School of Mathematics, Meteorology
and Physics

University of
Reading

Department of Mathematics

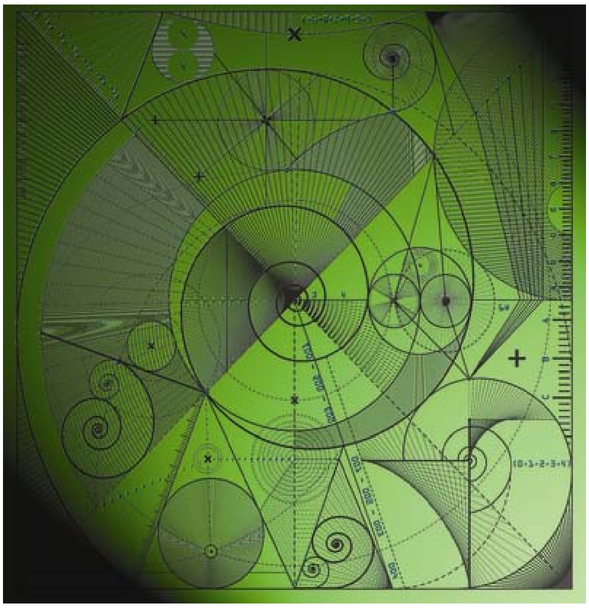
Preprint MPS_2010-06

18 March 2010

Model order reduction for discrete unstable control systems using a balanced truncation approach

by

C. Boess, N.K. Nichols and A. Bunse-Gerstner



Model order reduction for discrete unstable control systems using a balanced truncation approach

C. Boess[*], N.K. Nichols[*] and A. Bunse-Gerstner[†]

Abstract

Mathematical modeling of problems occurring in natural and engineering sciences often results in a very large dynamical system. Efficient techniques for model order reduction are required, therefore, to reduce the complexity of the system. Almost all such techniques require the dynamical system to be asymptotically stable. Balanced truncation is a well-known and approved model reduction method. There already exists a simple approach for applying this technique to unstable systems, but it does not capture the full behavior of the system successfully. In this paper, we propose a new model reduction method based on α-bounded balanced truncation, which can be applied to unstable systems independently of the number of unstable poles. We establish that this new method computes a low order approximation to the full order system such that the corresponding error system is close to being optimal with respect to a well-defined norm for unstable systems. Moreover, we prove a global error bound for the error system. In numerical experiments with unstable test models we compare the new α-bounded balanced truncation method with the standard extension of balanced truncation for unstable systems. The results show the superior performance of the α-bounded method.

Keywords: Model order reduction, unstable systems, balanced truncation, control

1 Introduction

Reduced order modeling is a crucial concept within the study of dynamical systems. The purpose of model order reduction is to reduce the order of the system substantially while still capturing its most important properties. Most of the known model reduction techniques are for asymptotically stable systems only, but in many fields of applications large unstable systems do occur and an order reduction is required. For example to be able to make a reliable weather forecast high resolution models of the atmosphere are indispensable. These models generally contain a large number of unstable modes. Moreover, the large dimensions of these unstable systems - usually about 10^7 unknowns are involved - require efficient techniques to reduce the order of the model considerably without losing essential information. In this paper we propose a new concept for reducing the order of discrete-time unstable systems while still capturing the most important information to match the input-output behavior of the original full order model. Similar results can be established for continuous-time systems (see [12]). Our focus is on a balanced truncation type method because this is an approved and reliable technique for reducing the order of dynamical systems. However, our new approach can also be applied within other model

[*] C.Boess@reading.ac.uk (corresponding author), N.K.Nichols@reading.ac.uk, Department of Mathematics, University of Reading, Caroline Boess acknowledges support by NCEO

[†] Bunse-Gerstner@math.uni-bremen.de, ZeTeM, Universitaet Bremen

reduction methods, including rational interpolation and Kyrlov subspace methods, see e.g. [12].

Originally, the balanced truncation method was proposed for asymptotically stable continuous-time systems by Moore in 1981 [14]. Pernebo and Silverman [16] extended the method to discrete-time systems in 1982. There already exist some extensions of the standard method to unstable systems. Most of these methods are based on an additive decomposition separating the asymptotically stable from the unstable part of the system. These techniques assume that unstable poles cannot be neglected when modeling the dynamics of a system, see e.g. [7, pp. 1177-1178], [15, 10, 19] and the references therein.

The main disadvantage of all methods based on this idea is that they are very limited when the system has a large number of unstable poles. A reduction of the full order system to a reduction order smaller than the number of unstable poles supplies a low order model which can only keep some of the unstable modes while the asymptotically stable part is ignored completely. Thus, this procedure cannot supply a good approximation of the input-output behavior of the whole full order system but only of its unstable part.

In this paper we propose a new approach to approximate discrete unstable control systems by systems of lower order using a balanced truncation technique. In contrast to existing approaches, this new method approximates the input-output behavior of the asymptotically stable as well as of the unstable part of the full order system, no matter how many unstable poles there are. The main idea is to extend the balanced truncation method to unstable systems by considering a different norm in which the error system is measured.

Usually, balanced truncation for asymptotically stable systems computes a low order system such that the output error is close to being optimal in the h_2-Hardy-norm [4, 5]. This norm is only well-defined for asymptotically stable systems. There exists an extension to unstable systems, the so-called $h_{2,\alpha}$-norm [9, 5]. We use this extended norm to define a new balanced truncation method for unstable systems where the output error is then close to being optimal in the $h_{2,\alpha}$-norm. Moreover, we derive a global error bound for the new balanced truncation method.

The outline of this paper is as follows. Section 2 gives a brief introduction to the model reduction method of balanced truncation for asymptotically stable discrete systems summarizing its most important properties. It also presents the main ideas of the standard extension to unstable systems. In Section 3 this is followed by the proposition of a new model reduction approach for unstable systems, the *α-bounded balanced truncation method*. Finally, Section 4 contains results of various numerical experiments using three different unstable discrete test models. We compare our new method of α-bounded balanced truncation with the already existing balanced truncation approach for unstable systems. The paper concludes with a summary of our results.

2 Reduced order modeling for discrete systems using balanced truncation

We investigate discrete linear time-invariant systems of the form

$$\mathcal{S} : \begin{cases} x_{i+1} &= Ax_i + Bu_i, \\ y_i &= Cx_i, \end{cases} \tag{1}$$

with state $x_i \in \mathbb{R}^n$, input $u_i \in \mathbb{R}^m$, output $y_i \in \mathbb{R}^p$, system matrix $A \in \mathbb{R}^{n\times n}$, input matrix $B \in \mathbb{R}^{n\times m}$, output matrix $C \in \mathbb{R}^{p\times n}$ and zero initial state $x_0 = 0$. A good and compact way to describe the input-output behavior of the system can be achieved by applying the

$\mathcal{Z}$-transform to the system (1):

$$\begin{aligned} zX(z) &= AX(z) + BU(z), \\ Y(z) &= CX(z), \end{aligned} \tag{2}$$

where $X(z), U(z), Y(z)$ are the $\mathcal{Z}$-transforms of x_i, u_i, y_i, respectively. Rewriting (2) we obtain

$$Y(z) = \big(C(zI - A)^{-1}B\big)\, U(z). \tag{3}$$

2.1 Definition
For a discrete linear system $\mathcal{S}$ of the form (1) the function

$$G(z) := C(zI - A)^{-1}B \tag{4}$$

is known as the transfer function.

Equation (3) shows that the transfer function relates inputs to outputs in frequency domain. In the following we consider discrete linear systems which are in general *unstable*.

2.2 Definition
A discrete linear system $\mathcal{S}$ of the form (1) is called asymptotically stable *if all eigenvalues of the system matrix A lie inside the unit disk $D := \{x \in \mathbb{C} \mid |x| < 1\}$.*

The dimension n of the system matrix A is known as the *order* of the dynamical system (1). We consider problems where the order is typically very large. Techniques to reduce the order of the system are indispensable. The main idea of model reduction methods is to approximate the system (1) by a system of much smaller order $k \ll n$:

$$\hat{\mathcal{S}} : \left\{ \begin{aligned} \hat{x}_{i+1} &= \hat{A}\hat{x}_i + \hat{B}u_i, \\ \hat{y}_i &= \hat{C}\hat{x}_i, \end{aligned} \right. \tag{5}$$

with reduced state $\hat{x}_i \in \mathbb{R}^k$, input $u_i \in \mathbb{R}^m$, output $\hat{y}_i \in \mathbb{R}^p$, reduced system matrix $\hat{A} \in \mathbb{R}^{k\times k}$, reduced input matrix $\hat{B} \in \mathbb{R}^{k\times m}$ and reduced output matrix $\hat{C} \in \mathbb{R}^{p\times k}$. The aim of model reduction is to find a low order system $\hat{\mathcal{S}}$ of order $k \ll n$ such that the response $\hat{y}_i$ of $\hat{\mathcal{S}}$ is as close as possible to the response y_i of the full order system $\mathcal{S}$.

One approach to finding a low order system $\hat{\mathcal{S}}$ that approximates the input-output behavior of the full order system $\mathcal{S}$ is to minimize the distance between the transfer functions of the full and low order system in a suitable norm:

$$\|G - \hat{G}\| = \min! \tag{6}$$

where G and $\hat{G}$ are the transfer functions of $\mathcal{S}$ and $\hat{\mathcal{S}}$, respectively.

The minimization of (6) will also assure that the outputs of the low order system are not too far from the outputs of the full order system due to the following relation between inputs and outputs in frequency domain:

$$\|Y(z) - \hat{Y}(z)\| = \| \Big(G(z) - \hat{G}(z)\Big)\, U(z)\| \leq \|G(z) - \hat{G}(z)\|\|U(z)\|, \tag{7}$$

where $Y(z)$, $\hat{Y}(z)$ and $U(z)$ are the $\mathcal{Z}$-transforms of y_i, $\hat{y}_i$ and u_i, respectively.

Before specifying a suitable norm for the minimization (6) we first have a closer look at the output. The output of the system (1) after ℓ time steps is given by:

$$y_\ell = CA^\ell \sum_{j=1}^{\ell} A^{-j}Bu_{j-1}.$$

This leads to the following description of the output in frequency domain:

$$\begin{aligned} Y(z) &= \sum_{\ell=0}^{\infty} y_\ell \, z^{-\ell} \\ &= C \sum_{\ell=0}^{\infty} \sum_{j=1}^{\ell} z^{-\ell} A^{\ell-j} B u_{j-1}. \end{aligned} \tag{8}$$

Thus, we see that $Y(z)$ is only a finite number if the absolute value of z is larger than the largest eigenvalue of A in absolute value. As a consequence, the inequality (7) is only well-defined for $|z| > \alpha$ where α is an upper bound for the largest eigenvalue of A in absolute value. This observation has to be taken into account when choosing an appropriate norm for the minimization (6).

For asymptotically stable systems $\alpha = 1$ is an upper bound for the absolute value of all eigenvalues. This justifies the use of the h_2-norm as defined in [9, 1]:

2.3 Definition
We consider the space

$$M^{(q,s)} := \{F : \mathcal{D}^C \to \mathbb{C}^{q \times s} \mid \ F \textit{ is holomorphic in } \mathcal{D}^C\}$$

where $\mathcal{D}^C$ denotes the complement of the closed unit circle. For any element $F \in M^{(q,s)}$ the corresponding h_2-norm is defined as:

$$\|F\|_{h_2} := \left(\frac{1}{2\pi} \sup_{|r|>1} \int_0^{2\pi} \textit{trace} \, \left[F^*(re^{-\imath\theta}) F(re^{\imath\theta}) \right] \, d\theta \right)^{\frac{1}{2}}.$$

It is crucial for this norm to be well-defined that the supremum is only considered over radii r which have an absolute value that is larger than all eigenvalues of A in absolute value. For asymptotically stable systems this is always fulfilled because all eigenvalues are smaller than one in absolute value. We will see in Section 3 how this insight motivates the use of a generalized h_2-norm when considering unstable systems.

We now focus on model reduction methods for asymptotically stable systems that minimize the difference between the transfer functions of the full and the low order model with respect to the h_2-norm:

$$\|G - \hat{G}\|_{h_2} = \min! \tag{9}$$

There already exist several approaches for computing a reduced order system $\hat{\mathcal{S}}$ such that (9) is minimized. Necessary conditions for such a minimum are established in [4]. It is not practicable to find the optimal reduced model matrices that satisfy these conditions, however, as large systems of nonlinear equations must be solved. Instead we concentrate on the method of balanced truncation - an approved technique for model reduction of asymptotically stable linear systems. Its main idea is to truncate the states of the system that are least influenced by the inputs and have least effects on the outputs. This is only possible if the system has been transformed to balanced form first. The balanced truncation method then computes a reduced order system which is close to being optimal in the sense that the h_2-norm difference of the transfer functions (9) is approximately minimized [4].

