Model Order Reduction by Approximate Balanced Truncation: A Unifying Framework

Modellreduktion durch approximatives Balanciertes Abschneiden: eine vereinigende Formulierung

Thomas Wolf*, Heiko K. F. Panzer, Boris Lohmann, Technische Universität München

* Correspondence author: thomas.wolf@tum.de

Summary A novel formulation of approximate truncated balanced realization (TBR) is introduced to unify three approaches: two iterative methods for solving the underlying Lyapunov equations – the alternating directions implicit (ADI) iteration and the rational Krylov subspace method (RKSM) – and a two-step procedure that performs a Krylov-based projection and subsequently direct TBR. The framework allows to compare the three methods with respect to solvability, fidelity, numerical effort, stability preservation and global error bounds, which suggests to merge ADI with the two-step procedure. ►►► **Zusammenfassung** Eine neue Formulierung des approximativen Balancierten Abschneidens vereinheitlicht drei Varianten: zwei iterative Lösungsverfahren der zugrundeliegenden Ljapunow-Gleichungen – Alternating Directions Implicit (ADI) Iteration und Rational Krylov Subspace Method (RKSM) – und ein Zwei-Schritt-Verfahren (ZSV), welches nacheinander eine Krylow-Projektion und Balanciertes Abschneiden durchführt. Die einheitliche Beschreibung ermöglicht den Vergleich hinsichtlich Lösbarkeit, Approximationsgüte, numerischem Aufwand, Stabilitätserhaltung und globaler Fehlerschranken, was eine Fusion von ADI und ZSV empfiehlt.

Keywords Model order reduction, truncated balanced realization, Krylov, ADI ►►►
Schlagwörter Modellordnungsreduktion, balanciertes Abschneiden, Krylov, ADI

1 Introduction

Model order reduction (MOR) aims to approximate large-scale dynamical systems by another model of reduced order [1]. For linear time invariant (LTI) systems of the form

$$\mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)\,, \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t)\,, \tag{1}$$

the dynamics are described by $\mathbf{E}, \mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times m}, \mathbf{C} \in \mathbb{R}^{p\times n}$; and $\mathbf{x}(t) \in \mathbb{R}^n$, $\mathbf{u}(t) \in \mathbb{R}^m$ and $\mathbf{y}(t) \in \mathbb{R}^p$ denote the states, inputs and outputs of the system, respectively. It is assumed that $\mathbf{E}$ is non-singular: $\det(\mathbf{E}) \neq 0$. With the usual abuse of notation, let $\mathbf{G}(s)$ denote the transfer function of system (1) in the Laplace domain as well as the dynamical system itself.

For model order reduction of (1), different approaches have been shown to be well-suited, such as the *Krylov subspace methods*, [1; 13], or the *Truncated Balanced Realization* (TBR), [1; 17]. TBR has been widely investigated and is known to yield a good approximation of $\mathbf{G}(s)$. Additionally, an a priori error bound permits to judiciously assign the reduced order. However, TBR suffers from high numerical effort compared to the Krylov subspace

methods. To be precise, the main costs are the solutions of two dual Lyapunov equations of the form

$$\mathbf{APE}^T + \mathbf{EPA}^T + \mathbf{BB}^T = \mathbf{0}, \tag{2}$$

$$\mathbf{A}^T\mathbf{QE} + \mathbf{E}^T\mathbf{QA} + \mathbf{C}^T\mathbf{C} = \mathbf{0}, \tag{3}$$

where $\mathbf{P}$ defines the *Controllability Gramian* and $\mathbf{E}^T\mathbf{QE}$ the *Observability Gramian* of system (1), [1]. Although good direct solvers are available for (2) and (3) – when n becomes large – their execution might be too time consuming or might even fail due to shortage of RAM.

A remedy is to employ algorithms that compute approximate solutions $\widehat{\mathbf{P}} \approx \mathbf{P}$ and $\widehat{\mathbf{Q}} \approx \mathbf{Q}$. Here, we consider two iterative methods, which is on the one hand the *alternating directions implicit* (ADI) iteration, that was adapted to large-scale systems in [16; 20]. Comprehensive analysis is available in [21; 23] and recent results can be found in [5–7; 22; 27]. On the other hand, the *rational Krylov subspace method* (RKSM) was introduced recently in [9] as a generalization of [25]. Further analysis of RKSM can be found in [4; 8; 29]. First attempts to link both methods in special cases are made in [5; 8; 11].

Another possibility to circumvent the direct solution of (2) and (3) is to use a two-step procedure [15], that first reduces the original system (1) to moderate order $k < n$ by the numerically efficient Krylov subspace methods. The resulting *intermediate* system, which is assumed to be a good approximation of (1), is subsequently further reduced by TBR to a reasonable final order. Although only heuristically motivated, the two-step procedure performs well, as reported e. g. in [15; 24].

This work introduces a novel framework for approximate TBR, that contains the three aforementioned methods: low-rank solution of the Lyapunov equations by ADI or RKSM and the two-step procedure. The unifying formulation is based on the construction of two virtual systems stemming from particular projections of (1). Both systems provide a degree of freedom to produce the outcome of the three methods, which allows their reasonable comparison.

Towards this aim, low-rank TBR is reviewed in Sect. 2 and ADI and RKSM in Sect. 3. The three methods are included into the new framework in Sect. 4. Their performance is compared in Sect. 5 and a technical example is given in Sect. 6.

2 Low-Rank Square-Root Method for Approximate TBR

Standard TBR computes a state transformation – called the balancing transformation – such that the transformed state variables x_i are equally controllable and observable. This is reflected in diagonal and equal Gramians: $\mathbf{P} = \mathbf{E}^T\mathbf{QE} = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$. The square roots of the eigenvalues of $\mathbf{PE}^T\mathbf{QE}$ are called the *Hankel singular values* (HSV) σ_i and are system invariants. Truncating those states that correspond to small HSVs results in a reduced system that retains only the best controllable/observable states. The balancing and the truncation can be performed in a single step, which leads to the *square-root* method, reviewed in the following.

2.1 Square-Root TBR

The basic course of action for square-root TBR is sketched in Algorithm 1; for details please refer to [1].

Algorithm 1 (Square-root method).

1. Solve the two dual Lyapunov equations (2), (3) for $\mathbf{P}, \mathbf{Q}$ and compute their Cholesky factors $\mathbf{P} =: \mathbf{RR}^T$, $\mathbf{Q} =: \mathbf{SS}^T$.
2. Compute the singular value decomposition (SVD)
$$\mathbf{S}^T\mathbf{ER} =: \mathbf{M}\boldsymbol{\Sigma}\mathbf{N}^T$$
3. Select the first q columns from $\mathbf{M}$ and $\mathbf{N}$, denoted by $\mathbf{M}_{(:,1:q)}$ and $\mathbf{N}_{(:,1:q)}$, respectively, and take the first q singular values $\boldsymbol{\Sigma}_{(1:q,1:q)} = \mathrm{diag}(\sigma_1, \ldots, \sigma_q)$ to define the projection matrices
$$\mathbf{V}_{Bal} := \mathbf{RN}_{(:,1:q)}\boldsymbol{\Sigma}^{-1/2}_{(1:q,1:q)} \quad \in \mathbb{R}^{n\times q},$$
$$\mathbf{W}_{Bal} := \mathbf{SM}_{(:,1:q)}\boldsymbol{\Sigma}^{-1/2}_{(1:q,1:q)} \quad \in \mathbb{R}^{n\times q}.$$
4. Compute the reduced system matrices of order q
$$\mathbf{A}_r := \mathbf{W}_{Bal}^T\mathbf{AV}_{Bal}, \qquad \mathbf{E}_r := \mathbf{W}_{Bal}^T\mathbf{EV}_{Bal},$$
$$\mathbf{B}_r := \mathbf{W}_{Bal}^T\mathbf{B}, \qquad \mathbf{C}_r := \mathbf{CV}_{Bal}.$$

The singular value decomposition (SVD) is assumed to arrange the singular values σ_i in descending order: $\sigma_1 \geq \ldots \geq \sigma_n$. The outcome of Algorithm 1 is a reduced system $\mathbf{G}_r(s)$ with the state-space representation

$$\mathbf{E}_r\dot{\mathbf{x}}_r(t) = \mathbf{A}_r\mathbf{x}_r(t) + \mathbf{B}_r\mathbf{u}(t), \quad \mathbf{y}_r(t) = \mathbf{C}_r\mathbf{x}_r(t), \tag{4}$$

in balanced form. By construction, $\mathbf{E}_r = \mathbf{I}_q$ is always identity, which follows from the properties of the SVD,

$$\mathbf{E}_r = \mathbf{W}_{Bal}^T\mathbf{EV}_{Bal} \tag{5}$$

$$= \boldsymbol{\Sigma}^{-1/2}_{(1:q,1:q)}\mathbf{M}^T_{(:,1:q)}\mathbf{S}^T\mathbf{ERN}_{(:,1:q)}\boldsymbol{\Sigma}^{-1/2}_{(1:q,1:q)} \tag{6}$$

$$= \boldsymbol{\Sigma}^{-1/2}_{(1:q,1:q)}\mathbf{M}^T_{(:,1:q)}\mathbf{M}\boldsymbol{\Sigma}\mathbf{N}^T\mathbf{N}_{(:,1:q)}\boldsymbol{\Sigma}^{-1/2}_{(1:q,1:q)} \tag{7}$$

$$= \boldsymbol{\Sigma}^{-1/2}_{(1:q,1:q)}\left[\mathbf{I}_q,\ \mathbf{0}_{q\times n-q}\right]\boldsymbol{\Sigma}\begin{bmatrix}\mathbf{I}_q\\ \mathbf{0}_{n-q\times q}\end{bmatrix}\boldsymbol{\Sigma}^{-1/2}_{(1:q,1:q)} \tag{8}$$

$$= \mathbf{I}_q, \tag{9}$$

where $\mathbf{0}_{i\times j}$ denotes the zero matrix of dimension $i \times j$. Algorithm 1 is stated for general state-space representations (1), however, the standard form is included by setting $\mathbf{E} = \mathbf{I}_n$.

