Discrete Multivariable System Approximations by Minimal Pade-Type Stable Models

YUVAL BISTRITZ, MEMBER, IEEE, AND URI SHAKED, MEMBER, IEEE

Abstract —The approximation of discrete-time multivariable high-order linear systems is considered. Reduced order models are derived by a generalized minimal partial realization algorithm. The derived models approximate the system in the Pade sense and the presented method overcomes some serious limitations of former multivariable reduction methods. The set of all different models of minimal order that approximate a mixed sequence of Markov and time moment matrices of a given length is characterized by the common structural properties of these models. A maximal set of free parameters for the above set of all models is determined. These parameters can assign values independently and can be used to satisfy further desired specifications. A procedure is presented to solve the problem of possible instability of Pade approximated models of a stable system. Stable models maybe chosen among the models of the same minimal order that differently emphasize the approximation of the steady state and the transient responses. When applicable, the free parameters can also be adjusted to yield stable models. Finally, a complementary systematic method is presented by which unstable models can be replaced by a stable model of the same order and with the same singular values that approximate, in the Pade sense, the magnitude of the high-order system.

Introduction

THE FAST development and usage of small digital computers and processors in the design and implementation of control systems increases the importance of reduced order modeling methods for discrete systems. Preferable reduction methods should have a good approximating quality and a relatively low cost of computational time and memory requirements. Analytical methods suggested for single input-single output (SISO) systems [1]–[5] have the above advantage over numerical minimization procedures, e.g., [6], [7]. Some early approaches suggested derivation of discrete reduced order models by continuous techniques via the bilinear or other homographic transformations [3], [4], [8], [9]. These methods may yield poor approximations, [4], [8], where the unsatisfactory quality of approximation stems from the deformation caused by these transformations. Approximations of a good quality were obtained by approximating the scalar z-transfer function in the Pade sense by solving the Pade set of equations [4] or by the continued fraction method [2]. The stability problem, by which stable systems may yield unstable Pade approximated models was resolved for SISO systems by deriving transfer functions in which stable denominators are first determined and the final models approximate the system only partially in the Pade sense [1], [10].

Manuscript received October 15, 1981; revised June 12, 1982.
The authors are with the Department of Electrical Engineering, Faculty of Engineering, Tel-Aviv University, Ramat-Aviv 69978, Israel.

The extension of these SISO techniques to multivariable systems is not trivial. The difficulties encountered are similar to those found also in the extensions of the continuous system reduction methods [5], [11]. Simple matrix versions of the scalar Pade equations or continued fraction expansions may frequently be unsolvable because of some singular conditions [5]. Even in the case where they have a solution, this solution may result in models of too high order that may even exceed, unnoticeably, the order of the given system [9], [11].

The present paper generalizes the idea of discrete system Pade approximation at two points [2], [4] to multivariable systems. The problem is defined as that of finding a model of minimal order that will match a given sequence of Markov and time moment matrices. In contrast to former approaches that derive models of predetermined order and may not have a solution (even in the SISO case), the so defined problem always has a solution. The problem is solved by a generalized minimal partial realization method which overcomes naturally the above-mentioned problem of the actual order of the reduced models. A formal as well as structural similarity between the discrete matching problem and the analogous continuous model reduction problem is established, followed by many of the results and the algorithms of [11]. These are given with slight modifications without further proof. They include conditions for the existence of a unique solution to the formulated problem or else, the characterization of the set of all possible solutions in terms of a maximal set of independent parameters.

The mathematical conditions for stability of discrete systems is well known to be different from its continuous counterpart. The possible instability of the reduced order approximations to a stable system receives, therefore, a special treatment. One obvious possibility is the choice of stable models among the various models of the same order that differently emphasize the steady-state and the transient responses. Another possibility, when applicable, is the adjustment of the above-mentioned free parameters to acquire the stability of the models. An additional systematic approach is developed to solve the stability problem. It presents a modification of the already derived

0098-4094/84/0400-0382$01.00

unstable model that yields a stable model of a same order and with the same singular values that approximate the magnitude of the system in a partial Pade sense and match its steady-state value. The solution of the stability problem results, therefore, both from the presented wide range of Pade approximated reduced order models that can all be derived by the unified approach as well as the combination, if necessary, of the Pade and the singular value approximations [17]–[21].

II. Minimal Reduced Order Models

A discrete-time high-order linear system is given by an $l \times m$ z-transform transfer function matrix $H(z)$ or equivalently by a triple of state-space realization matrices (A_H, B_H, C_H), $A_H \in R^{N \times N}$, $B_H \in R^{N \times m}$, $C_H \in R^{l \times N}$ where N is the order of the system and

$$H(z) = C_H(zI - A_H)^{-1}B_H. \tag{1}$$

Reduced order approximating models are required that match some first time moments and first Markov matrices of the system. Such models were shown in the case of continuous-time systems, to yield satisfactory approximations of both the steady-state and the transient responses of the system [5], [11].

The Markov matrices for the system are given by

$$Y_{i+1} = C_H A_H^i B_H, \qquad i = 0,1,2 \cdots. \tag{2}$$

These matrices have the same form both for continuous and discrete systems. In the discrete case, however, they have additional interpretation as the sequence $\{Y_{i+1}\}$ forms the unit impulse response of the system,

$$h(k) = Y_{k+1} \tag{3}$$

where $h(k)$ is the inverse z transform of $H(z)$.

The time moments for the system are given by the discretized version of

$$\int t^i h(t)\, dt$$

references [2], [3], [22], namely,

$$L_{i+1} = \sum_{k=0}^{\infty} k^i h(k) = \sum_{k=0}^{\infty} Y_{k+1} k^i, \qquad i = 0,1,2,\cdots. \tag{4}$$

The exact definition of the required reduced order model in terms of the Markov and time moment matrices is the following:

A First Statement of the Problem: Given a high-order system (A_H, B_H, C_H) find for a sequence of p time moments $\{L_i, i = 1,\cdots,p\}$ and of Markov matrices $\{Y_i, i = 1,\cdots,q\}$, where $p + q = r$, a model (A, B, C), $A \in R^{n \times n}$, $B \in R^{n \times m}$, $C \in R^{l \times n}$ of a minimal order, that satisfies

$$\hat{Y}_i = Y_i, \qquad i = 1,\cdots,q \tag{5a}$$

and

$$\hat{L}_i = L_i \qquad i = 1,\cdots,p \tag{5b}$$

where

$$\hat{Y}_{i+1} = CA^iB, \qquad i = 0,1,2,\cdots \tag{5c}$$

$$L_{i+1} = \sum_{k=0}^{\infty} \hat{Y}_{k+1} k^i, \qquad i = 0,1,2,\cdots. \tag{5d}$$

This is the same definition as the one used for continuous system [11], where the relative choice of p and q, $p + q = r$, are known to determine the respective weight of the approximation of the steady-state and the transient responses.

