International Journal of Systems Science, 2002, volume 33, number 10, pages 847–854

Computation of reduced-order models of multivariable systems by balanced truncation

J. S. GARCIA† and J. C. BASILIO‡*

Previous algorithms to obtain reduced-order models by balanced truncation in a single step either require a very specific way to solve a pair of Lyapunov equations or are suitable only for scalar or symmetric MIMO systems. In this paper, model reduction is revisited and an algorithm to obtain a reduced order model in one step only is proposed. As in the previous algorithms, the key point is to construct two rectangular matrices whose smaller dimensions are equal to the number of Hankel singular values to be kept in the lower model. Unlike the one-step algorithms available in the literature, the algorithm proposed here does not make any restriction to the way the Lyapunov equations necessary to obtain the controllability and observability gramians are solved. Furthermore, since the algorithm only relies on singular value decomposition, it is expected to be robust.

1. Introduction

Balanced realization (Moore 1981) has been proved crucial in model reduction (Glover 1984) and also in the computation of H_∞ optimal controllers in the 1984 approach (Doyle 1984). The idea behind its use in model reduction is to measure the degree of controllability and observability of the system modes and then to discard those modes which are weakly controllable or observable. The computation of reduced-order models by balanced truncation for non-minimal order systems was initially carried out in three steps: (1) computation of a minimal realization for the system; (2) construction of a similarity transformation that relates the state-space realization obtained in step (1) to a balanced realization (Moore 1981, Laub *et al*. 1987 and references therein); and (3) for a given error bound, balanced truncation is deployed to reduce the system order (Glover 1984). This three-step approach has the drawback that a minimal order realization has to be found whose computation is known to be problematic. To avoid such a difficulty, Tombs and Postlethwaite (1987) proposed an algorithm to compute a lower-order model in a single step. Despite its usefulness, this algorithm has the disadvantage that it requires a particular way (Hammarling 1982) to solve the Lyapunov equations necessary to find the controllability and observability gramians of the system. More recently, Aldhaheri (1991) has proposed another single-step algorithm based on the computation of the eigenvectors associated with the largest eigenvalues (in modules) of the cross-gramian W_{co} (Fernando and Nicholson 1982, 1985). The drawbacks of this approach are that the cross-gramian requires the system to be either scalar or symmetric (in the multivariable case), and the realization obtained is not balanced. In common, Tombs and Postlethwaite (1987) and Aldhaheri (1991) have the fact that the reduced-order model is obtained via pre- and post-multiplication of the state matrix by rectangular matrices.

In this paper, an algorithm is proposed to obtain a reduced-order model for a non-minimal state-space realization, whose key point, as in Tombs and Postlethwaite (1987) and Aldhaheri (1991), is the construction of two rectangular matrices whose smaller dimension corresponds to the number of Hankel singular values to be kept in the lower model. Differently from the previous algorithms, it does not make any restriction on the way the Lyapunov equations, necessary to compute the gramians, are solved, and is suitable

Received 25 July, 2000. Revised 1 May 2002. Accepted 31 May 2002.

† Instituto Militar de Engenharia, Departamento de Engenharia Elétrica, Praça General Tibúrcio, 80 – Praia Vermelha, 22.290-270 Rio de Janeiro R.J. Brazil.

‡ Universidade Federal do Rio de Janeiro, Escola de Engenharia, Departamento de Eletrotécnica, Cidade Universitária, Ilha do Fundão, 21.945-970 Rio de Janeiro, R.J., Brazil.

* To whom correspondence should be addressed. e-mail: basilio@-vishnu.coep.ufrj.br

International Journal of Systems Science ISSN 0020–7721 print/ISSN 1464–5319 online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0020772021000017308

for both scalar and multivariable systems. Furthermore, since the algorithm relies only on singular value decomposition, it is expected to be robust.

This paper is structured as follows. In Section 2, the problem of finding reduced-order models by balanced truncation is reviewed and, in the sequel, the problem of obtaining directly a balanced realization for the reduced-order model of a given non-minimal order realization is formulated. Some preliminary mathematical results are presented in Section 3. The main result is given in Section 4, where a rectangular matrix and its right-inverse are constructed. The balanced reduced-order system will be obtained by appropriate pre- and post-multiplications by these rectangular matrices. The paper results are summarized in Section 5, where an algorithm is presented. In Section 6, the results are illustrated by means of a numerical example. Finally, conclusions are drawn is Section 7.

2. Problem formulation

Assume that a $p \times m$ stable transfer matrix $G(s)$ has the following state–space realization:

$$G(s) = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right], \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$ and m, n, $p \in \mathbb{N}^*$ with $(A, B)/(C, A)$ are possibly uncontrollable/unobservable. In addition, let W_c and W_o denote, respectively, the controllability and observability gramians of (1), i.e. W_c and W_o are solutions of the following Lyapunov equations:

$$AW_c + W_cA^{\mathrm{T}} = -BB^{\mathrm{T}} \tag{2}$$

$$A^{\mathrm{T}}W_o + W_oA = -C^{\mathrm{T}}C. \tag{3}$$

It is well known that since $G(s)$ is stable, then W_c and W_o are positive semidefinite. It is also known that although the gramians are not invariant under similarity transformation, their product is invariant in the sense that the eigenvalues of the product of the gramians remain the same no matter what state–space realization is being used. In the control literature, the square roots of the non-zero eigenvalues of W_cW_o are usually referred to as the Hankel singular values of $G(s)$.