An important feature of TBR is an a priori error bound in the $\mathcal{H}_\infty$ norm, based on the truncated HSVs σ_i, [1]:

$$\|\mathbf{G} - \mathbf{G}_r\|_\infty \leq 2\sum_{i=q+1}^{n}\sigma_i. \tag{10}$$

2.2 Low-Rank Square-Root TBR

The direct (also called dense) solvers [26] for the two dual Lyapunov equations in Algorithm 1 are restricted to small or medium scale systems. Therefore, when n becomes large, indirect methods are required for the approximate solution of (2) and (3). The approximations $\widehat{\mathbf{P}} \approx \mathbf{P}$ and $\widehat{\mathbf{Q}} \approx \mathbf{Q}$ are typically represented by low-rank factorizations: let $\operatorname{rank}(\widehat{\mathbf{P}}) = k_B$ and $\operatorname{rank}(\widehat{\mathbf{Q}}) = k_C$, then the low-rank Cholesky factorizations read as

$$\widehat{\mathbf{P}} =: \mathbf{Y}\mathbf{Y}^T, \quad \widehat{\mathbf{Q}} =: \mathbf{Z}\mathbf{Z}^T, \tag{11}$$

where $\mathbf{Y} \in \mathbb{R}^{n \times k_B}$ and $\mathbf{Z} \in \mathbb{R}^{n \times k_C}$. In order to perform TBR with the approximations (11), we assume $q \le k_B, k_C \le n$, and replace the Cholesky factors $\mathbf{R}$ and $\mathbf{S}$ in Algorithm 1 by the low-rank Cholesky factors $\mathbf{Y}$ and $\mathbf{Z}$, respectively. This yields the low-rank square-root method in Algorithm 2; for details please refer to [21].

Algorithm 2 (Low-rank square-root method).

1. Find low-rank Cholesky factors $\mathbf{Y}$ and $\mathbf{Z}$ such that

   $$\widehat{\mathbf{P}} =: \mathbf{Y}\mathbf{Y}^T \approx \mathbf{P}, \quad \widehat{\mathbf{Q}} =: \mathbf{Z}\mathbf{Z}^T \approx \mathbf{Q}.$$

2. Compute the singular value decomposition (SVD)

   $$\mathbf{Z}^T\mathbf{E}\mathbf{Y} =: \mathbf{M}\mathbf{\Sigma}\mathbf{N}^T$$

3. Select the first q columns from $\mathbf{M}$ and $\mathbf{N}$, denoted by $\mathbf{M}_{(:,1:q)}$ and $\mathbf{N}_{(:,1:q)}$, respectively, and take the first q singular values $\mathbf{\Sigma}_{(1:q,1:q)} = \operatorname{diag}(\sigma_1, \ldots, \sigma_q)$ to define the projection matrices

   $$\mathbf{V}_{Bal} := \mathbf{Y}\mathbf{N}_{(:,1:q)}\mathbf{\Sigma}^{-1/2}_{(1:q,1:q)} \quad \in \mathbb{R}^{n \times q},$$
   $$\mathbf{W}_{Bal} := \mathbf{Z}\mathbf{M}_{(:,1:q)}\mathbf{\Sigma}^{-1/2}_{(1:q,1:q)} \quad \in \mathbb{R}^{n \times q}.$$

4. Compute the reduced system matrices of order q

   $$\mathbf{A}_r := \mathbf{W}_{Bal}^T\mathbf{A}\mathbf{V}_{Bal}, \qquad \mathbf{E}_r := \mathbf{W}_{Bal}^T\mathbf{E}\mathbf{V}_{Bal},$$
   $$\mathbf{B}_r := \mathbf{W}_{Bal}^T\mathbf{B}, \qquad \mathbf{C}_r := \mathbf{C}\mathbf{V}_{Bal}.$$

Here again $\mathbf{E}_r = \mathbf{I}_q$ holds. Algorithm 2 can be applied for arbitrary low-rank Cholesky factors $\mathbf{Y}$, $\mathbf{Z}$, but we consider two methods for their computation – RKSM and ADI – reviewed in the following.

3 Approximate Solution of Lyapunov Equations

The low-rank factors $\mathbf{Y}$, $\mathbf{Z}$ of both RKSM and ADI span rational Krylov subspaces, reviewed in the following.

3.1 Krylov Subspaces

A block Krylov subspace is generally defined as:

$$\mathcal{K}^j(\mathbf{A}, \mathbf{B}) = \operatorname{Range}\left\{\mathbf{B}, \mathbf{A}\mathbf{B}, \ldots, \mathbf{A}^{j-1}\mathbf{B}\right\}. \tag{12}$$

Towards rational input Krylov subspaces, a complex valued expansion point $s_i \in \mathbb{C}$ and a desired multiplicity $m_i \in \mathbb{N}^+$ has to be selected:

$$\mathcal{K}^{m_i}_{s_i} := \mathcal{K}^{m_i}\left((\mathbf{A}-s_i\mathbf{E})^{-1}\mathbf{E}, (\mathbf{A}-s_i\mathbf{E})^{-1}\mathbf{B}\right). \tag{13}$$

For the ease of presentation it is assumed that all directions defining a block Krylov subspace are linearly independent. This implies that the block Krylov subspace (13) has full column rank $m_i \cdot m$. If the assumption does not hold, deflated block Krylov subspaces should be employed [12]. However, they can be incorporated into the presented framework in a straightforward way, which we omit here for the ease of presentation.

Let $\mathbf{V}$ denote the basis of the union of arbitrary input block Krylov subspaces:

$$\operatorname{Range}(\mathbf{V}) \supseteq \mathcal{K}^{m_i}_{s_i}, \quad i = 1, \ldots, n_{\mathcal{K}}. \tag{14}$$

Due to numerical reasons, the basis $\mathbf{V}$ of a rational Krylov subspace is often computed as an orthogonal matrix, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, or such that $\mathbf{V}^T\mathbf{E}\mathbf{V} = \mathbf{I}$. Both ways are typically accomplished by the Arnoldi algorithm [1].

From the concept of duality, the rational output Krylov subspaces readily follow by replacing $\mathbf{A}, \mathbf{E}$ and $\mathbf{B}$ with $\mathbf{A}^T, \mathbf{E}^T$ and $\mathbf{C}^T$, respectively, in (13). A basis of the union of arbitrary output Krylov subspaces will be denoted as $\mathbf{W}$ in the following.

For the remainder of this paper, it is assumed, that a set of expansion points, denoted by the vector $\mathbf{s} := [s_1, s_2, \ldots, s_{n_{\mathcal{K}}}]$, is given. Furthermore, let the vector $\mathbf{m} := [m_1, m_2, \ldots, m_{n_{\mathcal{K}}}]$ denote the associated multiplicities. Using $\mathbf{s}$ and $\mathbf{m}$ in the Arnoldi algorithm then defines the rational input and output Krylov subspaces $\mathbf{V} \in \mathbb{R}^{n \times k_B}$ and $\mathbf{W} \in \mathbb{R}^{n \times k_C}$, respectively. (The dimensions k_B and k_C depend on $\mathbf{s}, \mathbf{m}, m$ and p, respectively). Of course, it is possible to use individual sets $\mathbf{s}$ and $\mathbf{m}$ for the input and output side. However, this case is excluded for the ease of presentation. Several works are available on the choice of expansion points $\mathbf{s}$, see e. g. [9; 10; 14; 20] to mention just a few.

3.2 Projective Model Order Reduction

Rational Krylov subspaces are commonly used for the reduction of dynamical systems. Assume $k_B = k_C =: k$, then the reduced dynamical system $\mathbf{G}_k(s)$ of order k has the state-space representation

$$\mathbf{E}_k\dot{\mathbf{x}}_k(t) = \mathbf{A}_k\mathbf{x}_k(t) + \mathbf{B}_k\mathbf{u}(t), \quad \mathbf{y}_k(t) = \mathbf{C}_k\mathbf{x}_k(t), \tag{15}$$

with $\mathbf{A}_k := \mathbf{W}^T\mathbf{A}\mathbf{V}$, $\mathbf{E}_k := \mathbf{W}^T\mathbf{E}\mathbf{V}$, $\mathbf{B}_k := \mathbf{W}^T\mathbf{B}$ and $\mathbf{C}_k := \mathbf{C}\mathbf{V}$, and stems from a projection of (1) such that the so-called *Petrov–Galerkin condition* is fulfilled [13]. The classical result for the projection by rational Krylov subspaces is that the reduced model (15) matches $2m_i$ (block) moments around the respective expansion points s_i, if s_i is neither an eigenvalue of $\mathbf{E}^{-1}\mathbf{A}$ nor an eigenvalue of $\mathbf{E}_k^{-1}\mathbf{A}_k$, [1; 12; 13]. Moments are defined as the coefficients of the Taylor series expansion of the transfer function around s_i. As the rational transfer function $\mathbf{G}_k(s)$ interpolates $\mathbf{G}(s)$ at the given frequencies s_i, this method for model reduction is referred to as *rational interpolation* or *rational Krylov*.

- 10.1524/auto.2013.1007

3.3 Two-Step Procedure for Approximate TBR

The two-step procedure for approximate TBR starts with rational interpolation: the projected system $\mathbf{G}_k(s)$ from (15) can be interpreted as an intermediate approximation of moderate order, that allows to perform square-root TBR (Algorithm 1) with direct solvers for the two dual Lyapunov equations. The outcome is a final reduced system of order q that can be considered as the reduction of (1) by approximate TBR, if the intermediate system represents a sufficiently good approximation. The prerequisite is equal dimensions of $\mathbf{V}$ and $\mathbf{W}$.

So far, the two-step procedure is a heuristic approach for model reduction: use the numerically efficient Krylov subspaces to reduce the original system to moderate but still high order and hope for a good approximation; subsequently, perform TBR of the intermediate system with direct methods to find a reduced model of reasonable order. However, the two-step procedure is justified by comparing it to RKSM and ADI in Sects. 5 and 6.