The response of a discrete linear system is represented by its Hankel matrix. Balanced truncation computes the reduced order system $\hat{\mathcal{S}}$ in such a way that the Hankel singular values of the full linear model are retained. We refer to [3, 18] for more computational details. A main advantage of this model reduction technique is that there exists a global bound for the error between the transfer function of the original and the low order model.

2.4 Theorem:
Let $\mathcal{S}$ be a system of the form (1) with corresponding transfer function G. Moreover, let $\hat{\mathcal{S}}$ with corresponding transfer function $\hat{G}$ be a reduced system of the form (5) with order $k < n$ that is computed using balanced truncation. Then the following bound for the error system holds:

$$\|G - \hat{G}\|_{h_\infty} \leq 2(\sigma_{r+1} + \ldots + \sigma_n), \tag{10}$$

where σ_i are the Hankel singular values of the original system. The h_∞-norm is defined as

$$\|G\|_{h_\infty} := \sup_{\theta \in [0, 2\pi]} \sigma_{max}\left(G(e^{i\theta})\right),$$

where σ_{max} denoted the largest singular value.

Proof: We refer to [9]. □

Balanced truncation requires the linear system (1) to be asymptotically stable. Otherwise the system cannot be transformed to balanced form. However, as briefly stated in the introduction, there exist extensions to unstable systems. They are based on an additive decomposition of the system into its asymptotically stable and its unstable part:

$$G = G_+ + G_-\,,$$

where G_+ and G_- are the transfer function of an asymptotically stable and an unstable subsystem, respectively. Once this additive stable-unstable decomposition of G is found then the original balanced truncation technique for asymptotically stable systems can be applied to G_+. In this procedure the unstable part G_- remains unchanged. Finally, the reduced stable part is recomposed with the unchanged unstable part. In general, this model reduction procedure for unstable systems can only work well if the system has a small number of unstable poles. The attempt to reduce the order of the system to an order smaller than the number of unstable poles leads to a low order system that only keeps a part of the unstable subsystem G_-. It is not even assured that at least the most dominant part of G_- should be kept. Additionally, the asymptotically stable part is ignored completely. For further details on this method we refer to [7, 15, 10, 19].

The following section proposes a new balanced truncation approach for unstable systems which takes into account the asymptotically stable as well as the unstable part of the full order system.

3 Balanced truncation for unstable α-bounded systems

An important property of balanced truncation for asymptotically stable systems is that it computes a low order system such that the h_2-norm difference of the transfer functions of the full and the reduced order systems (9) is close to being optimal. As the h_2-norm is only defined for asymptotically stable systems we cannot aim to get the same result when considering unstable systems. However, we are able to derive a similar property for our new method for unstable systems. As mentioned in the previous section the common h_p-norms are only well-defined if all eigenvalues of the system matrix lie inside the unit circle. Moreover, we have shown that using the inequality (7) as a basis for the approximation of the original system (1) is only reasonable for $|z| > \alpha$, where α is an upper bound for the largest eigenvalue in absolute value. This insight motivates a natural generalization of standard h_p-norms to unstable systems as proposed in [9].

3.1 Definition ($h_{p,\alpha}$-norms)
Let α be a real positive number. For any element

$$F \in \mathcal{M}_\alpha^{(p,m)} := \{F : \bar{\mathcal{D}}_\alpha^C \to \mathbb{C}^{p\times m} | F \textit{ is holomorphic in } \bar{\mathcal{D}}_\alpha^C\},$$

where $\bar{\mathcal{D}}_\alpha^C$ is the complement of the closed circle around the origin with radius α, the corresponding $h_{2,\alpha}$- and $h_{\infty,\alpha}$-norms are defined as:

$$\begin{aligned} \|F\|_{h_{2,\alpha}} &:= \left(\frac{1}{2\pi} \sup_{|r|>\alpha} \int_0^{2\pi} \textit{trace } \left[F^*(re^{-\imath\theta}) F(re^{\imath\theta}) \right] \, d\theta \right)^{\frac{1}{2}} \\ &= \left(\frac{1}{2\pi} \int_0^{2\pi} \textit{trace } \left[F^*(\alpha e^{-\imath\theta}) F(\alpha e^{\imath\theta}) \right] \, d\theta \right)^{\frac{1}{2}} \end{aligned}$$

and

$$\begin{aligned} \|F\|_{h_{\infty,\alpha}} &:= \sup_{z\in\bar{\mathcal{D}}_\alpha^C} \sigma_{max}\left(F(z)\right) \\ &= \sup_{\theta\in[0,2\pi]} \sigma_{max}\left(F(\alpha e^{\imath\theta})\right), \end{aligned}$$

where σ_{max} denotes the largest singular value.

We note that the special case of the $h_{p,\alpha}$-norm where α is equal to one supplies the standard h_p-norm. The main advantage of the $h_{p,\alpha}$-norm is that it is well-defined for unstable systems if the value of α is chosen such that all eigenvalues of the system matrix A of the system (1) lie inside a disk around the origin with radius α.

3.2 Definition (α-boundedness)
Let $\alpha \in \mathbb{R}$ be a positive number. Then a discrete control system $\mathcal{S}$ of the form (1) is called α-bounded *if all eigenvalues of the system matrix A lie inside a disk around the origin with radius α, i.e.*

$$\lambda \textit{ eigenvalue of } A \Rightarrow \lambda \in D_\alpha$$

with $D_\alpha := \{x \in \mathbb{C} \mid |x| < \alpha\}$.

We note that for $\alpha = 1$ the concept of α-boundedness is equivalent to asymptotic stability. For a regular discrete (in general) unstable system of the form (1) it is always possible to find real positive numbers α such that the system is α-bounded. In general α-bounded systems are not asymptotically stable. Thus, the standard h_p-norm is not well-defined, but the $h_{p,\alpha}$-norm is.

Using this generalized norm for unstable systems, we now derive a new *α-bounded balanced truncation method.* To determine a suitable α, a rough knowledge of the eigenstructure of the system matrix A is needed. This can be achieved using a simple iterative method for computing the largest eigenvalue in absolute value, such as the Arnoldi method, or using the concept of Gershgorin circles, see e.g. [8, 17, 2, 6]. Once a suitable α is determined the following shift of the original system is considered.

3.3 Lemma
For any linear discrete, reachable and observable α-bounded system $\mathcal{S}$ of the form (1) we consider the shifted system

$$\mathcal{S}_\alpha : \left\{ \begin{array}{rcl} x_{i+1}^{(\alpha)} &=& A_\alpha x_i^{(\alpha)} + B_\alpha u_i, \\ y_i^{(\alpha)} &=& C_\alpha x_i^{(\alpha)}, \end{array} \right. \tag{11}$$

with $A_\alpha := A/\alpha$, $B_\alpha := B/\sqrt{\alpha}$ and $C_\alpha := C/\sqrt{\alpha}$. Let G and G_α be the corresponding transfer functions of (1) and (11), respectively. Then the following properties hold:

(i) $\mathcal{S}_\alpha$ is asymptotically stable.

(ii) The h_2-norm of $\mathcal{S}_\alpha$ is equal to the $h_{2,\alpha}-$norm of $\mathcal{S}$:

$$\|G_\alpha\|_{h_2} = \|G\|_{h_{2,\alpha}}.$$

(iii) The $h_\infty-$norm of $\mathcal{S}_\alpha$ is equal to the $h_{\infty,\alpha}-$norm of $\mathcal{S}$:

$$\|G_\alpha\|_{h_\infty} = \|G\|_{h_{\infty,\alpha}}.$$

Proof:

(i) It is a well-known result from linear algebra that the eigenvalues of the matrix A_α are the eigenvalues of A divided by α. This implies the statement.

(ii) It holds that

$$\begin{aligned} G_\alpha(e^{\imath\theta}) &= C_\alpha \left(e^{-\imath\theta} I - A_\alpha\right)^{-1} B_\alpha \\ &= \frac{C}{\sqrt{\alpha}} \left(e^{-\imath\theta} I - \frac{A}{\alpha}\right)^{-1} \frac{B}{\sqrt{\alpha}} \\ &= \frac{C}{\sqrt{\alpha}} \left(\frac{1}{\sqrt{\alpha}}(\alpha e^{-\imath\theta} I - A)\frac{1}{\sqrt{\alpha}}\right)^{-1} \frac{B}{\sqrt{\alpha}} \\ &= C(\alpha e^{-\imath\theta} I - A)^{-1} B \\ &= G(\alpha e^{\imath\theta}), \end{aligned}$$

and thus,

$$\begin{aligned} \|G_\alpha\|_{h_2} &= \left(\frac{1}{2\pi}\int_0^{2\pi} \text{trace } \left[G_\alpha^*(e^{-\imath\theta})G_\alpha(e^{\imath\theta})\right] \, d\theta\right)^{\frac{1}{2}} \\ &= \left(\frac{1}{2\pi}\int_0^{2\pi} \text{trace } \left[G^*(\alpha e^{-\imath\theta})G(\alpha e^{\imath\theta})\right] \, d\theta\right)^{\frac{1}{2}} \\ &= \|G\|_{h_{2,\alpha}}. \end{aligned}$$

(iii) Then it also holds that

$$\begin{aligned} \|G_\alpha\|_{h_\infty} &= \sup_{\theta\in[0,2\pi]} \sigma_{max}\left(G_\alpha(e^{i\theta})\right) \\ &= \sup_{\theta\in[0,2\pi]} \sigma_{max}\left(G(\alpha e^{i\theta})\right) \\ &= \|G\|_{h_{\infty,\alpha}}, \end{aligned}$$

where σ_{max} denotes the largest singular value.

□

Given a discrete linear system $\mathcal{S}$ of the form (1) our new balanced truncation approach for unstable systems can be stated as follows:

3.4 Algorithm (α-bounded balanced truncation)

(I) *Determine a suitable real positive α such that the system $\mathcal{S}$ is α-bounded.*

(II) *Shift the α-bounded system $\mathcal{S}$ to its asymptotically stable form $\mathcal{S}_\alpha$ as described in Lemma 3.3.*

(III) *Apply the original balanced truncation method for asymptotically stable systems to the shifted system $\mathcal{S}_\alpha$ which is asymptotically stable. This supplies the reduced system*

$$\hat{\mathcal{S}}_\alpha : \begin{cases} \hat{x}_{i+1}^{(\alpha)} &= \hat{A}_\alpha \hat{x}_i^{(\alpha)} + \hat{B}_\alpha u_i, \\ \hat{y}_i^{(\alpha)} &= \hat{C}_\alpha \hat{x}_i^{(\alpha)}. \end{cases} \tag{12}$$

(IV) *Shift the reduced system back:*

$$\hat{\mathcal{S}} : \begin{cases} \hat{x}_{i+1} &= \hat{A}\hat{x}_i + \hat{B}u_i, \\ \hat{y}_i &= \hat{C}\hat{x}_i, \end{cases} \tag{13}$$

with $\hat{A} := \alpha\hat{A}_\alpha$, $\hat{B} := \sqrt{\alpha}\hat{B}_\alpha$ and $\hat{C} := \sqrt{\alpha}\hat{C}_\alpha$.

Lemma 3.3 shows that the balanced truncation method for α-bounded systems supplies an approximation that is close to being optimal in the $h_{2,\alpha}-$norm. Thus, this new technique provides a good approach for extending standard model reduction methods for asymptotically stable systems to unstable systems. In the following theorem we derive an explicit error bound for α-bounded balanced truncation.

3.5 Theorem:
Let $\mathcal{S}$ be an α-bounded system of the form (1) with corresponding transfer function G. Moreover, let $\hat{\mathcal{S}}$ with corresponding transfer function $\hat{G}$ be a reduced order system of order $k < n$ that is computed using α-bounded balanced truncation as stated in Algorithm 3.4. Then the following bound for the error system holds:

$$\|G - \hat{G}\|_{h_{\infty,\alpha}} \leq 2\left(\sigma_{r+1}^{(\alpha)} + \ldots + \sigma_n^{(\alpha)}\right),$$

where $\sigma_{r+1}^{(\alpha)}, \ldots, \sigma_n^{(\alpha)}$ are the neglected Hankel Singular values of the α-shifted system (11).