The mathematical representation of (5a)–(5d) can be brought to a formal agreement with the corresponding representation of the continuous case by defining the shifted transform plane

$$\eta = z - 1 \tag{6a}$$

and correspondingly,

$$F_H = A_H - I_{N \times N} \qquad F = A - I_{n \times n}. \tag{6b}$$

The transfer function matrix $H(\eta) = C_H(\eta I - F_H)^{-1}B_H$ can be expanded about $\eta = \infty$ and $\eta = 0$ by

$$H(\eta) = \sum_{i=1}^{\infty} M_i \eta^{-i} \tag{7a}$$

and

$$H(\eta) = -\sum_{i=1}^{\infty} T_i \eta^{i-1} \tag{7b}$$

where

$$M_i = C_H F_H^{i-1} B_H, \qquad i = 1,2,\cdots \tag{8a}$$

and

$$T_i = C_H F_H^{-i} B_H, \qquad i = 1,2,\cdots \tag{8b}$$

whereas for (7b) and (8b) to be well defined the assumption that the system (A_H, B_H, C_H) has no eigenvalues of A_H at $z = 1$ is made. It is clear from the linear relations (6b) that the model matches q Markov matrices of (5a) if in the shifted plane it matches $\{M_i, i = 1,\cdots,q\}$. Similarly, it follows from the relations

$$T_{i+1} = -\left.\frac{d^i H(z)}{i!\,dz^i}\right|_{z=1}, \qquad i = 0,1,2,\cdots \tag{9a}$$

and

$$L_{i+1} = (-1)^i \left(z\frac{d}{dz}\right)^i H(z)\bigg|_{z=1}, \qquad i = 0,1,2,\cdots \tag{9b}$$

that the matching of $T_i, i = 1,\cdots,p$ leads to the matching of p time moments L_i [22]. Applying these relations, the model reduction problem can be restated as follows.

A Second Statement of the Problem: Given a high-order system (A_H, B_H, C_H) find for the first terms of the transfer function matrix $H(z)$ expansions about $z = 1$ and $z = \infty$, $\{T_i, i = 1,\cdots,p\}$ and $\{M_i, i = 1,\cdots,q\}$, respectively, a triple of matrices (F, B, C), $F \in R^{n \times n}$, $B \in R^{n \times m}$, $C \in R^{l \times n}$ with a minimal value of n that satisfies

$$CF^{i-1}B = M_i, \qquad i = 1,\cdots,q \tag{10a}$$

and

$$CF^{-i}B = T_i, \qquad i = 1, \cdots, p. \tag{10b}$$

Using the second statement, the problem actually becomes the minimal Pade approximation (m.p.a.) problem that has been extensively treated in [11]. Thus applying the linear relation $A = F + I$ between the required reduced order models (A, B, C) and the m.p.a. solutions (F, B, C) many important results can be directly adopted from [11].

Pade approximations of $G(z)$ at $z = 1$ or both at $z = 1$ and $z = \infty$ have been suggested before for SISO systems $(m = l = 1)$ in [2] and [4]. These references apply continued fraction expansions about $z = 1$ and $z = \infty$ or equivalent Pade equations. It follows from (9a) and (9b) that the methods of [2] and [4] match the first scalar time moments and Markov parameters of the system, in spite of incorrect relation between T_i and the time moments that is given in [2], as indicated in [22]. The extension of these scalar methods to multivariable systems $(m, l > 1)$ may not always have a solution due to inherent limitations such as the requirement for $m = l$. If this extension has a solution it may yield models of order up to $m \cdot l \cdot n$. The Pade "reduced" order model may thus even exceed unnoticeably the order of the high-order system. A similar phenomenon has been indicated in [11] for the continuous analogous case. One of the contributions of this paper is that it removes the above limitations (such as $m = l$). Using a state-space approach and requiring models of minimal orders the above difficulty with the order of the reduced model is also removed.

We start by discussing some structural properties of discrete models that will enable us to apply in the discrete case results from [11] and [12] concerning conditions for uniqueness and nonuniqueness of the solutions and their derivations.

Let (A, B, C) be a model of order n that solves the (p, q) reduced order model problem. We denote by $I_n = \{i_1, \cdots, i_n\}$ the indexes of the first independent rows in $U(A, C)$, the observability matrix of the pair (A, C), and by $J_n = \{j_1, \cdots, j_n\}$ the indexes of the first n independent columns in $V(A, B)$, the reachability matrix of the pair (A, B). We have from the linear relation $F = A - I$ that I_n and J_n also yield the respective indices for $U(F, C)$ and $V(F, B)$. Thus as it has been shown in [11] that m.p.a. shares the same sets I_n and J_n for all p and q, $p + q = r$. We obtain that I_n and J_n are structural indexes that are common for all the discrete reduced order models that match a sequence of r matrices of the high-order system for any $p + q = r$. The sets I_n and J_n are directly related to the more frequently referred to observability and reachability indexes, $\nu = \{\nu_1, \cdots, \nu_l\}$ and $\mu = \{\mu_1, \cdots, \mu_m\}$, of the models. The index ν_k represents the number of terms in the arithmetical sequence $\{i_k, i_k + l, i_k + 2l, \cdots\}$ that are included in I_n for all $k = 1, \cdots, l$ ($\nu_k = 0$ for an empty intersection). The index μ_k is related similarly to J_n via the sequence $\{j_k, j_k + m, j_k + 2m, \cdots\}$ for $k = 1, \cdots, m$. The relations

$$\sum_{i=1}^{l} \nu_i = \sum_{j=1}^{m} \mu_j = n$$

are also well established. The structural properties of the discrete reduced order models of a given system are thus as follows:

Proposition 1: The stated model reduction problem has always a solution. All the sequences of length r yield, for any $p + q = r$, models of a common order n and they all have the sets of observability and reachability indexes ν and μ (or equivalently, the same sets I_n and J_n).

