A NEW ALGORITHM FOR MODAL APPROACH TO REDUCED-ORDER MODELING

by

Felix F. Wu and N. Narasimhamurthi

Memorandum No. ERL-M613

12 November 1976

ENGINEERING RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A NEW ALGORITHM FOR MODAL APPROACH TO REDUCED-ORDER MODELING

Felix F. Wu and N. Narasimhamurthi

Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory,
University of California, Berkeley, California 94720

ABSTRACT

The conventional approach to modal reduction is based on the diagonization of the coefficient matrix A. It requires the costly computations of the eigenvalues and eigenvectors of the large A matrix. We present a new approach to modal reduction, along with a computationally quite efficient and numerically rather stable algorithm. Our algorithm utilizes elementary transformations and avoids the direct computation of the eigenvalues and eigenvectors of A.

Research sponsored by the National Science Foundation Grant ENG 75-21747.

I. INTRODUCTION

In many practical applications, the number of state variables of the system may be very large, for example, in electric circuits [1] and power systems [2]. Simplified models of reduced order may have to be used. Lal and Van Valkenburg [3] have reviewed various existing reduced-order modeling techniques in a recent paper. Interested readers are referred to that paper for additional references. Davidson [4] suggested a reduced-order model which retains only the contributions to the response by the modes associated with small eigenvalues. This is known as the modal reduction approach. The modal reduction method has been applied to the construction of power system dynamic equivalents [5, 6, 7] for use in stability calculations and dynamic simulations. It is found, however, that the computational savings are not very satisfactory [8]. The reason is that the computational schemes for modal reduction based on the conventional approach require the very costly computations of the eigenvalues and eigenvectors of the original large matrix. Recently Kokotovic [9] proposed a method for obtaining reduced-order model from the solution of a Ricatti equation. His procedure does not guarantee that the reduced-order system has the desired modes. We present a new approach to modal reduction in this paper, along with a computationally quite efficient and numerically rather stable algorithm. Our algorithm is based on elementary matrix transformations and it avoids the costly computations of the eigenvalues and eigenvectors of the large matrix of the original system.

In Section II we use the conventional approach to describe the modal reduction technique. In Section III we present a theorem, which is the foundation of our new approach to modal reduction, and provide some

motivations which lead to our modal reduction algorithm. Our modal reduction algorithm is described in Section IV. Schemes to further reduce the total computations are discussed in Section V. In Section VI we consider the special case such that the original system matrix is symmetrical and modify the modal reduction algorithm to take advantage of the symmetry. In the last section we consider an alternative criterion for mode-retention and present a reduction algorithm.

II. MODAL REDUCTION

1. Conventional approach.

Consider a linear time-invariant system representation

$$\dot{x} = Ax+Bu \tag{1}$$

$$y = Cx \tag{2}$$

where $x\in\mathbb{R}^n$, $u\in\mathbb{R}^m$, $y\in\mathbb{R}^p$, and the dimensions of the matrices A, B, C are n×n, n×m, n×p, respectively.

Assume that the eigenvalues of the matrix A are distinct. Let λ_1, $\lambda_2,\ldots,\lambda_n$ be the eigenvalues of A and e_1, $e_2,\ldots,e_n$ be the corresponding eigenvectors. Let us partition the eigenvalues into two sets $\{\lambda_1,\lambda_2,\ldots\lambda_{n_1}\}$ and $\{\lambda_{n_1+1},\ldots,\lambda_{n_1+n_2}\}$ where $n_1+n_2=n$. The fundamental matrix Q, consisting of the eigenvectors $\{e_1,e_2,\ldots,e_n\}$, and its inverse P, can be partitioned accordingly.

$$Q \triangleq \left[e_1 \vdots e_2 \vdots \cdots e_n\right] = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \tag{3}$$

$$P \triangleq Q^{-1} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

Where Q_{11} and P_{11} are $n_1 \times n_1$, Q_{22} and P_{22} are $n_2 \times n_2$, Q_{12} and P_{12} are $n_1 \times n_2$, Q_{21} and P_{21} are $n_2 \times n_1$. Let us partition x into x_1 and x_2, where $x_1 \in \mathbb{R}^{n_1}$ and $x_2 \in \mathbb{R}^{n_2}$ and then partition A, B, and C accordingly.

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{5}$$

$$B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \tag{6}$$

$$C = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \tag{7}$$

Now if we make a coordinate transformation

$$\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{8}$$

Then (1) and (2) becomes

$$\begin{bmatrix} \dot{\xi}_1 \\ \dot{\xi}_2 \end{bmatrix} = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} P_{11}B_1 + P_{12}B_2 \\ P_{21}B_1 + P_{22}B_2 \end{bmatrix} U \tag{9}$$

$$y = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \tag{10}$$

where $\Lambda = \text{diag}\,(\lambda_1, \lambda_2, \ldots \lambda_{n_1})$ and $\Lambda = \text{diag}\,(\lambda_{n_1+1}, \ldots, \lambda_{n_1+n_2})$ are diagonal matrices.

The transformation (8) changes the basis into one formed by the eigenvectors $\{e_1, e_2, \ldots e_n\}$ of A. The components of ξ are the coordinates of the vector relative to the new basis $\{e_1, e_2, \ldots e_n\}$. Each coordinate of ξ is called a <u>mode</u> of the system (1). From eq. (9) it is seen that

the free system (zero input, u = o) in the new coordinate system, is decoupled i.e., each mode can be excited independent of the other modes.

Suppose that we neglect the modes corresponding to ξ_1, then we need only to solve the following reduced-order system (11) to obtain the mode-reduced output $\hat{y}$.

$$\dot{\xi}_2 = \Lambda_2 \xi_2 + (P_{21}B_1 + P_{22}B_2)U \tag{11}$$

$$\hat{y} = [C_1 \; C_2] \begin{bmatrix} Q_{12} \\ Q_{22} \end{bmatrix} \xi_2 \tag{12}$$

The foregoing procedure is known as the modal reduction.

2. Criterion for mode retention.

In practical applications [4 - 7], the modal reduction is applied to such cases that all the eigenvalues of A have nonpositive real parts. Consider the eigenvalues which are far from the origin, e.g., $\lambda = -\sigma + j\omega$. They either have large (in magnitude) real part σ or imaginary part ω or both. If σ is large, it can be seen from the solution of (9)-(10) that the contribution of the mode corresponding to such an eigenvalue dies out fast because of the $e^{-\sigma t}$ factor. If ω is large, since practical systems normally are low-pass filters[1], the contribution of such a mode in the steady state will likely to be very small. Indeed the modes corresponding to eigenvalues close to the origin roughly determines the type of the response which the system will have. Based on these considerations the decision in practice is to neglect the modes corresponding to eigenvalues with large magnitude and to retain the modes corresponding to eigenvalues of small magnitude.

[1] The elements of the transfer function matrix $G(s) = C(sI-A)^{-1}B$ normally are strictly proper rational functions, hence each element has the property of a low-pass filter, i.e., $G_{k\ell}(j\omega) \to 0$ as $\omega \to \infty$.

Comment: Conputational schemes for modal reduction based on the conventional approach require the calculations of the eigenvalues and eigenvectors of the large matrix A, which is a computationally costly task. Let us now contemplate on the possibility of computational improvements. It is clear that the essential part of the modal reduction is to decouple the modes to be retained from the rest of the modes. However, the conventional approach using the coordinate transformation (8) does far more than that. It actually decouples each and every mode, as evidenced by the diagonal matrix in eq. (9). Therefore it is reasonable to say that the computations involved in the conventional approach is more than necessary and improvements are possible. Indeed, we are going to present a new approach to the problem, along with an algorithm which avoids the costly computations of the eigenvalues and eigenvectors of A.