Proof: Let G_e be the transfer function of the error system

$$\mathcal{S}_\mathbf{e} : \begin{cases} \begin{bmatrix} x_{i+1} \\ \hat{x}_{i+1} \end{bmatrix} &= \underbrace{\begin{bmatrix} A & 0 \\ 0 & \hat{A} \end{bmatrix}}_{=:A_e} \begin{bmatrix} x_i \\ \hat{x}_i \end{bmatrix} + \underbrace{\begin{bmatrix} B \\ \hat{B} \end{bmatrix}}_{=:B_e} u_i, \\ \begin{bmatrix} y_i & \hat{y}_i \end{bmatrix} &= \underbrace{\begin{bmatrix} C & -\hat{C} \end{bmatrix}}_{=:C_e} \begin{bmatrix} x_i \\ \hat{x}_i \end{bmatrix}. \end{cases} \tag{14}$$

By definition it holds that

$$\|G - \hat{G}\|_{h_{\infty,\alpha}} = \|G_e\|_{h_{\infty,\alpha}}.$$

Using Lemma 3.3 we obtain

$$\|G_e\|_{h_{\infty,\alpha}} = \|G_{e,\alpha}\|_{h_\infty},$$

where $G_{e,\alpha} = \frac{1}{\sqrt{\alpha}} C_e \left(zI - \frac{1}{\alpha}A_e\right)^{-1} \frac{1}{\sqrt{\alpha}} B_e$ is the transfer function of the α-shifted error system. By definition then

$$\|G_{e,\alpha}\|_{h_\infty} = \|G_\alpha - \hat{G}_\alpha\|_{h_\infty},$$

where G_α, $\hat{G}_\alpha$ are the transfer functions of the α-shifted systems $\mathcal{S}_\alpha$, $\hat{\mathcal{S}}_\alpha$, respectively. Because the systems $\mathcal{S}_\alpha$ and $\hat{\mathcal{S}}_\alpha$ are asymptotically stable and $\hat{\mathcal{S}}_\alpha$ is the result of applying balanced truncation to $\mathcal{S}_\alpha$ the error bound (10) holds. Therefore:

$$\|G_\alpha - \hat{G}_\alpha\|_{h_\infty} \leq 2\left(\sigma_{r+1}^{(\alpha)} + \ldots + \sigma_n^{(\alpha)}\right),$$

where $\sigma_{r+1}^{(\alpha)}, \ldots, \sigma_n^{(\alpha)}$ are the Hankel singular values of G_α.
Then the statement of the theorem follows with $\|G_\alpha - \hat{G}_\alpha\|_{h_\infty} = \|G - \hat{G}\|_{h_{\infty,\alpha}}$. □

To summarize, we state that our new technique for balanced truncation of unstable systems computes a low order system that approximates the full order system well. It is close to being optimal with respect to the $h_{2,\alpha}$-norm and the error in the $h_{\infty,\alpha}$-norm is bounded by twice the sum of the neglected Hankel singular values of the α-shifted system.

In the following section we compare our new α-bounded balanced truncation method with the commonly used approach for treating unstable systems.

4 Numerical experiments

We now perform numerical experiments to illustrate the benefit of the new α-bounded model reduction method in comparison with the standard balanced truncation approach for unstable systems. For these experiments we consider three different unstable discrete linear test models. In Subsections 4.1 and 4.2 the focus is on two simple discrete systems of the form

$$\mathcal{S}^{(\mathbf{k})} : \left\{ \begin{array}{lcl} x_{i+1} & = & A^{(k)} x_i + B^{(k)} u_i, \\ y_i & = & C^{(k)} x_i, \end{array} \right. \quad \text{for } k \in \{1, 2\} \tag{15}$$

with zero initial states $x_0 = 0$. The simplicity provides a direct insight into the dynamics of the system. A more realistic test model derived from discretized shallow water equations is then investigated in Subsection 4.3. It is an approved test model within meteorology because it retains key properties of the model equations used by operational weather forecasting centers.

4.1 First simple test model

The first test model $\mathcal{S}^{(\mathbf{1})}$ is chosen to be a multiple-input, single-output system of the form (15), with a real diagonal matrix $A^{(1)} = \text{diag}\{\lambda_1, \ldots, \lambda_{30}\}$ of dimension 30 times 30. The input matrix $B^{(1)} \in \mathbb{R}^{30\times 30}$ is the identity matrix and the output matrix $C^{(1)} \in \mathbb{R}^{1\times 30}$ is a row vector which contains only ones. The eigenvalues λ_i, $i = 1 \ldots 30$, of $A^{(1)}$ are all real and lie inside as well as outside the unit circle. The distribution of the eigenvalues is shown in Figure 1. We note that a considerable part of the system is unstable: 17 eigenvalues lie outside the unit circle (see Appendix A.1.1 for the eigenvalues of $A^{(1)}$).

We have chosen this rather simple test model because it reveals the relation between inputs and outputs in an obvious way. If we choose the input u_i as the j-th unit impulse, i.e.

$$u_i = \left\{ \begin{array}{ll} e_j & \text{for } i = 0, \\ 0 & \text{for all } i > 0, \end{array} \right.$$

where e_j is the j-th canonical unit vector, then the state and the output at time $t_i > 0$ are given by

$$\begin{array}{lcl} x_i & = & \lambda_j^{i-1} e_j, \\ y_i & = & \lambda_j^{i-1}, \end{array}$$

respectively. Thus, the impulse response y_i is a power of the eigenvalue λ_j (the j-th diagonal entry of the system matrix $A^{(1)}$). The state vector x_i only has components in the direction of the corresponding j-th eigenvector e_j.

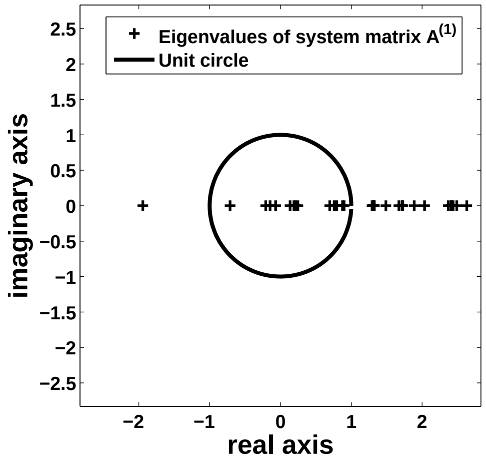


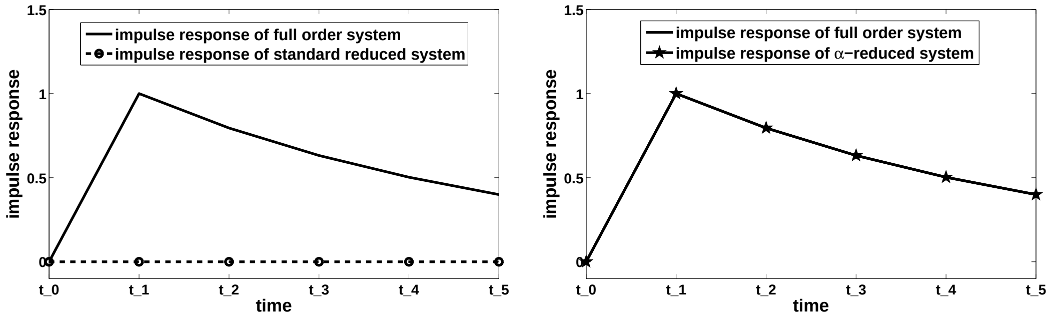
Figure 1: Eigenvalues of system matrix $A^{(1)}$ of first simple test model

In the numerical experiments we consider a time window $[t_0, t_N]$ which consists of five to 20 time steps. Such a relatively small time window is chosen because this is the interesting (transient) period in the case of unstable systems. In many applications unstable discrete systems are derived from nonlinear systems by linearization. To obtain a good approximation to the full nonlinear system it is essential to repeat the linearization process every few time steps. Thus, only a small to medium size time window of the linearized system is generally of interest.

The aim of this numerical section is to compare the new α-bounded balanced truncation approach (proposed in Algorithm 3.4) with the standard balanced truncation method for unstable systems (described in Section 2). For the numerical computation of stable-unstable decompositions and of balanced realizations of asymptotically stable systems we use the MATLAB routines `stabsep.m` and `balreal.m`, respectively, as implemented in the Control Toolbox of MATLAB Release R2009a [13].

We now investigate the impulse responses of the full and the reduced order systems computed by the two different model reduction methods. Our model $\mathcal{S}^{(1)}$ is a multiple-input, single-output system with 30 input channels. For such a system the impulse response at time t_i is a matrix of outputs. The j-th column of this matrix contains the response of the system at time t_i to an input vector that is the j-th unit impulse $\delta_j := \delta(t)e_j$. Thus, the impulse response consists of 30 different components. For each of these we investigate the approximation to the output of the full order system by the output of the low order systems computed by standard balanced truncation and by our new α-bounded approach.

Figure 2 shows the outputs of the first impulse response, i.e. the outputs of the systems where the input is the first unit impulse $\delta_1 = \delta(t)e_1$, over the time window $[t_0, t_5]$. In Figure 2(a) we see the approximation of the output of the full order system (solid line) by the output of the low order system of reduction order $k = 10$ computed by the standard balanced truncation method (dashed line with circles). In comparison, Figure 2(b) shows the output of the full order system (solid line) together with the output of the low order system of reduction order $k = 10$ computed by the α-bounded balanced truncation method for $\alpha = 12$ (solid line with stars). We note that the solid line with stars is nearly invisible in the latter case because it lies on top of the solid line. This shows that the output of the α-reduced system approximates the output of the full order system so well that the two output lines are indistinguishable. In contrast, the standard balanced truncation method computes an output that is zero at all time steps and therefore contains no information at all on the response of the full order system.



(a) Approximation using standard method (b) Approximation using α-bounded method

Figure 2: Comparison of first impulse responses of full and reduced systems of order $k = 10$ using standard balanced truncation (a) as well as α-bounded balanced truncation for $\alpha = 12.0$ (b)

Figure 3 shows the corresponding error plot in logarithmic scale over the time window $[t_1, t_5]$. For illustration purposes the initial time t_0 is omitted in the figure. This is reasonable because all outputs at the initial time t_0 (and thus also the output errors) are zero, no matter which low order model is considered. The dashed line with circle shows the error in the standard balanced truncation method. We see that its order of magnitude is 10^0. In comparison the error in the α-bounded method (solid line with stars) is of order of magnitude 10^{-12} to 10^{-15}.

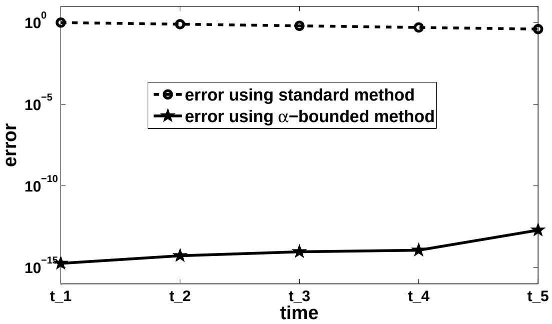
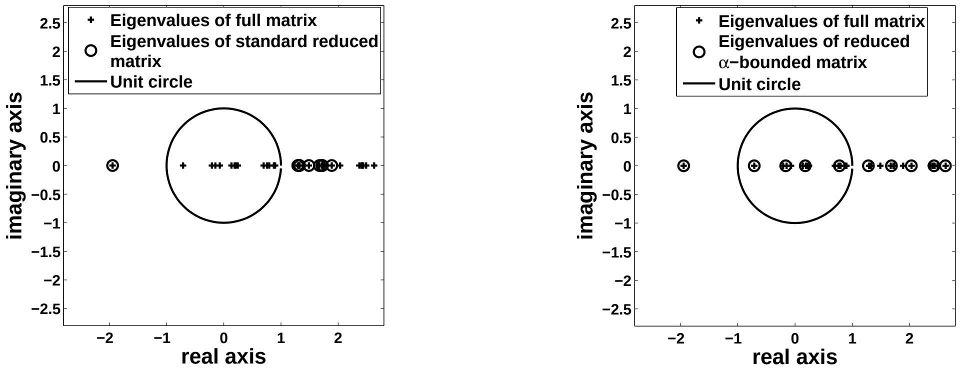


Figure 3: Comparison of errors (logarithmic scale) in the first impulse response of reduced system of order $k = 10$ using standard balanced truncation *(dashed line with circles)* and α-bounded balanced truncation for $\alpha = 12.0$ *(solid line with stars)*

To understand the reason that the new α-bounded method performs so much more accurately than the standard approach, we investigate the eigen-structure of the system matrices of the different low order systems. Figure 4 compares the eigenvalues of the full order system matrix (crosses) with those of the low order matrix computed by the standard method (Figure 4(a), circles) and those of the low order matrix computed by α-bounded balanced truncation (Figure 4(b), circles). We see that the α-bounded approach matches eigenvalues outside as well as inside the unit circle while the standard approach only keeps some of the eigenvalues outside the unit circle, but none inside.

