Peter Benner Patrick Kürschner Jens Saak

An Improved Numerical Method for Balanced Truncation for Symmetric Second Order Systems

Max Planck Institute Magdeburg Preprints

MPIMD/12-20 **November 8, 2012**

Impressum:

Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg

Publisher:
Max Planck Institute for Dynamics of Complex Technical Systems

Address:
Max Planck Institute for Dynamics of Complex Technical Systems
Sandtorstr. 1
39106 Magdeburg

www.mpi-magdeburg.mpg.de/preprints

An Improved Numerical Method for Balanced Truncation for Symmetric Second Order Systems

Peter Benner*, Patrick Kürschner*, Jens Saak*

November 8, 2012

Abstract

We consider balanced truncation model order reduction for symmetric second order systems. The occurring large-scale generalized and structured Lyapunov equations are solved with a specially adapted low-rank ADI type method. Stopping criteria for this iteration are investigated and a new result concerning the Lyapunov residual within the low-rank ADI method is established. We also propose a goal oriented stopping criterion which tries to incorporate the balanced truncation approach already during the ADI iteration. The model reduction approach using the ADI method with different stopping criteria is evaluated on several test systems.

Keywords model order reduction; balanced truncation; alternating directions implicit; Lyapunov equations; second-order systems

AMS subject classification 15A18, 15A24, 93A15, 93C, 93D20

1 Introduction

This work is an extension of [1] which appeared in the proceedings of the 2012 MathMod conference in Vienna 2012. It is concerned with model reduction for linear, time-invariant control systems in second order form:

$$M\ddot{x}(t) + D\dot{x}(t) + Kx(t) = Bu(t), \tag{1a}$$

$$y(t) = C^{\mathrm{v}}\dot{x}(t) + C^{\mathrm{p}}x(t) \tag{1b}$$

*Research Group Computational Methods in Systems and Control Theory (CSC), Max Planck Institute for Dynamics of Complex Technical Systems Magdeburg, Sandtorstr. 1, 39106 Magdeburg, Germany, {benner, kuerschner, saak}@mpi-magdeburg.mpg.de.

with large and sparse coefficient matrices $M,\ D,\ K \in \mathbb{R}^{n\times n}$, an input matrix $B \in \mathbb{R}^{n\times m}$ position and, respectively, velocity output matrices and $C^{\mathrm{p}},\ C^{\mathrm{v}} \in \mathbb{R}^{p\times n}$. The time dependent vectors $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$ and $y(t) \in \mathbb{R}^p$ are usually referred to as state, input and output vector. We mainly consider a special class of (1), namely symmetric second order systems which arise, e.g., in elastic mechanical multibody systems [2] or circuit simulation [3]. In modern computer simulations of such systems the demand for a highly accurate description of realistic phenomena leads, in one way or another, to a very large state space dimension. For a numerically efficient time-domain solution, model order reduction approaches are employed which drastically reduce the dimension of the state space.
Here we focus on the balanced truncation [4] model order reduction method as a system theoretic approach. Some more recent adaptions take the second order structure of (1) into account [3, 5, 6]. Balanced truncation requires the solution of large-scale Lyapunov equation. For this purpose we discuss a specially tailored version of the low-rank ADI (LR-ADI) iteration [7, 8]. We strongly focus on termination criteria of this iterative method.

This paper is organized as follows: The next section defines symmetric second order systems and gives some applications where they naturally arise. In Section 3 the main ideas behind balanced truncation for first and second order systems are recalled. The approximate solution of the occurring large-scale Lyapunov equation within the LR-ADI method is investigated in Section 4. There we give a novel expression for the Lyapunov residual in the ADI method in general. We also apply another stopping criterion proposed in [9]. We further review a modification of the LR-ADI which exploits the structure introduced by (1). Some perspectives on the handling of non-standard coefficients D in (1), e.g., the case when D is given by $D = \tilde{D} + uu^T$, are also discussed. The balanced truncation approaches as well as the ADI type methods for solving the Lyapunov equations are evaluated with some test examples in Section 5 and Section 6 concludes.

Notation used in this paper: $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex numbers, and $\mathbb{R}_-$, $\mathbb{C}_-$ refer to the set of strictly negative numbers and the open left half plane. $\mathbb{R}^{n\times m}$, $\mathbb{C}^{n\times m}$ are $n \times m$ real and complex matrices, respectively. For a complex quantity $X = \mathrm{Re}\,(X) + \jmath\,\mathrm{Im}\,(X)$, $\mathrm{Re}\,(X)$, $\mathrm{Im}\,(X)$ denote its real and imaginary part, where $\jmath$ is the imaginary unit. The complex conjugate of a complex quantity X is denoted by $\overline{X} = \mathrm{Re}\,(X) - \jmath\,\mathrm{Im}\,(X)$. The absolute value of $\xi \in \mathbb{C}$ is denoted by $|\xi|$. The matrix A^T is the transpose of a real $n \times m$ matrix, and $A^H = \overline{A}^T$ is the complex conjugate transpose of a complex matrix. The inverse of A is denoted by A^{-1} and expressions of the form $x = A^{-1}b$ should always be understood as solving the linear system of equations $Ax = b$ for x. The identity matrix of dimension n is indicated by I_n. If not stated otherwise $\|\cdot\|$ is the Euclidean vector, or subordinate matrix norm.

2 Symmetric Second Order Systems

In the remainder we restrict ourselves to the case when M, D, K in (1) are symmetric. If they are in addition positive definite the matrix polynomial $\lambda^2 M + \lambda D + K$ has eigenvalues only in $\mathbb{C}_-$ which implies that the second order system is asymptotically stable. This is a sufficient condition for the applicability of balanced truncation approaches and is hence assumed in the following.

Definition 2.1. *A system of the form (1) with symmetric M, D, K is called symmetric second order system of the first kind if*

$$C^p = \pm B^T, \quad C^v = 0,$$

and of the second kind if

$$C^p = 0, \quad C^v = \pm B^T.$$

2.1 Applications

One important application, where symmetric second order systems of the first kind frequently occur, is vibration analysis of elastic mechanical structures. There M, D, K refer to mass, damping and stiffness of a finite element model of the structure under consideration. The states in $x(t)$ are usually the nodal displacements of the finite element model. The condition $C^{\mathrm{p}} = B^T$ occurs, e.g., in elastic multibody systems [2].

The second kind appears in the simulation of RLCK circuits [3], where M, D, K are conductance, capacitance and susceptance matrices, and $x(t)$, $u(t)$, $y(t)$ contain nodal voltages, input currents and output voltages, respectively. For example, in the design of microchips, the high dimension of the dynamical system comes from the sheer high number of resistors, conductors and inductors in modern circuit designs, or because nonlinear circuit elements such as diodes and transistors are replaced by large sub-networks of linear elements.

2.2 Transformations to First Order Systems

In order to apply most modern model order reduction techniques, including balanced truncation, the starting point is to rewrite (1) formally into an equivalent generalized state space system of first order

$$\begin{aligned} E\dot{z}(t) &= Az(t) + Gu(t), \\ y(t) &= Lz(t), \end{aligned} \tag{2}$$

with $E, A \in \mathbb{R}^{2n\times 2n}$, $G \in \mathbb{R}^{2n\times m}$, $L \in \mathbb{R}^{p\times 2n}$ and the augmented generalized state vector $z(t) = [x(t)^T, \dot{x}(t)^T]^T$. This is closely related to rewriting the associated quadratic matrix polynomial $\lambda^2 M + \lambda D + K$ into a linear pencil $\lambda E - A$ [10]. Hence we will refer to this transformation as linearization. The two most standard linearizations are

the *first companion linearization*

$$E_1 = \begin{bmatrix} N & 0 \\ 0 & M \end{bmatrix}, \; A_1 = \begin{bmatrix} 0 & N \\ -K & -D \end{bmatrix}, \; G_1 = \begin{bmatrix} 0 \\ B \end{bmatrix}, \; L_1 = \begin{bmatrix} C^{\mathrm{p}} & C^{\mathrm{v}} \end{bmatrix} \tag{3a}$$

and the *second companion linearization*

$$E_2 = \begin{bmatrix} D & M \\ F & 0 \end{bmatrix}, \; A_2 = \begin{bmatrix} -K & 0 \\ 0 & F \end{bmatrix}, \; G_2 = \begin{bmatrix} B \\ 0 \end{bmatrix}, \; L_2 = \begin{bmatrix} C^{\mathrm{p}} & C^{\mathrm{v}} \end{bmatrix}, \tag{3b}$$

where F, N are arbitrary nonsingular $n \times n$ matrices. Common choices for those are $N = I_n$ or $N = -K$ in (3a) and $F = I_n$ or $F = M$ in (3b). Note that it is possible to transform (3b) into (3a) and vice versa via a generalized state-space transformation

$$U_l A_2 U_r = A_1, \; U_l E_2 U_r = E_1, \; U_l G_2 = G_1, \; L_2 U_r = L_1 \tag{4}$$

with

$$U_l = \begin{bmatrix} 0 & NF^{-1} \\ I_n & -DF^{-1} \end{bmatrix}, \; U_r = I_{2n}, \tag{5}$$

see also [6, Section 3.1].