III. A NEW APPROACH

1. Foundation

A necessary consequence of decoupling certain modes from the rest of the modes through a coordinate transformation is that the coefficient matrix A will be transformed into a block triangular form. Therefore, let us consider a transformation $\eta = Tx$ such that $\tilde{A} = TAT^{-1}$ is block upper triangular, $\begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}$, i.e., η_2 is decoupled from η_1, where $\eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}$ (See Fig. 1). Theorem 1 relates the coordinates η and the coordinates of the individual modes ξ. The theorem provides the foundation for our new approach to modal reduction.

Theorem 1 Consider a linear time-invariant system representation.

$$\begin{cases} \dot{x} = Ax + Bu & (13) \\ y = Cx & (14) \end{cases}$$

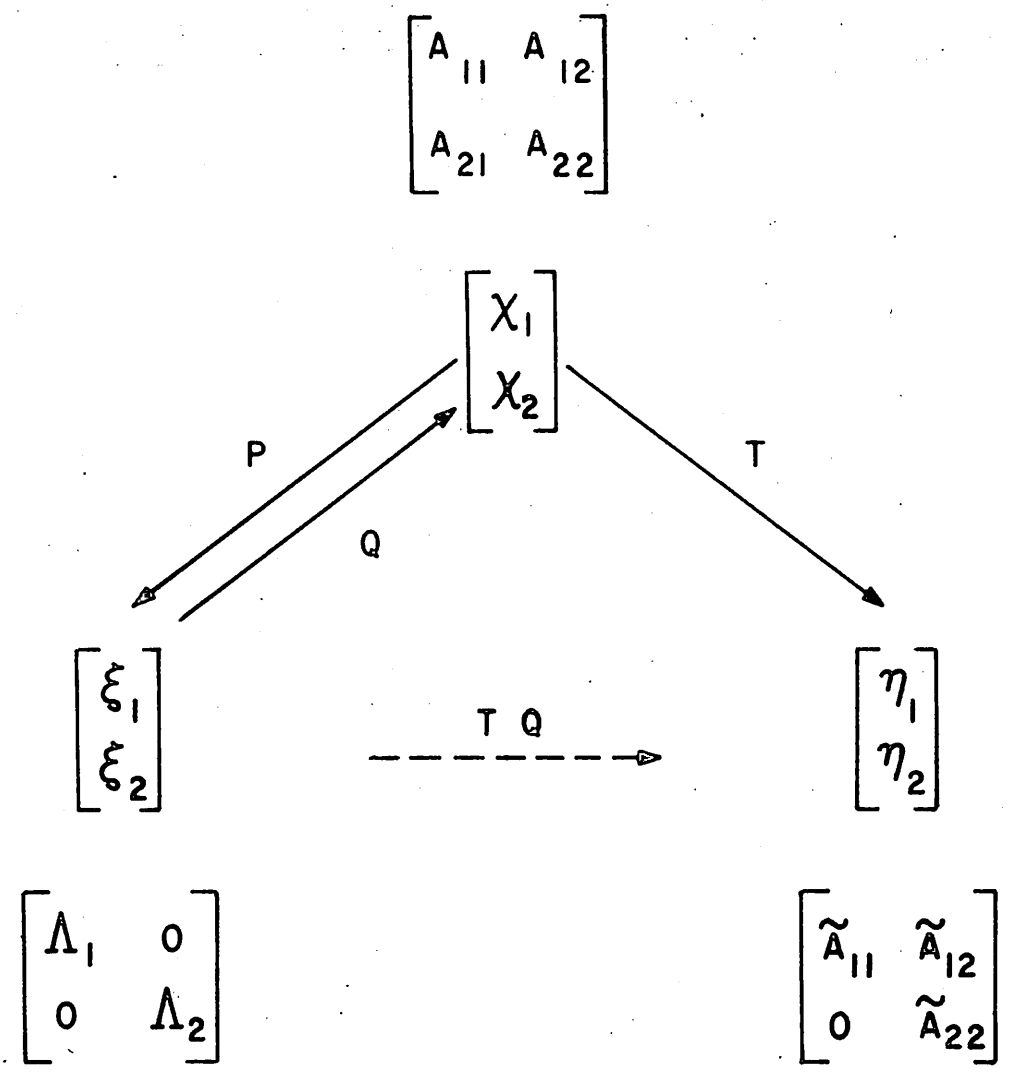

Fig. 1 Relationships among the coordinate systems x, ξ and η

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$. Assume that the eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ of A are distinct.

Suppose that a similarity transformation T, i.e.,

$$\eta = Tx \tag{15}$$

transforms (13) into the form

$$\begin{bmatrix} \dot{\eta}_1 \\ \dot{\eta}_2 \end{bmatrix} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix} u \tag{16}$$

where $\tilde{A}_{11}$ is $n_1 \times n_1$, $\tilde{A}_{22}$ is $n_2 \times n_2$, with $n_1 + n_2 = n$, and $\eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}$.

Let the eigenvalues of $\tilde{A}_{11}$ be $\{\lambda_1, \ldots, \lambda_{n_1}\}$, the eigenvalues of $\tilde{A}_{22}$ be $\{\lambda_{n_1+1}, \ldots, \lambda_{n_1+n_2}\}$[2], and the eigenvectors of A corresponding to the eigenvalues $\{\lambda_{n_1+1} \cdots, \lambda_{n_1+n_2}\}$ be $\{e_{n_1+1}, \ldots, e_{n_1+n_2}\}$.

Then η_2 is related to ξ_2, the coordinates of the vector x relative to the set of eigenvectors $\{e_{n_1+1}, \ldots e_{n_1+n_2}\}$ corresponding to the eigenvalues of $\tilde{A}_{22}$, by a nonsingular trasnformation $\tilde{Q}_{22}$, i.e.,

$$\eta_2 = \tilde{Q}_{22} \xi_2 \tag{17}$$

In fact, $\tilde{Q}_{22}$ is the fundamental matrix of $\tilde{A}_{22}$.

Furthermore, the mode-reduced output $\hat{y}$, i.e., the output with contributions only from the modes ξ_2, as defined in eq. (12), is given by:

$$\hat{y} = [\tilde{C}_1 \ \tilde{C}_2] \begin{bmatrix} \tilde{Q}_{12} \ \tilde{Q}_{22}^{-1} \\ I \end{bmatrix} \eta_2 \tag{18}$$

where $\tilde{C} = [\tilde{C}_1 \ \tilde{C}_2] = CT^{-1}$, $\tilde{Q}_{12}$ is the solution of the matrix equation

---

[2] The eigenvalues of A are identical to the eigenvalues of $\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}$. The eigenvalues of $\tilde{A}$ are the union of the eigenvalues of $\tilde{A}_{11}$ and the and the eigenvalues of $\tilde{A}_{22}$.

$$\tilde{A}_{11}\tilde{Q}_{12}-\tilde{Q}_{12}\Lambda = -\tilde{A}_{12}\tilde{Q}_{22} \tag{19}$$

where $\Lambda_2 \triangleq \text{diag}(\lambda_{n_1+1},\ldots,\lambda_{n_1+n_2})$.

