A posteriori error bounds for discrete balanced truncation

Chahlaoui, Younes

2009

MIMS EPrint: **2009.12**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

A POSTERIORI ERROR BOUNDS FOR DISCRETE BALANCED TRUNCATION*

YOUNES CHAHLAOUI†

Abstract. Balanced truncation of discrete linear time-invariant systems is an automatic method once an error tolerance is specified and yields an a priori error bound, which is why it is widely used in engineering for simulation and control. We present some new insight into this method. We derive a discrete version of Antoulas's $\mathcal{H}_2$-norm error formula [1, p.218] and show how to adapt it to some special cases. This error bound is an a posteriori computable upper bound for the $\mathcal{H}_2$-norm of the error system defined as the system whose transfer function corresponds to the difference between the transfer function of the original system and the transfer function of the reduced system. The main advantage of our results is that we use the information already available in the balanced truncation algorithm in order to compute the $\mathcal{H}_2$-norm instead of computing one gramian of the corresponding error system. There is always a computational restriction on solving high-dimensional Stein equations for gramians. The a posteriori bound gives insight into the quality of the reduced system and can be used to solve many problems accompanying the order reduction operation.

Key words. model reduction, balanced truncation, gramians, Stein equations, $\mathcal{H}_2$-norm.

AMS subject classifications. 15A18,15A24,65F15,65F35,65P99,93C05,93C55.

1. Introduction. Modeling real world physical processes gives rise to mathematical systems of increasing complexity. Good mathematical models have to reproduce the original process as precisely as possible but the computing time and the storage resources needed to simulate the mathematical model are limited. As a consequence, there must be a tradeoff between accuracy and computational constraints. One often has to deal with systems that have an unacceptably high level of complexity. It is then desirable to approximate such systems by systems of lower complexity. This is the model reduction problem.

Balanced truncation is one of the best known method for model reduction of linear systems [4, 7, 8, 9]. It is characterized by the principle of projection of dynamics. Balanced truncation is widely used in practice for three main reasons. First, for a reasonably small system order, say a few hundred, it gives a satisfactory approximation in the majority of cases without having to solve a complicated minimization problem or having to choose a set of essential system parameters first. Second, this approximation can be obtained at relatively reasonable computational cost. Third, an a priori upper bound for the error between the original plant and the reduced-order model exists for the $\mathcal{H}_\infty$-norm, the preferred measure of approximation accuracy in engineering. Recently, an a posteriori error bound for balanced truncation was presented by Antoulas [1]. Here we will derive a discrete version of this error bound and show how to adapt it to some special cases. This error bound is a computable upper bound for the $\mathcal{H}_2$-norm of the error system defined as the system whose transfer function corresponds to the difference between the transfer function of the original system and the transfer function of the reduced system. The main advantage of our results is that we use the information already available in the balanced truncation algorithm in order to compute the $\mathcal{H}_2$-norm instead of computing one gramian of the

*This work was supported by EPSRC grant EP/E050441/1.

†Centre for Interdisciplinary Computational and Dynamical Analysis (CICADA), School of Mathematics, The University of Manchester, UK. (Younes.Chahlaoui@manchester.ac.uk, http://www.maths.manchester.ac.uk/∼ chahlaoui/).

corresponding error system. There is always a computational restriction on solving high-dimensional Lyapounov equations for gramians.

The a posteriori bound gives insight into the quality of the reduced system and can be used to solve many problems accompanying the order reduction operation. For example in the problem of choosing the reduced order, the purpose of the model determines the "acceptable" order reduction in an implicit way; an explicit criterion for acceptable reduced order is hard to give, as we need to analyze a priori the dynamics involved in order to obtain some sort of dynamics ranking. For systems of reasonable orders (in general a few hundred), this analysis can be done at a reasonable cost, including the problem of finding an appropriate value of the reduced order by the use of the Hankel singular values [1]. But for large-scale models this pre-treatment is prohibitive. Our error formulas and bounds could be implemented into the loop of the model order-reduction method in order to check if the chosen reduced order is the best choice or needs to be modified before stopping the reduction algorithm.

For large scale problems one has to use iterative methods to find an adequate approximation. In this respect, ideas based on balanced reduction methods are interesting since they offer the possibility to perform order selection during the computation of the projection spaces and not in advance. However, serious drawbacks of balanced truncation (and all direct methods in general) are that it ignores any sparsity of the system, and that it is not easy to parallelize (note however the work of Benner and al. [2, 3] who parallelize some traditional model reduction methods). Its use is therefore limited if large, sparse systems have to be reduced.

In this paper we consider discrete-time systems

$$\mathcal{S} \quad \left\{ \begin{array}{rcl} Ex_{k+1} & = & Ax_k + Bu_k \\ y_k & = & Cx_k \end{array} \right. \tag{1.1}$$

with input $u_k \in \mathbb{R}^m$, state $x_k \in \mathbb{R}^N$ and output $y_k \in \mathbb{R}^p$, and $m, p \ll N$. The input sequence is assumed to be square-summable, i.e., $u_k \in l_2^m$ [10], and we assume that the matrices A, B, and C are of appropriate dimensions. We will assume also the system (1.1) to be stable (i.e., all eigenvalues of the matrix A are strictly inside the unit circle). The transfer function corresponding to the system $\mathcal{S}$ is

$$H(z) = C(zI - A)^{-1}B.$$

This paper is organized as follow. We introduce first in Section 2 the principle of projection of dynamics. In Section 3, we review the balanced truncation method and give some new insight into the principle behind it. Section 4 is the main contribution of this paper. It is dedicated to the presentation of the new error formulas and some new a posteriori bounds of the $\mathcal{H}_2$ norm of the error system corresponding to the balanced truncation method. We also discuss some features of these formulas and bounds. We end this section by the presentation of some special cases for which the bounds are better. We finish with some further discussion and concluding remarks in Section 5.

2. Projection of dynamics. Let T be a desired coordinate (similarity) transformation of the system $\mathcal{S}$ and consider the following partition of the transformed system matrices:

$$T^{-1}ET = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}, \quad T^{-1}AT = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad T^{-1}B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix},$$

$$CT = \left[\begin{array}{cc} C_1 & C_2 \end{array}\right],$$

where $E_{11}, A_{11} \in \mathbb{C}^{n\times n}, B_1 \in \mathbb{C}^{n\times m}, C_1 \in \mathbb{C}^{p\times n}$, and $n \ll N$. Then the system

$$\hat{\mathcal{S}} \quad \left\{\begin{array}{rcl} E_{11}\hat{x}_{k+1} & = & A_{11}\hat{x}_k + B_1 u_k, \\ \hat{y}_k & = & C_1\hat{x}_k, \end{array}\right.$$

is an nth order truncation of $\mathcal{S}$. The truncation is obtained by applying the projection

$$\Pi = \left[\begin{array}{c} I_n \\ 0 \end{array}\right]\left[\begin{array}{cc} I_n & 0 \end{array}\right]$$

to the transformed system. The combination of applying a similarity transformation and a subsequent truncation is often referred to as *projection of dynamics* (also known as transform and truncate). Let $\Pi_l, \Pi_r \in \mathbb{C}^{N\times n}$ satisfy $\Pi_l^*\Pi_r = I_n$.[1] The projected system $\hat{\mathcal{S}}$ matrices are obtained as follows:

$$E_{11} = \Pi_l^* E\Pi_r, \quad A_{11} = \Pi_l^* A\Pi_r, \quad B_1 = \Pi_l^* B, \quad C_1 = C\Pi_r.$$