For symmetric second order systems, a reasonable choice is to use (3b) with $F = M$ for the first kind, since then E_2, A_2 are symmetric and $G_2 = \pm L_2^T$ such that (2) is a symmetric first order system. Similarly, for the second kind a good choice is to take (3a) with $N = -K$, since again E_1, A_1 are symmetric and $G_1 = \pm L_1^T$. These symmetric linearizations will play an important role later.

2.3 Reachability and Observability Gramians

Under the assumption that the generalized first order state-space system (2) is asymptotically stable, the generalized Lyapunov equations

$$APE^T + EPA^T = -GG^T, \tag{6a}$$

$$A^T QE + E^T QA = -L^T L. \tag{6b}$$

have unique, positive semidefinite solutions P and Q. There, P is the *reachability Gramian*, whereas $E^T QE$ is referred to as *observability Gramian*. These quantities will be the main ingredients for the balanced truncation model order reduction approaches described in Section 3. The magnitudes of the square roots of the eigenvalues of $PE^T QE$ are system invariants and referred to as *Hankel singular values* (HSVs). They give a measurement of how good states can be observed and controlled (see, e.g., [11]).

Let us briefly discuss the relation of these Gramians, respectively the HSVs, to different linearizations of (1) and implications to symmetric second order systems. Note that due to (4) we have that (6) with E_1, A_1, G_1, L_1 is equivalent to

$$A_2 U_r P_1 U_r^T E_2^T + E_2 U_r P_1 U_r^T A_2^T = -G_2 G_2^T,$$
$$A_2^T U_l^T Q_1 U_l E_2 + E_2^T U_l^T Q_1 U_l A_2 = -L_2^T L_2,$$

such that the solutions $P_1,\ Q_1$ of the first companion linearization are related to $P_2,\ Q_2$ of the second companion linearization via

$$P_2 = P_1, \quad Q_2 = U_l^T Q_1 U_l. \tag{7}$$

This implies for the observability Gramians that it holds

$$E_2^T Q_2 E_2 = E_2 U_r = U_r E_1^T U_l^{-T} U_l^T Q_1 U_l U_l^{-1} E_1 U_r = E_1^T Q_1 E_1,$$

and hence the HSVs of both linearizations are identical as already proved in (4).

Now for a symmetric second order system of the first kind with $F = M$, we have in the second companion form $E_2 = E_2^T,\ A_2 = A_2^T$ and $L_2 = \pm G_2^T$ such that it holds additionally $P_1 = P_2 = Q_2$. A similar result holds for the second kind in first companion form with $N = -K$. To conclude, for symmetric second order systems the knowledge of either the reachability or the observability Gramian is sufficient to determine the HSVs, provided the symmetry and the relations (7) are exploited.

3 Balanced truncation

3.1 First order systems

Here we briefly discuss the main ideas to carry out balanced truncation for asymptotically stable first order systems (2) with invertible E. Note that balanced truncation was originally designed for the case $E = I_n$ [4], but the modification to the generalized case is straightforward [12, 13]. Balanced truncation is guided by the identification of those states requiring the most energy to control, and which yield the least energy through observation. This energy interpretation is equivalent to asking for states which are hard to control and to observe. This information is revealed by the HSVs, such that the solutions P and Q of (6) are the main requirements for balanced truncation. In practice, balanced truncation is often carried out with Cholesky-like factorizations of $P,\ Q$ or, as it is usually the case for large-scale systems, use low-rank representations $P \approx \tilde{Z}_c \tilde{Z}_c^T$ and $Q \approx \tilde{Z}_o \tilde{Z}_o^T$ with $\operatorname{rank}(\tilde{Z}_c),\ \operatorname{rank}(\tilde{Z}_o) \ll n$. The computation of such low-rank factors $\tilde{Z}_c,\ \tilde{Z}_o$ is studied in Section 4. In any case, a singular value decomposition with decreasingly ordered singular values

$$\tilde{Z}_o^T E \tilde{Z}_c = X \Sigma Y^T = [X_1,\ X_2] \operatorname{diag}(\Sigma_1,\ \Sigma_2) [Y_1,\ Y_2]^T, \tag{8}$$

reveals the $r \ll n$ dominant singular values $\sigma_1 \geq \ldots \geq \sigma_r > 0$ in the block $\Sigma_1 \in \mathbb{R}^{r \times r}$. The left and right truncation matrices are then computed by

$$T_l = \tilde{Z}_o X_1 \Sigma_1^{-\frac{1}{2}} \text{ and } T_r = \tilde{Z}_c Y_1 \Sigma_1^{-\frac{1}{2}} \tag{9}$$

and the reduced system of order r is given by

$$\begin{aligned} \dot{\tilde{z}}(t) &= \tilde{A}\tilde{z}(t) + \tilde{G}u(t), \\ \tilde{y}(t) &= \tilde{L}\tilde{z}(t), \end{aligned}$$

where $\tilde{A} := T_l^T A T_r \in \mathbb{R}^{r\times r}$, $\tilde{G} := T_l G \in \mathbb{R}^{r\times m}$ and $\tilde{L} := L T_r \in \mathbb{R}^{p\times r}$. Note that $T_l^T E T_r = I_r$ holds by construction (e.g., [13]). The most important features of balanced truncation are the guaranteed stability preservation and the readily available error bound

$$\|y - \tilde{y}\|/\|u\| \leq 2 \sum_{j=r+1}^{n} \sigma_j =: \rho, \tag{10}$$

where the σ_j are the singular values from the neglected block Σ_2. One can easily monitor when ρ drops below a specified tolerance to adaptively determine the reduced order model dimension.

The disadvantage of this approach is that if (2) is a first order representation of a second order system, the second order structure will be lost in the reduced order model. This especially means that the reduced states no longer have physical interpretations. In the next section we investigate how the second order structure can be preserved.

3.2 Symmetric Second Order systems

Balanced truncation can be modified in order to generate a reduced order model which is also in second order form, see e.g., [6, 3, 5]. Consider the partition of the Gramians P and Q according to the structure present in the equivalent generalized first order system:

$$P = \begin{bmatrix} P_{\mathrm{p}} & P_{1,2} \\ P_{1,2}^T & P_{\mathrm{v}} \end{bmatrix}, \; Q = \begin{bmatrix} Q_{\mathrm{p}} & Q_{1,2} \\ Q_{1,2}^T & Q_{\mathrm{v}} \end{bmatrix},$$

where $P_{\mathrm{p}},\ Q_{\mathrm{p}} \in \mathbb{R}^{n\times n}$ and $P_{\mathrm{v}},\ Q_{\mathrm{v}} \in \mathbb{R}^{n\times n}$ are called *position*, *velocity* reachability and observability Gramians, respectively.

For symmetric systems it is enough to consider only P_{p} and P_{v} [2, 1], or respectively their Cholesky factorizations $P_{\mathrm{p}} = Z_{\mathrm{p}} Z_{\mathrm{p}}^T$, $P_{\mathrm{v}} = Z_{\mathrm{v}} Z_{\mathrm{v}}^T$. As for first order systems, low-rank representations are usually used in a large-scale setting. Following the approach in [6], one can choose between four possible singular value decompositions

$$\begin{aligned} Z_\alpha^T M Z_\beta &= X_{\alpha\beta} \Sigma_{\alpha\beta} Y_{\alpha\beta}^T \\ &= [X_{\alpha\beta,1}, X_{\alpha\beta,2}] \begin{bmatrix} \Sigma_{\alpha\beta,1} & 0 \\ 0 & \Sigma_{\alpha\beta,2} \end{bmatrix} [Y_{\alpha\beta,1}, Y_{\alpha\beta,2}]^T, \end{aligned} \tag{11}$$

where the subscripts $\alpha,\ \beta \in \{\mathrm{p},\ \mathrm{v}\}$ denote whether the position or velocity blocks of P are used. Similar to the first order case, the $\Sigma_{\alpha\beta,1} \in \mathbb{R}^{r\times r}$ block contains the largest singular values. Depending on the choice of α, β, they are referred to as

- position-position (PP) if $\alpha = \beta = \mathrm{p}$,
- velocity-velocity (VV) if $\alpha = \beta = \mathrm{v}$,
- velocity-position (VP) if $\alpha = \mathrm{v}, \beta = \mathrm{p}$, and

- position-velocity (PV) singular-values if $\alpha = \mathrm{p}, \beta = \mathrm{v}$.