The triple of matrices that solves the m.p.a. problem of (10) for $p = 0$ is recognized as the conventional minimal partial realization that was basically treated by Tether and Kalman [13], [14] and recently in [12].[1] From the one to one relation between the m.p.a. solutions (F, B, C) and the required models (A, B, C) and from Proposition 1, the uniqueness conditions of the minimal partial realization theory [13], [14] can be applied to the present reduced order modelling problem as follows.

Proposition 2: A reduced order model (5a)–(5d) is unique (in the sense of the equivalent class of similarity transformations) iff $\alpha + \beta < p + q = r$ where

$$\alpha = \max_{i=1\cdots l} \nu_i \qquad \text{and} \qquad \beta = \max_{j=1\cdots m} \mu_j. \tag{11}$$

Once the existence of solutions is established we describe an algorithm that can be used to derive their models. It may presently be assumed that the sequence under consideration corresponds to a unique solution. It is shown hereafter how, in the complementary case of nonuniqueness, the set of all solutions can also be derived by this algorithm. The algorithm is basically the algorithm of [15] that was used in [11] for the continuous case.

The Algorithm:

(i): Given the sequence

$$\{G_1, G_2, \cdots, G_r\} = \{T_p, T_{p-1}, \cdots, T_1, M_1, M_2, \cdots, M_q\} \tag{12a}$$

form the incomplete Hankel matrix

$$K(p, q) = \begin{bmatrix} G_1 & G_2 & \cdots & G_r \\ G_2 & & & \\ \vdots & & & \\ G_r & & & \end{bmatrix} \tag{12b}$$

where the matrix $K(p, q)$ is considered as being completed by some unknown matrices G_{r+1}, G_{r+2}, which are assumed to acquire values that do not influence the internal dependencies between the rows and between the columns as one determined by $\{G_1, \cdots, G_r\}$ [12].

Let $n = \rho K(p, q)$ denote the rank of $K(p, q)$. The rank of the incomplete Hankel matrix of (12b) is defined as the minimal rank that complies with the numerically specified parts. This rank determines the order of all the minimal models that match the sequences $\{G_1, \cdots, G_r\}$. We denote

[1] Tether and Kalman were the first to note that minimal partial realization may be useful also for reduced order modeling. The vast literature on Pade model reduction methods (considering mainly the continuous-time case, cf. [11]), has not realized, in general, the advantage of minimal partial realization over the extensively treated Pade equation and continued fraction techniques.

by I_n and J_n the indexing of the first n independent rows and columns, respectively, of $K(p,q)$. As the matrix $K(p,q)$ involves rows and columns which are not completely specified, it is noted that $I_n(J_n)$ represents the maximal number of first rows (columns) such that each row (column) does not depend on its preceeding rows (columns), where the dependency is tested only along columns (rows) that correspond to entries for which the considered row (column) is specified. The derivation of I_n, J_n, and n from an incomplete Hankel matrix is discussed in more detail in [11], [12]. It is evident that I_n and J_n form the structural indexes mentioned in Proposition 1.

(ii) Once I_n and J_n were determined, construct the following submatrices:

K_A: The $n\times n$ submatrix formed by the intersection of rows I_n with columns $\{j_1+m, j_2+m,\cdots,j_n+m\}$ where $J_n=\{j_1,\cdots,j_n\}$.

K_B: The $n\times m$ submatrix formed by the intersection of rows I_n with columns $\{1,2,\cdots,m\}$.

K_C: The $l\times n$ submatrix formed by the intersection of rows $\{1,2,\cdots,l\}$ with columns J_n.

(iii) Apply on $[K_B, K_A]$, a Gauss row elimination that brings the columns of $[K_B, K_A]$ that correspond to J_n to the nth identity matrix $I_{n\times n}$.

(iv) Denoting the resulting matrix by $[\tilde{K}_B, \tilde{K}_A]$ it follows that (F,B,C) of (10) is given by $(\tilde{K}_A, \tilde{K}_B, K_C\tilde{K}_A^p)$ or that the required model (A,B,C) is given by

$$A=\tilde{K}_A+I_{n\times n},\quad B=\tilde{K}_B,\quad C=K_C\tilde{K}_A^p. \tag{13}$$

In the case where $\alpha+\beta>r$ there exist, by Proposition 2, many models for each choice of p and q. In this case the submatrices K_A, K_B, K_C contain positions in $K(p,q)$ that are numerically undetermined. These positions may be viewed as entries of some unknown extension sequence $\{G_{r+1},\cdots,G_{\alpha+\beta}\}$. Denoting by g_{ijk} the (i,j) entry of the unknown matrix G_k, this set of numerically unknown entries is exactly determined in number and positions by the common structural properties of all the sequences of length r. This set is shown in [12], using the invariants in [23], to be given by

$$P=\{g_{ijk}\mid k=r+1,\cdots,\nu_i+\mu_j>r;\; i=1,\cdots,l;\; j=1,\cdots,m\}. \tag{14}$$

Carrying these parameters along in the algorithm given above, a solution (A,B,C) is obtained that includes also certain combinations of the parameters in P. In this way (A,B,C) represents the set of all different reduced order models that match the given p time moments and q Markov matrices with the following efficient characterization.

Proposition 3 [12]: The set of parameters P represents the maximal set of independent parameters for the set of all minimal reduced order models that match r time moment and Markov matrices of the high-order system.

The last proposition indicates that P is the maximal set of parameters that may each assign values freely and always yield a reachable and observable model that matches the given sequence of matrices. These parameters may, therefore, be adjusted to meet additional desired specifications. Some aspects of the method are illustrated in the following example.