Corollary. If a similarity transformation $\bar{T}$ transforms Eqs. (13), (14) into the form

$$\begin{bmatrix}\dot{\bar{\eta}}_1\\ \dot{\bar{\eta}}_2\end{bmatrix} = \begin{bmatrix}\bar{A}_{11} & 0\\ 0 & \bar{A}_{22}\end{bmatrix}\begin{bmatrix}\bar{\eta}_1\\ \bar{\eta}_2\end{bmatrix} + \begin{bmatrix}\bar{B}_1\\ \bar{B}_2\end{bmatrix} u \tag{20}$$

$$y = [\bar{C}_1 \ \bar{C}_2]\begin{bmatrix}\bar{\eta}_1\\ \bar{\eta}_2\end{bmatrix}$$

then the mode-reduced output $\hat{y}$ is given by

$$\hat{y} = \bar{C}_2\bar{\eta}_2 \tag{21}$$

Remarks

1) Note that in Eq. (16), η_2 is decoupled from η_1, i.e.,

$$\dot{\eta}_2 = \tilde{A}_{22}\eta_2+\tilde{B}_2 u \tag{22}$$

Theorem 1 asserts that η_2 and ξ_2 are related through a coordinate transformation $\tilde{Q}_{22}$, where ξ_2 corresponds to the modes of A associated with the eigenvalues of $\tilde{A}_{22}$. Representations (22) and (11) are thus algbraically equivalent [10, pp. 155-158]. In other words, any transformation T which transforms A into an upper triangular matrix $\tilde{A} = \begin{bmatrix}\tilde{A}_{11} & \tilde{A}_{12}\\ 0 & \tilde{A}_{22}\end{bmatrix}$ decouples the modes associated with the eigenvalues of $\tilde{A}_{22}$ from the remaining modes. This remarkable implication frees us from the dependence on the coordinate transformation (8) using the eigenvectors for model reduction. Theorem 1 indeed opens up a new direction.

There will be further discussions of the class of transformations T in secs. 2 and 3.

2) In Theorem 1, the mode-reduced output $\tilde{y}$ is obtained from the solution of (22), the eigenvectors $\tilde{Q}_{22}$ of $\tilde{A}_{22}$, and the solution of the n_2 linear equations in (19), using the expression (18). The eigenvalues and the eigenvectors of $\tilde{A}_{22}$ are needed for the computation, however it should be pointed out that the dimension of the reduced-order system n_2 is usually much smaller than n. Substantial savings in computation can be expected. The development of our algorithm is based on Theorem 1. However in section V we will discuss a scheme which further transforms A into a block diagonal form so that the Corollary can be applied. For symmetrical A matrix, discussed in Sec. VI, the Corollary can be directly applied.

2. The search for T

In view of Theorem 1, the problem of modal reduction now reduces to the search for a similarity transformation T such that:

(i) $\tilde{A} = TAT^{-1}$ is block upper triangular, viz.

$$\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}$$

(ii) The eigenvalues of $\tilde{A}_{22}$ correspond to those modes that we want to retain.

Clearly for different criteria for mode retention, different similarity transformations will be used. The criterion for mode retention used in this paper, except in Sec. VII, is to retain those modes associated with small eigenvalues.

We do not expect any process in finite steps to give us the desired

transformation T. The reason is simple: For if it were the case, it would imply that one could find eigenvalues in finite steps and it would also imply one could solve roots of a general polynomial in finite steps. This is known to be not possible. Therefore we believe that we should look for an iterative scheme such that at each iteration a similarity transformation transforms the matrix A into a form which will eventually become block upper triangular. One candidate for such a form is the Hessenberg form[3] (Fig. 2). Note that when a subdiagonal element of a Hessenberg matrix becomes zero, we have a block upper triangular matrix.

Next we have to determine the exact scheme to be used at each iteration so that the iterative process will converge and when it converges, the matrix in the lower right block will have the eigenvalues that we want to retain.

3. Geometric motivations

In this section we explain the geometric motivations which lead to the development of our algorithm. We shall use a three-dimensional case to illustrate the basic ideas involved.

Let $\{\lambda_1,\lambda_2,\lambda_3\}$ be the (distinct) eigenvalues of the 3×3 matrix A and $\{e_1,e_2,e_3\}$ be the corresponding eigenvectors. Suppose the magnitudes of λ_1 and λ_2 are much larger than that of λ_3. Suppose furthermore that b is a vector such that

---

[3] A matrix $A \triangleq (a_{ij})$ is said to be in Hessenberg form, or a Hessenberg matrix, if $a_{ij}=0$ for $i \geq j+2$. The elements $a_{i+1,i}$ are called subdiagonal elements.

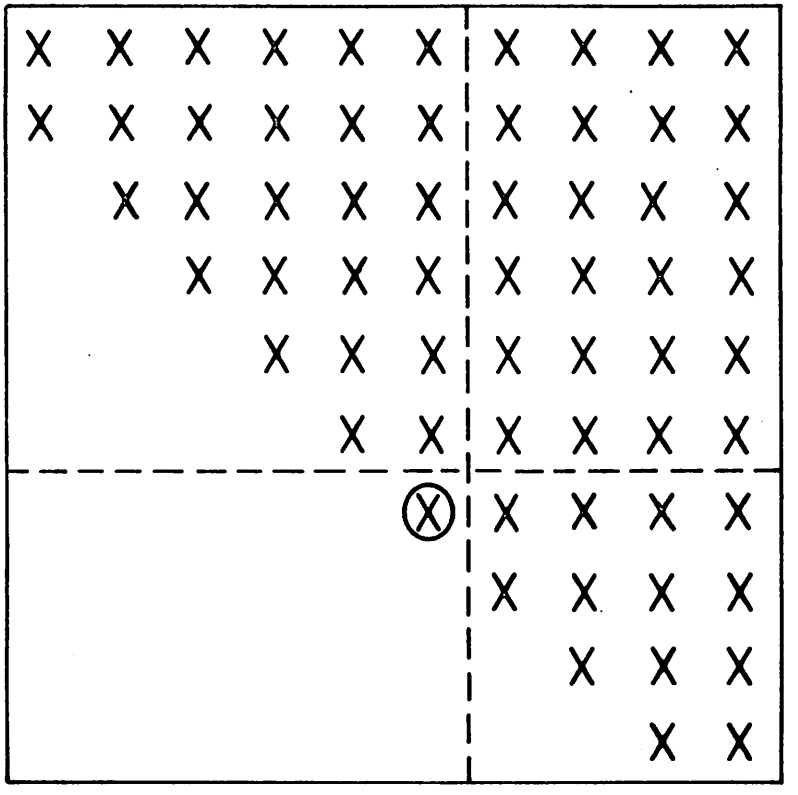

Fig. 2 An upper Hessenberg matrix becomes block upper triangular when a subdiagonal element becomes zero.

$$b = \sum_{i=1}^{3} \phi_i \, i \, , \quad \phi_i \neq 0 \text{ for } i=1,2,3 \tag{23}$$

Let us consider the following sequence of vectors:

$$\{ \quad b, \qquad Ab, \qquad A^2b, \qquad A^3b, \ldots \} \tag{24}$$

Since $A^k b = \sum_{i=1}^{3} \phi_i \lambda_i^k e_i$, i.e., the components along the eigenvectors associated with large eigenvalues get enhanced everytime on premultiplying by A, the normalized vector $\frac{A^k b}{\| A^k b \|}$ will converge to the subspace spanned by the vectors e_1 and e_2, i.e.,

$$\frac{A^k b}{\| A^k b \|} \to sp\{e_1, e_2\} \tag{25}$$

On the other hand, let $\mathcal{T}$ be the transformation which transforms the pair (A,b) into a Hessenberg matrix H and $t=(1,0,0)^T$, i.e.,

