BRIEF ON ORDER-REDUCTION FOR MODELS: A CRITICAL SURVEY

NGUYEN THUY ANH, NGUYEN NGOC SAN

Abstract. A review on different methods for obtaining reduced-order models for complex high-order systems is briefly made. A critical comparison is made of the extent to which the models obtained from the optimal projection equations adopting state - optimization method proposed by authors are seen retaining the physical significance of the original modeled states.

1. INTRODUCTION

In most practical situation of system control, a fairly complex, high-order often makes a difficulty in understanding the behavior of the system as well as in controlling the system can be accomplished with a great easy by using suitably selected low-order model having the important characteristics of the model.

During the last 40 years, a great deal of research work has been carried out for solving the order-reduction problem, as would be evident from the fact that more than 500 research papers have been so far published proposing different approaches on the subjects. However, practically all of the proposed approaches are seen to belong to one of three main groups. The first groups of methods (approaches) attempts to retain the important eigenvalues of the system-model and then obtain the corresponding parameters of the low-order model in such a manner that the response of the low-order model to certain inputs is a close approximation to that of the original model. The earliest methods of order-reduction for models proposed by Marshall [27], Davison [9], Mitra [28] and Aoki [3] are belonged to the first group. However, Hickin and Sinha [15] have shown that the first three methods may be regarded to be special cases of the aggregation method proposed by Aoki [3].

The second group is based on an optimum manner indifferent of the eigenvalues location of the original model. Anderson [1] has proposed a geometric approach based on orthogonal projection for obtaining a low-order model minimizing square errors in the time-domain. Sinha and Pile (1971) have proposed a method utilizing least squares fit with the samples of the response. Other criteria for optimization have also been studied for the mentioned purpose i.e. Wilson [45], Sinha and Berezail (1971), Bandler et al. [4], Hyland and Bernstein [17], Nath and San [31], San [39], Methods for obtaining optimum low-order models in the frequency domain have been proposed by Langholz and Bishtritz [21], Elliott and Wolovich [10], Haddad and Berntein (1985).

The third group of methods is based on matching some other properties of the responses. Chen and Shieh [6] have shown that if the continued fraction expansion of a transfer function was truncated it led to a low-order model with step response matching closely that of the original model of the system. The main attraction of this approach lies on the fact that computation is simpler as compared with methods in the earlier groups. The method of matching time moments proposed by Gibarillo and Lees [13] is an other interesting one to the problem. Later Shemash (1974) has shown that these method are equivalent and classified as Pade' approximation. An important drawback of the methods using pade' approximation is that low-order models obtained may sometimes turn out to be unstable even though the original one is stable. This led to the development of the Routh approximation method by Hutton and Friedlan [16] for a single-input-single-output model, a multivariable version of which was developed later by Sinha et al. (1980). Another solution is to combine aggregation with the matching of time-moments to ensure the stability, Hickin and Sinha [15]. Bistritz and Lanholz [5] have proposed a method for obtaining stable Chebyshev Pade' approximation.

However, there are still some methods which do not belong to any of the above mentioned groups. The singular perturbation methods proposed by Sannuti and Kokotovic [37] and balanced matrix method is particularly useful whenever the system has two time-scale property. The states are partitioned into "slow" and "fast" parts and the reduction is achieved by setting the derivatives of the "fast" states to zero so that "fast" states can be eliminated. An important advantage of the method lies on that the physical nature of the problem is preserved. Further, if necessary, the effect

of the "fast" modes, which were neglected, can be determined by returning to the original system, since all are not lost.

However, there are different difficulties with the method. The main difficulty lies on the fact that how to determine an appropriate partitioning of the state vector modes "fast" and "slow" are seldom decoupled making difficulty in deciding which state is related to which mode. Furthermore, "fast" and "slow" are relatively defined only.

The balanced matrix method is based on simultaneous diagonalization to the same diagonal matrix of the controllability and observability Gramians. The mentioned diagonalization permits the original systems to be transformed to an equivalent "internally balanced" system. A low-order model is then obtained by deleting the least significant eigenvalues contributing to the input/output relationship of the transformed system, the internally balanced method has been also adopted to the "closed loop" thinking system problem, Jonekheere and Silverman [20], Mustafa and Glover [30].

In the author's opinion, the purpose of order-reduction models is relegated to utilization of the low-order models for preliminary understanding and controlling the original system. The importance of reduction hence, lies rather on preserving the physical significance of the modeled states in the reduced version than on achieving the absolute minimum error. For this purpose, a provision would be made for keeping information about the desired states of the original states in the reduced version.

The object of this paper is to describe a brief survey of the basic methods. This will be followed by a presentation of recent work, one is of Hyland and Bernstein [17] and the other is of the author, San [39], both are seen belonging to the second group developed on an optimum principle. From the first-order necessary conditions for the optimization in the L_2 problem, Hyland and Bernstein have found the existence of an optimal projection which couples two modified Lyapunov equation in the optimal projection equations (OPEQ). The significance of developing the first-order necessary conditions for the L_2 optimization of model reduction problem in the OPEQ form lies on the question of multi-extrema since the coupling effect resulting within OPEQ can be seen to be the effect of an additional constrained-condition to the L_2 problem. A sufficient condition is obtainable on satisfying the L_2-limit and pre-assigned H_∞ bound. However, a lot of difficulties of the OPEQ developed by Hyland an Berntein especially on the computation point of view is removed by adopting the state-optimization approach proposed by one of the author, San [39]. The author has found that there exists a nonsimilarity transformation between the original model and reduced one and this nonsimilarity transformation can be factorized in terms of a partial isometry. An optimal projector has formed leading to a new form of OPEQ in the standard-like form of the Lyapunov equations. The effect of factorization has been seen as if an other additional constrained-condition were used for decoupling the modified Lyapronov equations. The significance of the author's method lies on the OPEQ developed preserving the desired state in the reduced version.

2. STATEMENT OF THE PROBLEM

Consider a linear, time invariant, multivariable system described by

$$
\begin{aligned}
x &= AX + Bu, \\
y &= Cx
\end{aligned}
\tag{1}
$$

where $x \in R^n$; $U \in R^p$; $y \in R^q$; $A \in R^{nxsn}$; $B \in R^{nxp}$; $C \in R^{qxn}$.