3.4 Rational Krylov Subspace Method

RKSM uses projections by rational Krylov subspaces to iteratively compute the low-rank square-roots $\mathbf{Y}$ and $\mathbf{Z}$ of the solutions of the two Lyapunov equations (2) and (3), respectively. We assume the sets $\mathbf{s}$ and $\mathbf{m}$ to be given, which uniquely define the Krylov subspaces $\mathbf{V}$ and $\mathbf{W}$. Then, $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{Q}}$ are generally defined by

$$\widehat{\mathbf{P}}_{RK} = \mathbf{V}\mathbf{P}_{RK}\mathbf{V}^T\,, \quad \widehat{\mathbf{Q}}_{RK} = \mathbf{W}\mathbf{Q}_{RK}\mathbf{W}^T\,, \tag{16}$$

where $\mathbf{P}_{RK} \in \mathbb{R}^{k_B \times k_B}$ and $\mathbf{Q}_{RK} \in \mathbb{R}^{k_C \times k_C}$ represent reduced Lyapunov solutions. RKSM determines these reduced solutions as follows: plugging $\widehat{\mathbf{P}}_{RK}$ and $\widehat{\mathbf{Q}}_{RK}$ into the original Lyapunov equations (2) and (3) leads to residuals $\mathbf{R}_P$ and $\mathbf{R}_Q$, respectively:

$$\mathbf{A}\widehat{\mathbf{P}}_{RK}\mathbf{E}^T + \mathbf{E}\widehat{\mathbf{P}}_{RK}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T =: \mathbf{R}_P\,, \tag{17}$$

$$\mathbf{A}^T\widehat{\mathbf{Q}}_{RK}\mathbf{E} + \mathbf{E}^T\widehat{\mathbf{Q}}_{RK}\mathbf{A} + \mathbf{C}^T\mathbf{C} =: \mathbf{R}_Q\,. \tag{18}$$

If equation (17) is multiplied by $\mathbf{V}^T$ and $\mathbf{V}$ from the left and right, respectively, and if equation (18) is multiplied by $\mathbf{W}^T$ and $\mathbf{W}$ from the left and right, respectively, two reduced Lyapunov equations result

$$\mathbf{A}_\mathbf{V}\mathbf{P}_{RK}\mathbf{E}_\mathbf{V}^T + \mathbf{E}_\mathbf{V}\mathbf{P}_{RK}\mathbf{A}_\mathbf{V}^T + \mathbf{B}_\mathbf{V}\mathbf{B}_\mathbf{V}^T = \mathbf{0}\,, \tag{19}$$

$$\mathbf{A}_\mathbf{W}^T\mathbf{Q}_{RK}\mathbf{E}_\mathbf{W} + \mathbf{E}_\mathbf{W}^T\mathbf{Q}_{RK}\mathbf{A}_\mathbf{W} + \mathbf{C}_\mathbf{W}^T\mathbf{C}_\mathbf{W} = \mathbf{0}\,, \tag{20}$$

where the following notation is employed

$$\mathbf{A}_\mathbf{V} = \mathbf{V}^T\mathbf{A}\mathbf{V}\,, \quad \mathbf{E}_\mathbf{V} = \mathbf{V}^T\mathbf{E}\mathbf{V}\,, \quad \mathbf{B}_\mathbf{V} = \mathbf{V}^T\mathbf{B}\,, \tag{21}$$

$$\mathbf{A}_\mathbf{W} = \mathbf{W}^T\mathbf{A}\mathbf{W}\,, \quad \mathbf{E}_\mathbf{W} = \mathbf{W}^T\mathbf{E}\mathbf{W}\,, \quad \mathbf{C}_\mathbf{W} = \mathbf{C}\mathbf{W}\,. \tag{22}$$

Solving (19), (20) for $\mathbf{P}_{RK}$, $\mathbf{Q}_{RK}$ by direct methods finally defines the approximations $\widehat{\mathbf{P}}_{RK}$ and $\widehat{\mathbf{Q}}_{RK}$ by (16). This approach implies that the residuals vanish in the subspaces $\mathbf{V}$ and $\mathbf{W}$: $\mathbf{V}^T\mathbf{R}_P\mathbf{V} = \mathbf{0}$ and $\mathbf{W}^T\mathbf{R}_Q\mathbf{W} = \mathbf{0}$, respectively. Therefore, the reduced Lyapunov equations (19), (20) result from orthogonal projections of the original Lyapunov equations; for details, please refer to [9].

The approximations (16) can be used for the low-rank square-root method by computing the Cholesky factorizations of the reduced Lyapunov solutions:

$$\mathbf{P}_{RK} = \mathbf{R}_{RK}\mathbf{R}_{RK}^T\,, \quad \mathbf{Q}_{RK} = \mathbf{S}_{RK}\mathbf{S}_{RK}^T\,. \tag{23}$$

The low-rank square-roots then are $\widehat{\mathbf{P}}_{RK} = \mathbf{Y}_{RK}\mathbf{Y}_{RK}^T$ and $\widehat{\mathbf{Q}}_{RK} = \mathbf{Z}_{RK}\mathbf{Z}_{RK}^T$, with

$$\mathbf{Y}_{RK} = \mathbf{V}\mathbf{R}_{RK}\,, \quad \mathbf{Z}_{RK} = \mathbf{W}\mathbf{S}_{RK}\,. \tag{24}$$

Approximate TBR with RKSM follows by plugging the low-rank Cholesky factors (24) into Algorithm 2.

3.5 Alternating Directions Implicit (ADI)

For the application of ADI, we assume again that the set of complex shifts $\mathbf{s}$ (here in the right half of the complex plane) with associated multiplicities $\mathbf{m}$ is given. Let $\widehat{\mathbf{P}}_0$ be an initial selection (e. g. $\widehat{\mathbf{P}}_0 = \mathbf{0}$) of the Lyapunov solution in (2), then basic ADI is defined by the following iteration (for the case $\mathbf{E} = \mathbf{I}_n$):

$$\begin{aligned}(\mathbf{A} - s_i\mathbf{I})\,\widehat{\mathbf{P}}_{i-\frac{1}{2}} &= -\mathbf{B}\mathbf{B}^T - \widehat{\mathbf{P}}_{i-1}\,(\mathbf{A}^T - s_i\mathbf{I})\,,\\ (\mathbf{A} - s_i\mathbf{I})\,\widehat{\mathbf{P}}_i^T &= -\mathbf{B}\mathbf{B}^T - \widehat{\mathbf{P}}_{i-\frac{1}{2}}\,(\mathbf{A}^T - s_i\mathbf{I})\,.\end{aligned} \tag{25}$$

Li et. al. observed in [16], that for the choice $\widehat{\mathbf{P}}_0 = \mathbf{0}$ the q-th iterate of (25) can be reformulated as a low-rank factorization $\widehat{\mathbf{P}}_{ADI} = \mathbf{Y}_{ADI}\mathbf{Y}_{ADI}^*$ with $\mathbf{Y}_{ADI} = [\mathbf{y}_1, \ldots, \mathbf{y}_{k_B}]$ where (now for arbitrary $\mathbf{E}$)

$$\begin{aligned}\mathbf{y}_1 &= \sqrt{2\,\mathrm{Re}(s_1)}\,(\mathbf{A} - s_1\mathbf{E})^{-1}\,\mathbf{B}\,,\\ \mathbf{y}_{i+1} &= \sqrt{\frac{\mathrm{Re}(s_{i+1})}{\mathrm{Re}(s_i)}}\left[\mathbf{I} + (s_{i+1} + \bar{s}_i)\cdot(\mathbf{A} - s_{i+1}\mathbf{E})^{-1}\,\mathbf{E}\right]\mathbf{y}_i\,,\\ &\qquad i = 2, \ldots, k_B\,.\end{aligned} \tag{26}$$

The iteration (26) generally delivers a complex valued $\mathbf{Y}_{ADI} \in \mathbb{C}^{n \times k_B}$; however, if the set $\mathbf{s}$ is closed under conjugation, the iteration (26) can be easily adjusted in order to deliver a real basis $\mathbf{Y}_{ADI} \in \mathbb{R}^{n \times k_B}$, [7].

Equation (26) specifies an iteration that is easy to implement. Together with a dual version for the output side, the iteration delivers the low-rank Cholesky factors $\mathbf{Y}_{ADI}$ and $\mathbf{Z}_{ADI}$, that can be plugged into Algorithm 2.

It was observed in [16] that the ADI basis $\mathbf{Y}_{ADI}$ spans a rational Krylov subspace; but only recently it was shown in [27], that $\widehat{\mathbf{P}}_{ADI} = \mathbf{Y}_{ADI}\mathbf{Y}_{ADI}^T$ can alternatively be generated via particular projections by rational Krylov subspaces. This procedure is crucial for incorporating ADI into the framework for approximate TBR in the next section, which is why it is reviewed in the following.

Let $\mathbf{V}$ be a basis of the input Krylov subspace with the expansion points $\mathbf{s}$ and multiplicities $\mathbf{m}$. The ADI solution (26) can be constructed by a projection onto the so-called $\mathcal{H}_2$ pseudo-optimal reduced system, [27]. This is achieved by a distinct projection matrix $\mathbf{W}_{ADI} \in \mathbb{R}^{n \times k_B}$ that leads to the reduced matrices of order k_B:

$$\begin{aligned}&\mathbf{A}_{ADI} := \mathbf{W}_{ADI}^T\mathbf{A}\mathbf{V}\,, \quad \mathbf{E}_{ADI} := \mathbf{W}_{ADI}^T\mathbf{E}\mathbf{V}\,,\\ &\mathbf{B}_{ADI} := \mathbf{W}_{ADI}^T\mathbf{B}\,.\end{aligned} \tag{27}$$

The $\mathcal{H}_2$ pseudo-optimum is characterized by the fact that the expansion points $\mathbf{s}$ are the mirror images of the eigenvalues of the matrix pair $\mathbf{E}_{ADI}, \mathbf{A}_{ADI}$, with respect to the imaginary axis. Therefore, a projection matrix $\mathbf{W}_{ADI}$ is sought-after, that places the reduced eigenvalues at $-\mathbf{s}$. Several approaches are possible to find such a $\mathbf{W}_{ADI}$; here we will follow the pole-placement approach due to Antoulas [2], by defining $\widehat{\mathbf{C}} \in \mathbb{R}^{m \times n}$ that satisfies

$$\widehat{\mathbf{C}}\mathbf{V} = \mathbf{0}\,. \tag{28}$$

One possibility is to choose

$$\widehat{\mathbf{C}} = \mathbf{B}^T \left(\mathbf{I} - \mathbf{V}\left(\mathbf{V}^T\mathbf{V}\right)^{-1}\mathbf{V}^T\right). \tag{29}$$

Then, for all s_i in $\mathbf{s}$, the following directions have to be contained in the span of $\mathbf{W}_{ADI}$

$$\operatorname{span}\left(\mathbf{W}_{ADI}\right) \supset \left[(\mathbf{A}^T + s_i\mathbf{E}^T)^{-1}\mathbf{E}^T\right]^{j-1} \cdot (\mathbf{A}^T + s_i\mathbf{E}^T)^{-1}\widehat{\mathbf{C}}^T\,, \quad j = 1, \ldots, m_i \tag{30}$$

This means that $\mathbf{W}_{ADI}$ has to span the output Krylov subspace with the virtual output vector $\widehat{\mathbf{C}}$ for the mirrored set of expansion points $-\mathbf{s}$ with multiplicities $\mathbf{m}$. The reduced system (27) then has its eigenvalues at $-\mathbf{s}$ with multiplicities $\mathbf{m}$. Please note, that alternative ways to compute (27) are the parameterization from [3] or the PORK algorithm from [28].