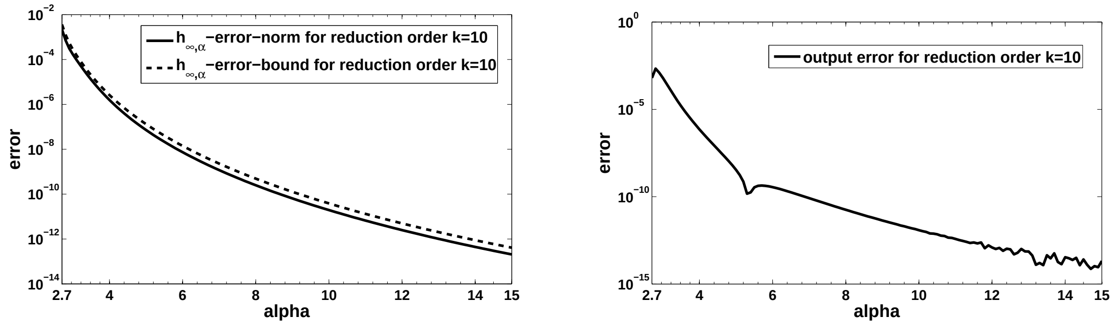
Thus, the failure of the standard method is not surprising. Because of the simple structure of this first test model we know that if the input vector u_i is chosen as the first unit impulse, then all state vectors x_i are multiples of the eigenvector e_1 associated with the eigenvalue $\lambda_1 \approx 0.8$. The reduced order model computed by the standard method neglects all directions of eigenvectors associated with asymptotically stable eigenvalues. Thus, the output of the standard low order system is not able to approximate the response of the full order system, which is a power of the asymptotically stable eigenvalue λ_1.



(a) Eigenvalues using standard method

(b) Eigenvalues using α-bounded method

Figure 4: Comparison of eigenvalues of full and reduced systems of order $k = 10$ using standard balanced truncation as well as α-bounded balanced truncation for $\alpha = 12.0$



(a) $h_{\infty,\alpha}$-error-norms and -bounds

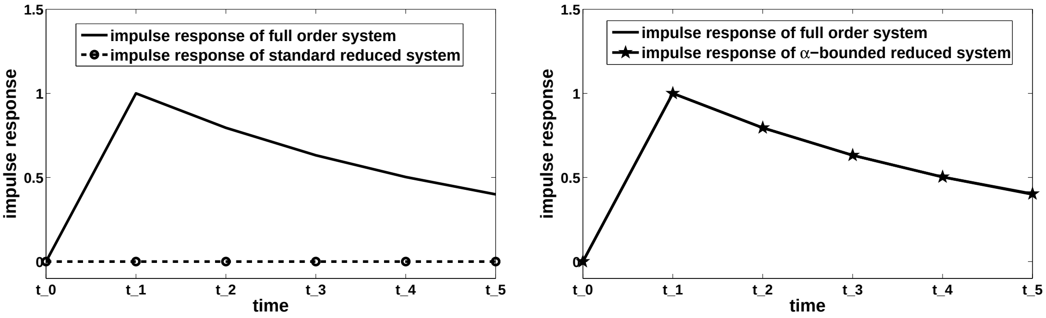
(b) Relative output errors

Figure 5: $h_{\infty,\alpha}$-error-norms and -bounds and relative output errors of the impulse response of α-bounded balanced truncation method of reduction order $k = 10$ for different values of α

As the α-bounded model reduction method is dependent on the variable α we investigate the effect of the choice of α on the quality of the approximation of the low order models. Figure 5 illustrates the change in the approximation error of α-bounded balanced truncation for reduction order $k = 10$ for different values of α. In Figure 5(a) it is shown that the $h_{\infty,\alpha}$-error-norm $\|\mathcal{S}^{(1)} - \hat{\mathcal{S}}^{(1)}\|_{h_{\infty,\alpha}}$ (solid line) decreases with increasing α. The dashed line is a plot of the theoretical error bound derived in Theorem 3.5. The figure validates the theoretical result that the actual $h_{\infty,\alpha}$-error-norm (solid line) is always below the error bound (dashed line). Figure 5(b) plots the behavior of the relative error norm e_{rel} of the first impulse output for different values of α. The relative error norm is defined as

$$e_{rel} := \frac{\|y - \hat{y}\|_2}{\|y\|_2},$$

where $y := [y_0, \ldots, y_5]$, $\hat{y} := [\hat{y}_0, \ldots, \hat{y}_5]$ are the vectors of outputs of the full and the low order systems, respectively, over the time window $[t_0, t_5]$. We see that the relative error is smaller than 10^{-3} for all values of α. As α increases it even falls below 10^{-12}. For all values of α the approximation to the output of the first impulse response computed by α-bounded balanced truncation is much more accurate than the standard reduction approach. However, to get a very good approximation with the α-bounded method it is recommended to choose α not too close to the largest eigenvalue in absolute value (which is approximately 2.63 in this test model).



(a) Approximation using standard method

(b) Approximation using α-bounded method

Figure 6: Comparison of first impulse responses of full and reduced systems of order $k = 5$ using standard balanced truncation (a) as well as α-bounded balanced truncation for $\alpha = 4.8$ (b)

We next test whether a further reduction to an order of $k = 5$ influences the quality of the approximations. Using the α-bounded method with $\alpha = 4.8$, we are able to reduce the system to a sixth of the order of the full order model while still capturing the most important information in the system response (see Figure 6(b)). In contrast, the standard approach computes a low order model that fails to approximate the impulse response of the full order system (see Figure 6(a)) for the same reason that the reduction to the larger order of $k = 10$ failed.

Figure 7 shows the corresponding error plot. As before, the initial time step is omitted for illustration purposes. We see that the error of the α-bounded method (solid line with stars) is two to eight orders of magnitude smaller than the error of the standard method (dashed line with circles).

We also examine the change of the error in α-bounded balanced truncation of reduction order $k = 5$ as a function of α. Figure 8(a) shows that the $h_{\infty,\alpha}$-norm decreases with increasing α (solid line) and that the actual $h_{\infty,\alpha}$-norm always stays below the theoretical error bound (dashed line). Thus, the theoretical result of Theorem 3.5 is validated numerically for reduction order $k = 5$, as well. For values of α varying between 2.7 and 15 the relative output error of the first impulse response for reduction order $k = 5$ has a minimum at $\alpha = 4.8$. The error norm is approximately of order 10^{-3} (see Figure 8(b)). This is a good result taking into account that we have reduced the order of the system from $n = 30$ to $k = 5$. Again the actual choice of α is not of major importance as long as we choose it not to be too close to the largest eigenvalue in absolute value.

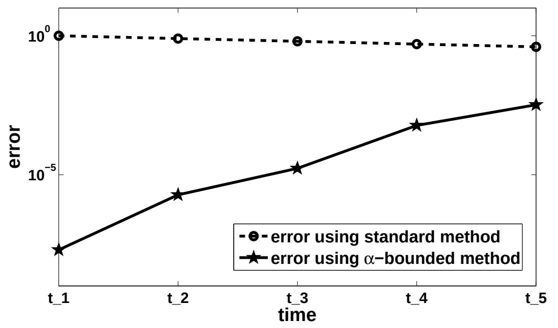
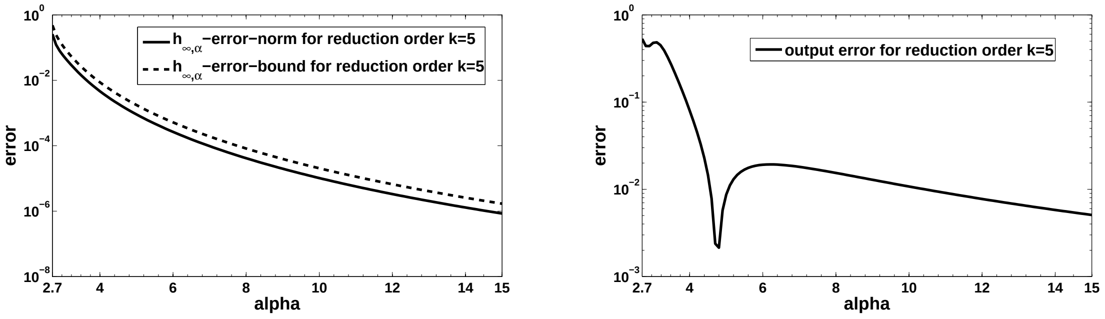


Figure 7: Comparison of errors (logarithmic scale) in the first impulse response of reduced system of order $k = 5$ using standard balanced truncation *(dashed line with circles)* and α-bounded balanced truncation for $\alpha = 4.8$ *(solid line with stars)*



(a) $h_{\infty,\alpha}$-error-norms and -bounds

(b) Relative output errors

Figure 8: $h_{\infty,\alpha}$-error-norms and -bounds and relative output errors of the impulse response of α-bounded balanced truncation method of reduction order $k = 5$ for different values of α

We have only considered the first impulse response thus far. This does not give a full picture of the behavior of a system with 30 input channels. All other 29 impulse responses have to be taken into account, as well.

The comparison of the relative error norms of all 30 components of the impulse response of the standard balanced truncation method with those of the α-bounded approach for reduction order $k = 10$ is summarized in Appendix A.1.2, Table 1. We see that the α-bounded approach supplies very good approximations to the outputs of the full order system for *all* components of the impulse response. The relative error has an order of magnitude between 10^{-12} and 10^{-14}. In contrast, the standard method does not supply good approximations for all impulse responses. In 20 of the 30 components of the impulse response the output does not approximate the output of the full order system at all. Only in 10 components do we obtain accurate results. In these cases the approximation is even slightly better than with the α-bounded method.

This can be explained as follows. The full order system has 13 asymptotically stable and 17 unstable poles. To be able to achieve a reduction order of $k = 10$ by using the standard approach the asymptotically stable part is truncated completely. The low order system is then defined as a subsystem of the unchanged unstable part of the full order model. This subsystem matches 10 of the unstable eigenvalues of the full order system, namely $\lambda_6, \lambda_8, \lambda_{10}, \lambda_{12}, \lambda_{14}, \lambda_{16}, \lambda_{20}, \lambda_{22}, \lambda_{28}$ and λ_{29}. Thus, whenever a component of the impulse response stimulates one of these 10 eigenvalues, then the standard approach will supply a low order system where the output matches the output of the full order system exactly (assuming the absence of rounding errors), see Appendix A.1.2, Table 1. For all remaining components of the impulse response (where none of these 10 eigenvalues is excited) the approximation obtained by the standard approach fails. Thus, considering the over all comparison of the two model reduction methods (including all input channels) we see the superiority of the α-bounded approach.

Table 2 (Appendix A.1.2) shows similar results for a reduction order of $k = 5$. Again the outputs of the low order system computed by the α-bounded method approximate the response of the full order system well for *all* impulse inputs. The relative output error lies between 10^{-2} and 10^{-5} for the responses to all unit impulse inputs. This is a good result taking into account that the order of the system is reduced from 30 to 5. In contrast, the standard method only supplies good approximations for 5 out of 30 impulse responses. Again these are exactly the 5 impulse responses which excite the 5 unstable modes which are matched by the standard method.

In these experiments we have only investigated a relatively small time window containing five time steps. We now conclude the investigation of the first test model $S^{(1)}$ by looking at a larger window which contains 20 time steps. In Figure 9 we see the relative error (in logarithmic scale) in the first impulse response using the standard approach for reduction order $k = 10$ (dashed line with circles) and using the α-bounded method for order $k = 10$ and $\alpha = 12$ (solid line with stars). The α-bounded approach performs very well for the first nine time steps with an approximation error lying between 10^{-5} and 10^{-15}. However, as time increases the error becomes larger, finally reaching an order of magnitude of 10^4 after 20 time steps. This reveals that the α-bounded method is especially designed to capture the behavior of the system at the initial time steps. In contrast, the error of the standard method is of order of magnitude 10^0 for the main part of the 20 time steps window becoming slightly more accurate towards the very end of the window.