Consider now the eigenvalue decomposition of W_cW_o (Zhou *et al.* 1996, p. 77):

$$W_cW_o = W\begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix}W^{-1},$$

where

$$\Sigma^2 = \begin{bmatrix} \Sigma_L^2 & & \\ & \Sigma_s^2 & \\ & & 0 \end{bmatrix},$$

and where $\Sigma_L^2 = \mathrm{diag}\ \{\sigma_1^2 I_{m1},\ \sigma_2^2 I_{m2}, \ldots, \sigma_{\bar{r}}^2 I_{m_{\bar{r}}}\}$ and $\Sigma_s^2 = \mathrm{diag}\ \{\sigma_{\bar{r}+1}^2 I_{m_{\bar{r}+1}},\ \sigma_{\bar{r}+2}^2 I_{m_{\bar{r}+2}}, \ldots, \sigma_{\bar{k}}^2 I_{m_{\bar{k}}}\}$, with $\sigma_i > 0$, $i = 1, 2, \ldots, \bar{k}$ and $\sigma_i > \sigma_j$, $i < j$. Note that Σ_L and Σ_s are formed, respectively, with the largest and smallest Hankel singular values of $G(s)$, i.e. those which are to be kept and discarded in the model reduction. Suppose that we are interested in obtaining a reduced-order model $\tilde{G}(s)$ for $G(s)$ such that the error between $\tilde{G}(s)$ and $G(s)$ is $\leq 2(\sigma_{\bar{r}+1} + \sigma_{\bar{r}+2} + \cdots + \sigma_{\bar{k}})$ in H_∞ sense, namely:

$$e = \|G - \tilde{G}\|_\infty \leq 2(\sigma_{\bar{r}+1} + \sigma_{\bar{r}+2} + \cdots + \sigma_{\bar{k}}). \tag{4}$$

Then, the problem of finding a balanced realization directly from the state–space representation (1) for the reduced order model $\tilde{G}(s)$ of $G(s)$ can be stated as follows: find rectangular matrices $T_L \in \mathbb{R}^{r\times n}$ and $T_L^\dagger \in \mathbb{R}^{n\times r}$, $r = m_1 + m_2 + \cdots + m_{\bar{r}}$, $T_L T_L^\dagger = I_r$ (I_r denoting the identity matrix of order r), such that

$$\tilde{G}(s) = \left[\begin{array}{c|c} \tilde{A} & \tilde{B} \\ \hline \tilde{C} & \tilde{D} \end{array}\right] = \left[\begin{array}{c|c} T_LAT_L^\dagger & T_LB \\ \hline CT_L^\dagger & D \end{array}\right] \tag{5}$$

with controllability and observability gramians being given by:

$$\Sigma_L = \mathrm{diag}\,\{\sigma_1 I_{m_1}, \sigma_2 I_{m_2}, \ldots, \sigma_{\bar{r}} I_{m_{\bar{r}}}\}. \tag{6}$$

3. Preliminary mathematical results

In this section, the following problem is considered: given a matrix $T_1 \in \mathbb{R}^{k\times n}$ and its right inverse $T_1^\dagger \in \mathbb{R}^{n\times k}$ $(k < n)$, with T_1 full row rank, i.e. $\rho(T_1) = k$, where $\rho(\cdot)$ denotes rank, find matrices $T_2 \in \mathbb{R}^{n-k\times n}$ and $T_2^\dagger \in \mathbb{R}^{n\times n-k}$ such that

$$\begin{bmatrix} T_1 \\ T_2 \end{bmatrix}[T_1^\dagger \quad T_2^\dagger] = I_n.$$

To solve this problem, a preliminary result is needed.

Lemma 1: *The matrix $I_n - T_1^\dagger T_1$ is diagonalizable, and has $n-k$ eigenvalues $= 1$ and k eigenvalues $= 0$.*

Proof: Since $T_1T_1^\dagger = I_k$ then (Horn and Johnson 1985, p. 53) $T_1^\dagger T_1$ has k eigenvalues $= 1$ and $n-k$ zero eigenvalues. This implies that $\rho(T_1^\dagger T_1) = k$ and consequently $\nu(T_1^\dagger T_1) = n-k$, where $\nu(.)$ denotes nullity. Thus, the dimension of the invariant subspace of $T_1^\dagger T_1$ associated with the zero eigenvalues is $n-k$. In addition, since $T_1T_1^\dagger = I_k$, then $(T_1^\dagger T_1)T_1^\dagger = I_kT_1^\dagger$, which shows that the columns of $T_1^\dagger$ generates the invariant subspace of $T_1^\dagger T_1$ associated with the eigenvalues which are equal to

1. These facts allow us to conclude that $T_1^\dagger T_1$ has the following eigenvalue decomposition:

$$T_1^\dagger T_1 = W \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} W^{-1}$$

for some full rank matrix W. Therefore:

$$I_n - T_1^\dagger T_1 = I_n - W \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} W^{-1} = W \begin{bmatrix} 0_k & 0 \\ 0 & I_{n-k} \end{bmatrix} W^{-1}, \quad (7)$$

which proves the lemma. □

The use of Lemma 1 leads to the following result.