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ 0 & h_{32} & h_{33} \end{bmatrix} \qquad t = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \tag{26}$$

and

$$H = \mathcal{T} A \mathcal{T}^{-1} \tag{27}$$

$$t = \mathcal{T} b \tag{28}$$

Geometrically this is simply a change of coordinate system. The sequence (24) in the new coordinate system is

$$\{ \quad \mathcal{T}b, \qquad \mathcal{T}Ab, \qquad \mathcal{T}A^2b, \qquad \mathcal{T}A^3b, \ldots \} \tag{29}$$

or

$$\{ \quad t, \qquad Ht, \qquad H^2t, \qquad H^3t, \ldots \} \tag{30}$$

Note that

$$t = \begin{bmatrix}1\\0\\0\end{bmatrix}, \quad Ht = \begin{bmatrix}x\\h_{21}\\0\end{bmatrix}, \quad H^2t = \begin{bmatrix}x\\x\\h_{32}h_{21}\end{bmatrix} \tag{31}$$

Therefore the subdiagonal elements of H have the following geometric interpretation: for example, h_{32} is proportional to the component of A^2b along a base vector in the new coordinate system which does not lie in the span of b and Ab.

New let us shift our attention to the next three vectors in the sequence (24). More specifically, let us change the coordinate system so that

$$Ab = \begin{bmatrix}1\\0\\0\end{bmatrix} \quad A^2b = \begin{bmatrix}x\\x\\0\end{bmatrix} \qquad A^3b = \begin{bmatrix}x\\x\\x\end{bmatrix} \tag{32}$$

This can be achieved if we transform the pair (H,Ht) into a Hessenberg matrix $H_1 = \begin{bmatrix}h'_{11} & h'_{12} & h'_{13}\\ h'_{21} & h'_{22} & h'_{23}\\ 0 & h'_{32} & h'_{33}\end{bmatrix}$ and $t_1=(1,0,0)^T$. Let $\mathcal{T}_1$ denote such a transformation. Then the sequence (24) can be written as

$$\{\mathcal{T}_1\mathcal{T}b,\ \mathcal{T}_1\mathcal{T}Ab,\quad \mathcal{T}_1\mathcal{T}A^2b,\ \mathcal{T}_1\mathcal{T}A^3b,\ldots\} \tag{33}$$

or

$$\{\mathcal{T}_1t\ ,\quad t_1\ ,\quad H_1t_1\ ,\quad H_1^2t_1\ ,\ldots\} \tag{34}$$

where

$$t_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \quad H_1t_1 = \begin{bmatrix}x\\h'_{21}\\0\end{bmatrix}, \quad H_1^2t_1 = \begin{bmatrix}x\\x\\h'_{32}h'_{21}\end{bmatrix} \tag{35}$$

Let us move to the next three vectors in the next round and the process is repeated. At the k-th iteration as k becomes large we know

$$\frac{A^k b}{\| A^k b \|}, \quad \frac{A^{k+1} b}{\| A^k b \|} \longrightarrow sp\{e_1, e_2\} \tag{36}$$

and

$$\frac{A^{k+2} b}{\| A^k b \|} \longrightarrow sp\{e_1, e_2\} \tag{37}$$

Therefore the component of $\frac{A^{k+2} b}{\| A^k b \|}$ along a base vector which does not lie in the span of $\frac{A^k b}{\| A^k b \|}$ and $\frac{A^{k+1} b}{\| A^k b \|}$ will approach zero. In other words, the (3,2)-th element of the Hessenberg matrix will converge to zero.

From the above intuitive geometric considerations it is seen that if we transform the pair $(H_k, H_k t_k)$ at each iteration to (H_{k+1}, t_{k+1}) where H_{k+1} is Hessenberg and $t_{k+1} = (1,0,0)^T$, the process will converge to an upper triangular matrix. Moreover the first two coordinates will correspond to the eigenvectors associated with large eigenvalues λ_1 and λ_2 and the last coordinate will correspond to the eigenvector associated with small eigenvalue λ_3, which is consistent with our mode-retention criterion.

The foregoing geometric motivations lead to the development of our modal reduction algorithm to be presented. The initialization and the convergence of the algorithm will be proved rigorously as theorems.

VI THE REDUCTION ALGORITHM

We will present an algorithm in section 1 which simultaneously transforms a matrix into a Hessenberg form and a vector to the form $(\tau,0,\ldots,0)^T$. The condition under which the algorithm works is stated in Theorem 2. The algorithm for obtaining a reduced-order system is outlined in section 2, with detailed discussions on various procedures in sections 3, 4, and 5. Theorem 3 guarantees that the algorithm can always be initialized. Theorem 4 guarantees the convergence of the algorithm. An acceleration scheme for the algorithm is discussed in section 6. The complete version of our modal reduction algorithm is summarized in sec. 7.

1. Transformation of (A,b) to (H,t).

Theorem 2. Given an $n\times n$ matrix $A = (a_{ij})$ and an n-vector $b = (\beta_1,\ldots,\beta_n)^T$. Suppose[4] that $\{b,Ab,\ldots,A^{n-1}b\}$ spans $\mathbb{R}^n$. Then there exists a non-singular matrix $\mathcal{T}$ such that

(i) $H \triangleq \mathcal{T}A\mathcal{T}^{-1}$ is a Hessenberg matrix

(ii) $t \triangleq \mathcal{T}b$ is of the form $(\tau,0,\ldots,0)^T$.

We give a constructive proof of Theorem 2 in the Appendix, which is based on the algorithm Ht(A,b) to be presented shortly.

The transformations and their inverses, which transform the pair (A,b) into (H,t), are performed on the matrices B and C, respectively, in the later application of the algorithm. For ease of later reference we include them in the following description of the algorithm.

[4] It is necessary that A has distinct eigenvalues [10, pp. 169-170].

Algorithm Ht(A,b)

Step 1 Find $\beta_{\underset{\sim}{i}}$, $\underset{\sim}{i}\in\{1,2,\ldots,n\}$, such that $|\beta_{\underset{\sim}{i}}|\geq|\beta_i|$ for all $i\in\{1,2,\ldots,n\}$. If $\beta_{\underset{\sim}{i}}=0$ go to step 9, else continue.

Step 2 Interchange rows 1 and $\underset{\sim}{i}$ of b, A and B and interchange columns 1 and $\underset{\sim}{i}$ of A and C.

Step 3 For i=2,...,n, do the following:

Subtract $(\frac{\beta_i}{\beta_1})\times$(row 1) from row i of A and B

Add $(\frac{\beta_i}{\beta_1})\times$(column i) to column 1 of A and C.

Step 4 For j=1,2,...,n-2, do steps 5-7.

Step 5 Find $a_{i'j}$, $i'\in\{j+1,\ldots,n\}$, such that $|a_{i'j}|>|a_{ij}|$ for all $i\in\{j+1,\ldots,n\}$. If $a_{i'j}=0$, go to step 9, else continue.

Step 6 Interchange rows (j+1) and i' of A and B, and interchange columns (j+1) and i' of A and C.

Step 7 For i=j+2,...,n, do the following:

Subtract $(\frac{a_{ij}}{a_{j+1,j}})\times$(row j+1) from row i of A and B

Add $(\frac{a_{ij}}{a_{j+1,j}})\times$(column i) to column (j+1) of A and C.

Step 8 Return H=A and $t = (\beta_1,0,\ldots,0)^T$

Step 9 Stop.