Example: Given a z-transfer function matrix of a reachable and observable system

$$H(z)=\begin{bmatrix}\dfrac{2.25(z-0.75)}{(z-0.95)(z-0.5)} & \dfrac{1.5(z-0.8)}{(z-0.9)(z-0.75)}\\ \dfrac{1.04(z-0.65)}{(z-0.95)(z-0.3)} & \dfrac{(z-0.7)}{(z-0.9)(z-0.35)}\end{bmatrix} \tag{15}$$

some first terms of expansion of $H(\eta)$, $\eta=z-1$. About $\eta=0$ and $\eta=\infty$ (7a), (7b) are given by

$$T_1=\begin{bmatrix}-22.5 & -12.\\ -10.4 & -20.\end{bmatrix}\quad T_2=\begin{bmatrix}405. & 108.\\ 193.14 & 266.67\end{bmatrix}$$

$$T_3=\begin{bmatrix}-8010. & 1032.\\ -3841.6 & -3111.1\end{bmatrix}\quad T_4=\begin{bmatrix}160020. & 10123.\\ 76807. & 34074\end{bmatrix}$$

$$M_1=\begin{bmatrix}2.25 & 1.5\\ 1.04 & 1.\end{bmatrix}\quad M_2=\begin{bmatrix}-6.675 & -0.225\\ -0.416 & 0.05\end{bmatrix}. \tag{16}$$

It is required to find a minimal order model that matches, say, T_1, T_2, and M_1. We first construct the Hankel matrix $K(2,1)$ of (12b):

$$K(2,1)=\left[\begin{array}{cc|cc|cc} 405. & 108. & -22.5 & -12. & 2.25 & 1.5\\ 193.14 & 266.67 & -10.4 & -20. & 1.04 & 1.\\ \hline -22.5 & -12. & 2.25 & 1.5 & & \\ -10.4 & -20. & 1.04 & 1. & & \\ \hline 2.25 & 1.5 & & & & \\ 1.04 & 1. & & & & \end{array}\right]. \tag{17}$$

The rank of $K(2,1)$ is $\rho K(2,1)=4$, thus the models that match T_1, T_2, and M_1 or any other three matrices $\{G_1,G_2,G_3\}$ of (12a), are of minimal order $n=4$. In this example the first four independent rows and columns of $K(2,1)$ are the four first rows and columns, $I_n=J_n=\{1,2,3,4\}$. Thus from the structural point of view this example is relatively simple. To illustrate the competence of the underlying method in cases of more complex structure [11], and [12] should be referred to. From I_n and J_n we find the observability and reachability indexes $\nu=\mu=\{2,2\}$ and thus $\alpha+\beta=4>3$. The solution is, therefore, not unique. Checking (14) for the free parameters we find $P=\{g_{114}, g_{124}, g_{214}, g_{224}\}$. These four free parameters participate in the submatrices K_A, K_B, K_C which are given by

$$K_c=\begin{bmatrix}405. & 108. & -22.5 & -12.\\ 193.14 & 266.67 & -10.4 & -20.\end{bmatrix} \tag{18a}$$

and

$$[K_B, K_A] = \begin{bmatrix} 405. & 108. & : & -22.5 & -12. & 22.5 & 1.5 \\ 193.14 & 266.67 & : & -10.4 & -20. & 1.04 & 1. \\ -22.5 & -12. & : & 22.5 & 1.5 & g_{114} & g_{124} \\ -10.4 & -20. & : & 1.04 & 1. & g_{214} & g_{224} \end{bmatrix}. \tag{18b}$$

Performing row elimination on $[K_B, K_A]$ and conveniently replacing independent combinations of the parameters in $\boldsymbol{P}$ by the four parameters $\{\pi_1, \pi_2, \pi_3, \pi_4\}$ we have

$$[\tilde{K}_B, \tilde{K}_A] = \begin{bmatrix} 1 & 0 & : & 0 & 0 & -0.0267 & -0.0082 \\ 0 & 1 & : & 0 & 0 & 0.0102 & -0.163 \\ 0 & 0 & : & 1 & 0 & \pi_1 & \pi_2 \\ 0 & 0 & : & 0 & 1 & \pi_3 & \pi_4 \end{bmatrix}$$

where $\pi_1, \pi_2, \pi_3, \pi_4$ may assign arbitrary values. The set of all the solution is (A, B, C) where $A = \tilde{K}_A + I$, $B = \tilde{K}_B$ and $C = K_C \tilde{K}_A^2$. The four parameters can be carefully chosen to satisfy further desired specification. Choosing for example values of too large magnitudes for these parameters draw the eigenvalues of A outside the unit circle which is not acceptable for the given stable system. The choice $\pi_1 = -0.6072$, $\pi_2 = -0.0178$, $\pi_3 = 0.1422$, and $\pi_4 = -0.2660$ yields, for example, a stable model that also matches M_2. The resulting model for this choice is

$$A = \begin{bmatrix} 1 & 0 & -0.0267 & -0.0008 \\ 0 & 1 & 0.1021 & -0.0163 \\ 1 & 0 & 0.3928 & -0.0178 \\ 0 & 1 & 0.1422 & 0.7340 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{19a,b}$$

$$C = \begin{bmatrix} 2.25 & 1.5 & -0.675 & -0.225 \\ 1.04 & 1. & -0.416 & 0.05 \end{bmatrix}. \tag{19c}$$

The transfer function matrix of this model is

$$G(z) = \frac{1}{D(z)} \begin{bmatrix} N_{11}(z) & N_{12}(z) \\ N_{21}(z) & N_{22}(z) \end{bmatrix} \tag{20}$$

where

$$N_{11}(z) = 2.25(z-0.9123)(z-0.8487)(z-0.6658)$$
$$N_{12}(z) = 1.5(z-0.9521)(z-0.8621)(z-0.4626)$$
$$N_{21}(z) = 1.04(z-0.9283)\left[(z-0.7993)^2+0.0215\right]$$
$$N_{22}(z) = (z-0.9523)(z-0.6380)(z-0.4865)$$

and

$$D(z) = (z-0.9523)(z-0.9059)(z-0.8211)(z-0.4475).$$

The responses (i, j) $i, j = 1,2$ of outputs i for a unit step at input j for the model (20) and the given system (15) are compared in Fig. 1.