Lemma 2: *Given T_1 and $T_1^\dagger$ satisfying the conditions above, there exist two matrices $T_2 \in \mathbb{R}^{(n-k)\times n}$ and $T_2^\dagger \in \mathbb{R}^{n\times(n-k)}$ such that:*

$$\begin{bmatrix} T_1 \\ T_2 \end{bmatrix} [T_1^\dagger \quad T_2^\dagger] = \begin{bmatrix} T_1 T_1^\dagger & T_1 T_2^\dagger \\ T_2 T_1^\dagger & T_2 T_2^\dagger \end{bmatrix} = I. \quad (8)$$

Proof: From (7), we may write:

$$I_n - T_1^\dagger T_1 = [W_1 \quad W_2] \begin{bmatrix} 0_k & 0 \\ 0 & I_{n-k} \end{bmatrix} \begin{bmatrix} V_1^{\mathrm{T}} \\ V_2^{\mathrm{T}} \end{bmatrix},$$

where W_1, $V_1 \in \mathbb{R}^{n\times k}$. W_2, $V_2 \in \mathbb{R}^{n\times(n-k)}$ and $W_1 V_1^{\mathrm{T}} + W_2 V_2^{\mathrm{T}} = I_n$. Defining $T_2 = V_2^{\mathrm{T}}$ and $T_2^\dagger = W_2$, we have:

(1) $T_1 T_1^\dagger = I_k$, by definition;

(2) $T_2 T_2^\dagger = V_2^{\mathrm{T}} W_2 = I_{n-k}$;

(3) $(I_n - T_1^\dagger T_1) T_2^\dagger = (I_n - T_1^\dagger T_1) W_2 = W_2 = T_2^\dagger$. Therefore $T_1(I_n - T_1^\dagger T_1) T_2^\dagger = T_1 T_2^\dagger$, which implies that $T_1 T_1^\dagger T_1 T_2^\dagger = 0$ or equivalently $T_1 T_2^\dagger = 0$.

(4) $T_2(I_n - T_1^\dagger T_1) = V_2^{\mathrm{T}}(I_n - T_1^\dagger T_1) = V_2^{\mathrm{T}} = T_2$. Thus, $T_2(I_n - T_1^\dagger T_1) T_1^\dagger = T_2 T_1^\dagger$ and proceeding as in (3) we obtain $T_2 T_1^\dagger = 0$, which completes the proof.

□

Remark 1: Note from Lemma 2 that given a full row rank rectangular matrix T_1 and its right-inverse $T_1^\dagger$, the construction of a square matrix T, whose first k rows are T_1, and its inverse T^{-1}, whose first k columns are $T_1^\dagger$, is not a matter of adding a bottom matrix T_2 whose rows are linearly independent on the rows of T_1. This is so because the right-inverse $T_1^\dagger$ is also given and, hence, as stated in the lemma, the rows of T_2 must lie in the left null space of $T_1^\dagger$, and the columns of its right-inverse $T_2^\dagger$ must lie in the right null space of T_1. This is achieved, as shown in the lemma, by taking, respectively, the eigenvectors and dual-eigenvectors of $I_n - T_1 T_1^\dagger$ associated with the unity eigenvalues. □

4. Main results

In this section, we will initially obtain an expression for two matrices $T_1 \in \mathbb{R}^{k\times n}$ and $T_1^\dagger \in \mathbb{R}^{n\times k}$ $(T_1 T_1^\dagger = I_k)$, which leads to a minimal realization in balanced form for the non-minimal state–space realization given in (1), i.e.

$$G(s) = \left[\begin{array}{c|c} A_b & B_b \\ \hline C_b & D_b \end{array}\right] = \left[\begin{array}{c|c} T_1 A T_1^\dagger & T_1 B \\ \hline C T_1^\dagger & D \end{array}\right], \quad (9)$$

where $T_1 A T_1^\dagger$ has all the controllable and observable modes of $G(s)$. At this point, it is important to find the relationship between the modes of a given realization and the eigenvalues of W_c and W_o, as far as controllability and observability are concerned. This is given by the following results.

Lemma 3: *Let*

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

be a state–space realization of a not-necessarily stable transfer matrix $G(s)$ and assume that there exists a symmetric matrix P,

$$P = P^* = \begin{bmatrix} P_1 & 0 \\ 0 & 0 \end{bmatrix},$$

solution to the Lyapunov equation

$$AP + PA^* + BB^* = 0,$$

with P_1 non-singular. If we partition (A, B, C, D) compatibly with P, i.e.

$$\left[\begin{array}{cc|c} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ \hline C_1 & C_2 & D \end{array}\right],$$

then

$$\left[\begin{array}{c|c} A_{11} & B_1 \\ \hline C_1 & D \end{array}\right]$$

is also a realization of $G(s)$. Moreover, (A_{11}, B_1) is controllable if A_{11} is stable.

Proof: See Zhou *et al.* (1996, pp. 72, 73). □

Lemma 4: *Let*

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

be a state–space realization of a not-necessarily stable transfer matrix $G(s)$ and assume that there exists a symmetric matrix Q,

$$Q = Q^* = \begin{bmatrix} Q_1 & 0 \\ 0 & 0 \end{bmatrix}$$

solution to the Lyapunov equation

$$QA + A^*Q + C^*C = 0,$$

with Q_1 non-singular. If we partition (A, B, C, D) compatibly with Q, i.e.

$$\left[\begin{array}{cc|c} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ \hline C_1 & C_2 & D \end{array}\right],$$

then

$$\left[\begin{array}{c|c} A_{11} & B_1 \\ \hline C_1 & D \end{array}\right]$$

is also a realization of $G(s)$. Moreover, (C_1, A_{11}) is observable if A_{11} is stable.