Remarks 1) Steps 2 and 6 are performed in order to have the largest element as the pivot at steps 3 and 7, respectively, for numerical stability. All the multipliers, i.e., $(\frac{\beta_i}{\beta_1})$, $(\frac{a_{ij}}{a_{j+1,j}})$, in the algorithm have magnitude less than or equal to 1.

2) The transformation in step 3 transforms the vector $b = (\beta_1,\beta_2,\ldots,\beta_n)^T$ into the form, $(\tau,0,\ldots,0)^T$. The transformation in step 7 for each j of step 4 eliminates the elements in the j-th column below the subdiagonal.

3) Steps 5-7 constitute the method of transforming a matrix to Hessenberg form by the elementary stabilized transformations [11, pp. 353-369]. There are other methods to perform the same task, e.g., Givens' method and Householder's method [11, pp. 345-353], which use orthogonal transformations. For assymmetric matrices, the total number of multiplications to reduce an n×n matrix to Hessenberg form is essentially $\frac{5}{6}n^3$, $\frac{10}{6}n^3$, and $\frac{10}{3}n^3$, respectively, by the elementary stabilized transformations, Givens' method, and Householder's method. Therefore we incorporate the method which has the least number of computations, i.e., by the elementary stabilized transformations. For symmetric A matrices, on the other hand, we will use Householder's method (Sec. VI).

2. Outline of the Reduction Algorithm

(1) Initialization

Select a vector b and apply the algorithm Ht(A,b) to produce a Hessenberg matrix H_0 and a vector of the form $t_0 = (\tau,0,\ldots,0)^T$, i.e.,

$$H_0 = \mathcal{J}_0 A \mathcal{J}_0^{-1}$$

$$t_0 = \mathcal{J}_0 b$$

At the same time obtain

$$B_0 = \mathcal{J}_0 B$$

$$C_0 = C\mathcal{J}_0^{-1}$$

Let $b_0 = \frac{1}{\alpha_0} t_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ where $\alpha_0 = \| t_0 \|$

(2) Iteration

Apply the algorithm $\underline{\text{Ht}}(H_k, H_k b_k)$ and denote the output as

$$H_{k+1} = \mathcal{J}_{k+1} H_k \mathcal{J}_{k+1}^{-1}$$

$$t_{k+1} = \mathcal{J}_{k+1} H_k b_k$$

At the same time obtain

$$B_{k+1} = \mathcal{J}_{k+1} B_k$$

$$C_{k+1} = C_{k+1} \mathcal{J}_{k+1}^{-1}$$

Let $b_{k+1} = \frac{1}{\alpha_{k+1}} t_{k+1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ where $\alpha_{k+1} = \| t_{k+1} \|$

(3) Termination

Terminate the iterations at k if $h_{n_1+1}^{(k)} \simeq 0$ for some n_1, where $h_i^{(k)}$, i=2,...,n, are the subdiagonal elements of H_k.

3. Initialization

The initialization procedure is actually carried out by the following algorithm $\underline{\text{IN1}}(A,B,C)$. Theorem 3 guarantees that the initialization procedure can indeed by carried out.

Algorithm IN1(A,B,C)

Step 1 Pick an arbitrary $b \neq (0,\ldots,0)^T$

Step 2 Call Ht(A,b)

If it exits from step 9 of Ht, go to step 3, else go to step 4.

Step 3 Let the current A,B,C be denoted by

$$\underset{\sim}{A} = \begin{bmatrix} \underset{\sim}{A}_{11} & \underset{\sim}{A}_{12} \\ 0 & \underset{\sim}{A}_{22} \end{bmatrix}, \quad B = \begin{bmatrix} \underset{\sim}{B}_1 \\ \underset{\sim}{B}_2 \end{bmatrix} \quad \underset{\sim}{C} = [\underset{\sim}{C}_1 \ \underset{\sim}{C}_2]$$

where $\underset{\sim}{A}_{11}$ is $j' \times j'$ and Hessenberg.

Call SIG($\underset{\sim}{A},\underset{\sim}{B},\underset{\sim}{C}$)

Return (A,B,C)

Step 4 Return $H_0=A$, $B_0=B$, $C_0=C$

Algorithm SIG($\underset{\sim}{A},\underset{\sim}{B},\underset{\sim}{C}$)

Step 1 Duplicate $A=\underset{\sim}{A}$, $B=\underset{\sim}{B}$, $C=\underset{\sim}{C}$

Step 2 Set $\sigma=1$

Step 3 Set $b = (1,0,\ldots,\sigma,\ldots,0)$, where σ is in the $(j'+1)$-th position

Step 4 Call Ht(A,b)

If it exits from step 9 of Ht go to step 5; else go to step 8

Step 5 Let the current A be denoted by

$$A' = \begin{bmatrix} A'_{11} & A'_{12} \\ 0 & A'_{22} \end{bmatrix}$$

where A'_{11} is $k \times k$ and Hessenberg.

If $k>j'$, go to step 7; else go to step 6.

Step 6 Set $\sigma=\sigma+1$, $A=\underset{\sim}{A}$, $B=\underset{\sim}{B}$, $C=\underset{\sim}{C}$, and go to step 3.

Step 7 If $k=n$, go to step 8; else set $\underset{\sim}{A}=A'$, $\underset{\sim}{B}=B'$, $\underset{\sim}{C}=C'$, and go to step 1.

Step 8 Return A,B,C.

Theorem 3. If the eigenvalues of A are distinct, then the algorithm IN1(A,B,C) will result in a pair (H_0,t_0) such that H_0 is Hessenberg and $t_0 = (\tau,0,\ldots,0)^T$.

Remark For a single input system we may pick the initial b to be the coefficient vector of the input. In such a case, when we come to step 3 of IN1 it means that in terms of the current coordinates all the components in $i>j'$ are in the uncontrollable subspace. They may be discarded if only the controllable and observable subspace is of interest. In general we may just as well pick the initial b to be $(1,0,\ldots,0)^T$.

4. Iteration

From the outline of the reduction algorithm, we see that the algorithm Ht is applied to the pair (H_k,H_kb_k) at each iteration. Note that H_k is a Hessenberg matrix and $H_kb_k = (\beta_1,\beta_2,0,\ldots,0)^T$ has only two nonzero components. We immediately notice that only one comparison is needed in Step 1 of the algorithm Ht and only one operation is needed to transform H_kb_k into the form $(\tau,0,\ldots,0)^T$. The corresponding operations on H_k, namely, subtracting $(\frac{\beta_1}{\beta_2})\times$row 1 from row 2 of H_k and adding $(\frac{\beta_i}{\beta_1})\times$(column 2) to column 1 of H_k, a matrix which differs from a Hessenberg matrix only in that a nonzero element in (3,1) The elimination of the (3,1)

using step 7 will introduce a nonzero element in (4,2). Hence, the operations required to transform this matrix to a Hessenberg form is an elimination process which "chases" down the sub-subdiagonal elements, $h_{i+2,i}$.

Taking the advantage of the special structures of H_k and $H_k b_k$, the algorithm $Ht(h_k, H_k b_k)$ to be performed at each iteration can be simplified. We denote the simplified algorithm by ITE.

Algorithm ITE(H_k, B_k, C_k)

Let the ij-th element of H_k be denoted by h_{ij}

Step 1 Let $\beta_1 = h_{11}$ and $\beta_2 = h_{21}$

If $|\beta_2| > |\beta_1|$, go to step 2, else go to step 3.

Step 2 Interchange β_1 and β_2

Interchange rows 1 and 2 of H_k and B_k

Interchange columns 1 and 2 of H_k and C_k.