The object of a reduction-order for model is to obtain the system of equations

$$
\begin{aligned}
x_r &= A_r x_r + B_r u, \\
y &= C_r x_r,
\end{aligned}
\tag{2}
$$

where $x_r \in R^r$; $r \leq n$; $y_z \in R^q$; $A_r \in R^{r \times r}$; $B_r \in R^{r \times p}$; $C_r \in R^{q \times r}$.

Such that y_r is a close approximation to y for all inputs $u(t)$ in some class of function.

3. ON THE BASIC METHODS

3.1. The aggregation method

Among the methods which retain the important eigenvalues of the original system in the reduced model, the most general is the aggregation method proposed by Aoli [3]. It is based on the intuitively

appealing relationship.

$$w = Kx \tag{3}$$

where K is an $r \times n$ constant projection matrix and is called the aggregation matrix. Equation (3) is called the aggregation law. Differentating both sides of equation (3) and substituting for x from equation (1), we obtain

$$w = KAx + LBu \tag{4}$$

Comparison of equation (4) with equation (2) yields the following relationships between the matrices in the two sets of state equations.

$$KA = A_r K,$$
$$KB = B_r, \tag{5}$$
$$C \approx C_r K,$$

where the last is only an approximate equality.

It is easily shown ([15]) that a nontrivial aggregation law exists if and only if the eigenvalues of A_r are a subset of the eigenvalues of A. The aggregation matrix is

$$K = D[[I_r \ 0]]V^{-1}, \tag{6}$$

where D is any nonsingular $r \times r$ and V is the modal matrix of A; the columns of V are the eigenvectors (generalized eigenvectors) of A. D is usually chosen so that K is a real matrix. In the particular case when D is an identity matrix, A_r will be obtained in the diagonal or Jordan form. The eigenvalues of A_r are those eigenvalues of A which correspond to the first r comuns of V. There is a natural choice for K, such that $K = V[[I_r \ 0]]^T D^{-1}$, (Lastman and Sinha 1983).

Some difficulties with the aggregation method are described below.

i) The determination of the reduced model requires the computation of the eigenvalues and eigenvectors of the A-matrix, which may have a large dimension. Although good methods for such computation exist, they require considerable amount of computer time.

ii) The d, c steady-state gain may not be preserved with the aggregation method, with the result that step-responses of the original system and the reduced-order model differ considerably. This mismatch in responses can be overcome by combining aggregation with moments matching [15].

iii) An important question in the aggregation method is how to select the eigenvalues. A rational approach is to consider the total impulse response energy in the output of the system and preserve those eigenvalues which contribute most. This criterion may also be utilized for deciding the most suitable order of the reduced model. Commault [8] also uses unit impulses to obtain a measure of the influence of each eigenvalue of A, for determining dominance. Another criterion for selection of eigenvalues to be retained is the contribution of that mode to the time moments of the system [47].

iv) Another shortcoming of the aggregation method (and, indeed, of most other methods of model reduction) is that the reduced states do not have a physical significance. This creates a problem in case the reduced model is to be considered in conjunction with other parts of the system, where the interconnections are through the states. This difficulty is overcome only partly by the singular perturbation method and a method proposed by Rozsa, Sinha and Lastman [36], Moore [29] also proposed internal dominant concept.

3.2. Methods based on matching moments

Another approach to model reduction is based on matching time moments of the impulse response of the original system with those of the reduced model. The continued-fraction expansion method of Chen and Shieh [6] was the earlied method of this type, although the connection with moment matching and Pade' approximation was realized later (Samash, 1974). The generalized version of this approach which is applicable to multivariable system has been proposed by Hichkin and Sinha [15].

For the system represented by equation (1), the transfer function matrix is given by

$$G(s) = c(sI - A)^{-1}B. \tag{7}$$

A formal Lauren series expansion of $G(s)$ in that case is given by

$$G(s) = \sum_{i=0}^{n} J_i S^{-(i+1)}, \tag{8}$$

where

$$J_i = C A^i B \tag{9}$$

are called the Markov parameters of the system and are invariant under linear transformation of the state.

If $G(s)$ has no poles at the origin of the s-plane, we can obtain the following Taylor series expansion

$$G(s) = \sum_{i=0}^{\omega} J_i S^{-(i+1)}, \tag{10}$$

where

$$J_i = C A^{-(i+1)} B. \tag{11}$$

It may be noted that if $g(t)$ is the inverse Laplace transform of $G(s)$ then

$$\int_0^{\infty} t^i g(t) = (-1)^{i+1} (i!) J_i. \tag{12}$$

where i is a positive integer.

In other word J_i are related to the time moments of the impulse matrix through a multiplicative constant.

Combining equations (9) and (10), it is noted that

$$J_{-i} = J_{i-1} \quad \text{for } i > 1 \tag{13}$$

so that the term "generalized Markov parameter" may be used to include J_i.

To determine a low-order model, the matrices A_r, B_r and C_r are obtained such that a number of generalized Markov parameters are identical to these of the original system. It may be noted that by matching time-moments one equates the steady-state responses to inputs in the form of power series (steps, ramps, parabolic functions, etc.). On the other hand, matching Markov parameters improves the approximation in the transient portion of the response.

The process of obtaining A_r, B_r and C_r is called partial realization, a block Hankel matrix is formed, consisting of the generalized Markov parameters.

$$H_{-ij}^{(k)} = \begin{vmatrix} J_{-k} & J_{-k+1} & \cdots & J_{-k+j-1} \\ J_{-k+1} & J_{-k+2} & \cdots & J_{-k+j} \\ \vdots & \vdots & & \vdots \\ J_{-ki-1} & J_{-k+1} & \cdots & J_{-k+i+j-2} \end{vmatrix}. \tag{14}$$

If $i > \alpha$ and $j > \beta$ are the observability and controllability indices of the system, respectively, then the rank of $H_{ij}(k)$ is n, and a minimal realization of order n is easily obtained following reduction to the Hermite normal form (Rozsa and Sinha 1974). If the process is stopped after r steps, where $r < n$, one gets a partial realization matching the first q generalized Markov parameters starting with J_k, where

$$q = [r/m] + [r/p]. \tag{15}$$

In view of the above, partial realization may be regarded as the generalization of Pade' approximation to the multivariable case.