In the case where it is required to match two matrices only we find that $I_2 = J_2 = \{1,2\} \to \nu = \mu = \{1,1\} \to \alpha + \beta = 2$. The problem has in this case a unique solution. For example, the matching of T_1 and T_2 yields the unique solution

$$A = \begin{bmatrix} 0.9440 & -0.0119 \\ 0.0015 & 0.9336 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 1.2408 & 1.0648 \\ 5.7896 & 2.5683 \end{bmatrix} \tag{21}$$

that represents a stable model with eigenvalues 0.9418 and 0.9359.

It may happen that a high order stable system yields unstable models where either no free parameters are available or the free parameters cannot be adjusted to reveal a stable choice. As an illustration for the first of these situations the problem of matching the first two Markov matrices for the above system yields the following unique solution:

$$A = \begin{bmatrix} 0.9261 & -0.4348 \\ 0.3391 & 1.5022 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 2.25 & 1.5 \\ 1.04 & 1.0 \end{bmatrix}. \tag{22}$$

The eigenvalues of A are $\lambda_1 = 1.6942$ and $\lambda_2 = 0.7342$ which indicate an unstable model. It should be noted at this place that matching of the transient response only is not recommended for a good practical approximation. Nevertheless, this model properly illustrates the possible instability of reduced order models derived for stable system. A method that overcomes the instability problem is treated in the next section where this last model will be further used for demonstration.

III. A Stabilization Method

One of the inherent shortcomings of model reduction by the method of Pade approximation is that it cannot guarantee a stable model to a stable system. Two partial remedies to this difficulty have been included in the preceeding section. In the cases where the resulting model of minimal order is not unique one may find an appropriate choice of the free parameters in $\boldsymbol{P}$ for which the resulting model is stable. In the case of uniqueness there exists a unique model for each of the r differently mixed sequences of $p > 0$ time moment and q Markov matrices $p + q = r$ from which a satisfactory stable model may be chosen. This section represents a new and complementary approach that is always applicable. By this approach the unstable model is modified such that the square magnitude of the modified stable model is the same as the corresponding square magnitude, $G^t(z^{-1})G(z)$ or $G(z)G^t(z^{-1})$, of the unstable model. The equalities between the square

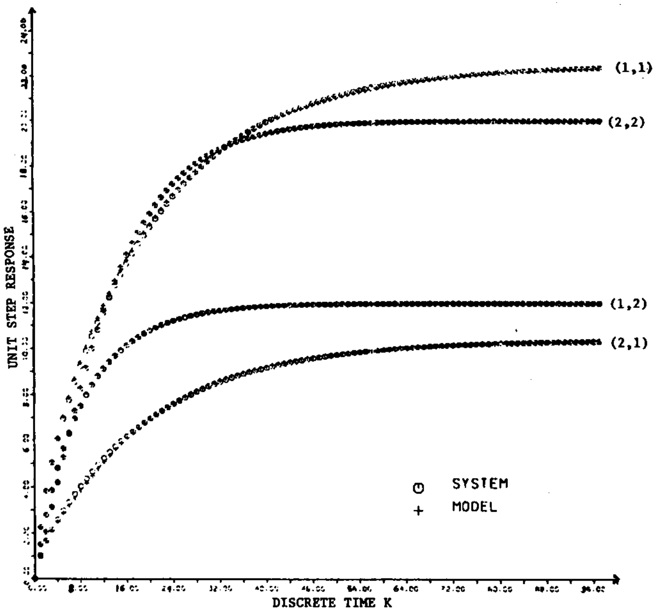

Fig. 1. Unit step responses for system (15) and model (20). (i, j) represents the response of output j to a unit step at input i, $i, j = 1,2$.

magnitudes imply that the modified stable model possesses the same singular values as the unstable model and as such it possesses most of the control features of the latter [17], [18]. The idea of using singular value as a tool for model reduction has received recently a growing attention [19]–[21], where models that match the dominant singular values of the Hankel matrix were found. The approach adopted in this section is to mix, when necessary, Pade approximation of the system with singular value approximation of its squared magnitude, similarly to the continuous case [11]. A model is first found that matches the required time moments and Markov matrices. This model, approximates the system transfer function matrix at $z=1$ and $z=\infty$ and thus its singular values are good approximations of the system singular values that dictate its steady-state response. If, for a stable system, an appropriate stable model cannot be chosen from the set of all models represented in Section II, we suggest a modification of an already derived model which satisfies the requirements of the stated problem but happens to be unstable. The modification replaces the unstable model by stable model of the same order that has the same magnitude and consequently the same singular values. This stabilized model also matches the steady state value of the high-order stable system response for a step input.

Let (A, B, C) be an unstable model, with the transfer function matrix $G(z) = C(zI - A)^{-1}B$, derived as a (p, q) minimal order model for a stable system. We show next the existence and the derivation of stable models (A_c, B_c, C_c) and (A_0, B_0, C_0) of the same minimal order whose transfer function matrices

$$G_c(z) = C_c(zI - A_c)^{-1}B_c \tag{23}$$

$$G_0(z) = C_0(zI - A_0)^{-1}B_0 \tag{24}$$

satisfy

$$G_c^t(z^{-1})G_c(z) = G^t(z^{-1})G(z) \tag{25a}$$

$$G_c(1) = G(1) \tag{25b}$$

and

$$G_0(z)G_0^t(z^{-1}) = G(z)G^t(z^{-1}) \tag{26a}$$

$$G_0(1) = G(1) \tag{26b}$$

respectively. Obviously, as $G(z)$ is a Pade approximation of $H(z)$ the modified models have squared magnitudes that match the first p terms of expansion about $z=1$ of the corresponding squared magnitudes $H^t(z^{-1})H(z)$ and $H(z)H^t(z^{-1})$. In addition, the model matches the steady-state value $H(1)$ of the high-order system response to a step input.