Thus transformation by T and truncation by Π are merged in the projection pair (Π_l, Π_r) as follows:

$$\Pi_r = T\left[\begin{array}{c} I_n \\ 0 \end{array}\right], \quad \Pi_l^* = \left[\begin{array}{cc} I_n & 0 \end{array}\right]T^{-1}.$$

It can be verified easily that this definition satisfies $\Pi_l^*\Pi_r = I_n$ and hence $\Pi_r\Pi_l^*$ is a projector.

A projection method is in fact a choice of two subspaces $\mathbb{P}_r, \mathbb{P}_l \subset \mathbb{C}^N$ of dimension n, so that $\hat{x}_k \in \mathbb{P}_r$ and the residual is orthogonal to $\mathbb{P}_l$. The columns of Π_r and Π_l form bases for $\mathbb{P}_r$ and $\mathbb{P}_l$, respectively:

$$\mathrm{Im}(\Pi_r) = \mathbb{P}_r, \quad \mathrm{Im}(\Pi_l) = \mathbb{P}_l.$$

If $\mathbb{P}_l = \mathbb{P}_r$, the projection is orthogonal, otherwise it is oblique. The choice of basis of $\mathbb{P}_r$ and $\mathbb{P}_l$ is not important in theory but unfortunately very important numerically. If we take any two other bases of these subspaces, say $\bar{\mathbb{P}}_r$ and $\bar{\mathbb{P}}_l$, then there exist two invertible matrices $X, Y \in \mathbb{C}^{n\times n}$ such that

$$\bar{\Pi}_r = \Pi_r X, \quad \bar{\Pi}_l = \Pi_l Y.$$

It is easy to see that for these two projector matrices we have:

$$\bar{H}(z) := C\bar{\Pi}_r(z\bar{\Pi}_l^* E\bar{\Pi}_r - \bar{\Pi}_l^* A\bar{\Pi}_r)^{-1}\bar{\Pi}_l^* B = C\Pi_r(z\Pi_l^* E\Pi_r - \Pi_l^* A\Pi_r)^{-1}\Pi_l^* B =: \hat{H}(z),$$

which means that the two reduced order models will be equivalent.

Indeed the main dilemma is how to find an adequate states transformation. This transformation should rank and sort the states in order to truncate question that one has to face in model reduction is how to choose these projection matrices. A special case of projection of dynamics is balanced truncation.

[1] The subscripts r and l refer to right and left, respectively

3. Balanced truncation. The method of balanced truncation of linear systems is well established for model reduction. It is a special case of the transform and truncate methods described above. It is a balancing then truncate method. It is based on a balanced realization[2] $\{T^{-1}AT, T^{-1}B, CT\}$ of the system. This realization has some desirable sensitivity properties with respect to poles, zeros, truncation errors in digital filter implementations, and so on [7, 10]. It is therefore recommended whenever the choice of a realization is not specified by the user.

For linear time-invariant systems, the approach requires standard matrix computations, and has been successfully used in control systems design. The main idea is to rewrite the system $\mathcal{S}$, which we suppose stable, controllable and observable[3] [7, 10], using a similarity transformation T called the balancing transformation and then use a truncation to obtain the reduced model. In this coordinate system one has [5]

$$T\mathcal{G}_c T^* = T^{-*}\mathcal{G}_o T^{-1} = \Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_N),$$

where the σ_i are the Hankel singular values of $\mathcal{S}$ and $\mathcal{G}_c$ and $\mathcal{G}_o$ are the controllability and observability gramians of $\mathcal{S}$ [10]. These gramians are solutions of the Stein equations

$$E\mathcal{G}_c E^T - A\mathcal{G}_c A^T - BB^T = 0, \quad E^T\mathcal{G}_o E - A^T\mathcal{G}_o A - C^T C = 0.$$

A natural question now arises: what is the use of balancing, i.e., diagonalizing $\mathcal{G}_c$ and $\mathcal{G}_o$? This can be explained using energy functions. The controllability and observability gramians measure to what degree each state is excited by an input, and each state excites future outputs, respectively. Given a stable linear system $\mathcal{S}$, it well known that for any state x

$$\epsilon_c(x) = (x^*\mathcal{G}_c^{-1}x)^{\frac{1}{2}}, \quad \epsilon_o(x) = (x^*\mathcal{G}_o x)^{\frac{1}{2}}$$

are respectively the smallest amount of energy needed to steer the system from 0 to x, and the largest amount of energy obtained by observing the output of the free system with the initial condition x. If we define the energy storage efficiency by

$$\epsilon(x_0) = \frac{x_0^*\mathcal{G}_o x_0}{x_0^*\mathcal{G}_c^{-1}x_0}, \tag{3.1}$$

then the maximization of $\epsilon(x_0)$ with respect to x_0 yields the generalized eigenproblem

$$\mathcal{G}_o x_0 = \mathcal{G}_c^{-1}\epsilon(x_0)x_0.$$

And so $\epsilon(x_0)$ takes an extremal value for x_0 an eigenvector of $\mathcal{G}_c\mathcal{G}_o$ (or equivalently a generalized eigenvector of the pair $(\mathcal{G}_o, \mathcal{G}_c^{-1})$). The extremal value of $\epsilon(x_0)$ corresponds thus to the maximal eigenvalue of $\mathcal{G}_c\mathcal{G}_o$ and hence to the square of the largest Hankel singular value σ_1 of the considered system. Another interpretation is that the transformation T solves the minimization problem

$$\min_T \operatorname{trace}\big(T\mathcal{G}_c T^* + T^{-*}\mathcal{G}_o T^{-1}\big).$$

[2] In the system theory context refers to a state space model implementing a given input-output behavior. For a linear time-invariant system specified by a transfer matrix, $H(z)$, a realization is any quadruple of matrices (A, B, C, D) such that $C(zI - A)^{-1}B + D = H(z)$.

[3] This means essentially that the gramians are full rank.

The minimum of this expression is $2\sum_{i=1}^{N}\sigma_i$, and a balancing transformation turns out to provide a minimizing similarity transformation T [1].

The balancing transformation T ensures that each state is as controllable as it is observable in the new coordinate system. It is also shown in [7] that for non-minimal systems the controllable subspace and the unobservable subspace are the image and the kernel of $\mathcal{G}_c$ and $\mathcal{G}_o$, respectively. And so, T transforms the observability and controllability ellipsoids to an identical ellipsoid aligned with principal axes along the coordinate axes as shown in Figure 3.1.