Step 3 Subtract $(\frac{\beta_2}{\beta_1})\times$row 1 from row 2 from H_k and B_k

Add $(\frac{\beta_2}{\beta_1})\times$column 2 to column 1 of H_k and C_k.

Step 4 For i=1,2,...,n-2, do steps 5-7.

Step 5 If $|h_{i+2,i}| > |h_{i+1,i}|$ go to step 6; else to step 7.

Step 6 Interchange rows (i+1) and (i+2) of H_k and B_k.

Interchange columns (i+1) and (i+2) of H_k and C_k.

Step 7 Subtract $(\frac{h_{i+2,i}}{h_{i+1,i}})\times$row(i+1) from row (i+2) of H_k and B_k.

Add $(\frac{h_{i+2,i}}{h_{i+1,i}})\times$column(i+2) to column (i+1) of H_k and C_k.

Step 8 Return $H_{k+1}=H_k$, $B_{k+1}=B_k$, $C_{k+1}=C_k$.

Remarks 1) The total number of multiplications performed on the elements of A in the algorithm ITE is essentially n^2. Compared with Ht, this is an order of magnitude less.

2) It is shown in the proof of Theorem 3 that the algorithm IN1 produces a vector t_0 which has nonzero component along every eigenvector of A. Consequently at least one of β_2 and β_1, and one of $h_{i+2,i}$ and $h_{i+1,i}$, for i=1,2,...,n-2, will be nonzero. (This can easily be proved using arguments similar to that in the proof of Theorem 2).

5. Termination

Theorem 4 guarantees the convergence of the reduction algorithm. Furthermore, it assures that when the algorithm converges the eigenvalues of the matrix on the lower right block correspond to the modes that we want to retain, namely, small eigenvalues.

Theorem 4 Given an n×n matrix A whose eigenvalues $\{\lambda_1,\lambda_2,\ldots,\lambda_n\}$ are distinct. Suppose that

$$|\lambda_i| >> |\lambda_{n_1+j}| \text{ for all } i\in\{1,2,\ldots,n_1\},\ j\in\{1,2,\ldots,n_2\} \tag{38}$$

where $n_1+n_2=n$. Then the application of the reduction algorithm will result in

$$\lim_{k\to\infty} h_{n_1+1}^{(k)} = 0 \tag{39}$$

where $\{h_2^{(k)},\ldots,h_n^{(k)}\}$ denote the subdiagonal elements of H_k. Furthermore, let $\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}$, where $\tilde{A}_{11}$ is $n_1\times n_1$ denote the matrix to which H_k converges, then the eigenvalues of $\tilde{A}_{22}$ are $\{\lambda_{n_1+1},\ldots,\lambda_{n_1+n_2}\}$.

Remark From the proof of Theorem 4 in the Appendix we see that the convergence of our modal reduction algorithm is of the order of $\left|\frac{\lambda_{n_1+j}}{\lambda_i}\right|$. Hence the algorithm will converge fast if A has two clusters of eigenvalues which are far apart.

6. Acceleration

Suppose that the eigenvalues of A are clustered into two groups, one centered around δ and the other σ, with $\delta<<\sigma$, the convergence of our modal reduction algorithm will be of the order of $\left|\frac{\delta}{\sigma}\right|$. If we can decrease this ratio faster convergence will result. Now note that

(i) A similarity transformation T which transforms $(A-\alpha I)$ into an upper block triangular form also transforms A into an upper block triangular form.

(ii) The eigenvalues of $(A-\alpha I)$ are $(\lambda_i-\alpha)$, $i=1,2,\ldots,n$, where λ_i, $i=1,2,\ldots,n$ are the eigenvalues of A.

In view of these two facts we know that we should in principle work on $(A-\alpha I)$ with α close to δ, in order to accerlerate convergence. In practice we do not know δ a priori, we have to estimate its value. We suggest the following practical way of estimating δ. If at the k-th iteration, an element in the subdiagonal, say $h_{r+1,r}$, becomes small enough, then we expect that clusters of eigenvalues are emerging, and the last $(n-r)\times(n-r)$ block are associated with the small eigenvalues. The trace of the matrix gives the sum of the eigenvalues. Based on these considerations we choose our estimate $\hat{\delta}$ of δ as

$$\hat{\delta} = \frac{1}{(n-r)} \sum_{i=r+1}^{n} h_{ii} \qquad (40)$$

Now let us consider the implementation of this acceleration scheme. Suppose at the k-th iteration we have a pair (H_k, b_k) and an estimate $\hat{\delta}$ of δ. In principle we would consider the pair $(H_k - \hat{\delta}I, b_k)$ instead, in order to accelerate convergence. We would transform the pair $(H_k - \hat{\delta}I, (H_k - \hat{\delta}I)b_k)$ to (H'_{k+1}, b'_{k+1}) by $\mathcal{J}_{k+1}$, i.e., $H'_{k+1} = \mathcal{J}_{k+1}(H_k - \hat{\delta}I)\mathcal{J}_{k+1}^{-1}$, and $b'_{k+1} = \mathcal{J}_{k+1}(H_k - \hat{\delta}I)b_k$. But $H'_{k+1} = \mathcal{J}_{k+1}H_k\mathcal{J}^{-1}_{k+1} - \hat{\delta}I$, and we have to add $\hat{\delta}I$ to get back our original matrix, i.e., $H_{k+1} = H'_{k+1} + \hat{\delta}I = \mathcal{J}_{k+1}H_k\mathcal{J}^{-1}_{k+1}$. Therefore for actual implementation of the acceleration scheme at the k-th step it suffices to simply transform the pair $(H_k, (H_k - \hat{\delta}I)b_k)$ to (H_{k+1}, t_{k+1}).

7. Summary

The complete version of our modal reduction algorithm MR(A,B,C) is summarized below.

Algorithm MR(A,B,C)

Step 1 Call IN1 (A,B,C)

Return H_0, B_0, C_0

Set k=0

Step 2 If $h_{n_1+1,n_1} \simeq 0$ for some n_1, go to Sept 4; else go to Step 3

Step 3 Call ITE (H_k,B_k,C_k)

Return H_{k+1}, B_{k+1}, C_{k+1}

Set k:=k+1, go to Step 2.

Step 4 Set $\tilde{A} = H_k = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}$, $\tilde{B} = B_k = \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix}$, $\tilde{C} = C_k = [\tilde{C}_1 \ \tilde{C}_2]$

where $\tilde{A}_{11}$ is $n_1 x n_1$.

Step 5 Find the eignevalues $\{\lambda_{n_1+1},\ldots,\lambda_{n_1+n_2}\}$ and eigenvectors q^j, $j = 1,\ldots,n_2$, of $\tilde{A}_{22}$.

Step 6 For $j = 1,\ldots,n_2$, solve

$(\tilde{A}_{11}-\lambda_{n_1+j}I)p^j = -\tilde{A}_{12}q^j$

for p^j.

Let $\tilde{Q}_{12} = [p^1 \ p^2 \ \ldots \ p^{n_2}]$ and $\tilde{Q}_{22} = [q^1 \ q^2 \ \ldots \ q^{n_2}]$

Step 7 Solve the reduced-order system

$\eta_2 = \tilde{A}_{22}\eta_2 + \tilde{B}_2 u$

and obtain the mode-reduced output

$\hat{y} = [\tilde{C}_1\tilde{Q}_{12}\tilde{Q}_{22}^{-1} + \tilde{C}_2]\eta_2$

Remarks 1) Note that after the termination of the iterations, the resultant $\tilde{A}_{22}$ in Step 4 is Hessenberg, which can be proceeded immediately, for example, by the QR-method [11,Chapter8], for the eigenvalue and eigenvector calculations in Step 5.