The main drawback of all such methods is that the stability of the reduced model is not guaranteed even though the original system is stable. Many methods have been developed to overcome this problem. The most notable in this group is the Routh approximation method, developed by Hutton and Friendland [16] for single-input-single-output systems. The multivariable version of this method Chen [7] and Sinha et al. [41] obtains the characteristic polynomial of the even and odd parts of the characteristic polynomial of the original system. This determines the matrices A_r and B_r in a canonical form. The C_r matrix is then obtained to match as many time moments as possible.

The main attraction of the methods matching time moments is that they require considerably less computation. The drawback of the methods stands on the fact that there exists no direct relationship between the states of the original system and those of the reduced-order model.

3.3. Aggregation with moment matching

Since the aggregated model retains the dominant eigenvalues of the original system, its stability is guaranteed if the system is stable. The steady-state responses of the model, however, do not necessarily match with those of the original system. This can be remedied by obtaining an aggregated model which matches as many time moments as possible. Not only does this give a better approximation to the response while retaining stability, but it also has the projective relationship between the states of the system and the reduce model. This is particularly useful if the latter is to be utilized for design with state-variable feedback.

The procedure for obtaining an aggregated model matching some of the time moments is straighforward. Since the matrix A_r in the diagonal (or Jordan) form is determined entirely by the eigenvalues to be retained in the low-order the matrix $b_R = kb$. The elements of the matrix C_r may now be selected to match as many time moments as possible.

3.4. The singular perturbation method

This is an attractive approach for model reduction since the physical nature of the model is preserved. It is based on partitioning the state vector into two parts, the "slow" and the "fast" part. Hence, equation (1) may be rewritten as (where x_2 represent the "fast" modes)

$$\begin{vmatrix} \dot{x}_1 \\ \dot{x}_2 \end{vmatrix} = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \begin{vmatrix} X_1 \\ X_2 \end{vmatrix} + \begin{vmatrix} B_1 \\ B_2 \end{vmatrix} u. \tag{16}$$

For a stable system, the "fast" modes decay much more quickly than the "slow" modes. So that after the transient period it is a reasonable approximation to set derivative of x_2 to zero. Consequently, it is possible to eliminate x_2 from equation (16) to obtain

$$x_1 = (A_{11} - A_{12}A_{22}^{-1}A_{21})x_1 + (B - A_{12}A_{22}^{-1}B_2)u. \tag{17}$$

The reduced model represented by equation (17) can now be solved to obtain the states immediately. If necessary, the effect of the "fast" modes, which were neglected, can also be determined by returning to equation (16).

The main difficulty with this method is the problem of deciding the proper partitioning of the state vector. It is further complicated by the fact that, in general, the states are not decoupled, so that it is difficult to relate a particular state to a particular mode.

3.5. The balanced method

This method is firstly proposed by Moore [29], then is further developed by Perenbo and Silverman [33] and Glover [14]. The method is based on the diagonalization simultaneously the controllability and observability Gramians of system described by equation (1). The Gramian for controllability and that for observability are defined as

$$W_c = \int_0^\infty e^{At} BB^T e^{A^T t} dt, \tag{18}$$

$$W_0 = \int_0^\infty e^{A^T t} C^T C e^{At} dt. \tag{19}$$

If A is stability matrix (each eigenvalue of A has negative real part) and system described by equation (1) is completely controllable and observable, then (18) and (19) are the unique real symmetric positive definite matrix solution of the Lyapunov equations

$$AW_c + W_c A^T = -BB^T, \tag{20}$$

$$A^T W_0 + W_0 A = -C^T C, \tag{21}$$

Since W_c is real, symmetric and positive definite, there exist an orthogonal matrix V_c and a diagonal matrix $A_c = \text{diag}(\mu_1, \mu_2, \ldots, \mu_n)$, where $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n > 0$, such that

$$V_c^T W_c V_c = A_c^2. \tag{22}$$

We now form the real symmetric positive definite matrix

$$W = (V_c \Lambda_c)^T W_0 (V_c \Lambda_c) \tag{23}$$

which can be diagonalized by $P^T W P = \Lambda$ where P is an orthogonal matrix and

$$\Lambda = \operatorname{diag}(\sigma_1, \sigma_2, \dots \sigma_n) \tag{24}$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$. Then the nonsingular matrix

$$S = V_c \Lambda_c P \Lambda^{-0,5} \tag{25}$$

has the property that

$$\begin{aligned}
(W_c(^* &= S^{-1} W_c (S^{-1})^T = \Lambda, \\
(W_0)^* &= S^{-1} W_0 = \Lambda,
\end{aligned} \tag{26}$$

where $(W_c)^*$ and $(W_0)^*$ are the controllability and observability Gramians of the transformed system

$$\begin{aligned}
x^* &= A^* x^* + B^* u, \\
y &= C^* x^*, \\
A^* &= S^{-1} A S, \qquad B^* S^{-1} B, \qquad C^* = CS.
\end{aligned} \tag{27}$$

The transformed system (27) is said to be internally balanced. If $\sigma_r \gg \sigma_{t+1}$ in (24) then there is an r-th order internally dominant subsystem in (27) [29, 33]. An r-th order reduced model can be obtained from (27); it is also balanced and has the form (2) with A_r equal to the upper left $r \times r$ block in A^*, B_r consists of rows 1 to r of B^*, C_r consists of columns 1 to r of C^*. Because A is a stability matrix, A_r is also a stability matrix [29].

Note that the balanced reduced-order model is obtained by deleting the least controllable and observable states from equation (27). Consequently, the states in the reduced-order model are an approximation to the first r states from equation (27). A comparison between the balanced matrix method and the aggregation method is given by Lastman et al. [22], where it was found, from computational examples, that a properly chosen aggregation model can compare favorably with an internally dominant balanced matrix model of the same order, provided the eigenvalues of the aggregated model are truly dominant. Worst-case error analysis of the balanced method of model reduction has been given in a more recent paper [25]. It was found that when the model and the full-scale system are balanced, the computation for obtaining the error bounds is simplified.