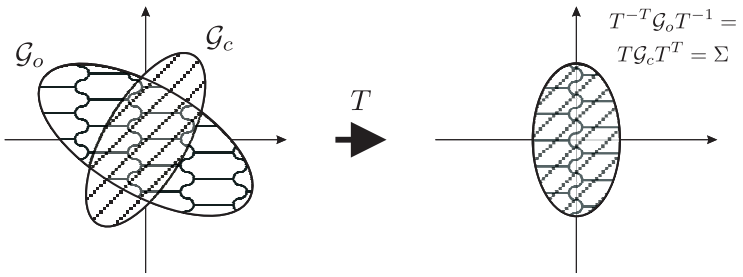

FIG. 3.1. *The effect of a balancing transformation T on the controllability and observability ellipsoids.*

After balancing the system, a reduced model is obtained by truncating the new state $x = (x_1, \ldots, x_N)^T$ to $\hat{x} = (x_1, \ldots, x_n)^T$, where $n \ll N$. The truncated states are the least controllable and observable states, corresponding to the smallest Hankel singular values and having little effect on the input/output behavior. This truncation is equivalent to projecting the system with a rank n projection $\Pi := \Pi_r \Pi_l^*$. The so-called truncation matrices Π_r and Π_l can be obtained from the Cholesky factorizations

$$G_c = S^*S, \quad G_o = R^*R,$$

where G_c and G_o are related to the gramians by $\mathcal{G}_c = G_c$, and $\mathcal{G}_o = E^*G_oE$. Compute the singular value decomposition

$$SE^*R^* = \left[\begin{array}{c|c} U_1 & U_2 \end{array}\right] \left[\begin{array}{c|c} \Sigma_1 & 0 \\ \hline 0 & \Sigma_2 \end{array}\right] \left[\begin{array}{c|c} V_1 & V_2 \end{array}\right]^* \tag{3.2}$$

where $\Sigma_1 = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$, $\Sigma_2 = \mathrm{diag}(\sigma_{n+1}, \ldots, \sigma_N)$ and define

$$\Pi_l = E^*R^*V_1\Sigma_1^{-1/2}, \quad \Pi_r = S^*U_1\Sigma_1^{-1/2}. \tag{3.3}$$

We can easily see that $\Pi_l^*\Pi_r = I_n$ (i.e., $\Pi = \Pi_r\Pi_l^*$ is a projector) and $\Pi_l^*\mathcal{G}_c\mathcal{G}_o\Pi_r = \Sigma_1^2$. It follows that the singular values σ_i of SE^*R^* are the (nonzero) Hankel singular values [10].

By this approach the gramians $\mathcal{G}_c$ and $\mathcal{G}_o$ (or equivalently the matrices G_c and G_o) are not needed to construct the projector $\Pi = \Pi_r\Pi_l^*$, but only the factors S and R, which can be obtained e.g. using Hammarling's method [6]. One then obtains the reduced model for the system $\mathcal{S} := \{E, A, B, C\}$ as $\hat{\mathcal{S}} := \{\pi_l^*E\pi_r, \pi_l^*A\pi_r, \pi_l^*B, C\pi_r\}$.

We summarize the procedure in the following algorithm.

Algorithm 1 Balanced truncation

Input the original system $\mathcal{S} \doteq \{E, A, B, C\}$ and a reduced order n.

• Solve the Stein equations

$$E\mathcal{G}_c E^T - A\mathcal{G}_c A^T - BB^T = 0, \quad E^T\mathcal{G}_o E - A^T\mathcal{G}_o A - C^T C = 0$$

for S and R where $\mathcal{G}_c = S^*S$ and $\mathcal{G}_o = R^*R$ are Cholesky factorizations.

• Compute the SVD $\quad SE^*R^* = U\Sigma V^*$.

• The projection matrices are given by

$$\Pi_l = E^*R^*V_n\Sigma_n^{-1/2}, \quad \Pi_r = S^*U_n\Sigma_n^{-1/2},$$

where $U_n = U(:, 1:n)$, $V_n = V(:, 1:n)$ and $\Sigma_n = \Sigma(1:n, 1:n)$.

• And the reduced order model is given by the matrices

$$\hat{E} = \Pi_l^* E \Pi_r, \quad \hat{A} = \Pi_l^* A \Pi_r, \quad \hat{B} = \Pi_l^* B, \quad \hat{C} = C\Pi_r.$$

An a priori error bound in the induced 2-norm can be given for the error between the original and the reduced system [10]

$$\sigma_{n+1} \leq \|\mathcal{S} - \hat{\mathcal{S}}\|_{\mathcal{H}_\infty} \leq 2(\sigma_{n+1} + \cdots + \sigma_N). \tag{3.4}$$

This result says that the $\mathcal{H}_\infty$-norm of the error system is bounded above by twice the sum of the neglected Hankel singular values.

More recently, a new result was derived by Antoulas [1, p. 218] for the $\mathcal{H}_2$ norm. It is a computable $\mathcal{H}_2$ norm of the error system which yields also a computable upper bound for this norm. A convenient way to determine the $\mathcal{H}_2$ norm is to use the formula:

$$\|\mathcal{S}\|_{\mathcal{H}_2}^2 = \operatorname{trace}(B^*\mathcal{G}_o B) = \operatorname{trace}(C\mathcal{G}_c C^*),$$

where $\mathcal{G}_c$ and $\mathcal{G}_o$ are respectively the controllability and observability gramians of the system.

If the original system is of order[4] N and the reduced order is n, the order of the error system will be $N + n$. To compute the $\mathcal{H}_2$ norm of the error system we have to solve again another Lyapunov equation for one gramian of this error system, and so the cost will be of the order of $(N+n)^3$ added to the cost of the model order reduction method. With Antoulas's formula, one needs only the gramian of the original system. This gramian is supposed to be available already by the balanced truncation method. So the cost will be only the cost of the double product of the gramian by the input matrix (or equivalently the output matrix) and its transpose, and the computation of the trace of that product.

In the following section we give a discrete-time version of this formula, and show how to adapt it to some special cases. The discrete-time version present some interesting features that we will discuss later.