For constructing the ADI basis $\mathbf{Y}_{ADI}$ by Krylov projections, the Controllability Gramian of system (27) has to be computed by solving the Lyapunov equation

$$\mathbf{A}_{ADI}\mathbf{P}_{ADI}\mathbf{E}_{ADI}^T + \mathbf{E}_{ADI}\mathbf{P}_{ADI}\mathbf{A}_{ADI}^T + \mathbf{B}_{ADI}\mathbf{B}_{ADI}^T = \mathbf{0}\,, \tag{31}$$

which yields the low-rank Gramian of the ADI method

$$\widehat{\mathbf{P}}_{ADI} = \mathbf{V}\mathbf{P}_{ADI}\mathbf{V}^T\,. \tag{32}$$

With the Cholesky factorization $\mathbf{P}_{ADI} =: \mathbf{R}_{ADI}\mathbf{R}_{ADI}^T$, the low-rank square-root can be identified

$$\mathbf{Y}_{ADI} = \mathbf{V}\mathbf{R}_{ADI}\,. \tag{33}$$

Remark 1. The calculation of the ADI basis (26) via (29)–(33) is an auxiliary way that helps to incorporate the ADI iteration into the framework in the next section. In practice, one would rather use the numerically efficient iteration (26).

Please note, that the low-rank Cholesky factor $\mathbf{Y}_{ADI}$ is not unique. Think e. g. of a column permutation in $\mathbf{Y}_{ADI}$, which, however, does not affect the approximation $\widehat{\mathbf{P}}_{ADI} = \mathbf{Y}_{ADI}\mathbf{Y}_{ADI}^T$. In this respect, $\mathbf{Y}_{ADI}$ as the output of the iteration (26) might be different from the computation by (33), although the approximate solution $\widehat{\mathbf{P}}_{ADI}$ is identical.

The low-rank square-root $\mathbf{Z}_{ADI}$ of the output side can be computed in a dual way: the Lyapunov equation (3) is projected onto the subspace of the dual matrix $\mathbf{V}_{ADI}$, orthogonally to the rational output Krylov subspace $\mathbf{W}$. This can guarantee pole-placement at the mirror images $-\mathbf{s}$ in a similar way as described above.

3.6 Procedure in Iterative Methods

We assumed the sets $\mathbf{s}$ and $\mathbf{m}$ to be given. In practical applications, however, $\mathbf{m}$ is often determined iteratively. A typical approach is to first accumulate a set $\mathbf{s}$ of distinct shifts by some heuristics or algorithms, see e. g. [9; 10; 14; 20; 23]. These shifts are then repeatedly used in the construction of the low-rank square-roots, leading to higher multiplicities. This iteration automatically determines the ranks k_B and k_C of $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{Q}}$, respectively, and can be characterized as follows:

- Compute $k_i \in \mathbb{N}^+$ new directions of the Krylov subspaces $\mathbf{V}$, $\mathbf{W}$ or of the ADI bases $\mathbf{Y}_{ADI}$, $\mathbf{Z}_{ADI}$.
- Evaluate an appropriate stopping criterion: this can be the norm of the residual in the Lyapunov equations (see e. g. (17) and (18)) or the relative change of the q leading singular values (in step 2 of Algorithm 2).
- If the desired accuracy is achieved, stop the algorithm; if not, restart by calculating k_i new directions.

3.7 Problem Setting

The freedom in the three methods for approximate TBR – ADI, RKSM and the two-step procedure – is the choice of appropriate shifts $\mathbf{s}$ with associated multiplicities $\mathbf{m}$ (possibly determined by iterative methods). Together, $\mathbf{s}$ and $\mathbf{m}$ not only uniquely define the output of the three methods, but also the column span of the bases $\mathbf{V}$ and $\mathbf{W}$ of the input and output Krylov subspaces, respectively. By comparing (33) with (24) it follows that the low-rank square-roots of ADI and RKSM span the same subspace. As will be shown in Theorem 1, the two-step procedure can be interpreted as a way to compute low-rank square-roots $\mathbf{Y}_{2step}$ and $\mathbf{Z}_{2step}$, that are plugged into Algorithm 2. Since $\mathbf{Y}_{2step}$ and $\mathbf{Z}_{2step}$ also span the rational Krylov subspaces, the low-rank Cholesky factors of all methods span the same subspace:

$$\operatorname{span}\left(\mathbf{Y}_{2step}\right) = \operatorname{span}\left(\mathbf{Y}_{RK}\right) = \operatorname{span}\left(\mathbf{Y}_{ADI}\right), \tag{34}$$

$$\operatorname{span}\left(\mathbf{Z}_{2step}\right) = \operatorname{span}\left(\mathbf{Z}_{RK}\right) = \operatorname{span}\left(\mathbf{Z}_{ADI}\right). \tag{35}$$

Since the only difference is the basis that each method creates, the problem that we tackle here can be summarized in the following question:

What is the difference between the bases of the low-rank Cholesky factors, and how does it affect the reduction by approximate TBR via Algorithm 2?

4 The Unifying Framework

The low-rank square-roots of RKSM are related to particular (possibly virtual) intermediate systems. Towards the unifying framework we define two generalized intermediate systems (one for the input and one for the output side), that contain all three methods.

4.1 Intermediate Systems

Let the sets $\mathbf{s}$ and $\mathbf{m}$ be fixed and compute the associated input and output Krylov subspaces $\mathbf{V} \in \mathbb{R}^{n \times k_B}$ and $\mathbf{W} \in \mathbb{R}^{n \times k_C}$, respectively. Further define two (possibly virtual)

projection matrices, $\widetilde{\mathbf{V}} \in \mathbb{R}^{n \times k_C}$ and $\widetilde{\mathbf{W}} \in \mathbb{R}^{n \times k_B}$, that shall be considered as the degrees of freedom in the framework. The choice of $\widetilde{\mathbf{V}}$ and $\widetilde{\mathbf{W}}$ then uniquely defines the transfer behavior of the TBR approximation by defining the two intermediate systems

$$\widetilde{\mathbf{A}}_B = \widetilde{\mathbf{W}}^T \mathbf{A} \mathbf{V}, \qquad \widetilde{\mathbf{E}}_B = \widetilde{\mathbf{W}}^T \mathbf{E} \mathbf{V}, \qquad \widetilde{\mathbf{B}} = \widetilde{\mathbf{W}}^T \mathbf{B}, \tag{36}$$

$$\widetilde{\mathbf{A}}_C = \mathbf{W}^T \mathbf{A} \widetilde{\mathbf{V}}, \qquad \widetilde{\mathbf{E}}_C = \mathbf{W}^T \mathbf{E} \widetilde{\mathbf{V}}, \qquad \widetilde{\mathbf{C}} = \mathbf{C} \widetilde{\mathbf{V}}, \tag{37}$$

where $\widetilde{\mathbf{A}}_B, \widetilde{\mathbf{E}}_B \in \mathbb{R}^{k_B \times k_B}$, $\widetilde{\mathbf{B}} \in \mathbb{R}^{k_B \times m}$ describe the intermediate input system, and $\widetilde{\mathbf{A}}_C, \widetilde{\mathbf{E}}_C \in \mathbb{R}^{k_C \times k_C}$, $\widetilde{\mathbf{C}} \in \mathbb{R}^{p \times k_C}$ the intermediate output system. If we assume asymptotic stability in both systems, the solutions $\widetilde{\mathbf{P}} \in \mathbb{R}^{k_B \times k_B}$ and $\widetilde{\mathbf{Q}} \in \mathbb{R}^{k_C \times k_C}$ of the Lyapunov equations

$$\widetilde{\mathbf{A}}_B \widetilde{\mathbf{P}} \widetilde{\mathbf{E}}_B^T + \widetilde{\mathbf{E}}_B \widetilde{\mathbf{P}} \widetilde{\mathbf{A}}_B^T + \widetilde{\mathbf{B}} \widetilde{\mathbf{B}}^T = \mathbf{0}, \tag{38}$$

$$\widetilde{\mathbf{A}}_C^T \widetilde{\mathbf{Q}} \widetilde{\mathbf{E}}_C + \widetilde{\mathbf{E}}_C^T \widetilde{\mathbf{Q}} \widetilde{\mathbf{A}}_C + \widetilde{\mathbf{C}}^T \widetilde{\mathbf{C}} = \mathbf{0} \tag{39}$$

exist and are unique. Their Cholesky factorizations

$$\widetilde{\mathbf{P}} = \widetilde{\mathbf{R}} \widetilde{\mathbf{R}}^T, \quad \widetilde{\mathbf{Q}} = \widetilde{\mathbf{S}} \widetilde{\mathbf{S}}^T \tag{40}$$

finally define the low-rank Cholesky factors

$$\widetilde{\mathbf{Y}} = \mathbf{V} \widetilde{\mathbf{R}}, \quad \widetilde{\mathbf{Z}} = \mathbf{W} \widetilde{\mathbf{S}}, \tag{41}$$

that can be used for approximate TBR by Algorithm 2.

4.2 Incorporating the Methods

We are now ready to state the main theorem, which presents how the degrees of freedom $\widetilde{\mathbf{V}}$ and $\widetilde{\mathbf{W}}$ have to be chosen to generate the three methods.