2) The resultant $\tilde{A}_{11}$ in Step 4 is Hessenberg, thus the coefficient matrices $(\tilde{A}_{11}-\lambda_{n_1+j}I)$, $j = 1,2,\ldots,n_2$, are all Hessenberg. This is one of the desirable forms for linear equations. Indeed the solution of the linear equations in Step 6 using Gaussian elimination or LU-decomposition method [11,Chapter4] will not introduce any fill-in to the elements below the subdiagonal, even if partial pivoting[5] [11,p.205] is employed. For Hessenberg matrices it has been shown that complete pivoting has no special merit [11,p.219]. It is possible to further reduce the total computations by transforming $\tilde{A}_{11}$ to a even more sparse form prior to Step 5, we will discuss these procedures in Sec. V.

3) If there are two clusters of eigenvalues having n_1 large ones and n_2 small ones (see Eq. (38)), the iterations in Step 2 will terminate as a result of the convergence to zero of the (n_1+1)-th subdiagonal element and we may thus discard the modes corresponding to the n_1 large eigenvalues. If, however, there are more than two clusters, say three clusters of eignevalues, with n_1 very large ones, n_2 large ones and n_3 small ones. Depending on the distances between these clusters, the iterations in Step 2 may first exhibit convergence at the (n_1+1)-th subdiagonal element.

[5] If at the r-th stage of elimination or decomposition, the pivot is selected from the elements in the first column of the square matrix of order (n-r+1) in the bottom right-hand corner, then it is called partial pivoting. On the other hand if pivot is chosen to be the element of maximum magnitude in the whole of this square matrix of order (n-r+1), then it is called complete pivoting.

Suppose we want to discard modes corresponding to the two clusters of n_1 very large and n_2 large eigenvalues, then we may proceed as follows. Let the return from the algorithm ITE(H_k, B_k, C_k) at this point be denoted by

$$H_k = \begin{bmatrix} H_{11} & H_{12} \\ 0 & H_{22} \end{bmatrix} \quad B_k = \begin{bmatrix} B_{k1} \\ B_{k2} \end{bmatrix} \quad C_k = [C_{k1} \quad C_{k2}]$$

where H_{11} is $n_1 x n_1$. We now take the triple (H_{22}, B_{k2}, C_{k2}) and apply Steps 1 and 2 of the algorithm MR. This time the iterations in Step 2 will terminate due to the convergence at the (n_2+1)-th subdiagonal element of H_{22}. Let the current return be denoted by

$$\begin{bmatrix} H_{\alpha\alpha} & H_{\alpha\beta} \\ 0 & H_{\beta\beta} \end{bmatrix} \begin{bmatrix} B_\alpha \\ B_\beta \end{bmatrix} \quad [C_\alpha \quad C_\beta]$$

where $H_{\alpha\alpha}$ is $n_2 x n_2$. The combining effect is that we have a similarity transformation which transforms the original (A,B,C) into

$$\left[\begin{array}{cc:c} H_{11} & H_{12} & H_{1\beta} \\ 0 & H_{\alpha\alpha} & H_{\alpha\beta} \\ \hdashline 0 & 0 & H_{\beta\beta} \end{array}\right] \begin{bmatrix} B_{k1} \\ B_\alpha \\ B_\beta \end{bmatrix} \quad [C_{k1} \quad C_\alpha \quad C_\beta]$$

where $[H_{1\alpha} \quad H_{1\beta}] = H_{12}$. Now we have an upper block triangular matrix with the lower right block corresponding to the n_3 modes to be retained.

4) Suppose we are interested in responses other than just zero-state response, the initial condition $\eta_2(0)$ is then needed in Step 7. This can be easily obtained by augmenting B with a column x(0). The return from Step 3 of the augmenting column in B will be $\begin{bmatrix} \eta_1(0) \\ \eta_2(0) \end{bmatrix}$.

V FURTHER COMPUTATIONAL REDUCTIONS

The linear equations in Step 6 of the modal reduction algorithm MR are solved n_2 times for different coefficient matrices $(\tilde{A}_{11} - \lambda_{n_1+j} I)$ having the same sparsity structure. The number of operations involved in solving each of the linear equations is related to the number of nonzero elements in $\tilde{A}_{11}$. Therefore, if prior to the entrance to Step 5 in MR we perform additional similarity transformations to bring $\tilde{A}_{11}$ to a more sparse and computationally more desirable form while maintaining $\tilde{A}$ as block upper triangular, then further reduction in computations will be possible. We will discuss transformations which bring $\tilde{A}_{11}$ to a tridiagonal matrix and a Frobenius matrix in Sections 1 and 2, respectively[6] [11,pp.395-409].

In view of the Corollary of Theorem 1, if a similarity transformation can be found which further transforms $\tilde{A}$ to a block diagonal form then the computations in Steps 5 and 6 of the algorithm MR can be avoided altogether We will discuss this approach in Sec. 3.

It should be pointed out that for the methods discussed in this section, large multipliers may have to be used. Therefore the reduction in total computation by the application of these methods will be attained at the risk of numerical instability. Among the three methods, from numerical point of view we recommend the third method (Sec. 3).

The methods discussed below start with a matrix $\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}$ where $\tilde{A}_{11}$ is Hessenberg. The algorithm TRID transforms $\tilde{A}_{11}$ into a tridiagonal form and the algorithm FROB transforms $\tilde{A}_{11}$ into a Frobenius form. The

[6] An nxn matrix $A = (a_{ij})$ is said to be a tridiagonal matrix if all the elements, except $a_{i+1,i}$, a_{ii}, and $a_{i,i+1}$, are zero. A is said to be a Frobenius matrix if all the elements, excep $a_{i+1,i}$ and a_{in}, are zero.

ij-th element of $\tilde{A}$ is denoted by a_{ij}.

1. Transform $\tilde{A}_{11}$ to tridiagonal form

Algorithm TRID(A,B,C)

Step 1 For $i = 1,2,\ldots,n_1-1$, do steps 2-4.

Step 2 If $a_{i,i+1} \neq 0$, go directly to step 4;
else go to step 3

Step 3 If $i = n_1-1$, go to step 5; else do the following:
Add row (i+2) to row i of $\tilde{A}$ and $\tilde{B}$,
Subtract column i from column (i+2) of $\tilde{A}$ and $\tilde{C}$.

Step 4 For $j = i+2,\ldots,n$, do the following

Add $(\frac{a_{ij}}{a_{i,i+1}})$ x row j to row (i+1) of $\tilde{A}$ and $\tilde{B}$

Subtract $(\frac{a_{ij}}{a_{i,i+1}})$ x column (i+1) from column j of $\tilde{A}$ and $\tilde{C}$.

Step 5 Return $\tilde{A}$, $\tilde{B}$, $\tilde{C}$.

Remarks 1) Steps 2-3 guarantee a nonzero pivot $a_{i,i+1}$ for $i = 1,2,\ldots,n-2$, since the subdiagonal elements of the original $\tilde{A}_{11}$ are nonzero.

2) The elimination process in step 4 is similar to step 7 in the algorithm Ht. Here we reduce the matrix to lower Hessenberg form instead.

3) The selection of the largest element as pivot (step 6 in Ht) is inconsistent with the preservation of the original upper Hessenberg form. As a consequence the multipliers $(\frac{a_{ij}}{a_{i,i+1}})$ in the algorithm TRID may have magnitude larger than one. This gives rise to the possibility of numerical instability. This is the risk one takes in order to utilize the tridiagonization of $\tilde{A}_{11}$ for reducing total computations. However the effect

of a small pivotal element does not propagate for more than three columns [11,pp.398-399].