Proof: See Zhou *et al.* (1996, p. 73). □

From Lemmas 3 and 4, it is possible to conclude that the number of non-hidden modes of $G(s)$ is equal to the number of non-zero eigenvalues of W_cW_o, the product of the controllability and observability gramians associated with the non-minimal realization (1). It is necessary therefore to compute the eigenvalues of W_cW_o, which can be done in a more robust way as follows.

Lemma 5: *The Hankel singular values of a stable $G(s)$ are identical to the non-zero singular values of $W_o^{1/2}W_c^{1/2}$.*

Proof: Let $\sigma(.)$, $\lambda(.)$ denote singular values and eigenvalues, respectively. Then:

$$\begin{aligned}\sigma(W_o^{1/2}W_c^{1/2}) &= \lambda^{1/2}(W_o^{1/2}W_c^{1/2}W_c^{1/2}W_o^{1/2}) \\ &= \lambda^{1/2}(W_o^{1/2}W_cW_o^{1/2}) = \lambda^{1/2}(W_cW_o).\end{aligned}$$

To complete the proof, note that the Hankel singular values are the square root of the non-zero eigenvalues of the product W_cW_o. □

Remark 2: At this point it is important to note that since W_c and W_o are symmetric positive semidefinite matrices, their square roots can be simply computed by finding the corresponding eigenvalue decomposition (or, equivalently, the singular value decomposition) and squaring down the eigenvalues (singular values). □

Besides being a robust way to compute the eigenvalues of W_cW_o, the singular value decomposition of $W_o^{1/2}W_c^{1/2}$ also plays an important role in the construction of the matrices T_1 and $T_1^\dagger$ as shown below.

Theorem 1: *Let $G(s)$ be a stable transfer matrix and assume that (1) is any state–space representation of $G(s)$ with gramians W_c and W_o. In addition, suppose that W_cW_o has the following eigenvalue decomposition (Zhou et al. 1996, p. 77):*

$$W_cW_o = W\begin{bmatrix}\Sigma^2 & 0 \\ 0 & 0\end{bmatrix}W^{-1}, \quad (10)$$

or, equivalently (Lemma 5), that $W_o^{1/2}W_c^{1/2}$ has the following singular value decomposition:

$$W_o^{1/2}W_c^{1/2} = [X_1 \quad X_2]\begin{bmatrix}\Sigma & 0 \\ 0 & 0\end{bmatrix}\begin{bmatrix}Y_1^{\mathrm{T}} \\ Y_2^{\mathrm{T}}\end{bmatrix} = X_1\Sigma Y_1^{\mathrm{T}}. \quad (11)$$

Defining

$$T_1 = \Sigma^{-1/2}X_1^{\mathrm{T}}W_o^{1/2} \text{ and } T_1^\dagger = W_c^{1/2}Y_1\Sigma^{-1/2}, \quad (12)$$

then the realization (9) is minimal, and has controllability and observability gramians both equal to Σ.

Proof: For T_1 and $T_1^\dagger$, defined in (12), find two matrices T_2 and $T_2^\dagger$ (Lemma 2) and construct a similarity transformation matrix T and its inverse T^{-1}, as follows:

$$T = \begin{bmatrix}T_1 \\ T_2\end{bmatrix} \quad \text{and} \quad T^{-1} = [T_1^\dagger \quad T_2^\dagger]. \quad (13)$$

Thus

$$G(s) = \left[\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \hat{D} \end{array}\right] = \left[\begin{array}{c|c} TAT^{-1} & TB \\ \hline CT^{-1} & D \end{array}\right]$$

has controllability and observability gramians, $\hat{W}_c$ and $\hat{W}_o$, respectively, given by:

$$\begin{aligned}\hat{W}_c &= \begin{bmatrix}T_1 \\ T_2\end{bmatrix}W_c[T_1^{\mathrm{T}} \quad T_2^{\mathrm{T}}] = \begin{bmatrix}T_1W_cT_1^{\mathrm{T}} & T_1W_cT_2^{\mathrm{T}} \\ T_2W_cT_1^{\mathrm{T}} & T_2W_cT_2^{\mathrm{T}}\end{bmatrix} \\ &= \begin{bmatrix}\Sigma & 0 \\ 0 & F\end{bmatrix} \\ \hat{W}_o &= \begin{bmatrix}(T_1^\dagger)^{\mathrm{T}} \\ (T_2^\dagger)^{\mathrm{T}}\end{bmatrix}W_o[T_1^\dagger \quad T_2^\dagger] \\ &= \begin{bmatrix}(T_1^\dagger)^{\mathrm{T}}W_oT_1^\dagger & (T_1^\dagger)^{\mathrm{T}}W_oT_2^\dagger \\ (T_2^\dagger)^{\mathrm{T}}W_oT_1^\dagger & (T_2^\dagger)^{\mathrm{T}}W_oT_2^\dagger\end{bmatrix} = \begin{bmatrix}\Sigma & 0 \\ 0 & L\end{bmatrix},\end{aligned}$$