Theorem 1. *For given sets* $\mathbf{s}$ *and* $\mathbf{m}$, *let* $\mathbf{V}$ *and* $\mathbf{W}$ *be the associated rational input and output Krylov subspaces, respectively. Let* $\widetilde{\mathbf{V}}$ *and* $\widetilde{\mathbf{W}}$ *determine the approximate truncated balanced realization via* (36) – (41) *and Algorithm 2. Then the following choices deliver the outcome of RKSM, ADI, and the two-step procedure:*

	Two-step	*RKSM*	*ADI*
$\widetilde{\mathbf{V}} :=$	$\mathbf{V}$	$\mathbf{W}$	$\mathbf{V}_{ADI}$
$\widetilde{\mathbf{W}} :=$	$\mathbf{W}$	$\mathbf{V}$	$\mathbf{W}_{ADI}$

Proof. The proof for RKSM is trivial: choosing $\widetilde{\mathbf{V}} := \mathbf{W}$, $\widetilde{\mathbf{W}} := \mathbf{V}$ and comparing (21), (22) with (36), (37) yields:

$$\widetilde{\mathbf{A}}_B = \mathbf{A}_\mathbf{V}, \qquad \widetilde{\mathbf{E}}_B = \mathbf{E}_\mathbf{V}, \qquad \widetilde{\mathbf{B}} = \mathbf{B}_\mathbf{V}, \tag{42}$$

$$\widetilde{\mathbf{A}}_C = \mathbf{A}_\mathbf{W}, \qquad \widetilde{\mathbf{E}}_C = \mathbf{E}_\mathbf{W}, \qquad \widetilde{\mathbf{C}} = \mathbf{C}_\mathbf{W}. \tag{43}$$

Therefore, $\widetilde{\mathbf{P}} = \mathbf{P}_{RK}$ and $\widetilde{\mathbf{Q}} = \mathbf{Q}_{RK}$, which proves the result. The proof for ADI is similar when following the computation via (29)–(33). It is left to prove the two-step procedure: choosing $\widetilde{\mathbf{V}} := \mathbf{V}$ and $\widetilde{\mathbf{W}} := \mathbf{W}$ leads to

$$\widetilde{\mathbf{A}}_B = \widetilde{\mathbf{A}}_C = \mathbf{A}_k, \qquad \widetilde{\mathbf{E}}_B = \widetilde{\mathbf{E}}_C = \mathbf{E}_k, \tag{44}$$

$$\widetilde{\mathbf{B}} = \mathbf{B}_k, \qquad \widetilde{\mathbf{C}} = \mathbf{C}_k, \tag{45}$$

with $\mathbf{A}_k, \mathbf{E}_k, \mathbf{B}_k$ and $\mathbf{C}_k$ from (15). Therefore, $\widetilde{\mathbf{P}} = \mathbf{P}_k$ and $\mathbf{E}^T \widetilde{\mathbf{Q}} \mathbf{E} = \mathbf{E}^T \mathbf{Q}_k \mathbf{E}$ are the Controllability and Observability Gramians, respectively, of the intermediate system from the two-step procedure. Their Cholesky factorizations $\mathbf{P}_k = \mathbf{R}_k \mathbf{R}_k^T$ and $\mathbf{Q}_k = \mathbf{S}_k \mathbf{S}_k^T$ define the low-rank square-roots $\mathbf{Y}_{2step} = \mathbf{V}\mathbf{R}_k$, $\mathbf{Z}_{2step} = \mathbf{W}\mathbf{S}_k$ that can be plugged into Algorithm 2. The resulting singular value decomposition

$$\mathbf{Z}_{2step}^T \mathbf{E} \mathbf{Y}_{2step} = \mathbf{S}_k^T \underbrace{\mathbf{W}^T \mathbf{E} \mathbf{V}}_{\mathbf{E}_k} \mathbf{R}_k = \mathbf{M} \boldsymbol{\Sigma} \mathbf{N}^T, \tag{46}$$

is the same SVD as the one in Algorithm 1 during the two-step procedure. Taking $\mathbf{A}_r$ as an example, we show that the choice $\widetilde{\mathbf{V}} := \mathbf{V}$, $\widetilde{\mathbf{W}} := \mathbf{W}$ together with Algorithm 2 results in the same reduced system as the two-step procedure including Algorithm 1:

$$\mathbf{A}_r = \mathbf{W}_{Bal}^T \mathbf{A} \mathbf{V}_{Bal} \tag{47}$$

$$= \boldsymbol{\Sigma}_{(1:q,1:q)}^{-1/2} \mathbf{M}_{(:,1:q)}^T \mathbf{Z}_{2step}^T \mathbf{A} \mathbf{Y}_{2step} \mathbf{N}_{(:,1:q)} \boldsymbol{\Sigma}_{(1:q,1:q)}^{-1/2} \tag{48}$$

$$= \boldsymbol{\Sigma}_{(1:q,1:q)}^{-1/2} \mathbf{M}_{(:,1:q)}^T \mathbf{S}_k^T \underbrace{\mathbf{W}^T \mathbf{A} \mathbf{V}}_{\mathbf{A}_k} \mathbf{R}_k \mathbf{N}_{(:,1:q)} \boldsymbol{\Sigma}_{(1:q,1:q)}^{-1/2} \tag{49}$$

Equation (49) reveals, that $\mathbf{A}_r$ in the framework is equal to the TBR of $\mathbf{A}_k$, which completes the proof. □

Remark 2. The two virtual intermediate systems (36) and (37) allow to unify the three methods: for the two-step procedure, (36) and (37) merge into one intermediate system (15); for RKSM, (36) and (37) become the orthogonally projected systems (21) and (22); and for ADI, (36) and (37) become the virtual intermediate systems (27) and its dual version.

4.3 Interpretation

How do the degrees of freedom $\widetilde{\mathbf{V}}$ and $\widetilde{\mathbf{W}}$ affect the reduction by approximate TBR, i. e. what is the difference between the reduced systems of the three methods? To answer this question, we consider the general case, i. e. $\widetilde{\mathbf{V}}$ and $\widetilde{\mathbf{W}}$ are arbitrary matrices of appropriate dimensions. Plugging general low-rank Cholesky factors $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Z}}$ from (41) into Algorithm 2 yields the generic SVD

$$\widetilde{\mathbf{Z}}^T \mathbf{E} \widetilde{\mathbf{Y}} = \widetilde{\mathbf{S}}^T \mathbf{W}^T \mathbf{E} \mathbf{V} \widetilde{\mathbf{R}} = \widetilde{\mathbf{S}}^T \mathbf{E}_k \widetilde{\mathbf{R}} =: \widetilde{\mathbf{M}} \widetilde{\boldsymbol{\Sigma}} \widetilde{\mathbf{N}}^T. \tag{50}$$

Following Algorithm 2 then defines the reduced matrices

$$\mathbf{A}_r = \widetilde{\boldsymbol{\Sigma}}_{(1:q,1:q)}^{-1/2} \widetilde{\mathbf{M}}_{(:,1:q)}^T \widetilde{\mathbf{S}}^T \mathbf{W}^T \mathbf{A} \mathbf{V} \widetilde{\mathbf{R}} \widetilde{\mathbf{N}}_{(:,1:q)} \widetilde{\boldsymbol{\Sigma}}_{(1:q,1:q)}^{-1/2}$$

$$= \widetilde{\mathbf{W}}_{Bal}^T \mathbf{A}_k \widetilde{\mathbf{V}}_{Bal} \tag{51}$$

$$\mathbf{E}_r = \widetilde{\boldsymbol{\Sigma}}_{(1:q,1:q)}^{-1/2} \widetilde{\mathbf{M}}_{(:,1:q)}^T \widetilde{\mathbf{S}}^T \mathbf{W}^T \mathbf{E} \mathbf{V} \widetilde{\mathbf{R}} \widetilde{\mathbf{N}}_{(:,1:q)} \widetilde{\boldsymbol{\Sigma}}_{(1:q,1:q)}^{-1/2}$$

$$= \widetilde{\mathbf{W}}_{Bal}^T \mathbf{E}_k \widetilde{\mathbf{V}}_{Bal} \tag{52}$$

$$\mathbf{B}_r = \widetilde{\boldsymbol{\Sigma}}_{(1:q,1:q)}^{-1/2} \widetilde{\mathbf{M}}_{(:,1:q)}^T \widetilde{\mathbf{S}}^T \mathbf{W}^T \mathbf{B} = \widetilde{\mathbf{W}}_{Bal}^T \mathbf{B}_k \tag{53}$$

$$\mathbf{C}_r = \mathbf{C} \mathbf{V} \widetilde{\mathbf{R}} \widetilde{\mathbf{N}}_{(:,1:q)} \widetilde{\boldsymbol{\Sigma}}_{(1:q,1:q)}^{-1/2} = \mathbf{C}_k \widetilde{\mathbf{V}}_{Bal} \tag{54}$$

with the projection matrices $\widetilde{\mathbf{V}}_{Bal} := \widetilde{\mathbf{R}} \widetilde{\mathbf{N}}_{(:,1:q)} \widetilde{\boldsymbol{\Sigma}}_{(1:q,1:q)}^{-1/2}$ and $\widetilde{\mathbf{W}}_{Bal} := \widetilde{\mathbf{S}} \widetilde{\mathbf{M}}_{(:,1:q)} \widetilde{\boldsymbol{\Sigma}}_{(1:q,1:q)}^{-1/2}$. Both $\widetilde{\mathbf{V}}_{Bal}$ and $\widetilde{\mathbf{W}}_{Bal}$ exhibit the structure of balancing and truncating projection matri-

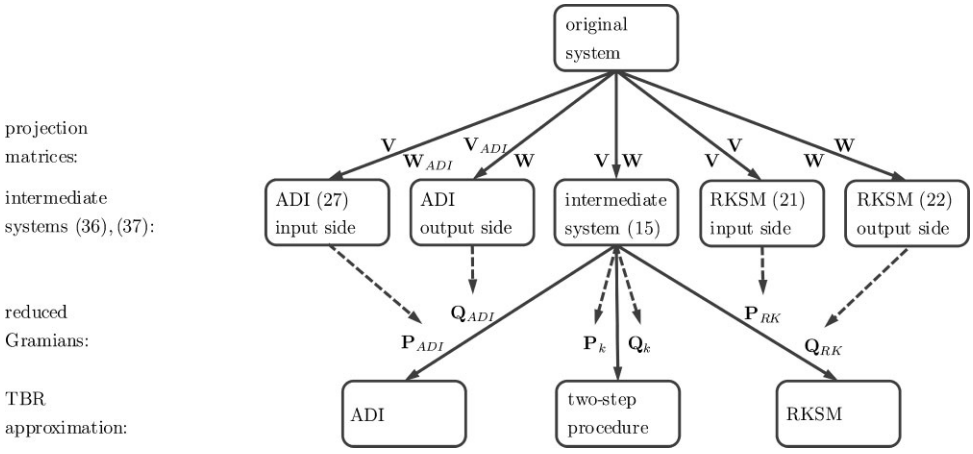

Figure 1 The unifying framework for approximate TBR.