4. $\mathcal{H}_2$ norm of the error system for balanced truncation. In this section we derive a computable a posteriori upper bound for the $\mathcal{H}_2$ norm of the error system for balanced truncation. For simplicity, let us assume henceforth that the matrix E

[4] Also called the McMillan degree.

is the identity (i.e., $E = I_N$), which means that we have a simple state-space model and the system $\mathcal{S}$ is already in balanced form, and partition the matrices A, B and C as follows:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix},$$

where $\hat{A} \doteq A_{11} \in \mathbb{C}^{n\times n}$, $\hat{B} \doteq B_1 \in \mathbb{C}^{n\times m}$ and $\hat{C} \doteq \mathbb{C}_1 \in C^{p\times n}$. Since the system $\mathcal{S}$ is balanced its controllability and observability gramians are diagonal and equal

$$\mathcal{G}_c = \mathcal{G}_o = \mathcal{G} = \begin{bmatrix} \mathcal{G}_1 & 0 \\ 0 & \mathcal{G}_2 \end{bmatrix}, \quad \text{where} \quad \mathcal{G}_1 \in \mathbb{R}^{n\times n}.$$

We have $\mathcal{G}_1 = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$ and $\mathcal{G}_2 = \operatorname{diag}(\sigma_{n+1}, \ldots, \sigma_N)$, where σ_i are the Hankel singular values. The unified gramian $\mathcal{G}$ then solves the Stein equations

$$A\mathcal{G}A^* - \mathcal{G} + BB^* = 0, \quad A^*\mathcal{G}A - \mathcal{G} + C^*C = 0. \tag{4.1}$$

To obtain the result, we consider the error system $\mathcal{S}_e$, defined as the system which has the transfer function $H_e(z) := H(z) - \hat{H}(z) = C(zI - A)^{-1}B - C_1(zI - A_{11})^{-1}B_1$, where $H(z)$ is the transfer function of $\mathcal{S}$ and $\hat{H}(z)$ is the transfer function of $\hat{\mathcal{S}}$. A realization of the system $\mathcal{S}_e$ is given by

$$\left\{ \begin{bmatrix} A & 0 \\ 0 & A_{11} \end{bmatrix}, \begin{bmatrix} B \\ -B_1 \end{bmatrix}, \begin{bmatrix} C & C_1 \end{bmatrix} \right\}. \tag{4.2}$$

The bound on the approximation error $\|\mathcal{S} - \hat{\mathcal{S}}\|_{\mathcal{H}_2} = \|\mathcal{S}_e\|_{\mathcal{H}_2}$ is obtained directly by bounding the $\mathcal{H}_2$ norm of $\mathcal{S}_e$. Let us first note that the controllability gramian $\mathcal{G}_{c_e}$ and the observability gramian $\mathcal{G}_{o_e}$ of $\mathcal{S}_e$ are given by

$$\mathcal{G}_{c_e} = \begin{bmatrix} \mathcal{G} & -Y \\ -Y^* & \hat{\mathcal{G}}_c \end{bmatrix}, \qquad \mathcal{G}_{o_e} = \begin{bmatrix} \mathcal{G} & Z \\ Z^* & \hat{\mathcal{G}}_o \end{bmatrix},$$

where $\hat{\mathcal{G}}_c$ and $\hat{\mathcal{G}}_o$ are the controllability and observability gramians of the reduced model $\hat{\mathcal{S}}$, respectively, which solve

$$A_{11}\hat{\mathcal{G}}_c A_{11}^* - \hat{\mathcal{G}}_c + B_1 B_1^* = 0, \qquad A_{11}^*\hat{\mathcal{G}}_o A_{11} - \hat{\mathcal{G}}_o + C_1^* C_1 = 0, \tag{4.3}$$

and where $Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$ and Y are solutions of

$$AYA_{11}^* - Y + BB_1^* = 0, \qquad A^*ZA_{11} - Z + C^*C_1 = 0. \tag{4.4}$$

The $\mathcal{H}_2$ norm of the error system is given by

$$\begin{aligned} \|\mathcal{S}_e\|_{\mathcal{H}_2}^2 &\doteq \operatorname{trace}\left(\begin{bmatrix} B^* & -B_1^* \end{bmatrix} \begin{bmatrix} \mathcal{G} & Z \\ Z^* & \hat{\mathcal{G}}_o \end{bmatrix} \begin{bmatrix} B \\ -B_1 \end{bmatrix} \right) \\ &= \operatorname{trace}\left(B^*\mathcal{G}B - 2B^*ZB_1 + B_1^*\hat{\mathcal{G}}_o B_1 \right) \\ &= \operatorname{trace}\left(B^*\mathcal{G}B - 2B_1^*Z_1B_1 - 2B_2^*Z_2B_1 + B_1^*\hat{\mathcal{G}}_o B_1 \right). \end{aligned} \tag{4.5}$$

Now, from (4.1), we obtain

$$A_{11}\mathcal{G}_1 A_{21}^* + A_{12}\mathcal{G}_2 A_{22}^* + B_1 B_2^* = 0,$$

and consequently

$$\text{trace}(-2B_2^*Z_2B_1) = \text{trace}(-2B_1B_2^*Z_2) = \text{trace}(2A_{11}\mathcal{G}_1A_{21}^*Z_2 + 2A_{12}\mathcal{G}_2A_{22}^*Z_2).$$

Substituting in (4.5) yields

$$\|\mathcal{S}_e\|_{\mathcal{H}_2}^2 = \text{trace}\Big(B^*\mathcal{G}B - 2B_1^*Z_1B_1 + 2A_{11}\mathcal{G}_1A_{21}^*Z_2 + 2A_{12}\mathcal{G}_2A_{22}^*Z_2 + B_1^*\hat{\mathcal{G}}_oB_1\Big).$$

From (4.4), we have

$$A_{11}^*Z_1A_{11} + A_{21}^*Z_2A_{11} - Z_1 + C_1^*C_1 = 0,$$

and consequently

$$\begin{aligned}\text{trace}(2A_{11}\mathcal{G}_1A_{21}^*Z_2) &= \text{trace}(2\mathcal{G}_1A_{21}^*Z_2A_{11}) \\ &= \text{trace}(-2\mathcal{G}_1A_{11}^*Z_1A_{11} + 2\mathcal{G}_1Z_1 - 2\mathcal{G}_1C_1^*C_1).\end{aligned}$$

Combining this with the definition of the $\mathcal{H}_2$ norm of $\mathcal{S}$ and $\hat{\mathcal{S}}$,

$$\|\mathcal{S}\|_{\mathcal{H}_2}^2 = \text{trace}(B^*\mathcal{G}B) = \text{trace}(C\mathcal{G}C^*),$$

and

$$\|\hat{\mathcal{S}}\|_{\mathcal{H}_2}^2 = \text{trace}\Big(B_1^*\hat{\mathcal{G}}_oB_1\Big) = \text{trace}\Big(C_1\hat{\mathcal{G}}_cC_1^*\Big),$$

gives

$$\begin{aligned}\|\mathcal{S}_e\|_{\mathcal{H}_2}^2 = \text{trace}\Big(&2A_{12}\mathcal{G}_2A_{22}^*Z_2 + C_2\mathcal{G}_2C_2^* - C_1\mathcal{G}_1C_1^* + C_1\hat{\mathcal{G}}_cC_1^*\Big) + \\ &\text{trace}(-2B_1B_1^*Z_1 - 2A_{11}\mathcal{G}_1A_{11}^*Z_1 + 2\mathcal{G}_1Z_1).\end{aligned}$$