where $F = T_2W_cT_2^{\mathrm{T}}$ and $L = (T_2^\dagger)^{\mathrm{T}}W_oT_2^\dagger$. Note that

$$\begin{aligned}FL &= T_2W_cT_2^{\mathrm{T}}(T_2^\dagger)^{\mathrm{T}}W_oT_2^\dagger \\ &= T_2W_c^{1/2}W_c^{1/2}T_2^{\mathrm{T}}(T_2^\dagger)^{\mathrm{T}}W_o^{1/2}W_o^{1/2}T_2^\dagger \\ &= T_2W_c^{1/2}(W_o^{1/2}T_2^\dagger T_2W_c^{1/2})^{\mathrm{T}}W_o^{1/2}T_2^\dagger = 0.\end{aligned}$$

since, according to Lemma 2 and equation (12),

$$\begin{aligned} W_o^{1/2} T_2^{\dagger} T_2 W_c^{1/2} &= W_o^{1/2}(I_n - T_1^{\dagger} T_1) W_c^{1/2} \\ &= W_o^{1/2} W_c^{1/2} - W_o^{1/2} T_1^{\dagger} T_1 W_c^{1/2} \\ &= W_o^{1/2} W_c^{1/2} - W_o^{1/2} W_c^{1/2} Y_1 \\ &\quad \times \Sigma^{-1/2} \Sigma^{-1/2} X_1^{\mathrm{T}} W_o^{1/2} W_c^{1/2} = 0 \end{aligned}$$

Let us now partition $\hat{A}$, $\hat{B}$ and $\hat{C}$ compatibly with T as follows:

$$G(s) = \left[\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \hat{D} \end{array}\right] = \left[\begin{array}{cc|c} \hat{A}_{11} & \hat{A}_{12} & \hat{B}_1 \\ \hat{A}_{21} & \hat{A}_{22} & \hat{B}_2 \\ \hline \hat{C}_1 & \hat{C}_2 & \hat{D} \end{array}\right],$$

where $\hat{A}_{11} = T_1 A T_1^{\dagger}$, $\hat{B}_1 = T_1 B$ and $\hat{C}_1 = C T_1^{\dagger}$. To prove that A_b, B_b and C_b, given in (9), are, respectively, equal to $\hat{A}_{11}$, $\hat{B}_1$ and $\hat{C}_1$, given above, note that since $FL = 0$ then one of the following possibilities must occur:

(1) Either L or F is identically zero. In this case, the result follows directly by application of Lemmas 3 and 4.

(2) The matrices L and F are both non-identically zero. This implies that the columns of L must lie in the right null space of F or, equivalently, the rows of F must lie in the left null space of L. This fact implies that F has the following eigenvalue decomposition:

$$F = U_F^{\mathrm{T}} \begin{bmatrix} \Lambda_f & 0 \\ 0 & 0 \end{bmatrix} U_F, \tag{14}$$

where $\Lambda_F = \mathrm{diag}\{\lambda_1, \lambda_2, \dots, \lambda_f\}$, $\lambda_i \geq \lambda_j > 0$, $i < j$ and $f < n - k$.

Consider now the similarity transformation:

$$\bar{T} = \begin{bmatrix} I_k & 0 \\ 0 & U_F \end{bmatrix} \quad \text{and} \quad \bar{T}^{-1} = \begin{bmatrix} I_k & 0 \\ 0 & U_F^{\mathrm{T}} \end{bmatrix}.$$

Therefore

$$\begin{aligned} G(s) &= \left[\begin{array}{c|c} \bar{T}\hat{A}\bar{T}^{-1} & T\hat{B} \\ \hline \hat{C}\bar{T}^{-1} & \hat{D} \end{array}\right] \\ &= \left[\begin{array}{cc|c} \hat{A}_{11} & \hat{A}_{12} U_F^{\mathrm{T}} & \hat{B}_1 \\ U_F \hat{A}_{21} & U_F \hat{A}_{22} U_F^{\mathrm{T}} & U_F \hat{B}_2 \\ \hline \hat{C}_1 & \hat{C}_2 U_F^{\mathrm{T}} & \hat{D} \end{array}\right] \end{aligned} \tag{15}$$

has gramians

$$\bar{W}_c^{(1)} = \bar{T}\hat{W}_c \bar{T}^{\mathrm{T}} = \begin{bmatrix} \Sigma & 0 \\ 0 & U_F F U_F^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Lambda_F & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$\begin{aligned} \bar{W}_o^{(1)} &= (\bar{T}^{-1})^{\mathrm{T}} \hat{W}_o \bar{T}^{-1} = \begin{bmatrix} \Sigma & 0 \\ 0 & U_F L U_F^{\mathrm{T}} \end{bmatrix} \\ &= \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \bar{L}_{11} & \bar{L}_{12} \\ 0 & \bar{L}_{21} & \bar{L}_{22} \end{bmatrix}. \end{aligned}$$