In a recent paper, Prakash and Rao [34] have proposed a modification to Moore's method where the reduced-order model is obtained by approximating the state of the weak subsystem of the balanced representation around zero frequency. This has the effect of reducing the spectral norm of the modeling error at low frequencies.

3.6. Models using optimum approximation

Instead of obtaining the reduced model preserving the important eigenvalues of the original system, one may like to relax this restriction and select the parameters of the reduced model of a specified order in such a manner that its response to a given input is an optimum approximation, in some sense, of the response of the original system to the same input. Alternatively, one may attempt optimum approximation of the frequency response.

Several authors have proposed different methods of time-domain approximation. Anderson [1] proposed a geometric approach, based on orthogonal projection, for obtaining the low-order model which minimizes the integral square error. Sinha and Pille (1971) proposed utilizing the matrix pseudoinverse for obtaining a reduced order model which minimizes the sum of the square of the samples of the error between the responses. Other criteria for obtaining the optimum approximation have also been proposed (Sinha et al., 1971; Banaler, et al. 1973; Bistritz and Lanholz 1979). Elliott and Wolovich [10] have proposed a frequency domain model reduction procedure which is also applicable to multivariable systems.

In general, the reduced order models obtaining using these methods provide a better match than that obtained by aggregation. The use of efficient numerical optimization methods usually reduces the computational requirements as well. However, there is no guarantee that these models

will be useful for obtaining a near-optimal design of the controller. Furthermore, the states of the reduced model are not directly related to the states of the original system.

3.7. Some remarks

a. Selection of states and/or eigenvalues to be retained

All the methods discussed above, with the exception of singular perturbation, suffer from the drawback that the physical significance of the states is lost. This creates a problem when the original system is a part of a larger system. Such a situation occurs quite often while modeling large-scale power larger system of interconnected subsystems for the study of dynamic stability.

In view of the above, a method has been proposed (Rozsa et al., 1982; Lastman et al., 1983) in which the components of the low-order model state vector formed a subset of the components of the original state vector, so that the states of the reduced model retain their physical significance. The choice of the state to be retained is made on the basis of the energy integral participation matrix. Although the final procedure is similar to singular perturbation, it differs in the way to obtain the reduced order model through actual of the "slow" (dominant) and "fast" (non-dominant) modes.

Like the aggregation method, the proposed method requires the computation of the eigenvalue of A. One can assume that A is a stability matrix. Let

$$A = V\Lambda U, \tag{28}$$

where the columns of V are the right eigenvectors of A and the rows of U are the left eigenvector of A. Corresponding to the eigenvalues of A, Λ is given by

$$\Lambda = \text{diag}\,(\lambda_1, \lambda_2, ..., \lambda_n). \tag{29}$$

The column vectors in V and U are scaled in such a manner that

$$VU = I. \tag{30}$$

In (29), the eigenvalues have order so that

$$0 > \text{Re}\,(\lambda_1) \geq \text{Re}\,(\lambda_2) \geq ... \geq \text{Re}\,(\lambda_n). \tag{31}$$

If one assumes that all the inputs are unit impulses, it can be shown that the total energy at the output is given by

$$E = \int_0^\infty y^T y\,dt = \text{tr}\,(PW), \tag{32}$$

where $P = C^T C$, W is an $n \times n$ matrix that has $-b^T U^T M_{ij} U b$ as its (i,j)-th entry, b is the sum of the columns of B, M_{ij} is an $n \times n$ matrix that has $v_\mu^{(i)} v_\nu^{(j)}/(\lambda_\mu + \lambda_\nu)$ as its (μ, ν)-th entry, and $v^{(i)}$ is the i-th row of V.

Accordingly, the most significant states of the original system, to be retained in the reduced model, correspond to the elements on the main diagonal of PW that have the largest magnitude. This also indicates the dominant eigenvalues relative to their ordering in equation (31).

Once the most significant states and dominant eigenvalues have been determined, a reduced-order model can be obtained. If the physical significance of the states need not be retained within the low-order model, then the reduction methods discussed earlier can be used. However, if the physical significance of the states must be retained then the reduction method given below can be used.

b. On the retaining of physical significance of the states

The physical significance of the state can be made retaining if the following partition of the state vector would be possible (Rozsa et al., 1982; Lastman et al., 1983).

$$x = \begin{bmatrix} w^T & z^T \end{bmatrix}, \text{ where } w(t) \in R^r, \, z(t) \in R^{n-r}, \tag{33}$$

$$\Lambda = \text{diag}\,(\Lambda_1, \Lambda_2), \text{ where } \Lambda_1 = \text{diag}\,(\lambda_1, ..., \lambda_r), \, \Lambda_2 = (\lambda_{r+1}, ..., \lambda_n), \tag{34}$$

$$U = \left[\begin{array}{c|c} U_{11} & U_{12} \\ \hline U_{21} & U_{22} \end{array}\right], \qquad A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right], \tag{35}$$

$$B = \begin{bmatrix} B_1 \\ \hline B_2 \end{bmatrix}, \qquad C = [C_1 \; C_2].$$

In the above, U_{11}, A_{11} are $r \times r$ matrices, B_1 is $r \times p$ and C_1 is $q \times r$ and the block have appropriate dimensions. In equation (35), one does not assume the eigenvalue ordering of equation (31); A_{11} consists of the most dominant eigenvalues as determined by the procedure described in the previous section just as the states x_r are the most significant states. The eigenvalues may have to be renumbered from equation (31) in order to obtain equation (34).

The reduced model is of the same form as in equation (2), with

$$A_r = A_{11} - A_{12} U_{22}^{-1} U_{21}, \tag{36}$$

$$B_r = B_1 - A_{12} U_{22}^{-1} \Lambda_2^{-1} (U_{21} B_1 + U_{22} B_2), \tag{37}$$

$$C_r = C_1 - C_2 U_{22}^{-1} U_{21}. \tag{38}$$

Lastman et al. (1983) have proved that the response of system described by equation (2) with A_r, B_r and C_r given in equation from (36) to (38) converges to the response of the system described by equation (1) for any input u as approaching to infinity. This, except, for a transient phase, the response of the reduced-order model obtained is a close approximation to the response of the original system for all quasi-state inputs.