ces in terms of the square-root method. It is remarkable, that this structure shows up for arbitrary choices of $\widetilde{\mathbf{V}}$, $\widetilde{\mathbf{W}}$; and from (51)–(54) it follows that $\widetilde{\mathbf{V}}_{Bal}$ and $\widetilde{\mathbf{W}}_{Bal}$ always project the same intermediate system (15) – which is an oblique projection of (1) by the Krylov subspaces $\mathbf{V}$ and $\mathbf{W}$. This means, that the overall reduced system of all three methods stems from a kind of balanced truncation of the same intermediate system (15). In other words, any method for approximate TBR presented here, virtually generates the intermediate system (15), and subsequently further reduces (projects) it with the matrices $\widetilde{\mathbf{R}}\widetilde{\mathbf{N}}_{(:,1:q)}$ and $\widetilde{\mathbf{S}}\widetilde{\mathbf{M}}_{(:,1:q)}$. (The additional matrix $\widetilde{\mathbf{\Sigma}}_{(1:q,1:q)}$ in (51)–(54) only defines a diagonal scaling matrix, that does not affect the transfer behavior of the overall reduced system.) To sum up, all three methods can be interpreted as a two-step procedure: compute the intermediate system (15) by an oblique projection with $\mathbf{V}$ and $\mathbf{W}$ and then further reduce it by some kind of balanced truncation. Therefore, the degrees of freedom $\widetilde{\mathbf{V}}$ and $\widetilde{\mathbf{W}}$ – i. e. the difference in the three methods – can solely affect how the intermediate system (15) is further reduced. These findings are summarized in Fig. 1. The influence of $\widetilde{\mathbf{V}}$ and $\widetilde{\mathbf{W}}$ on the reduced system is discussed in Sect. 5.

Remark 3. Please note, that approximate TBR by ADI will always be calculated by the iteration (26) and the subsequent execution of Algorithm 2. Therefore, the intermediate system (15) – that originates from an oblique projection by Krylov subspaces – will not be constructed explicitly; it will be rather virtually generated during the procedure, as shown above. The same conclusion holds for approximate TBR by RKSM.

4.4 $\mathcal{H}_2$ optimal shifts

A special case of high interest is the choice of $\mathcal{H}_2$ optimal shifts/expansion points $\mathbf{s}_{\mathcal{H}_2}$, $\mathbf{m}_{\mathcal{H}_2}$. The concept of $\mathcal{H}_2$ optimality will be briefly reviewed in the following. For the ease of presentation, we concentrate on single input and output, $m = p = 1$, although the results can be generalized to multiple inputs and outputs.

Definition 1. *A reduced dynamical system $G_k(s)$ of order k is called (locally) $\mathcal{H}_2$ optimal if*

$$G(-\lambda_i) = G_k(-\lambda_i)\,, \quad i = 1,\ldots,k \tag{55}$$
$$G'(-\lambda_i) = G_k'(-\lambda_i)\,, \quad i = 1,\ldots,k \tag{56}$$

where λ_i denote the poles of $G_k(s)$ and $G'(s)$ denotes the derivative with respect to s.

Therefore, an $\mathcal{H}_2$ optimal reduced model interpolates the original system and its derivative at the mirror images of the reduced poles. It can be shown, that (55) and (56) are first-order necessary conditions for minimal error in the $\mathcal{H}_2$ norm, [14]. Neglecting condition (56) leads to the concept of pseudo-optimality.

Definition 2. *A reduced dynamical system $G_k(s)$ of order k is called $\mathcal{H}_2$ pseudo-optimal if for all poles λ_i, $i = 1,\ldots,k$ of $G_k(s)$*

$$G(-\lambda_i) = G_k(-\lambda_i)\,. \tag{57}$$

$\mathcal{H}_2$ pseudo-optimality is necessary for $\mathcal{H}_2$ optimality and can be enforced for arbitrary sets $\{\lambda_i\} := \{\lambda_1,\ldots,\lambda_k\}$ by projective model reduction, e. g. via (28)–(30). Therefore, as shown in Sect. 3.5, the virtual ADI system (27) is $\mathcal{H}_2$ pseudo-optimal.

The first-order necessary conditions (55) and (56) cannot be enforced for arbitrary sets of reduced poles. However, an iterative algorithm – the *Iterative Rational Krylov Algorithm* (IRKA) – converges to such a set of locally optimal shifts; see [14] for details. We are now ready to state a connection between the ADI method and the two-step procedure for $\mathcal{H}_2$ optimal shifts.

Lemma 1. *Assume that IRKA has been executed in order to find k $\mathcal{H}_2$ optimal shifts. If these shifts are used for computing both the Krylov subspaces $\mathbf{V}$, $\mathbf{W}$ and the ADI bases $\mathbf{Y}_{ADI}$, $\mathbf{Z}_{ADI}$, then the approximate Gramians $\widehat{\mathbf{P}}$ and $\mathbf{E}^T\widehat{\mathbf{Q}}\mathbf{E}$ by ADI and the two-step procedure are equal. Therefore, the resulting reduced systems by approximate TBR are equal.*

Proof. The transfer behavior of an $\mathcal{H}_2$ pseudo-optimal reduced system is uniquely defined by $2k$ values: k reduced eigenvalues are fixed and k interpolation constraints (57) are fulfilled. Therefore, the virtual intermediate systems (27) of ADI and also its dual version are equal to the intermediate systems from the two-step procedure (15). This is due to the fact that the intermediate system (15) using IRKA shifts is $\mathcal{H}_2$ optimal and therefore, also $\mathcal{H}_2$ pseudo-optimal, and that the (virtual) ADI systems (from both input and output side) are $\mathcal{H}_2$ pseudo-optimal, see [27]. □

Remark 4. Lemma 1 holds only for k $\mathcal{H}_2$ optimal shifts. Allowing higher multiplicities, as e. g. in iterative methods, leads to an intermediate system (15) that is not $\mathcal{H}_2$ optimal anymore. Then, Lemma 1 does not apply anymore, and the ADI method and the two-step procedure will result in different reduced systems, as shown by the numerical example in Sect. 6.

Remark 5. A closely related result was independently observed in [5; 8; 11]: ADI and RKSM are equal for particular sets of pseudo-optimal shifts, that can be found by a slightly modified IRKA.

5 Discussion

The unified formulation (51)–(54) allows to compare the three methods for approximate TBR with respect to different aspects in the following.

5.1 Existence and Solvability

The preservation of stability is crucial in model order reduction methods. However, in the presented framework, two aspects of stability have to be distinguished: the preservation of stability in the final reduced system and preservation of stability in the intermediate systems (36) and (37). We first tackle the second question, which is in fact a question of solvability: the intermediate solutions (40) admit a Cholesky factorization (which is required by Algorithm 2) if and only if they are positive definite. This in turn is true if and only if the intermediate matrix pairs $\widetilde{\mathbf{A}}_B, \widetilde{\mathbf{E}}_B$ and $\widetilde{\mathbf{A}}_C, \widetilde{\mathbf{E}}_C$ have their generalized eigenvalues in the open left half of the complex plane. Therefore, a reduced model by approximate TBR can be found if and only if both intermediate systems (36) and (37) preserve stability of the original system.

The ADI method guarantees solvability by construction: the eigenvalues of $\mathbf{A}_{ADI}, \mathbf{E}_{ADI}$ are the mirror images of the expansion points $\mathbf{s}$, which are most appropriately chosen in the right half of the complex plane.

RKSM relies on the direct solution of the intermediate Lyapunov equations (19), (20) that do not necessarily have positive definite solutions. However, a sufficient condition for preserving asymptotic stability by orthogonal projections is: $\mathbf{E} > \mathbf{0}$ and $\mathbf{A} + \mathbf{A}^T < \mathbf{0}$. (For example, second-order systems can be efficiently transformed into a first-order state space realization that fulfills $\mathbf{E} > \mathbf{0}$ and $\mathbf{A} + \mathbf{A}^T < \mathbf{0}$, [18].) Therefore, RKSM guarantees solvability for original systems that fulfill $\mathbf{E} > \mathbf{0}$ and $\mathbf{A} + \mathbf{A}^T < \mathbf{0}$, because the intermediate Lyapunov equations (19) and (20) stem from orthogonal projections.

In contrast, the intermediate system in the two-step procedure originates from an oblique projection, for which asymptotic stability cannot be guaranteed in general.

Another prerequisite of the two-step procedure – in contrast to the other methods – is equality in the dimensions of input and output Krylov subspaces: $k_B = k_C$. This however is not restrictive, as it can be resolved by the user: assume e. g. more inputs than outputs, $m > p$; applying the same sets $\mathbf{s}$ and $\mathbf{m}$ for both input and output side results in bases $\mathbf{V}$, $\mathbf{W}$ of different dimensions $k_B > k_C$. But then, k_C can easily be aligned to k_B by recycling some expansion points to compute additional vectors in $\mathbf{W}$ (resulting in higher multiplicities).

5.2 Stability Preservation

The second aspect of stability preservation is: do the methods guarantee stable reduced systems (output of Algorithm 2), if the original system is stable? For ADI and RKSM, preservation of stability *cannot* be guaranteed, as confirmed by the technical example in Sect. 6.

In contrast, the two-step procedure guarantees stable reduced systems – once a solution exists – because the reduced system is a balanced truncation of the intermediate system (15), and TBR is known to preserve stability. Therefore, if the intermediate system (15) is stable – i. e. the two-step procedure is solvable – also the final reduced system is stable.

5.3 Numerical Effort in Iterative Methods

We briefly compare the numerical effort for the different approaches. Once a low-rank Cholesky factor is found, all three methods run Algorithm 2, including an SVD as the main cost. In large-scale settings, however, the main effort for all three methods is usually the solution of n-by-n linear systems of equations, in order to compute the Krylov subspaces or ADI bases.

The difference is that RKSM and the two-step procedure additionally solve two intermediate Lyapunov equations by direct methods. (An efficient implementation of the two-step procedure might take advantage of the equal coefficient matrices $\mathbf{A}_k, \mathbf{E}_k$ in the two dual Lyapunov equations, resulting in a slightly faster two-step procedure compared to RKSM.) In contrast, ADI implicitly "solves" the intermediate pseudo-optimal Lyapunov equations by the iteration (26) and therefore requires less numerical effort. This however, might be negligible for the case $k \ll n$, because then the direct solution of two Lyapunov equations of order k takes marginal time, compared to the solution of $2k$ systems of equation of order n-by-n.