This yields four different pairs of matrices which perform the reduction:

$$\begin{aligned} \tilde{T}_{r,\alpha\beta} &:= Z_\alpha Y_{\alpha\beta,1} \Sigma_{\alpha\beta,1}^{-\frac{1}{2}}, \\ \tilde{T}_{l,\alpha\beta} &:= Z_\beta X_{\alpha\beta,1} \Sigma_{\alpha\beta,1}^{-\frac{1}{2}}. \end{aligned} \tag{12}$$

For $\alpha = \beta = \mathrm{p}$ the approach is called position-position, for $\alpha = \beta = \mathrm{v}$ velocity-velocity, for $\alpha = \mathrm{v}$, $\beta = \mathrm{p}$ velocity-position, and for $\alpha = \mathrm{p}$, $\beta = \mathrm{v}$ position-velocity second order balanced truncation, respectively. All four pairs of these transformation matrices are summarized in Table 1. The four different possible reduced order models in second order form are then consequently given by

$$\begin{aligned} \tilde{M}_{\alpha\beta}\ddot{\tilde{x}}(t) + \tilde{D}_{\alpha\beta}\dot{\tilde{x}}(t) + \tilde{K}_{\alpha\beta}\tilde{x}(t) &= \tilde{B}_{\alpha\beta}u(t), \\ \tilde{y}(t) &= \tilde{C}_{\alpha\beta}^{\mathrm{p}}\tilde{x}(t) + \tilde{C}_{\alpha\beta}^{\mathrm{v}}\dot{\tilde{x}}(t), \end{aligned}$$

with

$$\begin{aligned} &\tilde{M}_{\alpha\beta} := \tilde{T}_{l,\alpha\beta}^T M \tilde{T}_{r,\alpha\beta},\ \tilde{D}_{\alpha\beta} := \tilde{T}_{l,\alpha\beta}^T D \tilde{T}_{r,\alpha\beta},\ \tilde{K}_{\alpha\beta} := \tilde{T}_{l,\alpha\beta}^T K \tilde{T}_{r,\alpha\beta} \in \mathbb{R}^{r\times r}, \\ &\tilde{B}_{\alpha\beta} := \tilde{T}_{l,\alpha\beta}^T B \in \mathbb{R}^{r\times m},\ \tilde{C}_{\alpha\beta}^{\mathrm{p}} := B^T \tilde{T}_{r,\alpha\beta},\ \tilde{C}_{\alpha\beta}^{\mathrm{v}} := B^T \tilde{T}_{r,\alpha\beta} \in \mathbb{R}^{m\times r}. \end{aligned} \tag{13}$$

It can easily be shown that the reduced mass matrices are always equal to the identity. The preservation of the second order structure comes at the price of the absence of an error bound like (10). However, as an alternative one may monitor the ratio of the entries in $\Sigma_{\alpha\beta}$ and determine the reduced dimension once

$$\frac{\sigma_{j,\alpha\beta}}{\sigma_{1,\alpha\beta}} \leq \epsilon \tag{14}$$

is fulfilled [13]. Stability is preserved in the position-position and velocity-velocity approaches [2, 1] which also preserve the symmetry of the system, i.e., all reduced coefficient matrices remain symmetric and positive definite and the reduced input

Table 1: Left and right transformations matrices of balanced truncation for second order systems.

Type	right transformation	left transformation
PP	$\tilde{T}_{r,\mathrm{pp}} := R_{\mathrm{p}} Y_{\mathrm{pp},1} \Sigma_{\mathrm{pp},1}^{-\frac{1}{2}}$	$\tilde{T}_{l,\mathrm{pp}} := R_{\mathrm{p}} X_{\mathrm{pp},1} \Sigma_{\mathrm{pp},1}^{-\frac{1}{2}}$
PV	$\tilde{T}_{r,\mathrm{pv}} := R_{\mathrm{p}} Y_{\mathrm{pv},1} \Sigma_{\mathrm{pv},1}^{-\frac{1}{2}}$	$\tilde{T}_{l,\mathrm{pv}} := R_{\mathrm{v}} X_{\mathrm{pv},1} \Sigma_{\mathrm{pv},1}^{-\frac{1}{2}}$
VP	$\tilde{T}_{r,\mathrm{vp}} := R_{\mathrm{v}} Y_{\mathrm{vp},1} \Sigma_{\mathrm{vp},1}^{-\frac{1}{2}}$	$\tilde{T}_{l,\mathrm{vp}} := R_{\mathrm{p}} X_{\mathrm{vp},1} \Sigma_{\mathrm{vp},1}^{-\frac{1}{2}}$
VV	$\tilde{T}_{r,\mathrm{vv}} := R_{\mathrm{v}} Y_{\mathrm{vv},1} \Sigma_{\mathrm{vv},1}^{-\frac{1}{2}}$	$\tilde{T}_{l,\mathrm{vv}} := R_{\mathrm{v}} X_{\mathrm{vv},1} \Sigma_{\mathrm{vv},1}^{-\frac{1}{2}}$

matrix is still the transpose of the reduced (position or velocity) output matrix. It can also be shown that the velocity-position and position-velocity reduced order models are adjoint to each other. Hence, they show the same frequency response plot in the spectral or Frobenius norm [2].

4 Solving Large-Scale Lyapunov Equations

Since balanced truncation approaches rely on the reachability and observability Gramians, they require the numerical solution of the large-scale generalized Lyapunov equations (6), where for the symmetric second order systems of interest (6a) and (6b) can be made identical by certain linearizations. Hence it is sufficient to consider one single matrix equation

$$AXE^T + EXA^T + GG^T = 0. \tag{15}$$

Since direct methods [14, 15, 16] for solving (15) have a cubic computational complexity and quadratic memory demands for finding the solution X, their usage is limited to small or moderately large n. In large-scale settings, the numerical rank of the solution X is, however, often very small [17, 18], such that it appears reasonable to approximate X via low-rank factors $ZZ^T \approx X$ with $Z \in \mathbb{R}^{n\times t}$, $\operatorname{rank}(Z) = t$, $t \ll n$. Note that the balanced truncation approaches can be carried out with low-rank factors instead of full Cholesky factors in (8),(9). Algorithms for computing such low-rank factors mainly divide into two classes: projection methods using Krylov subspaces and low-rank Smith or ADI type iterations, where we focus on the latter one in this work. In the next subsection we review the generalized low-rank ADI iteration for solving (15) and give a brief motivation why we prefer ADI methods over Krylov subspace methods. Since in the context of second order systems, the occurring coefficient matrices E, A, G are structured, a modified structure exploiting version of the low-rank ADI iterations is discussed in Subsection 4.2. Some remarks on non-standard choices for the matrix D in (1) are given in Subsection 4.3. Stopping criteria are discussed in the subsequent subsection, where we give a new result for the Lyapunov residual in low-rank ADI methods, but also consider an approach based on the (Hankel) singular values which represents a goal oriented termination [9] in balanced truncation.

4.1 The low-rank ADI iteration

For standard Lyapunov equations ($E = I_n$), the alternating directions implicit (ADI) method [19, 20] iteratively computes approximations X_j, $j \geq 1$, of the solution X. Introducing the low-rank representations $X_j = Z_j Z_j^H$ into this iteration leads, after a couple of clever basic manipulations, to the low-rank ADI iteration (LR-ADI) [8, 21, 7]. It can easily be adapted to the general case (15) which yields the generalized LR-ADI (G-LR-ADI) [13, 12] given in Algorithm 1. In the remainder we always assume $X_0 = Z_j Z_j^H = 0$. The shift parameters $\{\mu_1, \ldots, \mu_{j_{\max}}\} \subset \mathbb{C}_-$ steer the convergence of the algorithm and are tightly related to the spectrum $\Lambda(A, E) = \Lambda(\lambda^2 M + \lambda D + K)$ via a rational optimization problem [20]. For large-scale matrices, the whole spectrum

Algorithm 1 Generalized Low-rank ADI iteration (G-LR-ADI)

Input: A, E and G as in (15) and shift parameters $\{\mu_1, \ldots, \mu_{j_{\max}}\}$.
Output: $Z = Z_{k_{\max}} \in \mathbb{C}^{n \times m j_{\max}}$, such that $ZZ^H \approx P$

1: **for** $k = 1, 2, \ldots, k_{\max}$ **do**
2: **if** $k = 1$ **then**
3: Solve $(A + \mu_1 E)V_1 = G$ for V_1.
4: $Z_1 = \sqrt{-2\,\mathrm{Re}\,(\mu_1)}V_1$.
5: **else**
6: Solve $(A + \mu_k E)\tilde{V} = EV_{k-1}$ for $\tilde{V}$.
7: $V_k = V_{k-1} - (\mu_k + \overline{\mu_{k-1}})\tilde{V}$.
8: Update low-rank factor $Z_k = [Z_{k-1}, \quad \sqrt{-2\,\mathrm{Re}\,(\mu_k)}V_k]$.
9: **end if**
10: **end for**

is usually unknown and therefore one uses, e.g., a small number of Ritz values which are used to solve the optimization problem in an approximate sense to get so called heuristic shift parameters [7]. Clearly, the linear systems in steps 3 and 7 cause the largest amount of computational work in the whole algorithm. We assume that we are able to solve them with either sparse direct [22, 23] or iterative solvers [24, 25].