4. THE OPTIMAL PROJECTION EQUATIONS

4.1. The work of Hyland and Bernstein [17]

Lemma 1. *Suppose $\hat{Q}, \hat{P} \in R^{n \times n}$ are non-negative definite. Then QP is semisimple. Furthermore, if rank of $\hat{Q}\hat{P} = r$ then there exists G, $\Gamma \in R^{r \times n}$ and positive semisimple $M \in R^{r \times r}$ such that*

$$\hat{Q}\hat{P} = G^T M \Gamma. \tag{39}$$

Proof. See Hyland and Bernstein [17]; also generalization version is available in paper of San [39].

Theorem 1. *Suppose $(A_r, B_r, C_r) \in @^+$ solves the optimal model-reduction problem. Then there exists non-negative definite matrices $\hat{Q}, \hat{P} \in R^{n \times n}$ such that, for some (G, M, Γ) factorization of $\hat{Q}, \hat{P}, A_r, B_r$ and C_r are given by*

$$A_r = \Gamma A G^T, \tag{41}$$

$$B_r = \Gamma B, \tag{42}$$

$$C_r = C G^T, \tag{43}$$

and such that with an optimal projection matrix $\sigma = G^T \Gamma$, the following condition are satisfied

$$\rho(\hat{O}) = \rho(\hat{P}) = \rho(\hat{Q}\hat{P}) = r, \tag{44}$$

$$\sigma[A\hat{Q} + QA^T + BVB^T] = 0, \tag{45}$$

$$[A^T \hat{P} + \hat{P}A + C^T RC]\sigma = 0, \tag{46}$$

where $\rho(\cdot)$ stands for the rank of $(\cdot)$. Equations (45) and (46) are seen to be coupled each to other by optimal projection matrix, hence they have been termed as modified Lyapunov equations, in which $\hat{Q}$ and $\hat{P}$ have been called the controllability and observability pseudo gramians.

Proof. See Hyland and Berstein [17].

Proposition 1. *Suppose (A_r, B_r, C_r) is extremal. Then the model-reduction error is given by:*

$$J_m(A_r, B_r, C_r) = 2 Tr[(QP - W_c W_o)A] = 2 Tr[(\hat{Q}\hat{P}A] + Tr[C^T RCW_c]$$

$$= 2 Tr[(\hat{Q}\hat{P}A] + Tr[BVB^T W_c]. \tag{47}$$

Proof. See Hyland and Bernstein [17].

With respect to the work of Hyland and Bernstein different discussions are made as follows.

a. The optimal projection equations are seen to be simpler form for the necessary conditions for the L_2-problem. The equation (45) and (46) can be computed only with some additional condition for decoupling them.

b. These optimal projection equations are found to be applicable to a controllable and observable system only. The fact hides behind it is that in the development of these equations, the constraint conditions have got no alternative but to be used in the augments form which demands the system to be an originally controllable and observable one. However, a system may actually consist of two other parts namely unobservable and uncontrollable ones. Hence, the developed OPEQ are found to have an impractical application. It is a desire to have the OPEQ independent of the nature of the system.

c. The common opinion is that the importance of reduction lies rather on preserving the physical significances of the modeled states in the reduction version than on achieving the absolute minimum error as the reduced version is utilized for further investigation of the system with respect to the modeled states. But the physical significances of the original states have been found losing their traces in the reduced versions obtainable with the developed OPEQ. So, the development of OPEQS a provision can be made for keeping information about desire states of the original system in the reduced version.

d. It is seen from the proof of the theorem, a high complexity of mathematics involved in the development of the result, especially some mathematical formulas for derivatives with respect to matrix variables are not familiar to scientist.

All the above comments leading to the work of San [39].

4.2. The work of San [39]

Lemma 2. *Let X_n, the state vector of a system, be given. Then there exists always an $r \times n$ matrix T, $\rho(T) = r < n$, which transforms the n states of the system to r independently specified states in n_r, the states of the system to r independently specified states in n_r, the state vector of the optimal reduced model. If the number q of the system outputs is less than r, then $T^+ x_r$ leads to the minimum norm among the least-squares criterion on the output-error for model reduction.*

Proof. See [39].

Lemma 3. *Let the full row rank $r \times n$ matrix T, $x \in R(T^T)$ and x_r be given. Then there exists an $r \times n$ partial isometryu E such that T is factorized as:*

$$T = GE = EH, \tag{48}$$

where G and H are $r \times n$ positive definite and $n \times n$ non negative definite matrix respectively.

Proof. See also in [39].

Theorem 2. *For a given linear, time-invariant system of the order n, there exists always a full row rank $r \times n$ transformation T on the states of the system such that optimal parameters of the reduced-order model are given as*

$$A_r = TAT^+; \quad B_r = TB; \quad C_r = CT^+.$$

Proof. See [39].

Theorem 3. *There exists an $r \times n$ partial isometry E and $n \times n$ nonnegative definite matrix H such that the optimal parameters of the reduced-order model are accordingly given by:*

$$A_r = EHAH^+E^T; \quad B_r = EHB; \quad C_r = CH^+E^T. \tag{50}$$

Further, there exists an $n \times n$ optimal projection matrix δ and two $n \times n$ nonnegative definite matrix Q and P such that if optimal model is to be controllable and observable, the following conditions are then to be satisfied:

$$\sigma\left[HA_n\hat{Q} + \hat{Q}A_n^T H + HB_n V_1 B_n^T H\right] = 0, \tag{51}$$

$$\left[H^+ A_n^T + \hat{P}^+ A_n H^+ H^+ H^+ C_n^T R_2 C_n H^+\right]\sigma = 0, \tag{52}$$

where $V_1 = Expectation\ (U_n\ U_n^T)$; R_2 is a $q \times q$ weighted matrix.

Proof. See [39].

Two modified Lyapunov equations (51) and (52) can be easily written as:

$$\sigma \left[H A \hat{Q} + \hat{Q} A^T H \sigma + \sigma H B_n V_1 B_n^T H \sigma \right] = 0, \tag{53}$$

$$\left[\sigma H^+ A^T + \hat{P}^+ A_n H^+ \sigma + \sigma H^+ C_n^T R_2 C_n H^+ \sigma \right] \sigma = 0. \tag{54}$$

Proposition 2. *Let optimal parameters of the reduced-order model obtained from equation (49). Then the cost function is*

$$J_m(A_r, B_r, C_r) = Tr \left[(I_n - H^+ \sigma H)^T (C_n^T R_2 C_n) H^+ (I_n - H^+ \sigma H) \right]. \tag{55}$$

Proof. See [39].