In iterative methods, however, the additional effort might be significant: for every evaluation of the con-

vergence criterion, the two dual Lyapunov equations of increasing dimension have to be solved. For a high number of iterations, this repeated solution considerably raises the numerical effort of RKSM and the two-step procedure compared to the ADI iteration.

5.4 Heuristic Evaluation of Performance

The intermediate system (15) (virtually) appears in all three methods, and it is further reduced by some kind of balanced truncation in (51)–(54), see Fig. 1. But which way of further reducing (15) generally yields the best approximation of the original system (1)?

Theorem 1 reveals, that the two-step procedure uses the Controllability and Observability Gramians of system (15) for the SVD (50). In this respect, the two-step procedure performs true TBR of (15). From an engineering point of view, this sounds reasonable: all available information on the transfer behavior of (1) is condensed in the Krylov subspaces $\mathbf{V}$, $\mathbf{W}$ and hence in the intermediate system (15). The best one can do for approximate TBR of (1) is then to apply TBR by direct methods to the intermediate system (15).

In contrast, ADI and RKSM use Gramians in the SVD (50), which belong to different intermediate systems (given by equations (19), (20) and (31), respectively). This results in some kind of "pseudo square-root TBR" of (15), because the Gramians of the virtual systems (21), (22) or (27) are employed for the balanced truncation of system (15).

To sum up, the two-step procedure conducts "true" square-root TBR, whereas ADI and RKSM conduct "pseudo" square-root TBR of (15). Therefore, the two-step procedure should usually lead to the best approximation of the reduced system from direct TBR of (1). Admittedly, this is a heuristic statement that cannot be proven in a mathematically rigorous way. However, the above considerations strongly suggest this conclusion, which is also supported by the technical example in Sect. 6 and the observations in [15; 24].

5.5 Possible Application of Error Bounds

A severe drawback of the low-rank square-root method is the loss of the a priori error bound (10): the bound relies on the knowledge of the truncated Hankel singular values; in low-rank TBR, depending on k_B and k_C, one can usually expect only a good approximation of the retained HSVs.

A remedy is to upper-bound all truncated HSVs by the first truncated one: $\sigma_i := \sigma_{q+1}$, $i = q+2, \ldots, n$. This however might introduce a huge overestimation. Furthermore, the exact σ_{q+1} might be larger than the approximated one. For that reason, the bound might not be rigorous, but with a decent convergence analysis this case can be practically excluded.

Nevertheless, rigorous global error bounds can be stated, based on the intermediate system (15). Although no general error bounds are available for Krylov-based projection methods, first attempts for special system classes can be found for example in [19]. In this respect, we now assume that some error bound γ_1 in the $\mathcal{H}_\infty$ norm is available for the intermediate system (15):

$$\|\mathbf{G} - \mathbf{G}_k\|_\infty \leq \gamma_1 \,. \tag{58}$$

We are interested in the overall error, i. e.

$$\mathbf{G}(s) - \mathbf{G}_r(s) = [\mathbf{G}(s) - \mathbf{G}_k(s)] + [\mathbf{G}_k(s) - \mathbf{G}_r(s)] \tag{59}$$

The advantage of the two-step procedure is that the error bound (10) can be applied to the second step

$$\gamma_2 := 2 \sum_{i=q+1}^{k} \sigma_i \geq \|\mathbf{G}_k - \mathbf{G}_r\|_\infty \,. \tag{60}$$

Due to the triangular inequality, taking norms in (59) leads to a rigorous bound on the $\mathcal{H}_\infty$ error,

$$\|\mathbf{G} - \mathbf{G}_r\|_\infty \leq \gamma_1 + \gamma_2 \,, \tag{61}$$

which solely applies to the two-step procedure, and only if a bound γ_1 for the first step is available.

If a global bound for RKSM or ADI is desired, one would have to compute the $\mathcal{H}_\infty$ norm of the error system in the second step $\|\mathbf{G}_k(s) - \mathbf{G}_r(s)\|_\infty$ by some direct method. This results in a higher effort than evaluating the bound (60) for the two-step procedure.

5.6 Résumé

The previous findings are summarized in Table 1, where the different methods are opposed to each other. RKSM suffers from almost the same problems as the two-step procedure and never outperforms it. Therefore, one should prefer the two-step procedure to RKSM in approximate TBR.

The two-step procedure preserves stability and generally yields the best approximation. Its drawbacks are the high numerical effort and that it might not be solvable (due to an unstable intermediate system), which in turn are the advantages of ADI: guaranteed existence of a solution and low numerical effort.

The benefits of both methods can be combined by the following approach: first apply ADI, because it guarantees a solution at low numerically effort; as soon as convergence occurs in iterative methods, explicitly construct the intermediate system (15) via projection by the

Table 1 Comparison of the different methods. (+: good, 0: medium, –: bad).

	Two-step	RKSM	ADI
Solvability	–	0	+
Stability preservation	+	–	–
Numerical Effort	–	–	+
Performance	+	0	0
Error Bound	+	0	0

ADI bases $\mathbf{V} := \mathbf{Y}_{ADI}$ and $\mathbf{W} := \mathbf{Z}_{ADI}$, and subsequently perform TBR by direct methods. The justification of this approach is that the numerical effort for one final TBR step is usually moderate and that the approximation is generally improved. In this respect, performing ADI with a final two-step procedure combines the benefits of both methods: numerical efficiency with good performance and most likely preservation of stability.

Remark 6. It is also reasonable to perform direct TBR of the intermediate system at arbitrary iteration cycles of the ADI method. This leads to a trade-off: on the one hand, one would like to perform many intermediate steps with direct TBR to achieve a good approximation and therefore, a fast decay in the convergence criterion; on the other hand, one would like to perform as few intermediate steps as possible to be numerically efficient. Giving a general compromise is challenging and out of the scope of the paper. However, this approach was already suggested in [21], where it is referred to as a *projection based acceleration technique* for ADI.

Remark 7. Another relevant approach for approximately solving Lyapunov equations is the *Krylov-plus-inverse-Krylov* (KPIK) method, suggested in [25]. This is equivalent to RKSM, but with a distinct choice of expansion points, namely for the input side

$$\mathcal{K}\left(\mathbf{A}^{-1}\mathbf{E},\, \mathbf{A}^{-1}\mathbf{B}\right) \cup \mathcal{K}\left(\mathbf{E}^{-1}\mathbf{A},\, \mathbf{E}^{-1}\mathbf{B}\right), \tag{62}$$

and for the output side

$$\mathcal{K}\left(\mathbf{A}^{-T}\mathbf{E}^{T},\, \mathbf{A}^{-T}\mathbf{C}^{T}\right) \cup \mathcal{K}\left(\mathbf{E}^{-T}\mathbf{A}^{T},\, \mathbf{E}^{-T}\mathbf{C}^{T}\right). \tag{63}$$

This means that KPIK applies the set of expansion points $\mathbf{s}_{KPIK} = [0,\, \infty]$ with multiplicities $\mathbf{m}_{KPIK} = [k/2,\, k/2]$. This is why KPIK was excluded in this work so far, because the expansion points 0 and ∞ are prohibited for the ADI method. Nevertheless, the comparison between RKSM and the two-step procedure, as carried out before, also applies for the Krylov bases (62) and (63): on the one hand, the two-step procedure generally leads to a better approximation than the low-rank Cholesky factors by (24); on the other hand, for systems that fulfill $\mathbf{E} > \mathbf{0}$ and $\mathbf{A} + \mathbf{A}^T < \mathbf{0}$, the existence of a solution is guaranteed by (24) but not by the two-step procedure.

6 Technical Example

As a first technical example we consider the "beam" from the SLICOT benchmark collection[1]. The model is a spatial discretization of a partial differential equation describing the dynamics of a clamped beam. The single input is a force at the free end and the single output the resulting displacement. The order is $n = 348$, and TBR by direct methods is performed for comparison.

[1] Y. Chahlaoui and P. Van Dooren. A collection of Benchmark Examples for Model Reduction of Linear Time Invariant Dynamical Systems. Slicot working note, 2002.

As all three methods intend to be an approximation of direct TBR, it is fair to compare them with the reduction by direct TBR – rather than with the original model. In order to contrast the three methods, we a priori fix the final reduced order q and also the dimension of the low-rank Cholesky factors $k_B = k_C = k$. The Lyapunov equations of the intermediate systems were solved in every iteration, starting with the qth iteration. All computations were performed in MATLAB on a dual-core processor with 3 GHz.

In the first use case, we set the final reduced order to $q = 8$ and applied Algorithm 1 for exact TBR, which took 1.63 seconds. We chose k logarithmically spaced expansion points in the interval [0.1, 10] and performed all three methods for $k = 8, \ldots, 26$. The $\mathcal{H}_2$ errors, relative to the reduced system computed by direct TBR, are shown in Fig. 2. (The relative $\mathcal{H}_\infty$ errors showed similar behavior.) Missing points of ADI correspond to unstable final reduced systems, and missing points of both other methods correspond to reductions that were not solvable, i. e. the intermediate systems were unstable. It can be seen, that the two-step procedure – if solvable – outperforms the other methods. For $k = 26$, the execution of ADI took 0.47 seconds, of RKSM 0.74 seconds and of the two-step prodecure 0.65 seconds. As suggested in Sect. 5.6, we combined ADI with a final step of direct TBR, which took 0.51 seconds altogether. This combines numerical efficiency with good performance.

The performance of ADI strongly depends on an appropriate choice of shifts, [21]; and it was recently

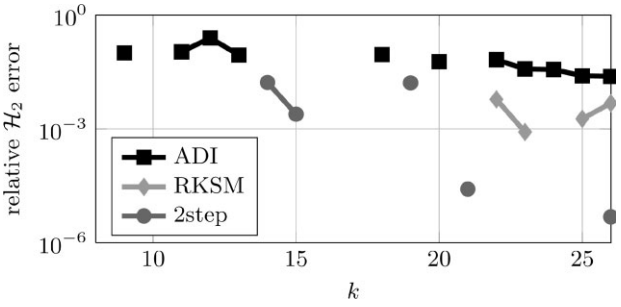

Figure 2 Relative $\mathcal{H}_2$ errors for beam model with $q = 8$ and k logarithmically spaced shifts.