The $(1,1)$ block of (4.1) gives

$$A_{11}\mathcal{G}_1A_{11}^* + A_{12}\mathcal{G}_2A_{12}^* - \mathcal{G}_1 + B_1B_1^* = 0,$$

from which it follows that

$$\text{trace}(-2B_1B_1^*Z_1 - 2A_{11}\mathcal{G}_1A_{11}^*Z_1 + 2\mathcal{G}_1Z_1) = \text{trace}(2A_{12}\mathcal{G}_2A_{12}^*Z_1).$$

Finally, we obtain

$$\begin{aligned}\|\mathcal{S}_e\|_{\mathcal{H}_2}^2 &= \text{trace}\left(C_2\mathcal{G}_2C_2^* + C_1(\hat{\mathcal{G}}_c - \mathcal{G}_1)C_1^* + 2A_{12}\mathcal{G}_2\begin{bmatrix} A_{12}^* & A_{22}^*\end{bmatrix}\begin{bmatrix} Z_1 \\ Z_2\end{bmatrix}\right) \\ &= \text{trace}(C_2\mathcal{G}_2C_2^*) + \text{trace}\Big(C_1(\hat{\mathcal{G}}_c - \mathcal{G}_1)C_1^*\Big) + 2\text{trace}\big(A_{12}\mathcal{G}_2\begin{bmatrix} A_{12}^* & A_{22}^*\end{bmatrix}Z\big)\end{aligned}$$

THEOREM 4.1. *Let* $\mathcal{S} = \left\{\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}\end{bmatrix}, \begin{bmatrix} B_1 \\ B_2\end{bmatrix}, \begin{bmatrix} C_1 & C_2\end{bmatrix}\right\}$ *be a balanced system and* $\hat{\mathcal{S}} = \{A_{11}, B_1, C_1\}$ *be the n-truncated model. The* $\mathcal{H}_2$ *norm of the error system is given either by*

$$\|\mathcal{S}_e\|_{\mathcal{H}_2}^2 = \text{trace}(C_2\mathcal{G}_2C_2^*) + \text{trace}\Big(C_1(\hat{\mathcal{G}}_c - \mathcal{G}_1)C_1^*\Big) + 2\text{trace}\big(A_{12}\mathcal{G}_2\begin{bmatrix} A_{12}^* & A_{22}^*\end{bmatrix}Z\big)$$

or

$$\|\mathcal{S}_e\|_{\mathcal{H}_2}^2 = \mathrm{trace}(B_2^*\mathcal{G}_2B_2) + \mathrm{trace}\Big(B_1^*(\hat{\mathcal{G}}_o - \mathcal{G}_1)B_1\Big) + 2\mathrm{trace}\big(A_{12}\mathcal{G}_2\begin{bmatrix} A_{12}^* & A_{22}^* \end{bmatrix}Y\big)$$

where $\mathcal{G}_2$ is the $(N-n)\times(N-n)$ trailing principal submatrix of the unified gramian of $\mathcal{S}$, $\hat{\mathcal{G}}_c$ and $\hat{\mathcal{G}}_o$ are respectively the controllability and the observability gramians of $\hat{\mathcal{S}}$, and Z and Y are the solutions of the Stein equations

$$A^*ZA_{11} - Z + C^*C_1 = 0, \quad AYA_{11}^* - Y + BB_1^* = 0.$$

The second formula is obtained if we used the C matrices instead of the B matrices in the definition of the $\mathcal{H}_2$ norm of the error system (4.5).

From the Cauchy–Schwarz inequality we obtain

$$|\mathrm{trace}(C_2\mathcal{G}_2C_2^*)| \le \sigma_{n+1}\|C_2\|_2^2, \quad \text{where} \quad \sigma_{n+1} = \|\mathcal{G}_2\|_2,$$

$$\Big|\mathrm{trace}\Big(C_1(\hat{\mathcal{G}}_c - \mathcal{G}_1)C_1^*\Big)\Big| \le \|\hat{\mathcal{G}}_c - \mathcal{G}_1\|_2\|C_1\|_2^2,$$

$$|\mathrm{trace}(2A_{12}\mathcal{G}_2A_{:2}^*Z)| \le 2\sigma_{n+1}\|A_{12}\|_2\|A_{:2}\|_2\|Z\|_2.$$

As Z is the solution of the Stein equation (4.4), it has the form

$$Z = \sum_{i=0}^{\infty}(A^*)^iC^*C_1(A_{11})^i,$$

and so

$$\|Z\|_2 \le \|C\|_2^2\sum_{i=0}^{\infty}\|A^i\|_2\|(A_{11})^i\|_2.$$

Moreover, the difference $E := \hat{\mathcal{G}}_c - \mathcal{G}_1$ satisfies the Stein equation

$$A_{11}^*EA_{11} - E + A_{21}^*\mathcal{G}_2A_{21} = 0, \tag{4.6}$$

which yields the formula

$$E = \hat{\mathcal{G}}_c - \mathcal{G}_1 = \sum_{i=0}^{\infty}(A_{11}^*)^iA_{21}^*\mathcal{G}_2A_{21}(A_{11})^i.$$

Finally, we have

$$\|\hat{\mathcal{G}}_c - \mathcal{G}_1\|_2 \le \sigma_{n+1}\sum_{i=0}^{\infty}\|(A_{11})^i\|_2^2\|A_{21}\|_2^2.$$

This analysis yields the following result.

LEMMA 4.2. *The $\mathcal{H}_2$ norm of the error system satisfies the a posteriori bound*

$$\sigma_{n+1} \le \|\mathcal{S}_e\|_{\mathcal{H}_2}^2 := \|\mathcal{S} - \mathcal{S}_n\|_{\mathcal{H}_2}^2 \le c\sigma_{n+1}\|C\|_2^2$$

where

$$c = 1 + 3\|A\|_2^2 \sum_{i=0}^{\infty} \|A^i\|_2 \|(A_{11})^i\|_2.$$

Another bound could be obtained as follows. Reconsider the Stein equations (4.4) and (4.6)

$$A^* Z A_{11} - Z + C^* C_1 = 0, \quad A_{11}^* E A_{11} - E + A_{21}^* \mathcal{G}_2 A_{21} = 0,$$

and let $A = UDU^{-1}$, and $A_{11} = U_1 D_1 U_1^{-1}$ be the eigenvalue decompositions of A and A_{11}. The Stein equations can be rewritten as

$$DU^{-1} Z U_1 D_1 - U^{-1} Z U_1 + U^{-1} C^* C_1 U_1 = 0,$$

and

$$D_1 U_1^{-1} E U_1 D_1 - U_1^{-1} E U_1 + U_1^{-1} A_{21}^* \mathcal{G}_2 A_{21} U_1 = 0.$$

From this, it can be easily seen that

$$\|Z\|_2 \leq \frac{\|C\|_2^2}{1 - \rho(A)\rho(A_{11})}, \quad \|E\|_2 \leq \frac{\sigma_{n+1}\|A_{21}\|_2^2}{1 - \rho(A_{11})^2}, \tag{4.7}$$

where $\rho(\cdot)$ denotes the spectral radius. We have

$$\rho(A) = \max_i |d_{ii}|, \quad \rho(A_{11}) = \max_i |\hat{d}_{ii}|,$$

where $D = (d_{ij})_{i,j=1}^N$ and $D_1 = (\hat{d}_{ij})_{i,j=1}^n$.