Numerical example. It is demonstrated in this example the advantages of the present method over the existing one. The system considered by many authors for theirs comparison is of fourth order having the parameters as

$$A_4 = \begin{vmatrix} 0 & 0 & 0 & -150 \\ 1 & 0 & 0 & -245 \\ 0 & 1 & 0 & -113 \\ 0 & 0 & 1 & -19 \end{vmatrix}, \qquad B_4 = \begin{vmatrix} 4 \\ 1 \\ 0 \\ 0 \end{vmatrix}, \qquad C_4 = \begin{vmatrix} 0 \\ 0 \\ 0 \\ 1 \end{vmatrix},$$

and a second order model is the subject.

Assuming due to some technical reasons one has a desire to retain a information about originally modeled states of the system in the reduced model as

$$x_r = \begin{vmatrix} x_{r1} \\ x_{r2} \end{vmatrix} = \begin{vmatrix} x_{n1} + 0.06x_{n3} \\ x_{n2} + 0.05x_{n3} + 0.01x_{n4} \end{vmatrix}. \tag{57}$$

Then

$$T = \begin{vmatrix} 1 & 0 & 0.06 & 0 \\ 0 & 1 & 0.05 & 0.01 \end{vmatrix} \tag{58}$$

which is a full row rank matrix. It is readily obtained

$$T^+ = \begin{vmatrix} -0.9964134 & -2.988867E\text{-}4 & 0.0597848 & -2.98867E\text{-}6 \\ -2.988867E\text{-}4 & 0.9998754 & 4.997583E\text{-}2 & 0.9998754E\text{-}2 \end{vmatrix}^T. \tag{59}$$

Corresponding to which, the best approximate to the original state vector of the system obtainable from the reduced one. If some one has a desire to get back as

$$x_{41} = 0.9964134x_{21} - 2.98867E\text{-}4x_{22},$$

$$x_{42} = -2.988867E\text{-}4x_{21} + 0.9998754x_{22},$$

$$x_{43} = 0.5978x_{21} + 4.997583E\text{-}2x_{22}, \tag{60}$$

$$x_{44} = -2.98867E\text{-}6x_{21} + 0.9998754E\text{-}2x_{22}.$$

The parameters of the reduced-order model are obtained according to Theorem 4.2 as

$$A_2 = \begin{vmatrix} 4.684152E\text{-}2 & -1.507612 \\ 1.070463 & -2.452044 \end{vmatrix}, \qquad B_2 = \begin{vmatrix} 4 \\ 1 \end{vmatrix}, \qquad C_2 = \begin{vmatrix} -2.988867E\text{-}4 \\ 0.9998754E\text{-}2 \end{vmatrix}^T. \tag{61}$$

This model has a pair of complex conjugate poles a real zero on the L.H.S. of the s-plane in the transfer function, which is

$$F(s) = \frac{8.803206E - 3xs + 3.986389E - 2}{s^2 + 2.405203xs + 1.498985}. \tag{62}$$

The impulse response of this reduced model can be computed and the error arised due to the reduction can be estimated. This error is not expected to be absolute minimum as no desired in this case.

It can thus be concluded the proposed method does not depend on both the nature and the eigenvalues of the original system. Hence, it is seen that the method enables filtering out the states related with high eigenvalues to preserve the desire states of the original system in the reduced model. However, the optimality in reduction is uniquely achieved due to the role of T^+. The minimum error resulted from the reduction, if it is a require, can also be made reachable by following a choice of T, for which different suggestions are given below.

Suggestion of choing T:

a. Relationship with the result obtained by Wilson [45].

Hyland Bernstein expressed Wilson's result for a narrow domain in optimal projection equations and related these equations with the results obtained by Moore [29] and Skelton [42]. Thus, it is sufficient to differentiate the present result with that obtained by Wilson.

The optimum reduced model obtained by Wilson are of the form:

Since the state space descriptions of the original system and reduced model are not unique, it follows that and also lead to a solution of the reduction problem, where are respective $n \times n$ and $r \times r$ arbitrary similarity noting that is a projector with rank r, a Π exists such that

$$\Pi^{-1}\theta_2\theta_1\Pi = \begin{pmatrix} I_0 & 0 \\ 0 & 0 \end{pmatrix} \tag{64}$$

It follows that (θ, Π) which can be verified by checking the four Moore-Penrose's condition for a pseudoinverse. The, by chosing Theorem 3.1 is obtained.

Different points to be noted on this occasion are in order. First, the result obtained by Wilson cannot be obtained for a system consisting of unstable, even uncontrollable part, while the present theorem does not deny for this. In the case of a controllable and observable where the result obtained by Wilson can be applied, the present result also agrees well with this result. However, the straight-forward, from the methodology point of view, is belonged to the state-optimization thinking. This is the cause for a simple mathematical treatment to be involved in tackling the problem. Second, the physical significances of the originally modeled states of the system are preserved in the reduced model obtainable with the present method. Further one can pay attention on particular combination of desired states in the reduced version for different purpose, while in the reduction achievable with result obtained by Wilson there is no trace of the original modeled states. Third is on the usage point of view. The result obtained by Wilson requires to carry an optimization after solving matrix nonlinear equations with a guess on initial values of the parameters of the reduced model. The convergence in soluting the said equations cannot be surely obtained. While, the present result is simple to be applied requiring only an optimization to be carried out in regard the choice of T for obtaining the absolute minimum. Less effort will be required for an such optimization if T is properly chosen.