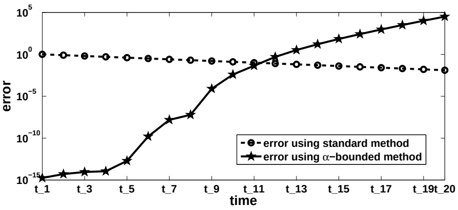


Figure 9: Comparison of errors (logarithmic scale) in the first impulse response of reduced system of order $k = 10$ using standard balanced truncation *(dashed line with circles)* and α-bounded balanced truncation for $\alpha = 12.0$ *(solid line with stars)* over a 20 time steps window

For capturing the transient motion of the system, the behavior of the reduced model over the initial time interval is most important. For applications where the unstable linear system is derived by linearizing a nonlinear model the first time steps are generally the most significant and for a good approximation of the original nonlinear system the linearization process has to be repeated every few time steps in any case. This is exactly where the strength of the α-bounded method lies: at the beginning of the time window the new approach is up to 15 orders of magnitude more accurate than the standard balanced truncation method.

We conclude the investigation of the first test model $\mathcal{S}^{(1)}$ by pointing out that the new α-bounded balanced truncation method supplies much better approximations to the input-output behavior of the full order system than the standard balanced truncation approach for unstable systems (as long as the time window is not chosen to be too large). The new method enables a reduction up to an order $k = 5$ while still capturing the most important information for all channels of the impulse response.

4.2 Second simple test model

The second test model $\mathcal{S}^{(\mathbf{2})}$ is chosen to be a single-input, single-output (SISO) system of the form (15), i.e. the input and the output matrices are a column and a row vector, respectively. The system matrix $A^{(2)} \in \mathbb{R}^{30\times 30}$ is a real dense matrix that has real and complex eigenvalues inside as well as outside the unit circle. The input matrix $B^{(2)} \in \mathbb{R}^{30\times 1}$ is the first canonical unit vector and the output matrix $C^{(2)} \in \mathbb{R}^{1\times 30}$ is, as in the previous example, a row vector which only contains ones. The distribution of the eigenvalues of $A^{(2)}$ is shown in Figure 10 (see also Appendix A.2).

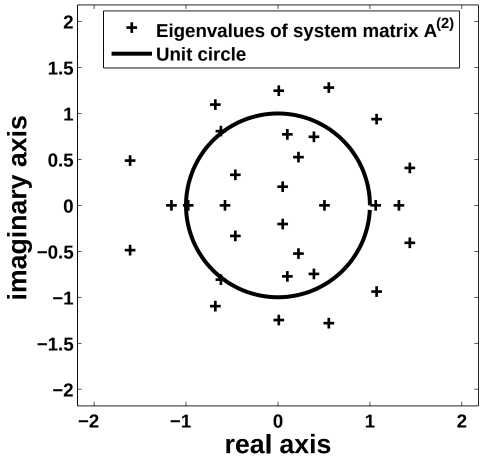
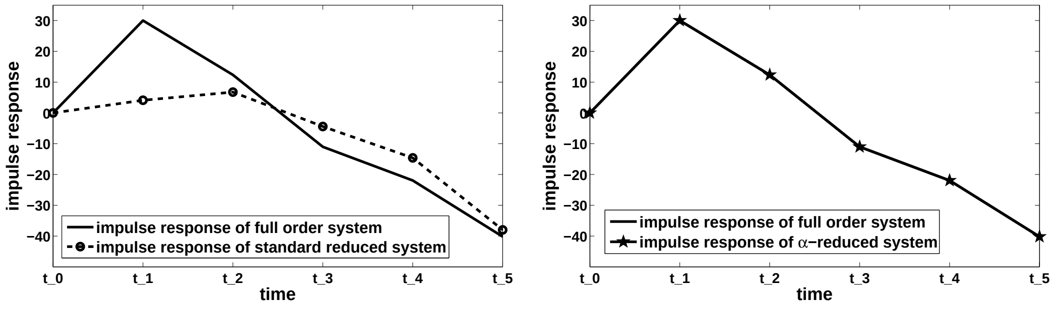


Figure 10: Eigenvalues of system matrix $A^{(2)}$ of the second simple test model

As for the previous test model we analyze the input-output behavior of the different low order systems by comparing the impulse responses. Here, the impulse response at time t_i is only a scalar because $\mathcal{S}^{(2)}$ is a SISO system.

In Figure 11 the impulse response of the full and low order order systems is plotted over the time window $[t_0, t_5]$. Figure 11(a) shows the impulse response of the full order system (solid line) and its approximation by the reduced system of order $k = 10$ computed by standard balanced truncation (dashed line with circles). Figure 11(b) shows the impulse response of the full order system (solid line) and its approximation by the reduced system of order $k = 10$ computed by α-bounded balanced truncation for $\alpha = 4.0$ (solid line with stars). We note that in the latter case the solid line and the solid line with stars lie on top of each other. While the approximation to the outputs by the new α-bounded approach is hardly distinguishable from the outputs of the original system, the outputs computed by the standard method are quite far away from the actual outputs of the full order system.

In Figure 12 we see that the error of the α-bounded method (solid line with stars) is on average about 14 orders of magnitude smaller than the error of the standard method (dashed line with circles). Thus, the clear superiority of the new α-bounded method also holds for the second test model. Figure 13 shows which eigenvalues of the full order model matrix are kept by the two different model reduction techniques. The standard balanced truncation method is capable of matching some of the eigenvalues outide the unit circle but none inside (Figure 13(a)) while the α-bounded approach also matches (approximately) an eigenvalue inside the unit circle (Figure 13(a)). This explains why the standard method cannot supply very accurate approximations of an output that is composed of a linear combination of both stable and unstable modes.



(a) Approximation using standard method

(b) Approximation using α-bounded method

Figure 11: Comparison of impulse responses of full and reduced systems of order $k = 10$ using standard balanced truncation (a) as well as α-bounded balanced truncation for $\alpha = 4.0$ (b)

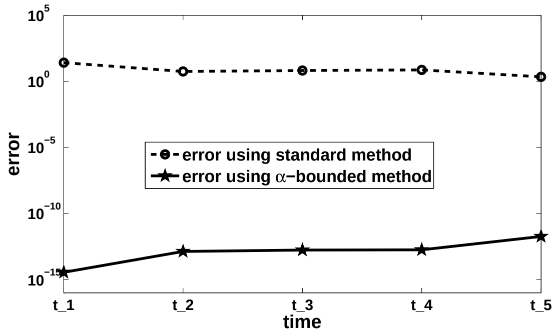
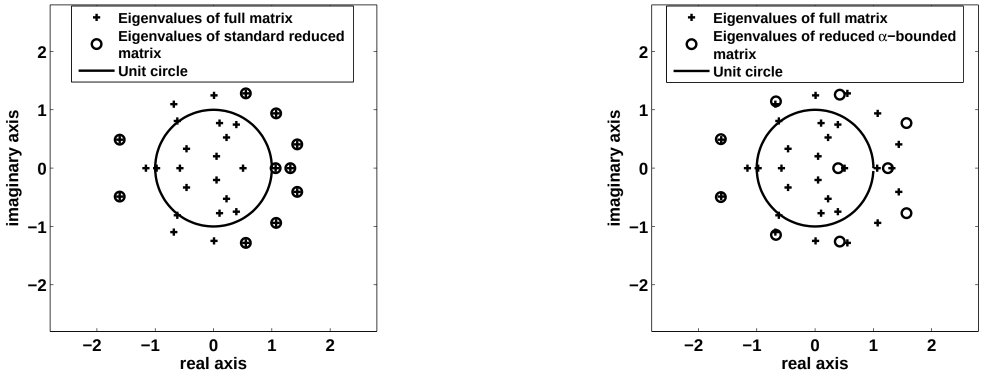


Figure 12: Comparison of errors (logarithmic scale) in the first impulse response of reduced system of order $k = 10$ using standard balanced truncation *(dashed line with circles)* and α-bounded balanced truncation for $\alpha = 4.0$ *(solid line with stars)*

We also examine the effect of different choices of α on the error norms of the low order approximations. Figure 14(a) illustrates that the actual $h_{\infty,\alpha}$-norm of the error system (solid line) is always smaller than the computed theoretical error bound (dashed line). In Figure 14(b) we see a plot of the relative error e_{rel} of the impulse response for different values of α. As long as α is chosen to be not too close to the largest eigenvalue of the system matrix in absolute value (which is approximately 1.68), then the output error e_{rel} becomes very small. For $\alpha > 3.5$ it even has the order of magnitude of the machine precision.

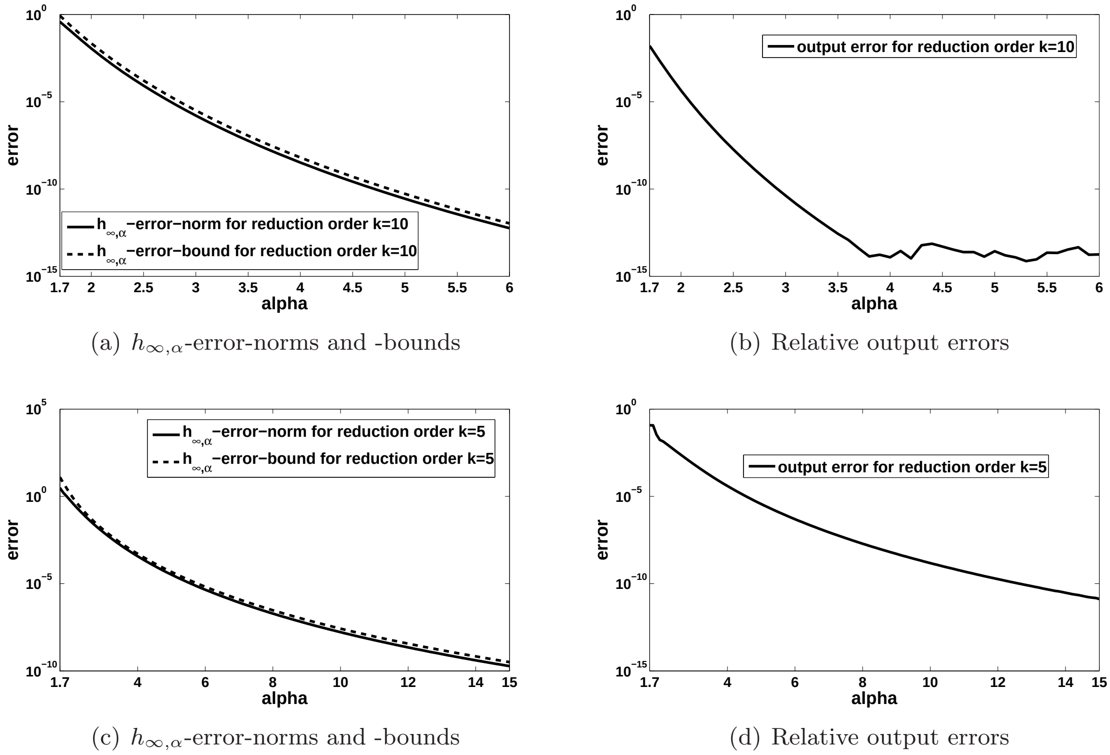
As for the first test model $\mathcal{S}^{(1)}$, we now investigate whether a further order reduction of the low order model is possible. We find that, using the α-bounded approach, a reduction up to a sixth of the order of the original system still supplies a good approximation. The most essential information of the input-output behavior is retained. Computing a low order system of order $k = 5$ using the α-bounded method, we obtain outputs that accurately approximate the outputs of the full order systems (see Figures 14(c), 14(d)). We note again that the actual choice of α is not significant as long as it is not too close to the largest eigenvalue in absolute value.



(a) Eigenvalues using standard method

(b) Eigenvalues using α-bounded method

Figure 13: Comparison of eigenvalues of full and reduced systems of order $k = 10$



(a) $h_{\infty,\alpha}$-error-norms and -bounds

(b) Relative output errors

(c) $h_{\infty,\alpha}$-error-norms and -bounds

(d) Relative output errors

Figure 14: $h_{\infty,\alpha}$-error-norms and -bounds and relative errors of impulse response of α-bounded balanced truncation method of reduction order $k = 10$, $k = 5$ for different α

4.3 Shallow water model

In addition to the two simple models we investigate, as a third test model $\mathcal{S}^{(\mathbf{3})}$, a 1-dimensional shallow water system which describes the flow of a fluid over an obstacle with rotation. The corresponding continuous shallow water equations are given by

$$\begin{aligned}
\frac{Du}{Dt} + \frac{\partial \phi}{\partial x} + g\frac{\partial \tilde{H}}{\partial x} - fv &= 0,\\
\frac{Dv}{Dt} + fu &= 0,\\
\frac{D \ln \phi}{Dt} + \frac{\partial u}{\partial x} &= 0,
\end{aligned}$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + (U_c + u)\frac{\partial}{\partial x}$$

and

$$\phi = gh,$$

where u denotes the departure of the velocity in the x-direction from a known constant forcing mean flow U_c, $\tilde{H} = \tilde{H}(x)$ is the height of the orography, f is the Coriolis parameter and g is the gravitational force. The model assumes that velocities u and v as well as the depth h do not vary in the y-direction. Moreover, the model states are periodic in the x-direction. The continuous equations are discretized using a two-time-level semi-implicit semi-Lagrangian integration scheme, following [11]. The discrete nonlinear system is then linearized by computing the Jacobian of the nonlinear system equations. The resulting discrete linear system is known as the *tangent linear model*.