Partitioning the realization (15) compatibly with $\bar{W}_c^{(1)}$, gives:

$$G(s) = \left[\begin{array}{ccc|c} \hat{A}_{11} & \bar{A}_{12} & \bar{A}_{13} & \hat{B}_1 \\ \bar{A}_{21} & \bar{A}_{22} & \bar{A}_{23} & \bar{B}_2 \\ \bar{A}_{31} & \bar{A}_{32} & \bar{A}_{33} & \bar{B}_3 \\ \hline \hat{C}_1 & \bar{C}_2 & \bar{C}_3 & \hat{D} \end{array}\right],$$

and applying Lemma 3 to the realization above, we obtain:

$$G(s) = \left[\begin{array}{cc|c} \hat{A}_{11} & \bar{A}_{12} & \hat{B}_1 \\ \bar{A}_{21} & \bar{A}_{22} & \bar{B}_2 \\ \hline \hat{C}_1 & \bar{C}_2 & \hat{D} \end{array}\right], \tag{16}$$

whose gramians are:

$$\bar{W}_c^{(2)} = \begin{bmatrix} \Sigma & 0 \\ 0 & \Lambda_F \end{bmatrix} \quad \text{and} \quad \bar{W}_o^{(2)} = \begin{bmatrix} \Sigma & 0 \\ 0 & \bar{L}_{11} \end{bmatrix}.$$

Note that

$$\Lambda_F \bar{L}_{11} = 0$$

since

$$\bar{W}_c^{(1)} \bar{W}_o^{(1)} = \begin{bmatrix} \Sigma^2 & 0 & 0 \\ 0 & \Lambda_F \bar{L}_{11} & \Lambda_F \bar{L}_{12} \\ 0 & 0 & 0 \end{bmatrix}.$$

and

$$\bar{W}_c^{(1)} \bar{W}_o^{(1)} = \bar{T}\hat{W}_c \hat{W}_o \bar{T}^{-1} = \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$

In addition, note that Λ_F is, by definition, full rank and hence, $\bar{L}_{11}$ must be identically zero.

Therefore:

$$\bar{W}_o^{(2)} = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}.$$

Finally, applying Lemma 4 to (16), gives equation (9), with $\hat{A}_{11} = A_b$, $\hat{B}_1 = B_b$ and $\hat{C}_1 = C_b$. □

Once a minimal realization for $G(s)$ in a balanced form has been obtained, model reduction by balanced truncation can be employed directly to realization (16). This can be done in accordance with the following theorem (Glover 1984).

Theorem 2: *Consider the balanced realization of $G(s)$ given in* (16) *and assume that its controllability and observability gramians are* $\Sigma = \text{diag}\{\Sigma_L,\ \Sigma_s\}$, *where* $\Sigma_L = \text{diag}\ \{\sigma_1 I_{m1},\ \sigma_2 I_{m2}, \ldots, \sigma_{\bar{r}} I_{m_{\bar{r}}}\}$ *and* $\Sigma_s = \text{diag}\ \{\sigma_{\bar{r}+1} I_{m_{\bar{r}+1}},\ \sigma_{\bar{r}+2} I_{m_{\bar{r}+2}}, \ldots, \sigma_{\bar{k}} I_{m_{\bar{k}}}\}$ *with* $\sigma_i > \sigma_j$, $i = 1, 2, \ldots, \bar{k}$, $i < j$ *and* m_i *being the multiplicity of* σ_i. *Partition* (16) *in accordance with* Σ, *namely:*

$$G(s) = \left[\begin{array}{cc|c} A_{b_{11}} & A_{b_{12}} & B_{b_1} \\ A_{b_{21}} & A_{b_{22}} & B_{b_2} \\ \hline C_{b_1} & C_{b_2} & D_b \end{array}\right].$$

Then the truncated system

$$\tilde{G}(s) = \left[\begin{array}{c|c} A_{b_{11}} & B_{b_1} \\ \hline C_{b_1} & D_b \end{array}\right] \tag{17}$$

is balanced and stable. Moreover

$$\|G - \tilde{G}\|_\infty \le 2(\sigma_{\bar{r}+1} + \sigma_{\bar{r}+2} + \cdots + \sigma_{\bar{k}}). \tag{18}$$

Proof: See Glover (1984) and Zhou *et al.* (1996). □

The direct application of Theorem 2 to equation (9) leads to the following result.

Corollary 1: *Let us partition the matrices T_1 and $T_1^\dagger$ in accordance with the gramian Σ, given in Theorem 2, as follows:*

$$T_1 = \begin{bmatrix} T_1 \\ T_s \end{bmatrix} \quad \text{and} \quad T_1^\dagger = [T_L^\dagger \quad T_s^\dagger]. \tag{19}$$

Then

$$\tilde{G}(s) = \left[\begin{array}{c|c} T_L A T_L^\dagger & T_L B \\ \hline C T_L^\dagger & D \end{array}\right]$$

is such that $\|G - \tilde{G}\|_\infty \le 2(\sigma_{\bar{r}+1} + \sigma_{\bar{r}+2} + \cdots + \sigma_{\bar{k}})$.