Assume that the eigenvalues of A that reside outside and inside the unit circle are, respectively, $(\lambda_1 \cdots \lambda_\zeta)$ and $(\lambda_{\zeta+1}, \cdots, \lambda_n)$. We denote the spectral decomposition of A by

$$A = UJV \tag{27a}$$

where

$$U = [U_1, U_2] \tag{27b}$$

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \tag{27c}$$

$$UV = I \tag{27d}$$

and

$$J = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix} \tag{27e}$$

and where J_1 and J_2 are the $\zeta \times \zeta$ and $(n-\zeta)\times(n-\zeta)$ Jordan blocks that correspond to the "unstable" and the "stable" eigenvalues of A, respectively.

Defining

$$P_c = U_1 Q_c^{-1} U_1^\dagger \tag{28a}$$

where $(\ )^\dagger$ denotes Hermitian transpose and $Q_c > 0$ is the solution of

$$Q_c - J_1^\dagger Q_c J_1 = -U_1^\dagger C^t C U_1 \tag{28b}$$

we have the following theorem.

Theorem 4: The eigenvalues of the matrix

$$A_c = A(I + P_c C^t C)^{-1} \tag{29a}$$

are

$$\{\lambda_1^{-1}, \cdots, \lambda_\zeta^{-1}, \lambda_{\zeta+1}, \cdots, \lambda_n\}$$

and the triple of matrices (A_c, B_c, C_c) satisfies (25a), (25b), where $B_c = B$ and

$$C_c = C\left[I + (A - I)^{-1}(A_c - A)\right]. \tag{29b}$$

For the derivation of (A_0, B_0, C_0) that satisfies (26) we define

$$P_0 = V_1^\dagger Q_0 V_1 \tag{30a}$$

where $Q_0 > 0$ solves

$$Q_0 - J_1 Q_0 J_1^\dagger = -V_1 BB' V_1^\dagger \tag{30b}$$

and the dual of theorem 4 becomes the following:

Theorem 5: The eigenvalues of the matrix

$$A_0 = (I + BB'P_0)^{-1} A \tag{31a}$$

are $\{\lambda_1^{-1}, \cdots, \lambda_\zeta^{-1}, \lambda_{\zeta+1}, \cdots, \lambda_n\}$ and the triple of matrices (A_0, B_0, C_0) satisfies (26a), (26b), where $C_0 = C$ and

$$B_0 = \left[I + (A_0 - A)(A - I)^{-1}\right] B. \tag{31b}$$

We shall prove theorem 4 and the proof of Theorem 5 will then follow by duality.

Proof: We show first that

$$P = U_1 Q_c^{-1} U_1^\dagger \tag{32}$$

solves the Ricatti equation

$$P = APA' - APC'[I + CPC']^{-1} CPA'. \tag{33}$$

For this we multiply (28b) from the left by $\overline{P} = Q_c^{-1}$ and from the right by $J^{-1}\overline{P}$ and denote $C_1 = CU_1$. We obtain that

$$(I + \overline{P} C_1^\dagger C_1) J_1^{-1} \overline{P} = \overline{P} J_1^\dagger$$

and, therefore, that

$$\begin{aligned}\overline{P} &= J_1 (I + \overline{P} C_1^\dagger C_1)^{-1} \overline{P} J_1^\dagger \\ &= J_1 \overline{P} J_1^\dagger - J_1 \left[I - (I + \overline{P} C_1^\dagger C_1)^{-1}\right] \overline{P} J_1^\dagger\end{aligned}$$

or

$$\overline{P} = J_1 P J_1^\dagger - J_1 \overline{P} C_1^\dagger (I + C_1 \overline{P} C_1^\dagger)^{-1} C_1 \overline{P} J_1^\dagger. \tag{34}$$

Denoting

$$\tilde{P} = \begin{bmatrix} \overline{P} & 0 \\ 0 & 0 \end{bmatrix} \tag{35a}$$

$$P = U \tilde{P} U^\dagger = U_1 Q_c^{-1} U_1^\dagger. \tag{35b}$$

We can replace (34) by

$$\tilde{P} = J \tilde{P} J^\dagger - J \tilde{P} U^\dagger C' (I + CU \tilde{P} U^\dagger C')^{-1} CU \tilde{P} J^\dagger \tag{36}$$

and obtain (33) by multiplying (36) by U and $U^\dagger$ from the left and from the right, respectively.

Denoting

$$K = APC'(I + CPC')^{-1} \tag{37}$$

where P is the solution of (33), it has been shown in [16] that this solution also satisfies

$$\left[I + C(zI - A)^{-1} K\right](I + CPC') \cdot \left[I + K'(z^{-1}I - A')^{-1} C'\right] = I \tag{38}$$

taking the inverses at both sides of (38) we obtain

$$I = \left[I + K'(z^{-1}I - A')^{-1}(I + CPC')^{-1} \cdot \left[I + C(zI - A)^{-1} K\right]^{-1}.$$

The substitution of this expression for I in

$$\phi(z) = B'(z^{-1}I - A')^{-1} C' I C (zI - A)^{-1} B \tag{39}$$

yields

$$\begin{aligned}\phi(z) &= B'(z^{-1}I - A')^{-1}\left[I + C'K'(z^{-1}I - A')^{-1}\right]^{-1} \\ &\quad \cdot C'(I + CPC')^{-1} C \left[I + (zI - A)^{-1} KC\right]^{-1} \\ &\quad \cdot (zI - A)^{-1} B \\ &= B'(z^{-1}I - A' + C'K')^{-1} C'(I + CPC')^{-1} \\ &\quad \cdot C(zI - A + KC)^{-1} B.\end{aligned} \tag{40}$$

Thus comparing (40) and (39) we find that

$$G'(z^{-1}) G(z) = \tilde{G}'(z^{-1}) \tilde{G}(z) \tag{41}$$

for

$$\tilde{G}(z) = \tilde{C}(zI - A_c)^{-1} B \tag{42a}$$

where

$$\tilde{C} = (I + CPC')^{-1/2} C \tag{42b}$$

$$A_c = A - KC \tag{42c}$$

and where it follows from the properties of the Ricatti Equation (33) that all the eigenvalues of A_c lie inside the unit circle and that they are given by $\{\lambda_1^{-1}, \cdots, \lambda_\zeta^{-1}, \lambda_{\zeta+1}, \cdots, \lambda_n\}$. An alternative expression for A_c is found by

$$\begin{aligned}A_c &= A - APC'(I + CPC')^{-1} C \\ &= A\left[I - PC'(I + PC'C)^{-1}\right]\end{aligned}$$

or

$$A_c = A(I + PC'C)^{-1}. \tag{29a}$$

The triple (A_c, B, C) satisfies (25a). To obtain a triple (A_c, B_c, C_c) that also satisfies (25b) we use the fact that $\tilde{C}$ appears in (40) only in the product $\tilde{C}'\tilde{C}$.