A time-invariant linear model that approximates the tangent linear model of the system is used in the experiments. It is a multiple-input, multiple-output (MIMO) system. Its system matrix $A^{(3)}$ and its input matrix $B^{(3)}$ are both of dimension 1500 × 1500. The output matrix $C^{(3)} \in \mathbb{R}^{750\times1500}$ is chosen such that every other point is observed. We refer to the first, second and third set of 500 components of the state vector as the u-, v- and ϕ-field, respectively.

This test model is only slightly unstable, i.e. only 10 of the 1500 eigenvalues lie strictly outside the unit circle and the absolute value of the largest eigenvalue is approximately 1.00013 (see Figure 15 for the distribution of the eigenvalues). However, the system is still an interesting test model because many of the asymptotically stable poles are so close to being unstable that it is impossible to separate them properly from the unstable poles.

We now investigate whether similar results to those for the two simple systems continue to hold for the shallow water test model $S^{(3)}$. This is a discrete MIMO system of order 1500 with 1500 input and 750 output channels. Obviously, it is not possible to consider all 1500 components of the impulse response. In the following we focus only on one representative component of the impluse response, its 250th, at time $t = t_5$.

Figure 16 shows the u-field vector components of the 250th impulse response after five time steps. The top figure (16(a)) contains a comparison of the impulse response of the full order system (solid line) and the low order system computed by standard balanced truncation of reduction order $k = 750$ (dashed line with circles). We see that the low order model does not approximate the full order system accurately. In contrast, Figure 16(b) shows the approximation to the full order impulse response (solid line) by the low order approximation using α-bounded balanced truncation of order $k = 750$ for $\alpha = 1.1$ (solid line with stars). Here the solid line and the solid line with stars lie on top of each other which shows the excellent performance of the α-bounded approach. The corresponding error (in logarithmic scale) is visualized in Figure 16(c). We see that the approximation error of the α-bounded method (solid line with stars) is of order 10^{-6} on average. This is three orders of magnitude smaller than the error of the standard method (dashed line with circles).

Very similar results hold for the ϕ-field vector components of the 250th impulse response as shown in Figure 17. The error of the α-bounded approach (solid line with stars) is approximately two orders of magnitude smaller than the error of the standard method (dashed line with circles). We omit to show the analog figure for the v-field where the true solution is almost zero everywhere. The errors in the standard and the α-bounded method are of the same orders of magnitude as for the u- and the ϕ-field.

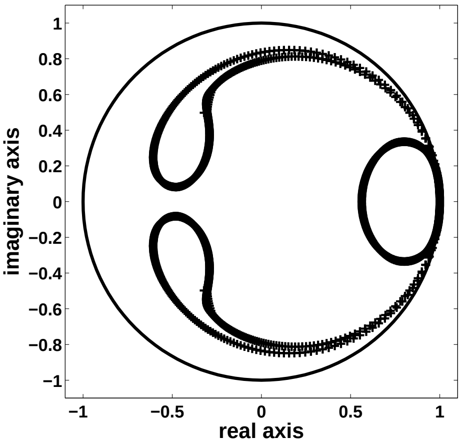
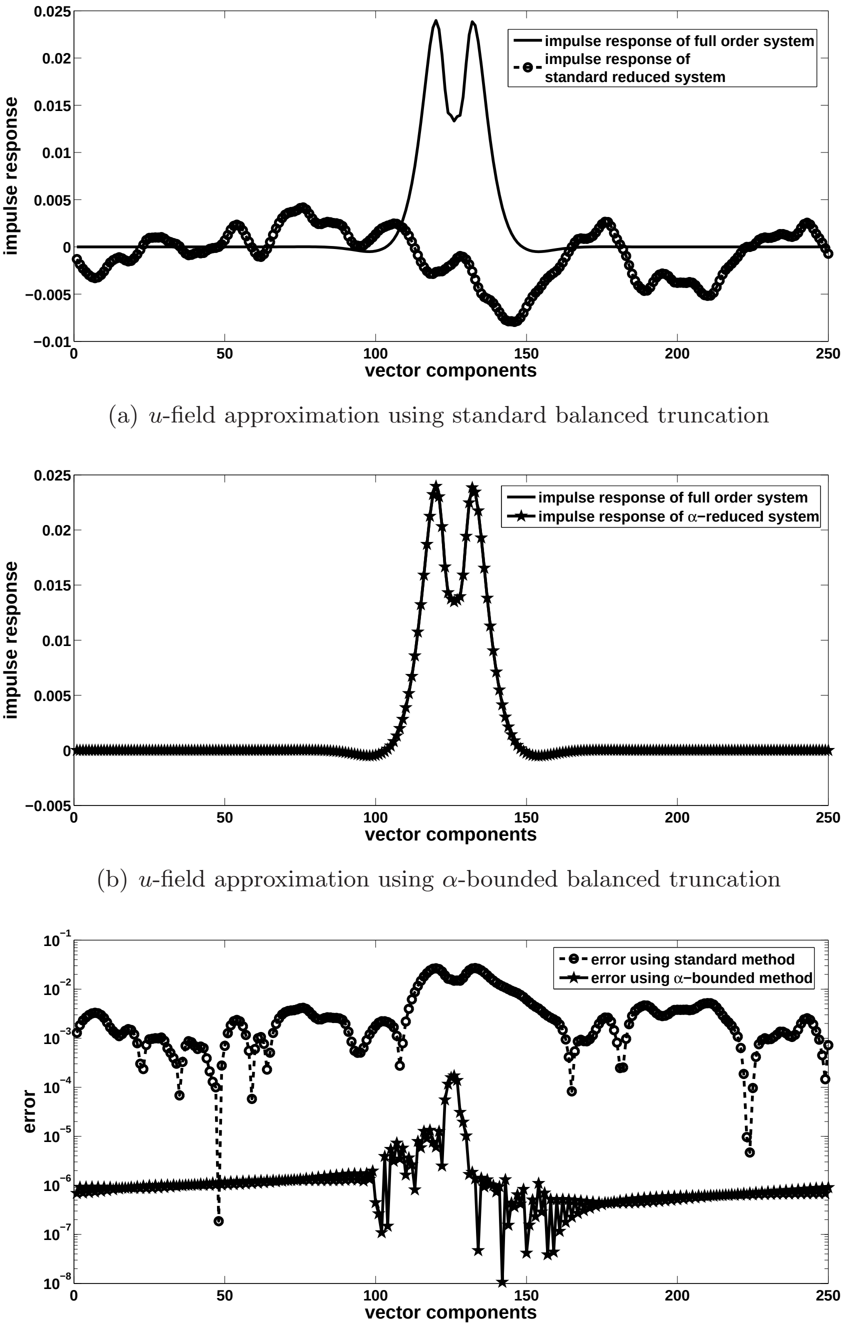


Figure 15: Eigenvalues of system matrix $A^{(3)}$ of the shallow water test model

We now investigate the effect of a further reduction to an order of $k = 150$, a tenth of the order of the full system. Figures 18 and 19 show the u- and the ϕ-field vector components of the 250th impulse response for $k = 150$ after five time steps, respectively. Despite the large reduction, the approximation error of the α-bounded method is still, on average, of order 10^{-4} (solid line with stars). This is two orders of magnitude smaller than the error of the standard method (dashed line with circles).

For the first test model $S^{(1)}$ we observed that the performance of the α-bounded method becomes quite poor when the time window consists of more than ten steps (see Figure 9). Figure 20 shows that this does not hold for the shallow water test model. After 20 time steps the errors in the u-,v- and ϕ-field components of the 250th impulse response of the α-bounded method for $\alpha = 1.1$ and order $k = 750$ (solid line with stars) are still on average two to three orders of magnitude smaller than the errors of the standard approach for order $k = 750$ (dashed line with circles).

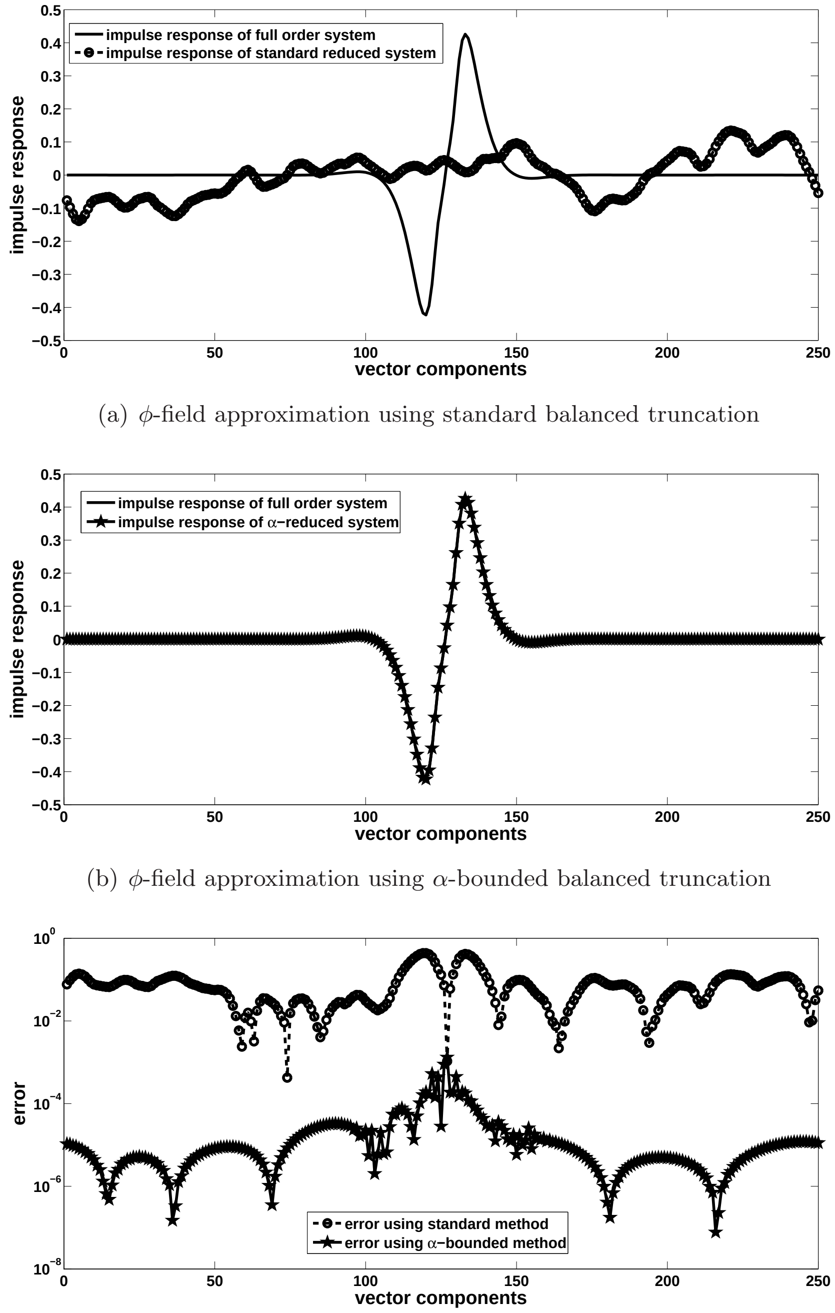


(a) u-field approximation using standard balanced truncation

(b) u-field approximation using α-bounded balanced truncation

(c) Errors in u-field approximations

Figure 16: Comparison of u-field vector components of $250th$ impulse response of full and reduced systems of order $k = 750$ using standard balanced truncation (a) as well as α-bounded balanced truncation for $\alpha = 1.1$ (b) after five time steps. Subfigure (c) shows the corresponding error plot in logarithmic scale.