The other large class of numerical methods for large Lyapunov equations are Krylov subspace projection methods, see e.g. [26] and the references therein. There, the large Lyapunov equation is projected onto a low-dimensional subspace $\mathcal{Y} = \mathrm{colspan}\,(Y) \subset \mathbb{R}^n$, $Y \in \mathbb{R}^{n \times k}$, and the resulting Lyapunov equation of small dimension is solved with direct methods for its solution $\hat{X} \in \mathbb{R}^{k \times k}$. The approximate solution of (15) is then obtained by $X \approx Y\hat{X}Y^T$. The various methods belonging to this class usually differ in the way $\mathcal{Y}$ is constructed. However, in order to ensure the unique solvability of the small matrix equation, one has to assume in the case $E = I_n$ that A is dissipative, i.e., $A + A^T$ is negative definite. For the generalized case this amounts to $AE^T + EA^T < 0$. In our case A, E, are block structured matrices representing a linearization of a second order system and it can be shown that, except for unnatural cases [27], the dissipativity condition is not fulfilled. Consider for instance the system with $M = 1$, $D = \alpha > 0$, $K = \beta > 0$. The matrix $AE^T + EA^T$ is indefinite for all $\beta \neq 1$. Thus, Krylov subspace methods may break down. ADI type methods do not require dissipative matrices, or matrix pairs, which makes them adequate for solving the generalized Lyapunov equation encountered in balanced truncation for second order systems.

4.2 Handling the Second Order Structure and Complex Shift Parameters

The main computational effort in Algorithm 1 lies in the solution of the linear systems in steps 3 and 7, where E, A, G are structured as in (3a) or (3b). As shown in [13, 27, 2, 1], G-LR-ADI can be rearranged to obtain the solution without explicitly forming the matrices of the linearization. By partitioning the increments of Algorithm

1 into an upper and a lower part according to the structure in the first order system (2),

$$V_k = \begin{bmatrix} V_k^{(\mathrm{p})} \\ V_k^{(\mathrm{v})} \end{bmatrix},$$

the linear system

$$(A + \mu E)V_k = EV_{k-1}$$

of dimension $2n \times 2n$, where E, A are as in the second companion linearization (3b) with $F = M$, is equivalent to

$$(\mu_k^2 M - \mu_k D + K)V_k^{(\mathrm{p})} = (\mu_k M - D)V_{k-1}^{(\mathrm{p})} - MV_{k-1}^{(\mathrm{v})} \tag{16}$$

$$V_k^{(\mathrm{v})} = V_{k-1}^{(\mathrm{p})} - \mu_k V_k^{(\mathrm{p})} \tag{17}$$

and similarly for (3), see e.g., [1, 2]. Clearly, solving the linear systems of dimension $n \times n$ is more efficient than using the blocked versions of dimension $2n \times 2n$. The rearrangement is usually referred to as second order LR-ADI (SO-LR-ADI) [13, 27]. SO-LR-ADI is illustrated in Algorithm 2, where the $n \times n$ linear systems involving the quadratic matrix polynomial in μ can be found in steps 3 and 6. By using the relations (4), it is immediately evident that another choice for the linearization will lead to the same expressions as above, such that symmetric second order systems of the second kind can be handled without any change of the core steps of Algorithm 2, but some differences occur for the Lyapunov residual.

The Steps 10 to 17 append the increment V_k to the previous low-rank solution factor such that it is a real matrix even if the current shift μ_k was complex. In the first case this is the same as in Algorithm 1. However, in the latter case it is assumed that μ_k, $\mu_{k+1} := \overline{\mu_k}$ holds, and Z is augmented by $2m$ new real columns. This construction is possible due to a result in [28] which allows the computation of real low-rank factors even if complex shift parameters are used. The main result there is that the increment with respect to μ_{k+1} can be constructed by

$$V_{k+1} = \operatorname{Re}(V_k) - \jmath \operatorname{Im}(V_k) + 2\tfrac{\operatorname{Re}(\mu_k)}{\operatorname{Im}(\mu_k)} \operatorname{Im}(V_k)$$

such that one implicitly carries out a double iteration step with respect to μ_k, $\overline{\mu_k}$ without explicitly solving the complex linear system for $\overline{\mu_k}$. This does not only reduce the amount of complex arithmetic operations and storage, but also always enables the generation of real low-rank factors which is beneficial for carrying out balanced truncation model order reduction. Note that the remaining complex arithmetic operations, which are mainly the remaining complex linear systems, can be avoided by working entirely on $\operatorname{Re}(V_k)$ and $\operatorname{Im}(V_k)$ which involves equivalent $2n \times 2n$ real linear systems [29]. However, it is often faster to solve the remaining $n \times n$ complex systems. If an iterative Krylov subspace method is used, one can exploit that due to the symmetry of M, D, K, the linear system in (16) is complex symmetric and employ solvers which make use of this property [30].

Algorithm 2 Second-Order Low-rank ADI iteration (SO-LR-ADI)

Input: M, D, K, B defining (1) and shift parameters $\{\mu_1, \ldots, \mu_{k_{\max}}\}$.
Output: Real low-rank solution factor $\hat{Z} \in \mathbb{R}^{2n \times mk_{\max}}$, such that $\hat{Z}\hat{Z}^T \approx X$ in (15).

for $k = 1, 2, \ldots, k_{\max}$ **do**
  if $k = 1$ **then**
    Solve $(\mu_k^2 M - \mu_k D + K)V_k^{(\mathrm{p})} = -B$ for $V_k^{(\mathrm{p})}$.
    Set $V_k^{(\mathrm{v})} = -\mu_k V_k^{(\mathrm{p})}$.
  else
    Solve $(\mu_k^2 M - \mu_k D + K)\hat{V}^{(\mathrm{p})} = (\mu_k M - D)V_{k-1}^{(\mathrm{p})} - MV_{k-1}^{(\mathrm{v})}$ for $\hat{V}^{(\mathrm{p})}$.
    Set $\hat{V}^{(\mathrm{v})} = V_{k-1}^{(\mathrm{p})} - \mu_k \hat{V}^{(\mathrm{p})}$.
    Construct iterate

$$V_k = \begin{bmatrix} V_k^{(\mathrm{p})} \\ V_k^{(\mathrm{v})} \end{bmatrix} = V_{k-1} - (\mu_k + \overline{\mu_{k-1}}) \begin{bmatrix} \hat{V}^{(\mathrm{p})} \\ \hat{V}^{(\mathrm{v})} \end{bmatrix}.$$

  end if
  if $\mathrm{Im}\,(\mu_k) = 0$ **then**
    $\hat{Z} = [\hat{Z},\ \sqrt{-2\mu_k}\,\mathrm{Re}\,(V_k)]$.
  else
    $\gamma = \sqrt{-2\,\mathrm{Re}\,(\mu_k)}$, $\delta = \frac{\mathrm{Re}\,(\mu_k)}{\mathrm{Im}\,(\mu_k)}$.
    $V_{k+1} = \overline{V_k} + 2\delta\,\mathrm{Im}\,(V_k)$.
    Update low-rank solution factor
    $\hat{Z} = \left[\hat{Z},\ \gamma(\mathrm{Re}\,(V_k) + \delta\,\mathrm{Im}\,(V_k)), \gamma\sqrt{(\delta^2+1)} \cdot \mathrm{Im}\,(V_k)\right]$.
    Set $k = k + 1$.
  end if
end for

4.3 Some Remarks on more General Damping Models

In this section we briefly point out some ideas regarding the treatment of the linear systems in Algorithm 2 for different constructions for the matrix D which we shall refer to in the following as *damping matrix*. A common choice to construct damping matrices is *proportional damping* given by the *Caughey series* [31]

$$D_{\mathrm{prop}} = \sum_{j=0}^{f} \xi_j M(M^{-1}K)^j = \xi_0 M + \xi_1 K + \xi_2 K M^{-1} K + \ldots, \tag{18}$$

where *Rayleigh damping*, i.e., $\xi_i = 0$, $i \geq 2$, is a common choice. For this easy case the damping matrix is sparse, the linear system can be handled without any particular care. Difficulties arise when some of the higher terms in (18) are kept. Forming the matrix $M^{-1}K$ is not feasible for large and sparse matrices. Except for some easy cases, e.g. when M is diagonal, this would deliver a large and dense matrix, such that the linear systems can not be solved rapidly by both direct or iterative solvers. If linear

systems with M can be solved efficiently, one could suggest to employ iterative solvers which only require one or a few matrix vector products with the coefficient matrix $\mu^2 M - \mu D + K$ in each iteration step, so that linear systems with M have to be solved in each step of the iterative solver. This, however, might easily become too expensive. Even more general variants of proportional damping of the form

$$D_{\text{prop}} = M f_1(M^{-1}K) + K f_2(M^{-1}K)$$

are considered in [32], where f_1, f_2 are analytic function in the neighborhood of $\Lambda(M^{-1}K)$. Another class of damping matrices leading to similar or even severe problems are given by *critical damping* [33]

$$D_{\text{crit}} = 2M^{\frac{1}{2}}\sqrt{M^{-\frac{1}{2}}KM^{-\frac{1}{2}}}M^{\frac{1}{2}}. \tag{19}$$

Of course, construction of this matrix is out of the question since it will also be large and dense. Moreover, in addition to the required linear systems for the above construction, computing the square root of a large sparse matrix might not be possible. One way out could be to consider these damping matrices as a matrix function $f(M)$ or $f(M^{-1}K)$, e.g., $f(x) = 1/x$ would formally correspond to M^{-1}. There are approaches in the literature which approximate a product $f(M)t$, $t \in \mathbb{R}^n$ for large and sparse matrices [34]. Since only matrix vector products are needed to run an iterative solver for linear systems it might be possible that this approach enables the efficient solution of (16). However, since an extensive study on how to treat these difficult damping versions is beyond the scope of this paper we restrict ourselves to cases where D can still be employed without the requirement to approximate matrix functions.