Choosing, therefore,

$$C_c = W\tilde{C} \tag{43a}$$

where

$$W = \left[I + C(I - A)^{-1} K\right](I + CPC')^{1/2}. \tag{43b}$$

We have that $C_c(I - A + KC)^{-1} = C(I - A)^{-1}$. From (38) we also have $C_c'C_c = \tilde{C}'\tilde{C}$. The triplet (A_c, B_c, C_c) satisfies, therefore, both (25a) and (25b) where $B_c = B$ and

$$C_c = \left[I + C(I - A)^{-1} K\right] C. \tag{44}$$

Equation (29b) readily follows from (44).

Example (continued)

We now replace the unstable model (22) by a stable one that satisfies (25a) and (25b) and Theorem 4. (The resulting model is not expected to match $H(1)$ because the unstable model was derived for $p = 0$). The eigenvector corresponding to $\lambda_1 = 1.6942$, $Au_1 = \lambda_1 u_1$ is

$$u_1 = (0.4951, -0.8746)'$$

solving the scalar equation (28b), with $(Cu_1)'Cu_1 = 0.1686$

$$q - \lambda_1^2 q = -0.1686$$

we get $q^{-1} = 11.0951$, thus P_c the solution of (33) is by (32),

$$P_c = u_1 q^{-1} u_1 = \begin{bmatrix} 2.7192 & -4.8037 \\ -4.8037 & 8.4859 \end{bmatrix}$$

and

$$[I + P_c C'C]^{-1} = \begin{bmatrix} -3.5010 & -3.6065 \\ 7.9512 & 7.3711 \end{bmatrix}^{-1} = \begin{bmatrix} 2.5682 & 1.2566 \\ -2.7703 & -1.2198 \end{bmatrix}$$

by which $A_c = A(I + P_c C'C)^{-1}$ is

$$A_c = \begin{bmatrix} 3.5829 & 1.6940 \\ -5.0324 & -2.2585 \end{bmatrix} \tag{45a}$$

whose eigenvalues can also be found by solving the characteristic equation

$$\lambda^2 - 1.3244\lambda + 0.4333 = 0$$

to be the expected values $\hat{\lambda}_1 = 0.5902 = \lambda_1^{-1}$ and $\hat{\lambda}_2 = 0.7342 = \lambda_2$.

The triple of matrices that satisfies, (25a), (25b) is (A_c, B_c, C_c) where A_c is given by (45a). The matrices $B_c = B$ and C_c derived by (29b), are given by

$$B_c = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{45b}$$

$$C_c = \begin{bmatrix} 0.7207 & 0.2752 \\ -1.7402 & -1.2274 \end{bmatrix}. \tag{45c}$$

This numerical example illustrates one of the attractive features of the stabilization method of this section. Although the modification is implicitely related to the solution of the discrete (singular) Ricatti Equation (33) in practice only a Lyapunov equations (28b) has to be solved. Moreover, the size of the solution matrices $Q > 0$ of the Lyapunov equation is only $\zeta \times \zeta$, where $\zeta < n$ is the number of unstable eigenvalues to be replaced by their stable mirrors.

IV. Conclusions

The model reduction problem of discrete time linear multivariable systems has been treated using the concept of generalized minimal partial realizations of mixed matrix sequences. The method presents a multivariable generalization of the approximation in the Pade sense of the steady-state and the transient responses suggested firstly for SISO systems by other techniques that are not adequate for multivariable systems. Former approaches asked for models of predetermined order and may not have solution even in the SISO case. The present method on the other hand finds models of minimal order that match a given sequence of p time moments and q Markov matrices and it is shown that this reduction problem always has a solution. Furthermore, the derived minimal order models converge and reproduce the original system from any sequence of large enough length for any p and q. The exact condition for the existence of a unique model or for the existence of many different models are given in terms of the common structural input and output properties of the models. In the latter case the set of all possible solutions is described by a maximal set of independent parameters that can be adjusted freely and used advantegously to satisfy additional desirable specifications.

The well-known stability problem encountered in any Pade-type model reduction problem is also treated. A range of possibilities to achieve stable models follow from the available choice among models that differently approximate the steady state and the transient responses of the system and from the freedom in assigning values to the obtained free parameters. A systematic complementary procedure is developed by which an already derived unstable model can always be replaced by a stable model of the same order. The stabilized model has the same square magnitude and it approximates, therefore, the singular values and the magnitude effect of the system.

The common structural properties of all possible minimal models that match a sequence of any p time moments and q Markov matrices of a given length r can be advantageously used for a significant computational saving in their derivations.

References

[1] Y. Shamash, "Analytical methods of reducing linear systems", *7th Ann. Pittsburgh Conf. on Modelling and Simulation*, pp. 305–309, April, 1977.

[2] ——, "Continued fraction methods for the reduction of discrete time dynamic systems," *Int. J. Contr.*, vol. 20, pp. 267–275, 1974.

[3] S. C. Chuang, "Homographic transformation for the simplification of discrete-time transfer functions by Pade approximations," *Int. J. Contr.*, vol. 22, pp. 721–728, 1975.

[4] ——, "Linear transformation for simplification of z-transfer functions by Pade approximations," presented at 3rd IFAC Symp. on Multivariable Technological Systems, Sept. 1974.

[5] M. Decoster and A. R. van Couwenberghe, "A comparative study of different reduction methods—part I and II," *J. Automatique*, vol. 17, pp. 67–74, 125–134, 1976, Belgium.

[6] T. Yahagi, "Optimal approximations of transfer of linear dynamical systems," *IEEE Trans. Circuits Syst.* vol. CAS-24, pp. 560–568, 1977.