b. Corresponding to the optimal parameters (49) of the reduced-order model (55) is

$$J_{2m} = \text{Trace}\left[(I_n - T^+T)^T \mathcal{H}\mathcal{L}(I_n - T^+T)\right], \tag{65}$$

where $\mathcal{H} = C_n^T R_2 C_n$, $\mathcal{L} = E(x_n x_n^T)$ are both $n \times n$ symmetric matrix. It is seen that due to symmetric property, indifferent of the nature of the original system, $\mathcal{H}$ and $\mathcal{L}$ are nonnegative definite matrices. Then there exists an $n \times n$ nonsingularity matrix Φ such that $\mathcal{H}$, $\mathcal{L}$ and $T^+ T$ are diagonalised. That is

$$\mathcal{H} = \Phi^{-1}\left[\begin{array}{c|c} \Lambda & 0 \\ \hline 0 & 0 \end{array}\right]\Phi, \quad \mathcal{L} = \Phi^{-T}\left[\begin{array}{c|c} l_r & 0 \\ \hline 0 & 0 \end{array}\right]\Phi, \quad T^+T = \Phi^{-1}\left[\begin{array}{c|c} \Lambda & 0 \\ \hline 0 & 0 \end{array}\right]\Phi, \tag{66}$$

where the dimension of positive diagonal $\Lambda_\mathcal{H}$ and that of $\Lambda_\mathcal{L}$ depend on $p(C_n)$ and the degree of the controllable part of the system.

As $\Lambda_\mathcal{H}\Lambda_\mathcal{L} = \Lambda_{\mathcal{H}\mathcal{L}}$, one gets

$$J_{2m} = \text{Trace}\left[\Lambda_{\mathcal{H}\mathcal{L}}(n-r)\right], \tag{67}$$

where $\Lambda_{\mathcal{H}\mathcal{L}}(n-r)$ is a nonnegative diagonal of order $n-r$ having the $n-r$ least eigenvalues of $\mathcal{H}\mathcal{L}$. Corresponding to which T to be chosen as

$$T^+T = I_n - \Phi^{-1}\left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & I_{n-r} \end{array}\right]\Phi, \tag{68}$$

where Φ is known, hence T can be computed.

It is clearly seen from (64) that when the order of the controllable and observable part of the system is r, $J_{2m} = 0$. The reduced model is exactly the said part of the system.

c. One tries for minimum cost function

$$J_{2m} = \int_0^\infty (y_n - C_n T^+ T x_n)^T R_2 (y_n - C_n T^+ T x_n)\, dt \tag{69}$$

to be the absolute minimum. Derivating the above equation with respect to T^+T and equating the result to zero one gets

$$C_n T^+ T x_n = y_n \quad \text{or} \quad T^+T = (C_n^T C_n)^+ C_n^T y_n x_n^T (x_n x_n^T)^+ \tag{70}$$

from which T^+T can be computed as x, C and y are known. Further, the absolute minimum can be easily seen obtainable only with an orthogonal projector. That is,

$$\int_0^\infty (y_n C_n T^+ T x_n)^T C_n T^+ x_r\, dt = 0. \tag{71}$$

Different ways for choosing T may also be suggested. For example, as $x_r = Tx$ and $y = Cx$ are of the same form, the difference in both x and y of only in the dimension. Thus, one may choose $T = [C_n^T \ Z^T]$, where Z is $(r-q) \times n$ matrix. Then $x_r = [y_n^T \ (Zx_n)^T]$. Performing T^+ and a few mathematical manipulation in substituting y, C, T^+ and x_r into the cost function, one may obtain Z as a function of the available data of the system.

5. CONCLUDING REMARKS

A comparison of the various methods developed so far to 1995 for obtaining reduced-order model for high-order, complex system shows their relative advantages and disadvantages. Where as the projective nature of the aggregation method appears as an advantage for controller designed for state feed back, the other methods prove better for different design criteria. All methods, except for two, suffer from the disadvantage that the physical nature of the states is lost. The preservation of the physical nature of the states retain in the reduced model is of great importance when the model is obtained for a subsystem which is a part of larger system. It is a hope that out of the method proposed by the author, some other promising approaches, especially the internally balanced method, can be modified to preserve that physical nature of the modeled states. However, to retain the data at the output side for a projective (trajectory) control problem, is to be considered, Nath and San (1991), which is also belonging to the optimum methods group. So far, the state-optimization approach proposed by the author have been further adopted to the closed-loop thinking, San and Anh (1996). The same is for balanced method. It is from the author's opinion that although great deal of research work has been carried out the research for the most suitable method for model order reduction is far from over, especially for non-linear system.

Hereby, the authors would like also to mention that even the concept of reduction for nonlinear system is yet to be defined. Without a clear definition various different type of nonlinearity happened to be in the reality physics leading to different treatment can not be thought of reduction for the simpler treatment purpose.

REFERENCES

[1] Anderson J. H., Geometrical approach to the reduction of dynamically systems, *Proc. I.E.E.* **114** (1967) 1014-1018.

[2] Anh N. T. and San N. N., Controller reduction by state - optimization approach, *Optimization* **37** (1996) 1-20.

[3] Aoki M., Control of large scale dynamic system by aggregation, *IEEE Trans. Auto. Control* **AC-13** (1968) 246-235.

[4] Bandler J. W., Markettons N. D., and Sinha N. K., Optimum system modeling using recent gradient methods, *Int. J. Systems Science* **4** (1973) 257-262.

[5] Bistritz Y. and Lanholz G., Model reduction by Chebyshev polynomial techniques, *IEEE Trans. Auto. Control* **AC-24** (1979) 714-747.

[6] Chen C. F. and Shieh L. S., Anovel approach to linear model simplification, *Int. J. Control* (1968) 561-570.

[7] Chen C. F., Model reduction of multivariable control systems by means of matrix continued fraction, *Int. J. Control* **20** (2) (1974) 225-238. 0

[8] Commault C., Optimal choice of modes for aggregation, *Automatica* **17** (1981) 397-399.

[9] Davison E. J., A method for simplifying linear dynamic systems, *IEEE Trans. Auto. Control* **AC-11** (1966) 93-101.

[10] Elliott H. and Wolovich W. A., A frequency domain model reduction procedure, *Automatica* **16** (1980) 167-177.

[11] EL - Nahas I., Sinha N. K., and Alden R. T. H., Reduction of high-order system with application controller design, *Proc. 19th Annual Allerton Conf. on Communication, Control and Computing* (Monticello, IL), 1981, 163-173.

[12] Fernando K. V. and Nicholson H., Singular perturbational model reduction of balanced system, *IEEE Trans. Auto. Control* **AC-27** (1982) 466-468.