It is also possible to combine one or more of the above damping approaches with *external* or *small-rank damping* [35, 36]

$$D_{\text{ext}} = FF^T \geq 0, F \in \mathbb{R}^{n\times h},\ h \ll n, \tag{20}$$

for instance,

$$D = D_{\text{prop}} + D_{\text{ext}}. \tag{21}$$

If D_{prop} can still be handled by sparse techniques, then the linear systems (16) can be solved by utilizing the Sherman-Morrison-Woodbury formula [37]:

$$Q(-\mu)^{-1} = \hat{Q}(\mu)^{-1} + \mu\hat{Q}(-\mu)^{-1}F\left[I_g - \mu F^T\hat{Q}(-\mu)^{-1}F\right]F^T\hat{Q}^{-1}(-\mu), \tag{22}$$

where $Q(\mu) := \mu^2 M + \mu D + K$ and $\hat{Q}(\mu) := \mu^2 M + \mu D_{\text{prop}} + K$.

4.4 Termination Criteria

Now we discuss when to stop Algorithms 1 or 2. At first we investigate the Lyapunov residual and afterwards the singular values of $Z_\alpha^H M Z_\beta$, $\alpha, \beta \in \{\text{p}, \text{v}\}$. We do not consider alternative stopping criteria based on, e.g., the relative change in the low-rank factor Z [21, 13].

4.4.1 Computing the Lyapunov Residual

A common termination criterion for algorithms solving particular equations is to stop when the residual, i.e., its norm, is sufficiently small. For large-scale Lyapunov equations this is difficult, since the residual

$$R_k = AX_kE^T + EX_kA^T + GG^T$$

is a large and dense matrix, such that even constructing and storing it is infeasible and computing, e.g., the spectral norm is expensive. Since R_k is a symmetric matrix, $\|R_k\|_2$ coincides with the spectral radius of R_k. Taking the low-rank structure $X_k = Z_kZ_k^T$ provided by G-LR-ADI into account, one could use a power iteration or a Lanczos process (see, e.g., [37]) to retrieve this eigenvalue [7]. This would require only matrix vector products with A, E, G, Z and their transposes. Unless the power iteration or Lanczos process converge in very few steps, this can still lead to a high portion of computational effort in a run of G-LR-ADI. Especially the power iteration tends to converge increasingly slow when the residual gets smaller. The next theorem gives a novel result on the structure of R_k within G-LR-ADI and enables a much cheaper computation of $\|R_k\|_2$. Note that it also applies to SO-LR-ADI since it is mathematically equivalent to G-LR-ADI. However, to ease the representation further up, we consider LR-ADI and the standard Lyapunov equation

$$\hat{A}X + X\hat{A}^T + \hat{G}\hat{G}^T = 0, \quad \hat{A} := E^{-1}A, \quad \hat{G} := E^{-1}G \tag{23}$$

which is equivalent to G-LR-ADI and the generalized matrix equation (15). We also work on the complex version of the algorithm.

Theorem 4.1. *The residual at step k of LR-ADI is of rank at most m and given by*

$$\hat{R}_k = \hat{A}Z_kZ_k^H + Z_kZ_k^H\hat{A}^T + \hat{G}\hat{G}^T = \hat{W}_k\hat{W}_k^H, \quad \hat{W}_k := (\hat{A} - \overline{\mu_k}I_n)V_k \in \mathbb{C}^{n\times m}.$$

If $\mu_k \notin \Lambda(A)$ for all k then the rank is exactly m.

Proof. It can be easily shown that for all $\mu \notin \Lambda(\hat{A})$ (23) is equivalent to the Stein equation (discrete-time Lyapunov equation)

$$X = \mathcal{C}(\hat{A}, \mu)X\mathcal{C}(\hat{A}, \mu)^H + X_1, \tag{24}$$

with $X_1 = Z_1Z_1^H$ being the first approximate solution of (23) obtained with Algorithm 1 and the complex Cayley type transformation

$$\mathcal{C}(\hat{A}, \mu) := (\hat{A} - \overline{\mu}I_n)(\hat{A} + \mu I_n)^{-1}. \tag{25}$$

Using this and [38, Lemma 5.2-3], [39, Lemma 3.5.1-2] one can show that the residual with respect to Algorithm 1 and (23) can be written as

$$\hat{R}_k = \hat{A}X_k + X_k\hat{A}^T + \tilde{G}\tilde{G}^T = \hat{A}(X_k - X) + (X_k - X)\hat{A}^T = \hat{W}\hat{W}^H \tag{26}$$

with

$$\hat{W} := \prod_{j=1}^{k} \mathcal{C}(\hat{A}, \mu_j)\hat{G}. \tag{27}$$

There, we used that $X_0 = 0$ in our setting. This already shows that $\operatorname{rank}(R_k) = m$. Now note that the increment V_k in step 7 can be expressed as

$$\begin{aligned} V_k &= (\hat{A} - \overline{\mu_{k-1}} I_n)(\hat{A} + \mu_k I_n)^{-1} V_{k-1} \\ &= (\hat{A} - \overline{\mu_{k-1}} I_n)(\hat{A} + \mu_k I_n)^{-1}(\hat{A} - \overline{\mu_{k-2}} I_n)(\hat{A} + \mu_{k-1} I_n)^{-1} V_{k-2} \\ &= (\hat{A} + \mu_k I_n)^{-1}(\hat{A} - \overline{\mu_{k-1}} I_n)(\hat{A} + \mu_{k-1} I_n)^{-1}(\hat{A} - \overline{\mu_{k-2}} I_n) V_{k-2} \\ &= \ldots = (\hat{A} + \mu_k I_n)^{-1} \prod_{j=1}^{k-1} \mathcal{C}(\hat{A}, \mu_j)\hat{G}, \end{aligned} \tag{28}$$

where we exploited that $\hat{A} \pm \mu I_n$ and $(\hat{A} \pm \xi I_n)^{-1}$ commute for all μ and $\xi \notin \Lambda(\hat{A})$. Comparing (27) and (28) yields the desired result. Now if any shift μ_k is an eigenvalue of $\hat{A}$, then the inverse in (25) still exists, but the matrix $\hat{A} - \overline{\mu} I_n$ has rank deficiency and thus the rank of R_k may drop below m. □

This novel result enables the computation of the spectral or Frobenius norm of the Lyapunov residual via $\|\hat{R}_k\| = \|\hat{W}_k \hat{W}_k^H\| = \|\hat{W}_k^H \hat{W}_k\| = \|\hat{W}_k\|^2$ which is much cheaper than the other approaches mentioned above. The only requirement is an extra product of $\hat{A}$ with V_k and the computation of the norm of a thin rectangular matrix. The result is exact compared to the approximate results obtained with power iteration and Lanczos. Moreover, the computational effort to compute $\|\hat{R}\|_2$ via this new approach stays essentially constant whereas using power or Lanczos iteration gets increasingly expensive since the number of columns in Z is growing. For generalized Lyapunov equations one has

$$R_k = E\hat{R}_k E^T = E\hat{W}\hat{W}^H E^T = W_k W_k^H, \quad W_k := (A - \overline{\mu_k} E)V_k.$$

By taking the block structure of A, E of our particular application into account this becomes

$$\begin{aligned} W_{k,1} &:= \begin{bmatrix} -\overline{\mu_k} N V_k^{(\mathrm{p})} + N V_k^{(\mathrm{v})} \\ -K V_k^{(\mathrm{p})} - D V_k^{(\mathrm{v})} - \overline{\mu_k} M V_k^{(\mathrm{v})} \end{bmatrix}, \\ W_{k,2} &:= \begin{bmatrix} -K V_k^{(\mathrm{p})} - \overline{\mu_k} D V_k^{(\mathrm{p})} - \overline{\mu_k} M V_k^{(\mathrm{v})} \\ -\overline{\mu_k} F V_k^{(\mathrm{p})} + F V_k^{(\mathrm{v})} \end{bmatrix}, \end{aligned} \tag{29}$$

for SO-LR-ADI and the first ($W_{k,1}$) and second ($W_{k,2}$) companion linearizations (3a), (3b).

4.4.2 Monitoring the Hankel Singular Values

Although the theorem above reveals an elegant and cheap way to compute the Lyapunov residual norm, it is not always the proper measure for stopping the iteration. It has been observed frequently that this quantity is not necessarily related to the accuracy of the reduced order model one wishes to compute using balanced truncation. For instance, even if the Lyapunov residual is still large, the reduced order model constructed with this low-rank factor Z can nevertheless be accurate.