[7] Y. Bistritz and G. Langholz, "Model reduction by best Chebyshev rational approximations in the complex plane," *Int. J. Contr.*, vol. 30, pp. 277–289, 1979.

[8] Y. P. Shih, and W. T. Wu, "Simplification of z-transfer functions by continued fractions," *Int. J. Contr.*, vol. 5, pp. 1089–1094, 1973. A comment on the above paper by K. J. Khatwani, and J. B. Bajwa, ibid, vol. 21, pp. 701–702, 1975.

[9] L. S. Shieh, K. M. Dadkhah, and R. E. Yates, "Reduction of z-transfer function matrices by means of the mixed method," *IEEE Trans. Ind. Electron Contr. Instrum.*, vol. IECI-23, pp. 371–375, 1976.

[10] Y. Shamash and D. Feinmesser, "Reduction of discrete time systems using a modified Routh array," *Int. J. Systems Sci.*, vol. 9, pp. 53–64, 1978.

[11] Y. Bistritz and U. Shaked, "Minimal Pade model reduction of multivariable systems," in *Proc. 1981 IEEE Conf. on Decision and Control* (San Diego, CA), pp. 672–675.

[12] Y. Bistritz, "Nested bases of invariants for minimal realizations of finite matrix sequences," *SIAM J. Contr Optimization*, vol. 21, pp. 804–821, 1983.

[13] A. J. Tether, "Construction of minimal linear state variable models from finite input-output data," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 427–436, 1970.

[14] R. E. Kalman, "On minimal partial realization of linear input-output maps," in *Aspects of Network and System Theory*, R. E. Kalman and M. DeClaris *et al.* New York: Holt Reinhart and Winston, 1971.

[15] L. M. Silverman, "Realization of linear dynamical systems," *IEEE Trans. Automat Contr.*, vol. AC-16, pp. 554–567, 1971.
[16] U. Shaked, "A transfer function approach to the linear discrete stationary filtering and the steady state discrete optimal problems," *Int. J. Contr.*, vol. 29, pp. 279–291, 1979.
[17] M. G. Safonov, A. J. Laub, and G. L. Hartman, "Feedback properties of multivariable systems; The role and use of the return difference matrix," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 47–65, 1981.
[18] I. Rosthethwaithe, J. M. Edmunds, and A. G. J. MacFarlane, "Principal gains and principal phases in the analysis of linear multivariable feedback systems, *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 32–46, 1981.
[19] B. C. Moore, "Principal component analysis in linear systems; Controlability, Observability and model reduction," *IEEE Trans Automat Contr.*, vol. AC-26, pp. 17–32, 1981.
[20] L. M. Silverman and M. Bettayeb, "Optimal approximation of linear systems," presented at Joint Automatic Control Conf. (San Francisco), 1980.
[21] S. Y. Kung and D. W. Lin, "Optimal Hankel-norm Model reductions: Multivariable systems," *IEEE Trans. Automat Contr.*, vol. AC-26, pp. 832–852, 1981.
[22] C. Hwang and Y. P. Shih, "On the time moments of discrete systems," *Int. J. Control*, vol. 34, pp. 1227–1228, 1981.
[23] O. H. Bosgra and A. J. J. Van der Weiden, "Input output invariants for linear multivariable systems," *IEEE Trans. Automat Contr.*, vol. AC-25, pp. 20–36, 1980.

Yuval Bistritz (S'79–M'79) was born in Romania in 1951. He received all his degrees from Tel-Aviv University, Israel. He received the B.Sc. degree in physics in 1972, his M.Sc. degree in electrical engineering in 1978 (summa cum laude, after also completing the requirements for a B.Sc. degree in electrical engineering), and the Ph.D. degree in electrical engineering in 1983.

From 1975 to 1978 he was a Research Engineer in the Israeli NRC in the area of simulation optimization and process control. Since 1979 he has been with the Department of Electronic, Communication, Control and Computer Systems, School of Engineering, Tel Aviv University, holding positions of Research Assistant, Instructor, and presently a Faculty Member position. His current research interests are realization, approximation, invariant description and stability of systems.

Dr. Bistritz was awarded a Chaim Weizmann Fellowship for Post-Doctoral scientific research.

Uri Shaked was born in Israel in 1943. He received the B.Sc and M.Sc degrees in physics from the Hebrew University, Jerusalem, Israel, in 1964 and 1966 and the Ph.D degree in applied mathematics-control theory from the Weizmann Institute of Science, Rehovot, Israel in 1974.

From 1974 to 1976 he was a Senior Visiting Fellow at the Control and Management Systems Division of the Engineering Department at Cambridge University, England. In 1976 he joined the Dept. of Electronic Communication Control and Computer Systems at Tel-Aviv University as a Lecturer. He was promoted to Senior Lecturer in 1977 and to an Associate Professor in 1982. He is currently a Visiting Professor at the Department of Electrical Engineering and Computer Sciences in the University of California at Berkeley. His research interest, in the area of automatic control, is in developing transfer function approach for the design of optimal control systems and in investigating the nearly singular optimal filtering and control problems.

Dr. Shaked was elected to a Royal Society Research Fellowship in 1975.

Bounds on the Tolerance Sensitivity of Resistive Networks

LUDO WEYTEN, MEMBER, IEEE

Abstract —**For the general class of resistive networks, lower bounds are derived on the summed absolute and squared element sensitivities of any dimensionless network function. As an application example, these results are used to obtain a lower bound on the linearity performance of digital-to-analog converters.**

I. Introduction

WE CONSIDER a network function F. The sensitivity of this network function to resistor variations is defined as

$$S_{R_i}^F = \frac{R_i}{F}\frac{\delta F}{\delta R_i}. \tag{1}$$

It is well known that the sum of sensitivities to all network elements is a function of F only, and is independent of the network configuration [1], [2]. The sum of absolute (or squared) sensitivities, on the contrary, does depend on the circuit and thus can be used to compare different networks realizing the same network function F. In a worst-case situation we are interested in the sum of absolute values.

Manuscript received March 2, 1982; revised March 15, 1983.
The author is with the Electronics Laboratory, University of Ghent, B-9000 Ghent, Belgium.

0098-4094/84/0400-0390$01.00