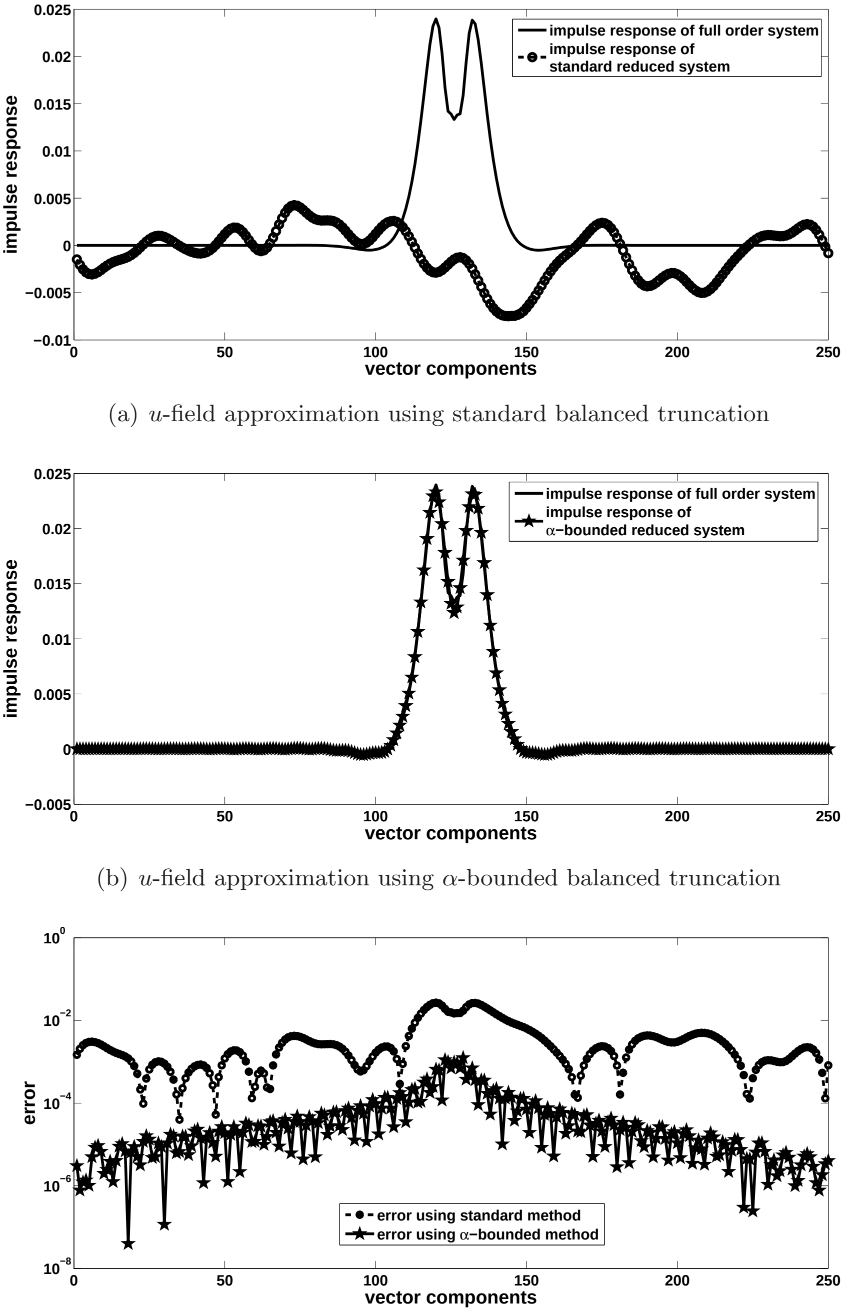


(a) ϕ-field approximation using standard balanced truncation

(b) ϕ-field approximation using α-bounded balanced truncation

(c) Errors in ϕ-field approximations

Figure 17: Comparison of ϕ-field vector components of $250th$ impulse response of full and reduced systems of order $k = 750$ using standard balanced truncation (a) as well as α-bounded balanced truncation for $\alpha = 1.1$ (b) after five time steps. Subfigure (c) shows the corresponding error plot in logarithmic scale.

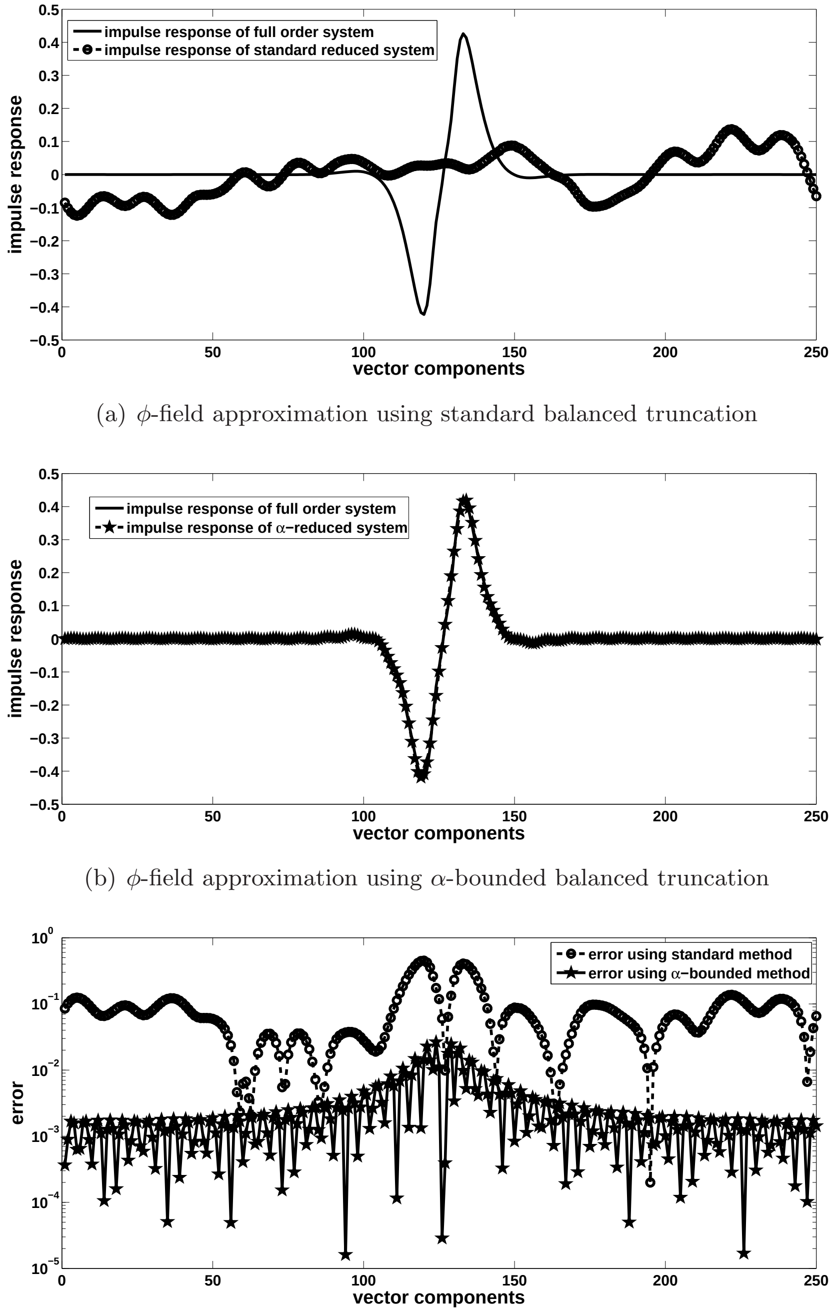


(a) u-field approximation using standard balanced truncation

(b) u-field approximation using α-bounded balanced truncation

(c) Errors in u-field approximations

Figure 18: Comparison of u-field vector components of $250th$ impulse response of full and reduced systems of order $k = 150$ using standard balanced truncation (a) as well as α-bounded balanced truncation for $\alpha = 1.1$ (b) after five time steps. Subfigure (c) shows the corresponding error plot in logarithmic scale.

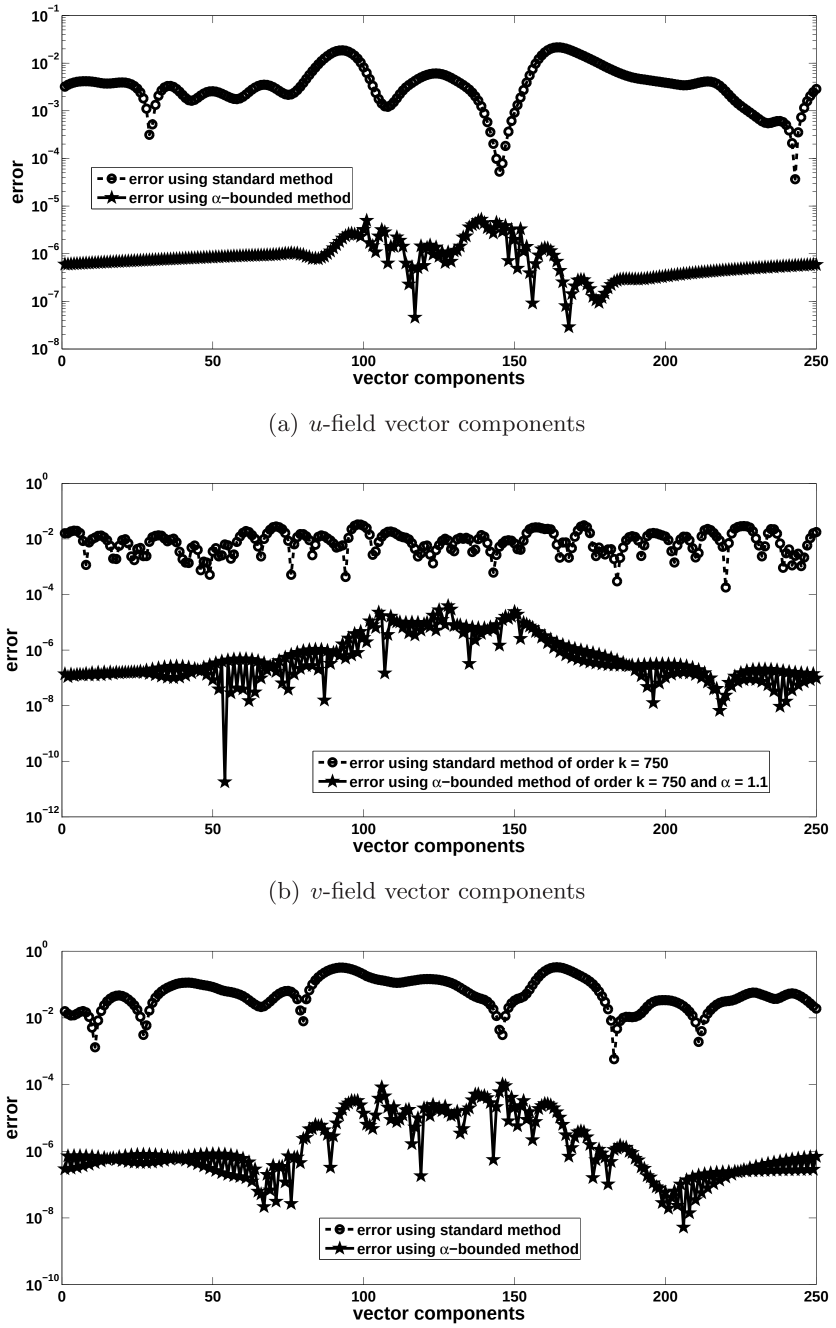


(a) ϕ-field approximation using standard balanced truncation

(b) ϕ-field approximation using α-bounded balanced truncation

(c) Errors in ϕ-field approximations

Figure 19: Comparison of ϕ-field vector components of $250th$ impulse response of full and reduced systems of order $k = 150$ using standard balanced truncation (a) as well as α-bounded balanced truncation for $\alpha = 1.1$ (b) after five time steps. Subfigure (c) shows the corresponding error plot in logarithmic scale.



(a) u-field vector components

(b) v-field vector components

(c) ϕ-field vector components

Figure 20: Comparison of errors in u-, v- and ϕ-field vector components of $250th$ impulse response of full and reduced system of order $k = 750$ using standard balanced truncation (dashed line with circles) and α-bounded balanced truncation for $\alpha = 1.1$ (solid line with stars) after 20 time steps.

4.4 Summary of numerical experiments

All the numerical experiments demonstrated the superiority of the α-bounded balanced truncation method over the currently used balanced truncation approach for unstable systems, especially over a short time window. This result is not very surprising. If the system has a considerable number of unstable poles, then the standard approach for unstable systems cannot supply a good approximation to the input-output behavior of the full order system. The reason is that essential or even all information of the asymptotically stable part of the full order system is lost (depending on the chosen reduction order). Thus, at the beginning of the time window (where the asymptotically stable part still influences the behavior of the system) we cannot expect the standard approach to supply good approximations.