In [9] a more problem oriented stopping criterion for (G-)LR-ADI was proposed which we are going to adapt to second order balanced truncation here. In balanced truncation, the quantities of interest are the dominant Hankel singular values which are revealed by the SVD (8). If both Gramians are identical, the data needed to compute this SVD are given in each step of G-LR-ADI and one can monitor the leading, say $r \ll n$, Hankel singular values with respect to their relative change compared to the previous ADI iteration. This suggests to stop the algorithm, e.g., when

$$\|\sigma^{(k)} - \sigma^{(k-1)}\|_\infty \quad < \quad \theta_\sigma \sigma_1^{(k)}, \tag{30}$$

where $\sigma^{(j)} = [\sigma_1^{(j)}, \ldots, \sigma_r^{(j)}]$ is the vector containing the leading r singular values at step j of G-LR-ADI and $0 < \theta_\sigma \ll 1$ a small constant. Of course, the above relative change can only be computed once the low-rank factor Z has at least r linearly independent columns which determines a startup phase for G-LR-ADI. Note that for the case $P \neq Q$ this approach can still be employed when both generalized Lyapunov equations are solved simultaneously with the dual G-LR-ADI [9, Algorithm 1].

Of course, this approach is also applicable for second order balanced truncation for symmetric systems, where according to (11) one has to monitor the singular values of $S_k := Z_{k,\alpha}^H M Z_{k,\beta}$, where $\alpha, \beta \in \{\mathrm{p}, \mathrm{v}\}$ determine if the first or last n rows of Z are used, depending on the chosen second order balanced truncation type (see Section 3). The matrix S_k can be accumulated efficiently via

$$\begin{aligned} S_k = Z_\alpha^H M Z_\beta &= [Z_{k-1,\alpha}, \gamma V_k^{(\alpha)}]^H M [Z_{k-1,\beta}, \gamma V_k^{(\beta)}] \\ &= \begin{bmatrix} S_{k-1} & \gamma Z_{k-1,\alpha} M V_k^{(\beta)} \\ \gamma (V_k^{(\alpha)})^H M Z_{k-1,\beta} & \gamma^2 (V_k^{(\alpha)})^H M V_k^{(\beta)} \end{bmatrix}, \end{aligned}$$

where further savings are possible if $\alpha = \beta$.

In any case, computing the $km \times km$ SVD in every step can easily become more expensive than the computation of the residual norm. One way to reduce the cost of this step, but also of the whole iteration, is to employ column compression techniques on Z [13]. Alternatively, since S_k is obtained by adding new columns $\gamma Z_{k-1,\alpha} M V_k^{(\beta)}$ and new rows $[\gamma (V_k^{(\alpha)})^H M Z_{k-1,\beta}, \gamma^2 (V_k^{(\alpha)})^H M V_k^{(\beta)}]$ to S_{k-1}, updating strategies for the SVD [40, 41] can be employed to reduce to cost for computing the SVD of S_k. There, either new rows or column are added to a matrix with an already known SVD and not both as it is the situation here, such that the proposed updating strategies might have to be modified. For most of our Examples the encountered SVDs were quite small and only minor savings were expected from this ideas such that we did not

pursue them further. Once the SO-LR-ADI terminates using this stopping criterion, the SVD of S_k can be directly used for constructing $\tilde{T}_{l,\alpha\beta}$, $\tilde{T}_{r,\alpha\beta}$ in (12).

5 Numerical Examples

Now we shall evaluate the second order balanced truncation approach and the SO-LR-ADI method as discussed in the two prevailing sections numerically. All experiments were carried out in MATLAB® 7.11.0 on a compute server using 4 Intel®Xeon®@2.67 GHz CPUs with 8 cores per CPU and 1 TB RAM.

We compare the accuracy of the balanced truncation approach by using the frequency response plots of the transfer function matrices of the original and reduced systems

$$\begin{aligned} H(s) &= (C_{\mathrm{p}} + sC_{\mathrm{v}})(s^2M + sD + K)^{-1}B, \\ \tilde{H}(s) &= (\tilde{C}_{\mathrm{p}} + s\tilde{C}_{\mathrm{v}})(s^2\tilde{M} + s\tilde{D} + \tilde{K})^{-1}\tilde{B}, \\ \tilde{H}_f(s) &= \tilde{L}(s\tilde{E} - \tilde{A})^{-1}\tilde{G}, \end{aligned}$$

where in our situation depending on the kind of symmetric second order system, either C_{p} or C_{v} is used and $\tilde{H}_f(s)$ denotes the transfer function matrix of the reduced system of first order. The value of s is chosen from an interval $\jmath[\omega_{\min},\ \omega_{\max}]$ along the imaginary axis for frequencies $\omega_{\min} < \omega_{\max} \in \mathbb{R}$. The relative errors of the reduction approaches

$$e(s) := \|H(s) - \tilde{H}(s)\| / \|H(s)\| \quad \text{and} \quad e_f(s) := \|H(s) - \tilde{H}_f(s)\| / \|H(s)\|$$

are considered as well.

The following test systems are used:

Example 5.1. *The first model* constitutes a Bernoulli beam, where a finite element discretization leads to a symmetric second order system of the first kind with dimension* $n = 3000$ *and* $m = 1$.

Example 5.2. *The next test system is taken from [42] and models a cantilever Timoshenko beam. Choosing 500 elements leads to* $n = 3000$. *It belongs to the first kind as well with* $m = 1$ *and* $C_p = -B^T$.

Example 5.3. *The Butterfly Gyro is also taken from the Oberwolfach Model Reduction Benchmark Collection† and models a vibrating mechanical gyroscope for the use in an inertia sensor. The original matrices arise from a finite element discretization and are of dimension* $n = 17361$. *The damping matrix is* $D = \xi_0 M + \xi_1 K$, $\xi_0 = 10^{-5}$, $\xi_1 = 10^{-6}$. *The system as such is not a symmetric one but comes with an output matrix* $C_p \in \mathbb{R}^{12\times n}$ *and we just take* $B = C_p^T$ *as input matrix.*

*Provided by C. Nowakowski, Institute of Engineering and Computational Mechanics, University Stuttgart.

†Available at `http://portal.uni-freiburg.de/imteksimulation/downloads/benchmark`.

Example 5.4. *The scalable triple chain oscillator [35] describes three coupled chains of masses interlinked with springs and dampers. The mass and stiffness matrices M and K are symmetric and of dimension $n = 150001$, whereas the damping matrix D is modeled as proportional plus small rank damping*

$$D = \xi_0 M + \xi_1 K + \xi_2 K M^{-1} K + \xi_3 K (M^{-1} K)^2 + U U^T,$$

where $\xi_0 = 0.2$, $\xi_i = 0.1$ $(i = 2, 3, 4)$ and $U \in \mathbb{R}^{n \times 10}$ contains 10 randomly selected columns of I_n. Since M is only a diagonal matrix the above Caughey series of length 4 can be computed explicitly. The result is a multidiagonal matrix of bandwidth 7 with a low-rank update. We also set $C_v^T = B \in \mathbb{R}^{n \times 5}$ random to make this a system of the second kind.

We employ the SO-LR-ADI method for solving the Lyapunov equations and test the different stopping criteria discussed in the previous section. On the one hand when the normalized residual norm is small enough, i.e. $\|R\|/\|BB^T\| < \theta_{\text{res}}$, $0 < \theta_{\text{res}} < 1$ and on the other hand when (30) is satisfied. The residual norm is computed via the low-rank structure of R_k proposed above and for comparison also via the older approach using a Lanczos iteration. Note that the SVDs or residuals norms are only computed after a complex pair of shifts has been processed completely since then the low-rank factor can be augmented with real columns according to Algorithm 2. For the shift parameter computations we use the heuristic approach [7], where J shifts are generated from k_+ and k_- Ritz values of $E^{-1}A$ and $A^{-1}E$, respectively. The values n, m and for later use the considered frequency interval $[\omega_{\min},\omega_{\max}]$, all relevant parameters for SO-LR-ADI, as well as the results for all examples are summarized in Table 2. In the upper part of Table 2 the normalized residual norm ($\|R_k\|/\|R_0\|$) was used as stopping criterion. There, θ_{res} denotes the tolerance used for the normalized residual, k_{iter} are the number of required SO-LR-ADI iterations, t_{ADI} denotes the time spent in the SO-LR-ADI without the computations regarding stopping criteria. The timings t_{res} and t_{Lan} give the times needed to compute $\|R_k\|$ with the novel relation (29) and via a Lanczos iteration, respectively, where we also give the average times over the complete run in brackets. The lower part of this table is devoted to the experiments when the singular value based criterion (30) is used. The quantities t_{ADI} and k_{iter} are as above, t_{SVD} denotes the time needed for computing the SVDs. The numbers in brackets are again the averages over all iterations, where the SVD computation was performed. The values θ_σ, $r_{\min}$, $\alpha\beta$ refer to the threshold used in (30), the minimum number of columns in Z, and the corresponding second order balanced truncation type.

From the timings it is evident that SO-LR-ADI is capable of solving the large-scale, sparse and structured generalized Lyapunov equations to the desired accuracy in a reasonable amount of time. Note that solving the Lyapunov equation of, for instance, Example 5.1 of dimension 6000 directly with the Bartels-Stewart algorithm [16] (`lyap(A,B*B',[],E)` command in MATLAB®) or with Hammarlings method [15] (`lyapchol(A,B,E)` command in MATLAB®) was not possible.