LEMMA 4.3. *The $\mathcal{H}_2$ norm of the error system satisfies the a posteriori bound*

$$\sigma_{n+1} \leq \|\mathcal{S}_e\|_{\mathcal{H}_2}^2 := \|\mathcal{S} - \mathcal{S}_n\|_{\mathcal{H}_2}^2 \leq c_1 \sigma_{n+1} \|C\|_2^2$$

where

$$c_1 = 1 + 3\frac{\|A\|_2^2}{1 - \rho(A)^2}.$$

Proof. Recall that

$$\begin{aligned}\|\mathcal{S}_e\|_{\mathcal{H}_2}^2 &= \operatorname{trace}(C_2 \mathcal{G}_2 C_2^*) + \operatorname{trace}\Big(C_1(\hat{\mathcal{G}}_c - \mathcal{G}_1)C_1^*\Big) + 2\operatorname{trace}(A_{12}\mathcal{G}_2 A_{:2}^* Z) \\ &\leq \|C_2\|_2^2 \|\mathcal{G}_2\|_2 + \|C_1\|_2^2 \|E\|_2 + 2\|A_{12}\|_2 \|\mathcal{G}_2\|_2 \|A_{:2}\|_2 \|Z\|_2\end{aligned}$$

And using the bounds (4.7) we have

$$\begin{aligned}\|\mathcal{S}_e\|_{\mathcal{H}_2}^2 &\leq \|C_2\|_2^2 \|\sigma_{n+1}\|_2 + \|C_1\|_2^2 \frac{\sigma_{n+1}\|A_{21}\|_2^2}{1 - \rho(A_{11})^2} + 2\|A_{12}\|_2 \|\mathcal{G}_2\|_2 \|A_{:2}\|_2 \frac{\|C\|_2^2}{1 - \rho(A)\rho(A_{11})} \\ &\leq \|C\|_2^2 \|\sigma_{n+1}\|_2 + \|C\|_2^2 \frac{\sigma_{n+1}\|A\|_2^2}{1 - \rho(A)^2} + 2\|A\|_2^2 \|\sigma_{n+1} \frac{\|C\|_2^2}{1 - \rho(A)^2}\end{aligned}$$

which gives the result. □

4.1. Discussion. First, notice that in Lemmas 4.2 and 4.3, the term $\|C\|_2^2$ could be replaced by $\|B\|_2^2$ as a result of Theorem 4.1. Moreover, as $\{A, B, C\}$ is balanced we have

$$A\Sigma A^T - \Sigma + BB^T = 0, \quad A^T\Sigma A - \Sigma + C^TC = 0,$$

where Σ is diagonal. We can see easily that $BB^T = C^TC$ and so $\|B\|_2 = \|C\|_2$. The discussion will focus then on the results with C.

In Theorem 4.1, the first term trace$(C_2\mathcal{G}_2C_2^*)$ is the $\mathcal{H}_2$-norm of the neglected subsystem of the original system; the second term trace$\Big(C_1(\hat{\mathcal{G}}_c - \mathcal{G}_1)C_1^*\Big)$ is the difference between the $\mathcal{H}_2$-norms of the reduced order system and the dominant subsystem of the original system; finally the third term trace$(A_{12}\mathcal{G}_2A_{:2}^*Z)$ is the inner product of the non-dominant block of the gramian with the Z (non square matrix) weighted by non-dominant submatrices of A. $\mathcal{G}_2$ is diagonal and its spectral norm is supposed to be negligible compared to the spectral norm of $\mathcal{G}_1$. Then as the first and last terms are proportional to $\mathcal{G}_2$, they will be very small; the mid-term has the major contribution to the value of the norm. As $E = \hat{\mathcal{G}}_c - \mathcal{G}_1$ is solution of the Stein equation

$$A_{11}^*EA_{11} - E + A_{21}^*\mathcal{G}_2A_{21} = 0,$$

if either the non dominant gramian $\mathcal{G}_2$ or the off-diagonal block of A are small (zero), then E will be small (zero). As a conclusion, the quality of the reduced model will be function of the smallness of the off-diagonal blocks of A and the smallness of σ_{n+1}, the largest neglected Hankel singular value. The last dependence is known but the first one is quite unusual. It can be interpreted as follow. The reduced order model will be a very good approximation of the original system if and only if firstly there is a gap between the kept Hankel singular values of the original system and the neglected ones and secondly if the truncated states have no major contribution to the dynamics of the other states.

In Lemmas 4.2 and 4.3, if the matrix A is close to normal we will have

$$\|A\|_2 \approx \rho(A) \approx \rho(A_{11}) < 1, \quad \lim_{i\mapsto\infty} \|A^i\|_2 = 0.$$

The two constants c and c_1 should be of the same order in this case. Note that usually the matrix A results from the finite-element method applied to a partial differential equation, which yields in general a matrix that is close to being normal or symmetric. In Lemma 4.2, the matrix Z is a non square matrix solution of a Stein equation. As Z is not symmetric, in some of our numerical tests, the trace of the term involving Z shows an imaginary term nevertheless neglectable. Moreover, the term trace$\Big(C_1(\hat{\mathcal{G}}_c - \mathcal{G}_1)C_1^*\Big)$ even very small could be negative sign. This is related to the still open problem of over-approximation and under-approximation of the gramians.

We end this discussion by discussing the utility of these formulas and bounds, and even more specifically the utility of the discrete case. First, a relationship between the discrete and continuous time $\mathcal{H}_2$ norms can be derived by introducing the relationship between discrete and continuous time gramians. One obtains

$$\|\mathcal{S}_c\|_{\mathcal{H}_2}^2 = \frac{1}{\sqrt{\Delta t}}\|\mathcal{S}_d\|_{\mathcal{H}_2}^2,$$

where $\mathcal{S}_c$ is a continuous system, $\mathcal{S}_d$ its discretization corresponding to the sampling time Δt. As a result of this formula, the discrete time $\mathcal{H}_2$ norm does not converge to the continuous time $\mathcal{H}_2$ norm when the sampling time approaches zero.

One key utility of the discrete case is that the spectral radii of the matrices A and A_{11} are smaller than 1. This is resulting from the stability of both systems: the original and the reduced. If A is close to be normal, this property will make both coefficients c and c_1 in Lemmas 4.2 and 4.3 reasonably small. For c, notice that the terms $\|A^i\|_2$ and $\|A_{11}^i\|_2$ will vanish very quickly as A has its spectral radius smaller than 1 and A_{11} is a sub matrix of A. Both coefficients c and c_1 are only functions of A and A_{11}. Moreover we can bound c as follows:

$$c \leq 1 + 3\|A\|_2^2 \sum_{i=0}^{\infty} \|A^i\|_2^2.$$

This leads to the conclusion that our error bounds are only functions of σ_{n+1}, the matrix A (its 2-norm and spectral radius) and the matrix C. Contrary to the continuous case [1] where one has to consider another residual system and computes its $\mathcal{H}_\infty$-norm. Moreover, the quality of the bound will be only function of the smallness of σ_{n+1} as the term $c\|C\|_2$ is constant and not function of the reduced order system.