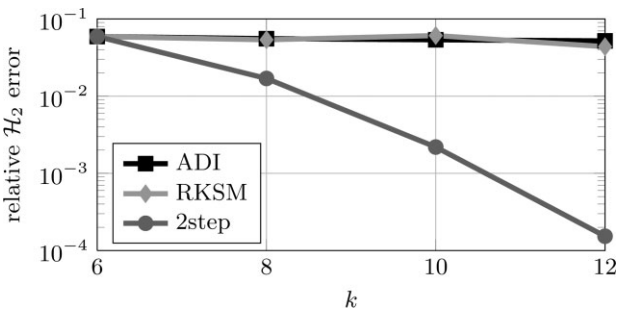

Figure 3 Relative $\mathcal{H}_2$ errors for beam model with $q = 6$ and 4 IRKA shifts.

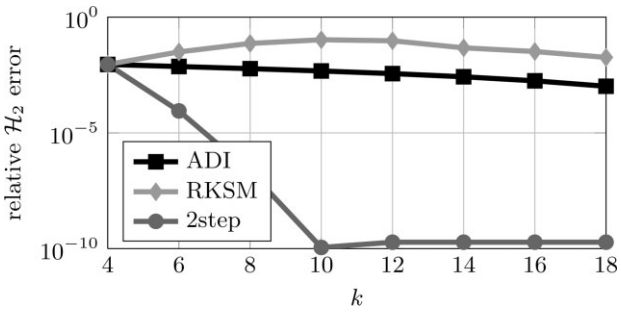

Figure 4 Relative $\mathcal{H}_2$ errors for steel profile with $q = 4$ and 2 IRKA shifts.

shown in [5], that the $\mathcal{H}_2$ optimal shifts are also optimal for ADI. In this respect, we ran IRKA, which took 1.79 seconds, to find 4 $\mathcal{H}_2$ optimal expansion points. Setting $q = 6$, we performed the three methods for the values $k = 6, 8, 10, 12$. The relative $\mathcal{H}_2$ errors are shown in Fig. 3. For $k = q = 6$ all three methods have to yield the same reduced system. By increasing k, the two-step procedure rapidly converges to the balanced truncation by direct methods, whereas the approximation is only marginally improved by ADI and RKSM.

As a second example we consider a semi-discretized heat transfer problem for optimal cooling of steel profiles[2] of the order $n = 1357$. We chose the model from the first input to the first output and set $q = 4$. Direct TBR by Algorithm 1 took 154 seconds. We ran IRKA to compute two $\mathcal{H}_2$ optimal shifts. The Algorithm converged in 9 iterations, which took 0.72 seconds. The three methods were performed for all even values of $k = 4, \ldots, 18$. The relative $\mathcal{H}_2$ errors are shown in Fig. 4. The two-step procedure rapidly converges to the reduction by direct TBR, whereas ADI converges slowly and RKSM does not converge at all. For $k = 18$, the execution of ADI took 0.14 seconds, of RKSM 0.32 seconds and of the two-step procedure 0.29 seconds. ADI together with a final step of direct TBR took 0.22 seconds. Therefore, the combination of ADI with the two-step procedure can outperform direct TBR of the original system.

7 Conclusions and Outlook

This work considers approximate truncated balanced realizations by three different approaches: the two-step procedure and the approximate solution of Lyapunov equations by RKSM or ADI. We presented a novel framework that unifies these three methods by introducing two projection matrices $\widetilde{\mathbf{V}}$ and $\widetilde{\mathbf{W}}$, that are able to trigger the respective solutions. The framework allows the reasonable comparison of the different methods and suggests strategies for their usage. A technical example confirmed the findings.

[2] Oberwolfach model reduction benchmark collection, Oct. 2003, Available online at http://portal.uni-freiburg.de/imteksimulation/downloads/benchmark.

References

[1] A. C. Antoulas. *Approximation of large-scale dynamical systems*. Advances in Design and Control, SIAM, Philadelphia, 2005.

[2] A. C. Antoulas. On pole placement in model reduction. *Automatisierungstechnik*, 55(9), 2007.

[3] A. Astolfi. Model reduction by moment matching for linear and nonlinear systems. *IEEE Transactions on Automatic Control*, 55(10):2321–2336, 2010.

[4] B. Beckermann. An error analysis for rational Galerkin projection applied to the Sylvester equation. *SIAM Journal on Numerical Analysis*, 49(6):2430–2450, 2011.

[5] P. Benner and T. Breiten. On optimality of interpolation-based low rank approximations of large-scale matrix equations. Max Planck Institute Magdeburg Preprint MPIMD/11–10v2, 2012.

[6] P. Benner, P. Kürschner, and J. Saak. An improved numerical method for balanced truncation for symmetric second order systems. Max Planck Institute Magdeburg Preprint MPIMD/12-20, 2012.

[7] P. Benner, P. Kürschner, and J. Saak. Real versions of low-rank adi methods with complex shifts. Max Planck Institute Magdeburg Preprint MPIMD/12-11, 2012.

[8] V. Druskin, L. Knizhnerman, and V. Simoncini. Analysis of the rational Krylov subspace and ADI methods for solving the Lyapunov equation. *SIAM Journal on Numerical Analysis*, 49(5):1875–1898, 2011.

[9] V. Druskin and V. Simoncini. Adaptive rational Krylov subspaces for large-scale dynamical systems. *Systems & Control Letters*, 60:546–560, 2011.

[10] R. Eid, H. Panzer, and B. Lohmann. How to choose a single expansion point in Krylov-based model reduction? Technical reports on automatic control, Institute of Automatic Control, Technical University of Munich, November 2009.

[11] G. M. Flagg and S. Gugercin. On the ADI method for the Sylvester equation and the optimal-H2 points. *Applied Numerical Mathematics*, 64:50–58, Feb 2013.

[12] R. W. Freund. Krylov-subspace methods for reduced-order modeling in circuit simulation. *Journal of Computational and Applied Mathematics*, 123:395–421, 2000.

[13] E. J. Grimme. *Krylov projection methods for model reduction*. PhD thesis, Uni. Illinois at Urbana Champaign, 1997.

[14] S. Gugercin, A. C. Antoulas, and C. A. Beattie. H2 model reduction for large-scale linear dynamical systems. *SIAM Journal on Matrix Analysis and Applications*, 30(2):609–638, 2008.

[15] M. Kamon, F. Wang, and J. White. Generating nearly optimally compact models from Krylov-subspace based reduced-order models. *IEEE Trans. Circuit and Systems*, 47(4):239–248, 2000.

[16] J.-R. Li and J. White. Low rank solution of Lyapunov equations. *SIAM Journal on Matrix Analysis and Applications*, 24(1):260–280, 2002.

[17] B. C. Moore. Principal component analysis in linear systems: Controllability, observability and model reduction. *IEEE Transactions on Automatic Control*, AC-26:17–32, 1981.

[18] H. Panzer, T. Wolf, and B. Lohmann. A strictly dissipative state space representation of second order systems. *at-Automatisierungstechnik*, 60:392–396, 2012.

[19] H. Panzer, T. Wolf, and B. Lohmann. H_2 and H_∞ error bounds for model order reduction of second order systems by Krylov subspace methods. In *European Control Conference*, Zurich, Switzerland, 2013.

[20] T. Penzl. A cyclic low-rank Smith method for large sparse Lyapunov equations. *SIAM Journal on Scientific Computing*, 21(4):1401–1418, 2000.

[21] J. Saak. *Efficient Numerical Solution of Large Scale Algebraic Matrix Equations in PDE Control and Model Order Reduction*. PhD thesis, Chemnitz Uni. of Technology, 2009.

[22] J. Saak, P. Benner, and P. Kürschner. A goal-oriented dual lrcf-adi for balanced truncation. In *Vienna Conference on Mathematical Modelling (MATHMOD)*, 2012.

[23] J. Sabino. *Solution of Large-Scale Lyapunov Equations via the Block Modified Smith Method*. PhD thesis, Rice Univ. Houston, 2007.
[24] B. Salimbahrami. A sequential reduction scheme in reduced order modelling of large scale systems. Technical Reports on Automatic Control TRAC-5, Institute of Automatic Control, TUM, 2010.
[25] V. Simoncini. A new iterative method for solving large-scale Lyapunov matrix equations. *SIAM Journal on Scientific Computing*, 29:1268–1288, 2007.
[26] D. C. Sorensen and Y. Zhou. Direct methods for matrix Sylvester and Lyapunov equations. *Journal of Applied Mathematics*, 2003(6):227–303, 2003.
[27] T. Wolf, H. Panzer, and B. Lohmann. ADI-Lösung großer Ljapunow-Gleichungen mittels Krylov-Methoden und neue Formulierung des Residuums. Talk given at the 2012 annual GMA Fachausschuss 1.30: Modellbildung, Identifikation und Simulation in der Automatisierungstechnik, VDI/VDE, Anif/Salzburg, Austria, 2012.
[28] T. Wolf, H. Panzer, and B. Lohmann. H2 pseudo-optimality in model order reduction by Krylov subspace methods. In *Proceedings of the European Control Conference*, Zurich, Switzerland, 2013.
[29] T. Wolf, H. Panzer, and B. Lohmann. On the residual of large-scale Lyapunov equations for Krylov-based approximate solutions. In *Proceedings of the American Control Conference*, Washington, DC, USA, 2013.

Received: January 31, 2013

Dipl.-Ing. Thomas Wolf is a research assistant at the Institute of Automatic Control of the Faculty of Mechanical Engineering at the Technische Universität München. His field of research includes model order reduction of large-scale systems and port-Hamiltonian model representation.

Address: Institute of Automatic Control, Technische Universität München, Boltzmannstr. 15, D-85748 Garching bei München, Phone: +49-(0)89-289-15592, e-mail: thomas.wolf@tum.de

Dipl.-Ing. Heiko K. F. Panzer is a research assistant at the Institute of Automatic Control of the Faculty of Mechanical Engineering at the Technische Universität München. His research interest includes model reduction of large-scale systems by Krylov subspace methods and parametric model order reduction by matrix interpolation.

Address: Institute of Automatic Control, Technische Universität München, Boltzmannstr. 15, D-85748 Garching bei München, Phone: +49-(0)89-289-15592, e-mail: panzer@tum.de

Prof. Dr.-Ing. habil. Boris Lohmann is Head of the Institute of Automatic Control of the Faculty of Mechanical Engineering at the Technische Universität München. His research activities include model order reduction, nonlinear, robust and optimal control as well as active vibration control.

Address: Institute of Automatic Control, Technische Universität München, Boltzmannstr. 15, D-85748 Garching bei München, Fax: +49-(0)89-289-15653, e-mail: lohmann@tum.de