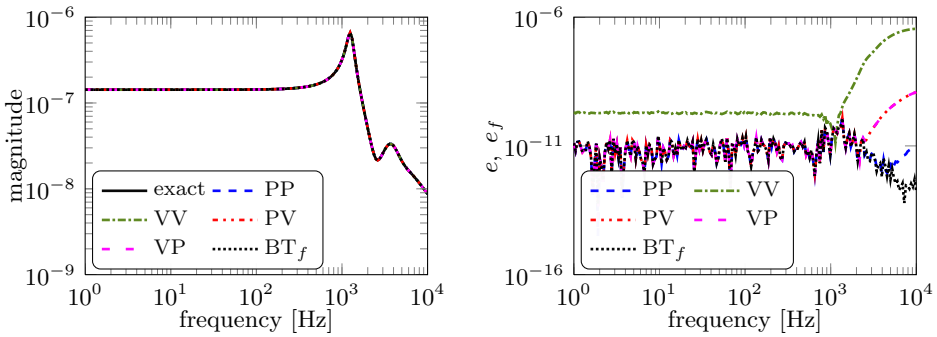
The timings t_{res}, t_{Lan} for the computations of $\|R\|$ clearly shows the superiority of our novel approach (29) over the older, traditional approach using a Lanczos itera-

Table 2: Specifications of the used examples as well as parameters and results of the SO-LR-ADI runs. All timings are given in seconds.

Example	5.1	5.2	5.3	5.4
$n,\ m$	3000, 1	3000, 1	17361, 12	150001, 5
$\omega_{\min}$,$\omega_{\max}$ [Hz]	1, 10^4	10^{-1}, 10^4	1, 10^5	10^{-3}, 10
$J,\ k_+,\ k_-$	30, 45, 45	60, 50, 80	40, 60, 50	60, 100, 80
θ_{res}	10^{-8}	10^{-8}	10^{-6}	10^{-6}
k_{iter}	57	354	102	429
t_{ADI}	0.46	2.96	149.54	1998.6
t_{res} (avg.)	0.064 (0.0015)	0.19 (0.001)	6.49 (0.12)	53.05 (0.126)
t_{Lan} (avg.)	0.45 (0.011)	2.93 (0.015)	94.39 (1.78)	6094.8 (1.5)
θ_σ	10^{-8}	10^{-8}	10^{-8}	10^{-8}
$r_{\min}$	20	35	50	400
$\alpha\beta$	PP	VV	PP	PV
k_{iter}	28	63	33	202
t_{ADI}	0.25	0.61	50.38	449.54
t_{SVD} (avg.)	0.013 (0.0016)	0.057 (0.0034)	2.65 (0.17)	864.37 (7.14)

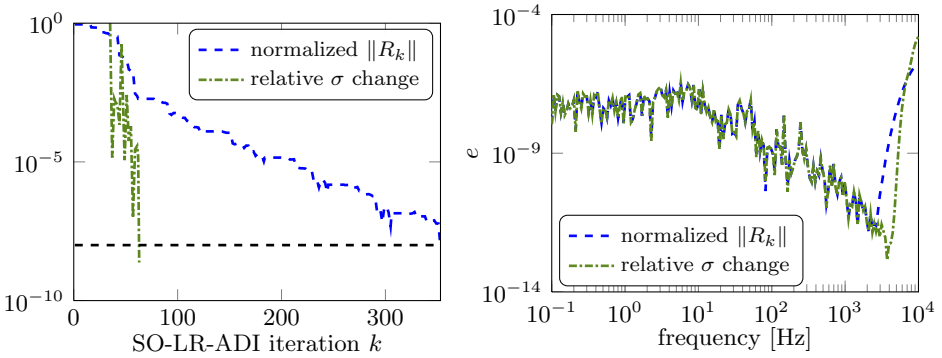
tion. With the latter one, a significantly large portion of CPU time t_{Lan} is spent on computing the Lyapunov residual norm. In some cases this is almost equal to the SO-LR-ADI execution time t_{ADI} (Examples 5.1 and 5.2) or even greater (Example 5.4). The average time when exploiting the symmetric low-rank structure of R is always much smaller than using Lanczos.

Using these approximate solutions of the Lyapunov equations within balanced truncation model order reduction still leads to very accurate reduced order models. Exemplary for Example 5.1, Figure 1 shows the Bode and relative error plots of the reduced systems of order $r = 20$ with respect to all four second order balanced truncations approaches, as well as, of the reduction to a first order system of twice the order $r_f = 40$. Apparently, there is no visible difference of exact and reduced models in the Bode plot. See [29] for a similar experiment. Next we investigate the singular value based termination criterion, i.e., the lower part of Table 2. Obviously, this criterion leads to a much lower number of required SO-LR-ADI steps compared to the residual based criterion for all examples. Clearly this also yields less time needed for SO-LR-ADI. This is especially outstanding for Example 5.2, where using the singular value based criterion required only 63 iteration which is much smaller than the 354 iterations when using the residual. The large number of iterations in the latter case can be explained by the fact that this system has a lot of eigenvalues very close to the imaginary axis which can deteriorate the convergence speed of ADI type methods. In Figure 2(a) the



(a) Bode plot (b) Relative error

Figure 1: Reduction results for the Bernoulli beam model resulting from the residual based stopping criterion.



(a) History (b) Relative errors

Figure 2: (a) History of normalized residual norm and relative change of singular change (30) of singular values for Example 5.2. The black dashed line refers to the tolerance of 10^{-8}. (b) Relative errors obtained from both approaches using VV second order balanced truncation.

residual norm and the relative singular value change are plotted against the iteration number k. Note that when the singular values stop the algorithm, the normalized residual is just of order 10^{-3}, such that the solution of the Lyapunov equation might be considered as very inexact. The question arises how this much earlier termination affects the accuracy of the model order reduction. The relative errors of the VV-reduced order models of order 35 obtained with both termination criteria are plotted in Figure 2(b). As before, the Bode plots of original and reduced order models were

Table 3: Largest relative errors of the reduced order models resulting from residual and singular value based termination of SO-LR-ADI.

	Example			
	5.1	5.2	5.3	5.4
residual based	$1.213 \cdot 10^{-10}$	$1.696 \cdot 10^{-6}$	$9.513 \cdot 10^{-7}$	$3.96 \cdot 10^{-4}$
$\sigma_{\alpha\beta}$ based	$1.216 \cdot 10^{-10}$	$1.641 \cdot 10^{-5}$	$1.033 \cdot 10^{-6}$	$2.2 \cdot 10^{-3}$

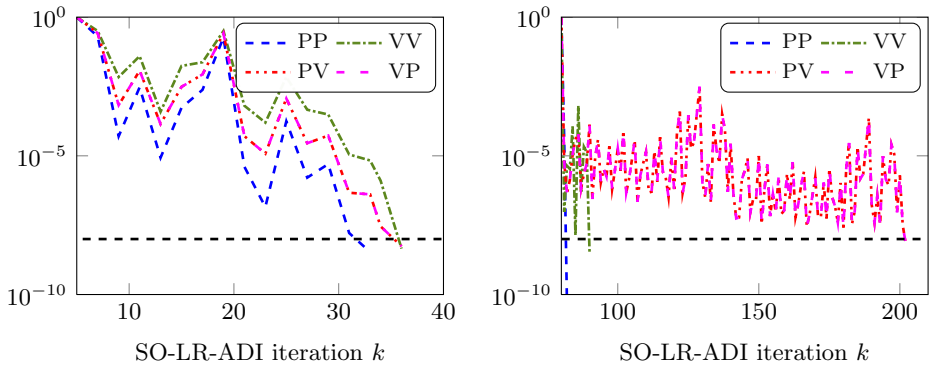
not distinguishable. Astonishingly, both approaches yield reduced order models of a comparable, satisfactory accuracy.

Table 3 summarizes the maximum relative errors from both termination approaches for the other examples using the second order balanced truncation type and a reduced order $r_{\min}$ as given in Table 2. For Example 5.2 the relative error of the model from the residual based stopping is only slightly better. However, the singular value based stopping approach gives this accuracy by requiring much less SO-LR-ADI iterations and hence, much less computational effort. This can also be observed for the other three test systems. It is therefore tempting to conclude that using (30) as stopping criterion is always more efficient than using the residual.

However, as it can already been conjectured from Figure 2(a), the relative change in the singular values shows an irregular behavior. Additionally, it depends on the values $r_{\min}$ and α, β. Which choice will lead to the best results is in general not known in advance. Figure 3 illustrates the irregular progress of (30) for all four possible choices of α, β for Examples 5.3 and 5.4. For Example 5.3 the history of (30) is similar for different choices of α, β corresponding to the designated second order balanced truncation variant. In contrast, all four choices yield a considerably different progress for Example 5.4. The PP, VV and PV settings lead to 82, 90 and 202 iterations of SO-LR-ADI, respectively. The irregular convergence is apparent for both examples. For all four test examples, however, using the singular value based termination lead always to a lower number of SO-LR-ADI iterations then the residual based termination. In similar experiments, the other way around occurred in a few cases when the threshold θ_σ was chosen too small. In Example 5.4 the computational costs for the SVD based criterion exceeded the costs of the main ADI iteration, so that using SVD updating strategies [40, 41] might be appropriate. We leave their actual implementation in this context for further work.