Our formula in 4.1 is (like the Antoulas's formula) computable. We use the data already available from balanced truncation and solve a Stein equation for a thin matrix which is still much less expensive than evaluating directly the $\mathcal{H}_2$-norm. The direct evaluation of the $\mathcal{H}_2$-norm, as for example the function `normh2` of MATLAB's Control System Toolbox, means that one has to compute the error system, find a realization of this error system, then solve a Lyapunov or a Stein equation for one gramian in order to evaluate the $\mathcal{H}_2$-norm.

4.2. A special case: square system. For square systems ($m = p$) one can define the cross gramian X of $\mathcal{S}$ as the solution of the Stein equation

$$AXA - X + BC = 0. \tag{4.8}$$

The $\mathcal{H}_2$ norm of the system $\mathcal{S}$ is given in this case by

$$\|\mathcal{S}\|_{\mathcal{H}_2}^2 = \operatorname{trace}(CXB).$$

In this case, the $\mathcal{H}_2$ norm of the error system $\mathcal{S}_e$ (4.2) is

$$\|\mathcal{S}_e\|_{\mathcal{H}_2}^2 = \operatorname{trace}\left(\begin{bmatrix} C & C_1 \end{bmatrix} \begin{bmatrix} X & Y \\ Z & -\hat{X} \end{bmatrix} \begin{bmatrix} B \\ -B_1 \end{bmatrix} \right), \tag{4.9}$$

where Y and Z are solutions of the Stein equations

$$AYA_{11} - Y + BC_1 = 0, \quad A_{11}ZA - Z - B_1C = 0, \tag{4.10}$$

and $\hat{X}$ is the cross gramian of the n reduced system by balanced truncation $\hat{\mathcal{S}}$. $\hat{X}$ is also solution of a Stein equations given by

$$A_{11}\hat{X}A_{11} - \hat{X} + B_1C_1 = 0. \tag{4.11}$$

THEOREM 4.4. *The $\mathcal{H}_2$ norm of the error system is given by*

$$\begin{aligned} \|\mathcal{S}_e\|_{\mathcal{H}_2}^2 = {} & \operatorname{trace}(C_2X_{22}B_2) + \operatorname{trace}\Big(C_1(\hat{X} - X_{11})B_1\Big) + \operatorname{trace}\big(A_{12} \begin{bmatrix} X_{21} & X_{22} \end{bmatrix} AY\big) \\ & - \operatorname{trace}\left(A_{21}ZA \begin{bmatrix} X_{12} \\ X_{22} \end{bmatrix} \right). \end{aligned}$$

Proof. To show this result we need to expand the formula (4.9) as

$$\begin{aligned}\|\mathcal{S}_e\|_{\mathcal{H}_2}^2 = \operatorname{trace}(&C_1X_{11}B_1 + C_2X_{21}B_1 + C_1Z_1B_1 + C_1X_{12}B_2 + C_2X_{22}B_2)\\ &+\operatorname{trace}\Big(C_1Z_2B_2 - C_1Y_1B_1 - C_2Y_2B_1 + C_1\hat{X}B_1\Big).\end{aligned} \tag{4.12}$$

From the (1,2) and (2,1) blocks of (4.8) we have respectively

$$B_2C_1Z_2 = \left(X_{21} - A_{21}X_{11}A_{11} - A_{22}X_{21}A_{11} - A_{21}X_{12}A_{21} - A_{22}X_{22}A_{21}\right)Z_2,$$

and

$$B_1C_2Y_2 = \left(X_{12} - A_{11}X_{11}A_{12} - A_{12}X_{21}A_{12} - A_{11}X_{12}A_{22} - A_{12}X_{22}A_{22}\right)Y_2.$$

Then from the second blocks of the equations (4.10) we have

$$\left(A_{11}X_{12}A_{22} - X_{12}\right)Y_2 = -X_{12}A_{21}Y_1A_{11} - X_{12}B_2C_1,$$

and

$$\left(X_{21} - A_{22}X_{21}A_{11}\right)Z_2 = X_{21}A_{11}Z_1A_{12} - X_{21}B_1C_2.$$

Collecting all this in the formula (4.12) we get

$$\begin{aligned}\|\mathcal{S}_e\|_{\mathcal{H}_2}^2 = \operatorname{trace}&\Big(C_1X_{11}B_1 + C_1Z_1B_1 + C_2X_{22}B_2 - C_1Y_1B_1 + C_1\hat{X}B_1\Big)\\ &-\operatorname{trace}(A_{21}X_{11}A_{11}Z_2 - A_{21}X_{12}A_{21}Z_2 - A_{22}X_{22}A_{21}Z_2 + A_{11}X_{11}A_{12}Y_2)\\ &+\operatorname{trace}(A_{12}X_{21}A_{12}Y_2 + A_{12}X_{22}A_{22}Y_2 - X_{12}A_{21}Y_1A_{11} + X_{21}A_{11}Z_1A_{12}).\end{aligned} \tag{4.13}$$

From the (1,1) block of (4.8) we have

$$B_1C_1 = X_{11} - A_{11}X_{11}A_{11} - A_{12}X_{21}A_{11} - A_{11}X_{12}A_{21} - A_{12}X_{22}A_{21}$$

Injecting this in (4.13) and using the first leading blocks of (4.10), i.e.,

$$Z_1 - A_{11}Z_1A_{11} - A_{11}Z_2A_{21} = -B_1C_1,$$

and

$$-Y_1 + A_{11}Y_1A_{11} + A_{12}Y_2A_{11} = -B_1C_1,$$

we get finally

$$\begin{aligned}\|\mathcal{S}_e\|_{\mathcal{H}_2}^2 &= \operatorname{trace}\Big(-C_1X_{11}B_1 + C_2X_{22}B_2 + C_1\hat{X}B_1 - A_{11}X_{12}A_{21}Z_1 - A_{12}X_{22}A_{21}Z_1\Big)\\ &\quad+\operatorname{trace}(A_{12}X_{21}A_{11}Y_1 + A_{12}X_{22}A_{21}Y_1 - A_{21}X_{12}A_{21}Z_2 - A_{22}X_{22}A_{21}Z_2)\\ &\quad+\operatorname{trace}(A_{12}X_{21}A_{12}Y_2 + A_{12}X_{22}A_{22}Y_2)\\ &= \operatorname{trace}\Big(C_1(\hat{X} - X_{11})B_1 + C_2X_{22}B_2 - A_{21}Z_1A_{11}X_{12} - A_{21}Z_1A_{12}X_{22}\Big)\\ &\quad-\operatorname{trace}(A_{21}Z_2A_{21}X_{12} - A_{21}Z_2A_{22}X_{22} + A_{12}X_{21}A_{11}Y_1 + A_{12}X_{22}A_{21}Y_1)\\ &\quad+\operatorname{trace}(A_{12}X_{21}A_{12}Y_2 + A_{12}X_{22}A_{22}Y_2)\\ &= \operatorname{trace}\left(C_1(\hat{X} - X_{11})B_1 + C_2X_{22}B_2 - A_{21}\begin{bmatrix} Z_1 & Z_2\end{bmatrix}\begin{bmatrix} A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix} X_{12}\\ X_{22}\end{bmatrix}\right)\\ &\quad+\operatorname{trace}\left(A_{12}\begin{bmatrix} X_{21} & X_{22}\end{bmatrix}\begin{bmatrix} A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix} Y_1\\ Y_2\end{bmatrix}\right),\end{aligned}$$