Proof: Note that the matrices $A_{b_{11}}$, B_{b_1} and C_{b_1} of (17) are obtained from A_b, B_b and C_b as follows:

$$A_{b_{11}} = [I_r \quad 0_{r\times(k-r)}] A_b \begin{bmatrix} I_r \\ 0_{(k-r)\times r} \end{bmatrix},$$

$$B_{b_1} = [I_r \quad 0_{r\times(k-r)}] B_b \quad \text{and} \quad C_{b_1} = C_b \begin{bmatrix} I_r \\ 0_{(k-r)\times r} \end{bmatrix},$$

where $r = \sum_{i=1}^{\bar{r}} m_i$ and $k = \sum_{i=1}^{\bar{k}} m_i$. Hence, substituting $A_b = T_1 A T_1^\dagger$, $B_b = T_1 B$ and $C_b = C T_1^\dagger$ in equation above, and noting that

$$T_L = [I_r \quad 0_{r\times(k-r)}] T_1 \text{ and } T_L^\dagger = T_1^\dagger \begin{bmatrix} I_r \\ 0_{(k-r)\times r} \end{bmatrix}$$

gives the result. □

Remark 3: Note that, from the definitions of T_1 and $T_1^\dagger$, given in (12), and of T_L and $T_L^\dagger$, given in (19), we may write:

$$T_1 = \begin{bmatrix} T_L \\ T_s \end{bmatrix} = \begin{bmatrix} \Sigma_L^{-1/2} & 0 \\ 0 & \Sigma_s^{-1/2} \end{bmatrix} \begin{bmatrix} X_L^{\mathrm{T}} \\ X_s^{\mathrm{T}} \end{bmatrix} W_o^{1/2}$$

$$T_1^\dagger = W_c^{1/2} [Y_L \quad Y_s] \begin{bmatrix} \Sigma_L^{-1/2} & 0 \\ 0 & \Sigma_s^{-1/2} \end{bmatrix},$$

and therefore

$$T_L = \Sigma_L^{-1/2} X_L^{\mathrm{T}} W_o^{1/2} \quad \text{and} \quad T_L^\dagger = W_c^{1/2} Y_L \Sigma_L^{-1/2}. \tag{20}$$

□

5. The algorithm

The results obtained in the previous section may be summarized in the following algorithm.

Algorithm 1: For a $p \times m$ stable, rational and proper transfer matrix $G(s)$ with a non-minimal state-space representation given by

$$G(s) = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right],$$

a reduced-order model $\tilde{G}(s)$, in balanced form, can be obtained as follows:

Step 1. Compute the observability and controllability gramians, W_c and W_o respectively, by solving the following Lyapunov equations:

$$A W_c + W_c A^{\mathrm{T}} = -B B^{\mathrm{T}}$$

$$A^{\mathrm{T}} W_o + W_o A = -C^{\mathrm{T}} C.$$

Step 2. Compute the singular value decompositions of W_c and W_o,

$$W_c = U_c \Lambda_c U_c^{\mathrm{T}} \quad \text{and} \quad W_o = U_o \Lambda_o U_o^{\mathrm{T}},$$

respectively, and find

$$W_c^{1/2} = U_c \Lambda_c^{1/2} U_c^{\mathrm{T}} \quad \text{and} \quad W_o^{1/2} = U_o \Lambda_o^{1/2} U_o^{\mathrm{T}}.$$

Step 3. Compute the singular value decomposition of the product $W_o^{1/2} W_c^{1/2}$ and partition it as follows:

$$W_o^{1/2} W_c^{1/2} = [X_L \quad X_s \quad X_2] \begin{bmatrix} \Sigma_L & 0 & 0 \\ 0 & \Sigma_s & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} Y_L^{\mathrm{T}} \\ Y_s^{\mathrm{T}} \\ Y_2^{\mathrm{T}} \end{bmatrix},$$

where $\Sigma_L = \text{diag}\ \{\sigma_1 I_{m_1},\ \sigma_2 I_{m_2}, \ldots, \sigma_{\bar{r}} I_{m_{\bar{r}}}\}$ and $\Sigma_s = \text{diag}\{\sigma_{\bar{r}+1} I_{m_{\bar{r}+1}},\ \sigma_{\bar{r}+2} I_{m_{\bar{r}+2}}, \ldots, \sigma_{\bar{k}} I_{m_{\bar{k}}}\}$ are formed with the Hankel singular values to be kept and discarded, respectively.

Step 4. Compute

$$T_L = \Sigma_L^{-1/2} X_L^{\mathrm{T}} W_o^{1/2} \quad \text{and} \quad T_L^{\dagger} = W_c^{1/2} Y_L \Sigma_L^{-1/2}.$$

Step 5. Obtain the reduced order model $\tilde{G}(s)$:

$$\tilde{G}(s) = \left[\begin{array}{c|c} T_L A T_L^{\dagger} & T_L B \\ \hline C T_L^{\dagger} & D \end{array}\right].$$

6. Example

With the view to illustrating the algorithm proposed in this paper, let us consider the MIMO system

$$G(s) = \frac{1}{(s+1)^2(s+1)} \times \begin{bmatrix} s+1 & (s+1)(2s+1) & s(s+1) \\ s+2 & (s+2)(s^2+5s+3) & s(s+2) \\ 1 & 2s+1 & s \end{bmatrix},$$

which has originally appeared in (Zhou *et al.* 1996, p. 82). From the Smith–McMillan form of $G(s)$, we can conclude that its poles are -2 and -1 (multiplicity 3) and its unique zero is -2. Therefore, any minimal order realization for $G(s)$ must have four states. An immediate state–space representation for $G(s)$ is as follows:

$$G(s) = \left[\begin{array}{ccccccc|ccc} -3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.5 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0.25 & 0 \\ 0 & 0 & -2 & 2 & 0 & 0 & 0 & 0 & 1.5 & 0.5 \\ 0 & 0 & -0.5 & 0 & 0 & 0 & 0 & 0.25 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & -4 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1.25 & 0 & 1 & 0 & 0.5 & 0.25 \\ 0 & 0 & 0 & 0 & -0.5 & 0 & 0 & 0.25 & 0.25 & 0 \\ \hline 2.0000 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

which is clearly non-minimal.