Moreover, the shallow water test model showed that the standard balanced truncation method fails not only for systems with large numbers of unstable poles, but also for systems that have only a few unstable poles, but large numbers of asymptotically stable modes that are very close to being unstable.

5 Conclusions

Model order reduction of unstable control systems is an important problem to be considered. However, most of the known and approved model reduction methods are for asymptotically stable systems only. The existing approaches for unstable systems are based on an additive decomposition of the system into its asymptotically stable and its unstable part. The model reduction procedure is then applied to the asymptotically stable subsystem while the unstable part remains unchanged. This procedure may only supply good approximations to the full order system if the number of unstable poles is rather small or if the asymptotically stable part of the system is of minor importance. These assumptions are rather restrictive. At the beginning of the time window, especially, the standard low order approximations are poor because at the initial time steps the asymptotically stable components (which are neglected in the standard approach) still have influence on the behavior of the system.

In this paper we have proposed a novel approach for model reduction for unstable systems using balanced truncation. The new *α-bounded balanced truncation* method is independent of the number of unstable poles. It equally takes into account the asymptotically stable as well as the unstable modes of the full order system within the reduction process. We were able to show that the new method is embedded in a theoretical framework very similar to that of the original balanced truncation method for asymptotically stable systems. While balanced truncation for asymptotically stable systems computes a low order system that is close to being optimal with respect to the h_2-norm, the α-bounded balanced truncation method supplies a low order system close to being optimal in the $h_{2,\alpha}$-norm. Moreover, we have proved a theoretical error bound for the new α-bounded approach based on neglected Hankel singular values.

In numerical experiments with two simple unstable test models we have shown that the new method computes a low order model that approximates the input-output behavior of the full order system very accurately. It is possible even to reduce the order up to a sixth of the order of the original system while still capturing the essential information in the response. Comparison with the standard balanced truncation approach for unstable systems demonstrated the superiority of the new α-bounded balanced truncation method, especially at the beginning of the time window.

In addition to the simple models, we have also investigated a more realistic test model derived from discretized and linearized shallow water equations. This model has only a small number of unstable poles and the largest unstable eigenvalue is only slightly larger than unity. At first sight this situation suggests that the standard approach should work well. However, this is not the case. Here the failure of the standard approach is caused by a large number of asymptotically stable poles which are so close to being unstable that it is impossible to separate them properly in the stable-unstable decomposition. The numerical experiments again showed the superiority of the new α-bounded method. For a reduction order of 150 which is a tenth of the order of the original system the error of the α-bounded method is still, on average, of order of magnitude 10^{-4}. This is two orders of magnitude smaller than the error in the standard approach.

A Appendix

A.1 First test model $\mathcal{S}^{(1)}$

A.1.1 Eigenvalues of first test model

The system matrix $A^{(1)}$ of the first test model $\mathcal{S}^{(1)}$ is a real diagonal matrix of dimension 30×30. The diagonal entries of the matrix are the following:

	eigenvalues of $A^{(1)}$
1	0.79503394
2	0.87585565
3	2.48969761
4	2.40903449
5	2.41719241
6	1.67149713
7	-0.20748692
8	1.71723865
9	2.63023529
10	1.48889377
11	2.03469301
12	1.72688513
13	0.69655908
14	1.29387147
15	0.21271720
16	1.88839563
17	-0.14707011
18	-0.06887046
19	0.19050131
20	-1.94428416
21	2.43838029
22	1.32519054
23	0.24507168
24	2.37029854
25	-0.71151642
26	0.89775755
27	0.75855296
28	1.31920674
29	1.31285860
30	0.13512008

A.1.2 Error norms of impulse responses of first test model

reduction order k=10	error norm standard method	error norm α-bounded method
1st impulse response	1.0000e+000	1.2426e-013
2nd impulse response	1.0000e+000	5.0860e-013
3rd impulse response	1.0000e+000	2.5934e-014
4th impulse response	1.0000e+000	3.1362e-014
5th impulse response	1.0000e+000	2.6209e-014
6th impulse response	0	1.4785e-014
7th impulse response	1.0000e+000	1.6157e-012
8th impulse response	0	1.1791e-014
9th impulse response	1.0000e+000	2.9874e-014
10th impulse response	0	1.2244e-013
11th impulse response	1.0000e+000	1.5842e-014
12th impulse response	0	1.5496e-014
13th impulse response	1.0000e+000	6.5983e-013
14th impulse response	0	3.5923e-014
15th impulse response	1.0000e+000	5.0539e-013
16th impulse response	0	4.2042e-014
17th impulse response	1.0000e+000	5.6336e-013
18th impulse response	1.0000e+000	2.0513e-012
19th impulse response	1.0000e+000	1.5261e-013
20th impulse response	0	1.6185e-014
21st impulse response	1.0000e+000	1.1253e-014
22nd impulse response	0	7.8109e-014
23rd impulse response	1.0000e+000	1.0147e-012
24th impulse response	1.0000e+000	5.6350e-014
25th impulse response	1.0000e+000	1.2073e-013
26th impulse response	1.0000e+000	5.5176e-013
27th impulse response	1.0000e+000	1.3319e-013
28th impulse response	0	6.4813e-014
29th impulse response	0	5.4033e-014
30th impulse response	1.0000e+000	7.7044e-013

Table 1: Test model $\mathcal{S}^{(\mathbf{1})}$: Comparison of relative output error norms of all components of the impulse response of standard balanced truncation with the α-bounded method for $\alpha = 12$ for reduction order $k = 10$

reduction order k=5	error norm standard method	error norm α-bounded method
1st impulse response	1.0000e+000	7.7524e-003
2nd impulse response	1.0000e+000	1.0094e-002
3rd impulse response	1.0000e+000	6.6808e-005
4th impulse response	1.0000e+000	5.9113e-004
5th impulse response	1.0000e+000	5.2918e-004
6th impulse response	1.0000e+000	7.1272e-005
7th impulse response	1.0000e+000	1.3513e-002
8th impulse response	1.0000e+000	5.5941e-004
9th impulse response	1.0000e+000	1.4342e-003
10th impulse response	1.0000e+000	2.6083e-003
11th impulse response	1.0000e+000	2.0784e-003
12th impulse response	1.0000e+000	6.5304e-004
13th impulse response	1.0000e+000	3.0403e-003
14th impulse response	0	6.5319e-003
15th impulse response	1.0000e+000	2.8598e-002
16th impulse response	1.0000e+000	1.7599e-003
17th impulse response	1.0000e+000	1.9727e-002
18th impulse response	1.0000e+000	2.5733e-002
19th impulse response	1.0000e+000	2.9324e-002
20th impulse response	0	6.1228e-004
21st impulse response	1.0000e+000	3.6335e-004
22nd impulse response	0	5.8604e-003
23rd impulse response	1.0000e+000	2.7304e-002
24th impulse response	1.0000e+000	8.7045e-004
25th impulse response	1.0000e+000	6.3947e-002
26th impulse response	1.0000e+000	1.0489e-002
27th impulse response	1.0000e+000	6.2358e-003
28th impulse response	0	5.9883e-003
29th impulse response	0	6.1243e-003
30th impulse response	1.0000e+000	3.0505e-002

Table 2: Test model $\mathcal{S}^{(\mathbf{1})}$: Comparison of relative output error norms of all components of the impulse response of standard balanced truncation with the α-bounded method for $\alpha = 12$ for reduction order $k = 5$

A.2 Second test model $\mathcal{S}^{(2)}$

The eigenvalues of $A^{(2)}$ and their absolute values are listed in the following table:

	eigenvalues of $A^{(2)}$	absolute values of eigenvalues
1	-1.60831292204084 + 0.487572318049913i	1.68059430575381
2	-1.60831292204084 - 0.487572318049913i	1.68059430575381
3	1.43492888457886 + 0.408037639927528i	1.49181621501992
4	1.43492888457886 - 0.408037639927528i	1.49181621501992
5	1.07250573368774 + 0.938788623072736i	1.42533947802054
6	1.07250573368774 - 0.938788623072736i	1.42533947802054
7	0.553666930493128 + 1.28161880222829i	1.39609950366969
8	0.553666930493128 - 1.28161880222829i	1.39609950366969
9	1.31663596919285 + 0.00000000000000i	1.31663596919285
10	0.00907874879630455 + 1.24764567318242i	1.24767870443096
11	0.00907874879630455 - 1.24764567318242i	1.24767870443096
12	-0.680989374647882 + 1.09716264135389i	1.29132195441956
13	-0.680989374647882 - 1.09716264135389i	1.29132195441956
14	1.06395319427490 + 0.00000000000000i	1.06395319427490
15	-0.621006290445976 + 0.808270698064141i	1.01928913175927
16	-0.621006290445976 - 0.808270698064141i	1.01928913175927
17	-1.15944187673225 + 0.00000000000000i	1.15944187673225
18	-0.976123239524686 + 0.00000000000000i	0.976123239524686
19	0.390824808288419 + 0.746041169810419i	0.842212240368055
20	0.390824808288419 - 0.746041169810419i	0.842212240368055
21	0.102412738805032 + 0.772584295090426i	0.779342583264842
22	0.102412738805032 - 0.772584295090426i	0.779342583264842
23	-0.575973176493380 + 0.00000000000000i	0.575973176493380
24	-0.463599216930749 + 0.332204742012911i	0.570336939496881
25	-0.463599216930749 - 0.332204742012911i	0.570336939496881
26	0.223072126844484 + 0.525128182130016i	0.570544285259345
27	0.223072126844484 - 0.525128182130016i	0.570544285259345
28	0.506111954360960 + 0.00000000000000i	0.506111954360960
29	0.0510992784147967 + 0.203544927135694i	0.209861081711660
30	0.0510992784147967 - 0.203544927135694i	0.209861081711660

References

[1] A.C. Antoulas. *Approximation of large-scale dynamical systems.* SIAM Publisher, Philadelphia, 2005.

[2] W.E. Arnoldi. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quarterly of Applied Mathematics*, 9:17–29, 1951.

[3] P. Benner, E.S. Quintana-Orti, and G. Quintana-Orti. Balanced truncation model reduction of large-scale dense systems on parallel computers. *Mathematical and Computer Modelling of Dynamical Systems*, 6:383–405, 2000.

[4] D.S. Bernstein, L.D. Davis, and D.C. Hyland. The optimal projection equation for reduced-order, discrete-timemodeling, estimation, and control. *J. Guidance, Control, and Dynamics*, 9:288–293, 1986.

[5] C. Boess. *Using model reduction techniques within the incremental 4D-Var method.* PhD thesis, Universitaet Bremen, 2008.

[6] S. Gerschgorin. ber die abgrenzung der eigenwerte einer matrix. *Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. Nauk*, 7:749–754, 1993.

[7] K. Glover. All optimal hankel norm approximation of linear multivariable systems and their L_∞ error bounds. *Int. J. Contr.*, 39:1115–1193, 1984.

[8] G.H. Golub and C.F. van Loan. *Matrix computations.* Johns Hopkins University Press, Baltimore, third edition, 1996.

[9] D. Hinrichsen and A.J. Pritchard. *Mathematical Systems Theory I.* Springer, Berlin, 2005.

[10] C.S. Hsu and D. Hou. Reducing unstable linear control systems via real schur transformation. *Electron. Lett.*, pages 984–986, 1991.

[11] D. Katz, A.S. Lawless, N.K. Nichols, M.J.P. Cullen, and R.N. Bannister. A comparison of potential vorticity-based and vorticity-based control variables. *Numerical Analysis Report*, 8/2005.

[12] D. Kubalinska. *Optimal interpolation-based model reduction.* PhD thesis, Universitaet Bremen, 2008.

[13] MATLAB. *Control Toolbox.* The MathWorks, Release R2009a.

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

[15] S.K. Nagar and S.K. Singh. An algorithmic approach for system dedcomposition and balanced realized model reduction. *J. of Franklin Inst.*, 341:615–630, 2004.

[16] L. Pernebo and L.M. Silverman. Model reduction via balanced state space representations. *IEEE Trans. Automatic Control*, 27:382–387, 1982.

[17] Y. Saad. *Numerical Methods for Large Eigenvalue Problems.* Manchester University Press, 1992.

[18] A. Varga. Balancing-free square-root algorithm for computing singular pertubation approximations. *Proc. 30th IEEE CDC, Brighton*, pages 1062–1065, 1991.

[19] A. Varga. Model reduction software in the slicot library. *Applied and Computational Control, Signals and Circuits*, 629:239–282, 2001.