which proves the result. □

In this theorem, the first term is the $\mathcal{H}_2$-norm of the neglected subsystem of the original system; the second term is the difference between the $\mathcal{H}_2$-norms of the reduced order system and the dominant subsystem of the original system; finally the third term is the difference of the inner product of the second block row of the cross gramian with Y and that of Z with the second block column of the cross gramian (each term weighted by the block off-diagonal terms of A and A).

Notice that the difference $\hat{X} - X_{11}$ satisfies the Stein equation

$$A_{11}(X_{11} - \hat{X})A_{11} - (X_{11} - \hat{X}) + A_{12}X_{21}A_{11} + A_{11}X_{12}A_{21} + A_{12}X_{22}A_{21} = 0,$$

if the cross gramian is block diagonal, i.e., $X_{12} = 0$ and $X_{21} = 0$. The first consequence of this assumption is that $X_{11} = \hat{X}$, hence the second term vanishes. As for the last term it becomes $A_{12}X_{22}A_{2:}Y - A_{21}ZA_{:2}X_{22}$. The previous theorem becomes

COROLLARY 4.5. *If the cross gramian is block diagonal, the $\mathcal{H}_2$ norm of the error system is given by*

$$\|\mathcal{S}_e\|^2_{\mathcal{H}_2} = \operatorname{trace}(C_2X_{22}B_2) + \operatorname{trace}\Big(C_1(\hat{X} - X_{11})B_1\Big) + \operatorname{trace}(A_{12}X_{22}A_{2:}Y - A_{21}ZA_{:2}X_{22}).$$

Using the same analysis as the previous section (for Lemmas 4.2 and 4.3) we obtain the following results.

LEMMA 4.6. *The $\mathcal{H}_2$ norm of the error system satisfies the following a posteriori bound*

$$\sigma_{n+1} \leq \|\mathcal{S}_e\|^2_{\mathcal{H}_2} := \|\mathcal{S} - \mathcal{S}_n\|^2_{\mathcal{H}_2} \leq c\sigma_{n+1}\|C\|_2\|B\|_2$$

where

$$c = 1 + 3\|A\|_2^2 \sum_{i=0}^{\infty} \|A^i\|_2 \|(A_{11})^i\|_2.$$

LEMMA 4.7. *The $\mathcal{H}_2$ norm of the error system satisfies the following a posteriori bound*

$$\sigma_{n+1} \leq \|\mathcal{S}_e\|^2_{\mathcal{H}_2} := \|\mathcal{S} - \mathcal{S}_n\|^2_{\mathcal{H}_2} \leq c_1\sigma_{n+1}\|C\|_2\|B\|_2$$

where

$$c_1 = \left(1 + 3\frac{\|A\|_2^2}{1 - \rho(A)^2}\right).$$

Here also our error bounds are only functions of the matrix A (its 2-norm and spectral radius) and the matrices B and C. We will illustrate later all this discussion in the numerical examples.

5. Concluding remarks. We have reviewed the most used projection based method in model reduction of linear time-invariant dynamical systems, balanced truncation. Moreover, we have presented computable error formulas and bounds for the response approximation. The advantage of these results is that we are using the already given results by balanced truncation and we don't need anything else. This has the feature that it can be included into the order reduction loop in order to improve

the quality of the reduced order model by choosing the optimal reduced order before ending the model reduction algorithm.

Many open questions remain. Particularly, what will be the expression for the error bounds if instead of balanced truncation method we use some other projection method? In other words, given two projection matrices $\Pi_r, \Pi_l \in R^{N\times n}$, with $\Pi_l^T\Pi_r = I_n$, the reduced-order system is given by $\{\Pi_l^T A\Pi_r, \Pi_l^T B, C\Pi_r\}$ and can we substitute A_{12} by $\Pi_l^T A\hat{\Pi}_r$, $\mathcal{G}_2$ by $\hat{\Pi}_l^T\mathcal{G}\Pi_r$, and so on, in all presented formulas, where $\hat{\Pi}_r$ and $\hat{\Pi}_l$ are defined from Π_l and Π_r?

Acknowledgements. I gratefully acknowledge the helpful remarks and suggestions of Nick Higham and Françoise Tisseur which significantly improved the presentation of this paper.

REFERENCES

[1] Antoulas, A. C. *Approximation of Large-Scale Dynamical Systems.* SIAM, Philadelphia, USA, 2005.

[2] Benner, P., Castillo, M., Quintana-Ortí, E. S., and Hernández, V. Parallel partial stabilizing algorithms for large linear control systems. *J. Supercomput. 615*, 2 (2000), 193–206.

[3] Benner, P., Quintana-Ortí, E. S., and Quintana-Ortí, G. Parallel Algorithms for Model Reduction of Discrete-Time Systems. *International Journal of System Sciences 34*, 5 (2003), 319–333.

[4] Enns, D. F. Model reduction with balanced realizations: An error bound and frequency weighted generalization. *Proc. of the IEEE Conference on Decision and Control* (1981), 127–132.

[5] Glover, K. All optimal Hankel norm approximations of linear multivariable systems and their $\mathcal{L}^\infty$-error bounds. *Internat. J. Control 39* (1984), 1115–1193.

[6] Hammarling, S. J. Numerical solution of the stable, non-negative definite Lyapunov equation. *Eds. R. V. Patel, A. J. Laub, and P. Van Dooren, Numerical Linear Algebra Techniques for Systems and Control, IEEE Press, New York, NY, USA* (1994), 500–516.

[7] Moore, B. C. Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Automat. Control 26* (1981), 17–31.

[8] Pernebo, L., and Silverman, L. M. Model reduction via balanced state space representations. *IEEE Trans. Automat. Control 27*, 2 (1982), 382–387.

[9] Safonov, M. G., and Chiang, R. Y. A Schur method for balanced-truncation model reduction. *IEEE Trans. Automat. Control 34(7)* (1989), 729–733.

[10] Zhou, K., Doyle, J. C., and Glover, K. *Robust and optimal control.* Prentice Hall, 1995.