6 Conclusion and Outlook

We reviewed structure preserving balanced truncation for symmetric second order systems. As in [13, 27, 2, 28] the occurring large-scale generalized Lyapunov equations were solved approximately with the SO-LR-ADI iterations. A new result showing the low-rank structure of the Lyapunov residual in low-rank ADI methods was given. It enables an efficient computation of the spectral norm of the residual which is much



(a) Example 5.3 (b) Example 5.4

Figure 3: History of relative change (30) of different singular values for Example 5.3 (a) and 5.4 (b). The black dashed line refers to the tolerance of 10^{-8}.

cheaper than existing approaches. Motivated by [9], another stopping criterion for the SO-LR-ADI was proposed which inhibits a stronger emphasis on the sought reduced order model. It is based on monitoring the relevant singular values during the SO-LR-ADI iteration. Numerical experiments with selected test systems confirm the superiority of the novel approach for computing the spectral norm of the Lyapunov residual over existing approaches. Using the singular value based stopping criterion lead in most cases to a significantly reduced number of required ADI iterations and still maintained a comparable accuracy of the reduced second order system in the end. Although this approach is more dependent on several additional parameters in contrast to the Lyapunov residual, it seems to be more efficient in the majority of cases for carrying out second order balanced truncation. The novel result on the Lyapunov residual might nevertheless still be very useful in applications apart from balanced truncation model order reduction, for instance solving large-scale algebraic Riccati equation with low-rank Newton-ADI type algorithms [21, 13, 39, 38].

References

[1] P. Benner, P. Kürschner, and J. Saak, *Improved second-order balanced truncation for symmetric systems*, MathMod 2012 Proceedings (2012) accepted.

[2] C. Nowakowski, P. Kürschner, P. Eberhard, and P. Benner, *Model Reduction of an Elastic Crankshaft for Elastic Multibody Simulations*, Z. Angew. Math. Mech. (2012) accepted.

[3] B. Yan, S.X.D. Tan, and B. McGaughy, *Second-Order Balanced Truncation for Passive-Order Reduction of RLCK Circuits*, IEEE Trans. Circuits Syst. II 55 (2008), pp. 942–946.

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

[5] Y. Chahlaoui, D. Lemonnier, A. Vandendorpe, and P.V. Dooren, *Second-order balanced truncation*, Linear Algebra Appl. 415 (2006), pp. 373–384.

[6] T. Reis and T. Stykel, *Balanced truncation model reduction of second-order systems*, Math. Comput. Model. Dyn. Syst. 14 (2008), pp. 391–406.

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

[8] J.R. Li and J. White, *Low Rank Solution of Lyapunov Equations*, SIAM J. Matrix Anal. Appl. 24 (2002), pp. 260–280.

[9] P. Benner, P. Kürschner, and J. Saak, *A Goal-Oriented Dual LRCF-ADI for Balanced Truncation*, MathMod 2012 Proceedings (2012) accepted.

[10] F. Tisseur and K. Meerbergen, *The Quadratic Eigenvalue Problem*, SIAM Review 43 (2001), pp. 235–286.

[11] A.C. Antoulas *Approximation of Large-Scale Dynamical Systems*, SIAM Publications, Philadelphia, PA, 2005.

[12] P. Benner, *Solving Large-Scale Control Problems*, IEEE Control Systems Magazine 14 (2004), pp. 44–59.

[13] J. Saak, *Efficient Numerical Solution of Large Scale Algebraic Matrix Equations in PDE Control and Model Order Reduction*, PhD thesis, TU Chemnitz, 2009 Available from http://nbn-resolving.de/urn:nbn:de:bsz:ch1-200901642.

[14] D. Sorensen and Y. Zhou, *Direct methods for matrix Sylvester and Lyapunov equations*, J. Appl. Math 2003 (2002), pp. 277–303.

[15] S. Hammarling, *Numerical Solution of the Stable, Non-negative Definite Lyapunov Equation*, IMA J. Numer. Anal. 2 (1982), pp. 303–323.

[16] R. Bartels and G. Stewart, *Solution of the Matrix Equation $AX + XB = C$: Algorithm 432*, Communications of the ACM 15 (1972), pp. 820–826.

[17] D. Sorensen and Y. Zhou, *Bounds on Eigenvalue Decay Rates and Sensitivity of Solutions to Lyapunov Equations*, TR02-07, Dept. of Comp. Appl. Math., Rice University, Houston, TX, 2002 Available online from http://www.caam.rice.edu/caam/trs/tr02.html.

[18] L. Grasedyck, *Existence of a low rank or H-matrix approximant to the solution of a Sylvester equation*, Numer. Lin. Alg. Appl. 11 (2004), pp. 371–389.

[19] E. Wachspress, *Iterative solution of the Lyapunov matrix equation*, Appl. Math. Letters 107 (1988), pp. 87–90.

[20] ———, The ADI Model Problem; Available from the author (1995), .

[21] P. Benner, J.R. Li, and T. Penzl, *Numerical Solution of Large Lyapunov equations, Riccati Equations, and Linear-Quadratic Control Problems*, Numer. Lin. Alg. Appl. 15 (2008), pp. 755–777.

[22] I. Duff, A. Erisman, and J. Reid *Direct methods for sparse matrices*, Clarendon Press, Oxford, UK, 1989.

[23] T.A. Davis *Direct Methods for Sparse Linear Systems (Fundamentals of Algorithms 2)*, SIAM, Philadelphia, PA, USA, 2006.

[24] Y. Saad *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, PA, USA, 2003.

[25] H.A. Van der Vorst *Iterative Krylov Methods for Large Linear Systems*, Cambridge University Press, Cambridge, 2003.

[26] V. Simoncini, *A new iterative method for solving large-scale Lyapunov matrix equations*, SIAM J. Sci. Comput. 29 (2007), pp. 1268–1288.

[27] P. Benner and J. Saak, *Efficient Balancing based MOR for Large Scale Second Order Systems*, Math. Comput. Model. Dyn. Sys. 17 (2011), pp. 123–143 DOI:10.1080/13873954.2010.540822.

[28] P. Benner, P. Kürschner, and J. Saak, *Efficient handling of complex shift parameters in the low-rank Cholesky factor ADI method*, Numerical Algorithms (2012) DOI:10.1007/s11075-012-9569-7.

[29] ———, *Avoiding complex arithmetic in the low-rank ADI method efficiently*, PAMM 12 (2012) accepted.

[30] R.W. Freund, *Conjugate Gradient-Type Methods for Linear Systems with Complex Symmetric Coefficient Matrices*, SIAM Sci. Stat. Comp. 13 (1992), pp. 425–448.

[31] T.K. Caughey, *Classical Normal Modes in Damped Linear Dynamic Systems*, Journal of Applied Mechanics 27 (1960), p. 269.

[32] S. Adhikari, *Damping modelling using generalized proportional damping*, Journal of Sound and Vibration 293 (2006), pp. 156 – 170.

[33] M. Paz *Structural Dynamics: Theory and Computation*, Chapman & Hall, 1997.

[34] M. Hochbruck and M.E. Hochstenbach, *Subspace extraction for matrix functions*, Preprint, Department of Mathematics, Case Western Reserve University, Cleveland, Ohio, USA, 2005. Submitted.

[35] N. Truhar and K. Veselic, *An efficient method for estimating the optimal dampers' viscosity for linear vibrating systems using Lyapunov equation*, SIAM J. Matrix Anal. Appl. 31 (2009), pp. 18–39.

[36] P. Benner, Z. Tomljanović, and N. Truhar, *Optimal damping of selected eigenfrequencies using dimension reduction*, Numerical Linear Algebra with Applications (2011).

[37] G. Golub and C. Van Loan *Matrix Computations*, third Johns Hopkins University Press, Baltimore, 1996.

[38] F. Feitzinger, T. Hylla, and E.W. Sachs, *Inexact Kleinman-Newton Method for Riccati Equations*, SIAM J. Matrix Anal. Appl. 31 (2009), pp. 272–288.

[39] T. Hylla, *Extension of inexact Kleinman-Newton methods to a general monotonicity preserving convergence theory*, PhD thesis, Universität Trier, 2011.

[40] M. Gu and S.C. Eisenstat, *A Stable And Fast Algorithm For Updating The Singular Value Decomposition*, Technical report, Department of Computer Science, Yale University, 1994.

[41] M. Brand, *Fast online SVD revisions for lightweight recommender systems*, in *SIAM International Conference on Data Mining*, 2003.

[42] H. Panzer, J. Hubele, R. Eid, and B. Lohmann, *Generating a Parametric Finite Element Model of a 3D Cantilever Timoshenko Beam Using MATLAB®*, Technical reports on automatic control, Institute of Automatic Control, Technische Universität München, 2009.