The next step towards obtaining a reduced model for $G(s)$ is to perform the singular value decomposition of $W_o^{1/2} W_c^{1/2} = X \Sigma Y^{\mathrm{T}}$. In doing so, we find out that $G(s)$ has the following Hankel singular values:

$$\mathcal{H} = \{1.34599705068499, 0.561442935324430, 0.229553347221876, 0.121751694629952,$$
$$0.994480795128365 \times 10^{-8}, 0.377508751606880 \times 10^{-8}, 0.137796554002854 \times 10^{-8}\}.$$

Note that if the last three Hankel singular values are discarded, then the error of approximation $\|G - \tilde{G}\|_\infty$ will be $\leq 3.0196 \times 10^{-8}$. Therefore, forming the matrices T_L ($T_L^{\dagger}$) with the first four columns (rows) of X (Y^{T}) we obtain the following reduced order model $\tilde{G}(s)$ of $G(s)$:

$$\tilde{G}(s) = \left[\begin{array}{cccc|ccc} -1.19912 & 1.17669 & -0.20410 & 0.08115 & 0.31978 & 1.71569 & 0.42683 \\ -0.22153 & -0.61453 & -0.23408 & 0.09276 & 0.66409 & -0.22411 & -0.44588 \\ -1.20966 & 0.99303 & -2.28659 & 0.50952 & -0.09831 & 0.85714 & 0.55265 \\ -0.23454 & -0.34857 & -0.25357 & -0.89976 & 0.10835 & 0.25235 & -0.37904 \\ \hline 0.77478 & -0.19715 & 0.82738 & -0.32752 & 0 & 0 & 0 \\ 1.58965 & -0.78316 & 0.06424 & 0.16676 & 0 & 1 & 0 \\ 0.31744 & -0.19452 & -0.60092 & -0.28986 & 0 & 0 & 0 \end{array}\right]$$

for which $W_c = W_o = \text{diag}\ \{1.34600, 0.56144, 0.22955, 0.12175\}$, and therefore the realization above is balanced. Note that the actual H_∞ error between $G(s)$ and $\tilde{G}(s)$ is approximately 1.5×10^{-15}.

Remark 4: A realization, with less states then that given above, could be obtained by following the steps of algorithm (1). Indeed, suppose that we are interested in keeping the first three Hankel singular values. In this case, proceeding according to algorithm (1), we obtain a three-state balanced realization for which the actual H_∞ error between $G(s)$ and $\tilde{G}(s)$ and the error bound (18) are approximately equal to 0.2435. □

7. Conclusions

The problem of model reduction by balanced truncation has been revisited and a simple algorithm presented. The reduced-order model is obtained by pre- and post-multiplication of the non-minimal order state–space realization by rectangular matrices. Moreover, there is no restriction on the way the Lyapunov equations, required to calculate the gramians, are computed. Other features of the algorithm are the realization obtained for the reduced model is in a balanced form, and the algorithm is expected to be robust, since it relies solely on singular value decompositions.

Acknowledgements

This work was partially supported by the Brazilian Research Council, CNPq (Grant No. 352810/96-3).

References

Aldhaheri, R. W., 1991, Model, order reduction via real Schur-form decomposition. *International Journal of Control*, **53**, 709–716.

Doyle, J. C., 1984, *Lecture Notes in Advances in Multivariable Control.* (Minneapolis: ONR/Honeywell Workshop).

Fernando, K. V., and Nicholson, H., 1982, Minimality of SISO linear systems. *Proceedings of the IEEE*, **70**, 1241–1242.

Fernando, K. V., and Nicholson, H., 1985, On the cross-Gramian for symmetric MIMO systems. *IEEE Transactions on Circuits and Systems*, **32**, 487–489.

Glover, K., 1984, All optimal hankel-norm approximations of linear multivariable systems and their l_∞-error bounds. *International Journal of Control*, **39**, 1115–1193.

Hammarling, S. J., 1982, Numerical solution of the stable, nonnegative definite Lyapunov equation. *I.M.A. Journal of Numerical Analysis*, **2**, 303–323.

Horn, R. A., and Johnson, C. R., 1985, *Matrix Analysis* (Cambridge: Cambridge University Press).

Laub, A. J., Heath, M. T., Paige, C. C., and Ward, R. C., 1987, Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms. *IEEE Transactions on Automatic Control*, **32**, 115–122.

Moore, B. C., 1981, Principal component analysis in linear systems: controllability, observability and model reduction. *IEEE Transactions on Automatic Control*, **26**,17–31.

Tombs, M. S., and Postlethwaite, I., 1987, Truncated balanced realization of a stable non-minimal state-space reduction. *International Journal of Control*, **46**, 1319–1330.

Zhou, K., Doyle, J. C., and Glover, K., 1996, *Robust and Optimal Control* (Upper Saddle River: Prentice-Hall).