[13] Gibarillo G. and Lees F. P., The reduction of complex transfer function models to simple models using the method of moments, *Cher. Eng. Science* **24** (1966) 85-93.

[14] Glover K., All optimal Hankel norm approximation of linear multivariable system and their L error bounds, *IEEE Trans. Auto. Control* **AC-29** (1984) 1115-1193.

[15] Hickin J. D. and Sinha N. K., Model reduction for linear multivariable system, *IEEE Trans. Auto. Control* **AC-25** (1980) 1121-1127.

[16] Hutton M. F. and Friedland B., Routh approximation for reducing order of linear time invariant systems, *IEEE Trans. Auto. Control* **AC-20** (1975) 329-337.

[17] Hyland D. C. and Bernstein D. S., The optimal projection equations for model reduction and the relationships among the methods of Wilson, Skelton, and Moore, *IEEE Trans. Auto. Control* **AC-30** (12) (1985) 1201-1211.

[18] Kabamba P. T. Balanced galns and their significance for L^2 model reduction, *IEEE Trans. Auto. Control* **AC-30** (1985) 690-693.

[19] King A. M., Desai U. B., and R. E. Skelton, A generalized approach to p-Markov covariance equivalent realizations for discrete systems, *Automatica* **24** (1988) 507-515.

[20] Jonekheere E. A. and Silverman L. M., A new set of invariants for linear systems - Application to deducted order compensator design, *IEEE Trans. Auto. Control* **AC-28** (10) (1993) 953-964.

[21] Lanholz G. and Bishtritz Y., Model reduction of dynamic systems over a frequency interval, *Proc. 16th Annual Allerton Cont. on Communication, Control and Computing* (Monticello, IL), 1978, 903-912.

[22] Lastman G. J., Sinha N. K., A comparison of the balanced matrix and the aggregation method of model reduction, *IEEE Trans. Auto. Control* **AC-30** (1985) 301-304.

[23] Lastman G. J., Sinha N. K., and Rozsa P., On the selection of states to be retained in a reduced-order model, *Proc. IEEE, Part D* **131** (1984) 15-22.

[24] Lastman G. J. and Sinha N. K., An error analysis of the balanced matrix method, with application to the selection of reduced-order modes, *Large Scale Systems* **9** (1985) 63-71.

[25] Lastman G. J. and Sinha N. K., Worst-case error analysis of the balanced matrix method of model reduction, *Can. I. Elect. and Comp. Eng.* **14** (1989) 18-23.

[26] Lucas T. N., Linear system reduction by impulse energy approximation, *IEEE Trans. Auto. Control* **30** (8) (1985) 784-786.

[27] Marshall S. A., An aproximate method for fedacing the order of a large systems, *Control Engineering* **10** (1966) 642-648.

[28] Mitra D., On the reduction of the complexty of linear dynamic model, *Rep. AEEW-R520, U.K. Atomic Energy Authority*, 1967.

[29] Moore B. C., Principal component analysis in linear systems: Controllability, observability and model reduction, *IEEE Trans. Auto. Control* **AC-26** (1981) 17-32.

[30] Mustafa S. and Glover K., Controller reduction by H balanced traction, *IEEE Trans. Auto. Control* **36** (6) (1991) 668-682.

[31] Nath N. G. and San N. N., An approach to linear model reduction, *Control and Cybernetics* **20** (2) (1991) 69-89.

[32] Nolan P. J., Sinha N. K., and Alden R. T. H., Eigenvalue sensitivities of power systems including network and shaft dynamics, *IEEE Trans. Power App. and Syst.* **PAS-95** (1976) 1381-1324.

[33] Perenbo I. and Silverman L. M., Model reduction via balanced state space representations, *IEEE Trans. Auto. Control* **AC-27** (1982) 382-387.

[34] Prakash R. and Rao S. V., Model reduction by low-frequency approximation of interally balanced representation, to presented at *IEEE Conf. Decision and Control* (Tampa, Florida), 1989.

[35] Rao S. V. and Lamba S. S., Eigenvalue assignment in linear optimal control systems via reduced models, *Proc. IEEE* **122** (1975) 197-201.

[36] Rozsa P., Sinha N., and Lastman G., On estimating state variable partition for model reduction, *Proc. 13th Annual Conf. on Modeling and Simulation* (Pittsburgh, PA), 1982, 251-260.

[37] Sannuti P. and Kokotovic, Near optimum design of linear systems using singular perturbation method, *IEEE Trans. Auto. Control* **AC-14** (1969) 15-21.

[38] Sanash Y., Stable reduced-order models using Pade-type approximations, *IEEE Trans. Auto. Control* **AC-14** (1969) 15-21.

[39] San N. N., State-optimization method for order reduction of linear models and of state estimators, *Optimization* **34** (1995) 341-357.

[40] San N. N and Nath N. G., On optimal projection equation for model reduction; input error approach, *Optimization* **31** (1994) 263-281.

[41] Sinha N. K., EI-Nahas I., and Alden R. T. H., Routh approximation of multivariable system, *Problems of Control and Information Theory* **11** (3) (1982) 420-425.

[42] Skelton R. E., Cost decomposition of linear systems with application to model reduction, *Int. J. Control* **32** (1980) 1031-1055.

[43] Skelton R. E. and Yousuff, Component cost analysis of large scale systems, *Int. J. Control* (1983) 285-297.

[44] Wagie D. A. and Skelton R. E., Model reduction and controller synthesis in the presence of parameter uncertainty, *Automatica* **22** (1986) 295-308.

[45] Wilson D. A., Optimum solution of model-reduction problem, *Proc. IEE* **117** (6) (1970) 1161-1165.

[46] Wu S. L. and Lee E. B., Model reduction via a quasi-Kalman decomposition, *IEEE Trans. Auto. Control* **AC-30** (8) (1985) 789-790.

[47] Zhao G. and Sinha N. K., Model selection in aggregated models, *Large Scale Systems* **4** (1983) 209-216.

Nguyen Thuy Anh - *Dept. of Electronic and Telecommunication,*
Hanoi Institute of Technology.

Nguyen Ngoc San - *Post and Telecommunication Institute of Technology.*