J. Matrix Anal. Appl., **24**, 260–280 (2002)

[LS00] Liu, W.Q., Sreeram, V.: Model reduction of singular systems. In: Proceedings of the 39th IEEE Conference on Decision and Control (Sydney, Australia, 2000), pages 2373–2378. IEEE (2000)

[LA89] Liu, Y., Anderson, B.D.O.: Singular perturbation approximation of balanced systems. Internat. J. Control, **50**, 1379–1405 (1989)

[Mar96] März, R.: Canonical projectors for linear differential algebraic equations. Comput. Math. Appl., **31**, 121–135 (1996)

[Moo81] Moore, B.C.: Principal component analysis in linear systems: controllability, observability, and model reduction. IEEE Trans. Automat. Control, **26**, 17–32 (1981)

[Pen98] Penzl, T.:. Numerical solution of generalized Lyapunov equations. Adv. Comput. Math., **8**, 33–48 (1998)

[Pen99] Penzl, T.: A cyclic low-rank Smith method for large sparse Lyapunov equations. SIAM J. Sci. Comput., **21**, 1401–1418 (1999/2000)

[Pen99b] Penzl, T.: Algorithms for model reduction of large dynamical systems. Preprint SFB393/99-40, Fakultät für Mathematik, Technische Universität Chemnitz, D-09107 Chemnitz, Germany, (1999) Available from `http://www.tu-chemnitz.de/sfb393/sfb99pr.html`.

[Pen00a] Penzl, T.: Eigenvalue decay bounds for solutions of Lyapunov equations: the symmetric case. Systems Control Lett., **40**, 139-144 (2000)

[Pen00b] Penzl, T.: LYAPACK Users Guide. Preprint SFB393/00-33, Fakultät für Mathematik, Technische Universität Chemnitz, D-09107 Chemnitz, Germany (2000) Available from `http://www.tu-chemnitz.de/sfb393/sfb00pr.html`.

[PS94] Perev, K., Shafai, B.: Balanced realization and model reduction of singular systems. Internat. J. Systems Sci., **25**, 1039–1052 (1994)

[Rud87] Rudin, W.: Real and Complex Analysis. McGraw-Hill, New York (1987)

[Rue84] Ruhe, A.: Rational Krylov sequence methods for eigenvalue computation. Linear Algebra Appl., **58**, 391–405 (1984)

[Saa96] Saad, Y.: Iterative Methods for Sparse Linear Systems. PWS Publishing Company, Boston, MA, 1996.

[SC89] Safonov, M.G., Chiang R.Y.: A Schur method for balanced-truncation model reduction. IEEE Trans. Automat. Control, **34**, 729–733 (1989)

[Sch95] Schüpphaus, R.: Regelungstechnische Analyse und Synthese von Mehrkörpersystemen in Deskriptorform. Ph.D. thesis, Fachbereich Sicherheitstechnik, Bergische Universität-Gesamthochschule Wuppertal. Fortschritt-Berichte VDI, Reihe 8, Nr. 478. VDI Verlag, Düsseldorf (1995) [German]

[Sok03] Sokolov, V.I.:. On realization of rational matrices. Technical Report 31-2003, Institut für Mathematik, Technische Universität Berlin, D-10263 Berlin, Germany (2003)

[SZ02] Sorensen, D.C., Zhou, Y.: Bounds on eigenvalue decay rates and sensitivity of solutions of Lyapunov equations. Technical Report TR02-07, Department of Computational and Applied Mathematics, Rice University, Houston, TX (2002)

[SS90] Stewart, G.W., Sun, J.-G.: Matrix Perturbation Theory. Academic Press, New York (1990)

[Sty02a] Stykel, T.: Analysis and numerical solution of generalized Lyapunov Equations. Ph.D. Thesis, Institut für Mathematik, Technische Universität Berlin, (2002)

[Sty02b] Stykel, T.: Numerical solution and perturbation theory for generalized Lyapunov equations. Linear Algebra Appl., **349**, 155–185 (2002)

[Sty02c] Stykel, T.: On criteria for asymptotic stability of differential-algebraic equations. Z. Angew. Math. Mech., **82**, 147–158 (2002)

[Sty03] Stykel, T.: Input-output invariants for descriptor systems. Preprint PIMS-03-1, The Pacific Institute for the Mathematical Sciences, Canada (2003)

[Sty04a] Stykel, T.: Balanced truncation model reduction for semidiscretized Stokes equation. To appear in Linear Algebra Appl. (2004)

[Sty04b] Stykel, T.: Gramian-based model reduction for descriptor systems. Math. Control Signals Systems, **16**, 297–319 (2004)

[Sty04c] Stykel, T.: Low rank iterative methods for projected generalized Lyapunov equations. In: De Moore, B., Motmans, B., Willems, J., Van Dooren, P., Blondel, V. (eds) Proceedings of the 16-th International Symposium on Mathematical Theory of Networks and Systems (Leuven, Belgium, July 5-9, 2004) (2004)

[TP84] Tombs, M.S., Postlethweite, I.: Truncated balanced realization of a stable non-minimal state-space system. Internat. J. Control, **46**, 1319–1330 (1987)

[Var87] Varga, A.: Efficient minimal realization procedure based on balancing. In: EL Moudni, A., Borne, P., Tzafestas, S.G. (eds) Proceedings of the IMACS/IFAC Symposium on Modelling and Control of Technological Systems (Lille, France, May 7-10, 1991), volume 2, pages 42–47 (1991)

[Var98] Varga, A.: Computation of coprime factorizations of rational matrices. Linear Algebra Appl., **271**, 83–115 (1998)

[Var00] Varga, A.: A Descriptor Systems Toolbox for MATLAB. In: Proceedings of the 2000 IEEE International Symposium on Computer Aided Control System Design (Anchorage, Alaska, September 25-27, 2000), pages 150–155 (2000). Available from `http://www.robotic.dlr.de/control/publications/2000/varga_cacsd2000p2.pdf`

[VLK81] Verghese, G.C., Lévy, B.C., Kailath, T.: A generalized state-space for singular systems. IEEE Trans. Automat. Control, **26**, 811–831 (1981)

[Wat00] Watkins, D.S.: Performance of the QZ algorithm in the presence of infinite eigenvalues. SIAM J. Matrix Anal. Appl., **22**, 364–375 (2000)

[YS81] Yip, E.L., Sincovec, R.F.: Solvability, controllability and observability of continuous descriptor systems. IEEE Trans. Automat. Control, **26**, 702–707 (1981)

[ZDG96] Zhou, K., Doyle, J.C., Glover, K.: Robust and Optimal Control. Prentice Hall, Upper Saddle River, NJ (1996)