4) Suppose we call the algorithm TRID before Step 5 of the modal reduction algorithm MR, i.e., replacing step 4 of MR by:

Step 4' Call TRID (H_k, B_k, C_k)

$$\text{Return } \tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}, \ \tilde{B} = \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix}, \ 0 = [\tilde{C}_1, \ \tilde{C}_2]$$

Then $\tilde{A}_{11}$ will be tridiagonal, and $\tilde{A}_{12}$ will have all elements, except possibly the last two rows, zero. The computations in forming $\tilde{A}_{12}q^j$ in Step 6 of MR will thus be reduced. The computations involved in solution of the n_2 linear equations in Step 6 with a more sparse tridiagonal for $(\tilde{A}_{11}-\lambda_{n_1+j}I)$ will be reduced substantially. Note that when the solution is carried out by Gaussian elimination or LU-decomposition, it will not introduce any fill-in if the elimination is performed in its natural order. If partial pivoting is employed fill-in will be introduced only along the super-superdiagonals $a_{i,i+2}$.

5) If numerical instability is not a problem, one could greatly reduce the total computation by modifying the algorithm as follows: (i) Tridiagonize H_0 right after initialization. (ii) Maintain H_k tridiagonal at each iteration. It turns out that the additional computation required to maintain H_k tridiagonal at each iteration is not very much. Since the computation at each iteration is related to the number of nonzero elements in H_k, this modification could reduce total computation considerably. Indeed the number of multiplications performed on the elements of A at each iteration to

transform $(H_k, H_k b_k)$ to (H_{k+1}, t) such that H_{k+1} remains to be tridiagonal would be approximately in the order of 14 n. However this scheme may result in poor numerical stability.

2. Transform $\tilde{A}_{11}$ to Frobenius form

Algorithm FROB $(\tilde{A}, \tilde{B}, \tilde{C})$.

Step 1 For $j = 1,2,\ldots,n_1-1$, do step 2

Step 2 For $i = 1,2,\ldots,j$, do the following

Substract $(\frac{a_{ij}}{a_{j+1,j}})$ x row (j+1) from row i of $\tilde{A}$ and $\tilde{B}$

Add $(\frac{a_{ij}}{a_{j+1,j}})$ x column i to column (j+1) of $\tilde{A}$ and $\tilde{C}$.

Step 3 Return $\tilde{A}$, $\tilde{B}$, $\tilde{C}$.

Remarks 1) We could similarly eliminate elements of $\tilde{A}_{12}$ except the last column as follows:

Step 3' For $j = n_1+1,\ldots,n-1$, do step 4'

Step 4' For $i = 1,2,\ldots,n_1$, do the following:

Substract $(\frac{a_{ij}}{a_{j+1,j}})$ x row (j+1) from row i of $\tilde{A}$ and $\tilde{B}$

Add $(\frac{a_{ij}}{a_{j+1,j}})$ x column i to column (j+1) of $\tilde{A}$ and $\tilde{C}$.

Step 5' Return $\tilde{A}$, $\tilde{B}$, $\tilde{C}$.

2) Since pivoting for size is inconsistent with the preservation of the original upper Hessenberg form, the multipliers $(\frac{a_{ij}}{a_{j+1,j}})$ in the algorithm may have magnitude larger than one. Consequently numerical instability is a risk one takes in applying this method. Although the effect of small

pivot will not appear in the final Frobenius form [11,p.407], it is quite common for a serious deterioration in the condition number to take place [11,pp.408-409] in passing to the Frobenius form.

3) The computations involved in solving the n_2 linear equations in step 6 of the algorithm MR will be greatly reduced once $\tilde{A}_{11}$ is in Frobenius form. Note that the diagonal elements of $(\tilde{A}_{11}-\lambda_{n_1+j}I)$, when $\tilde{A}_{11}$ is in Frobenius form, are the eigenvalues of $\tilde{A}_{22}$, which are small, hence the use of partial pivoting is advised. In this case the Gaussian elimination or LU-decomposition will introduce fill-in only in the superdiagonal $a_{i,i+1}$.

3. Transform A to block diagonal form.

The following algorithm BDF, will transform a block upper triangular matrix A into a block diagonal matrix, as asserted in Theorem 5.

Algorithm BDF (A,B,C)

Step 0 Select a row vector $c = (c_1,\ldots,c_{n_1},0\ldots0)$ so that $c(\tilde{A}-\lambda_{n_1+1}I)$ has at least a nonzero element in the first n_1 components.

Step 1 $k = 1$

Step 2 $c:=c(\tilde{A}-\lambda_{n_1+k}I)$

Step 3 Let $c = (\gamma_1,\gamma_2,\ldots,\gamma_n)$

Find $\gamma_{j'}$, $j' \in \{1,2,\ldots,n_1\}$ such that $|\gamma_{j'}| \geq |\gamma_j|$ for all $j \in \{1,2,\ldots,n_1\}$

Step 4 Interchange rows 1 and j' of $\tilde{A}$ and $\tilde{B}$ and interchange columns 1 and j' of $\tilde{B}$ and $\tilde{C}$.

Step 5 For $j = 2,3,\ldots,n$, do the following:

Add $(\frac{\gamma_j}{\gamma_1})$ x row j to row 1 of $\tilde{A}$ and $\tilde{B}$

Subtract $(\frac{\gamma_j}{\gamma_1})$ x column 1 from column j of $\tilde{A}$ and $\tilde{C}$.

Step 6 For $i = 1,2,\ldots,n_1-1$, do steps 7-9.

Step 7 Find $a_{ij'}$, $j' \in \{i+1,\ldots,n_1\}$, such that

$$|a_{ij'}| \geq |a_{ij}| \quad \text{for all } j \in \{i+1,\ldots,n_1\}$$

If $a_{ij'} = 0$ go to step 14;

else continue

Step 8 Interchange rows (i+1) and j' of $\tilde{A}$ and $\tilde{B}$, and interchange columns (i+1) and j' of $\tilde{A}$ and $\tilde{C}$.

Step 9 For $j = i+2,\ldots,n$, do the following

Add $(\frac{a_{ij}}{a_{i,i+1}})$ x row j to row (i+1) of $\tilde{A}$ and $\tilde{B}$

Subtract $(\frac{a_{ij}}{a_{i,i+1}})$ x column (i+1) from column j of $\tilde{A}$ and $\tilde{C}$.

Step 10 Find $a_{n_1 j'}$, $j' \in \{n_1+1,\ldots,n\}$, such that

$$|a_{n_1 j'}| \geq |a_{n_1 j}| \quad \text{for all } j \in \{n_1+1,\ldots,n\}$$

If $a_{n_1 j'} = 0$ go to step 15;

else go to step 11.

Step 11 Interchange rows (n_1+1) and j' of $\tilde{A}$ and $\tilde{B}$, and interchange columns (n_1+1) and j' of $\tilde{A}$ and $\tilde{C}$.

Step 12 For $j = n_1+2,\ldots,n$; do the following .

Add $(\frac{a_{n_1 j}}{a_{n_1,n_1+1}})$ x row j to row (n_1+1) of $\tilde{A}$ and $\tilde{B}$

Substract $\left(\frac{a_{n_1 j}}{a_{n_1, n_1+1}}\right)$ x column (n_1+1) from column j of $\tilde{A}$ and $\tilde{C}$.

Step 13 If $k = n_2$, go to Step 15;

else c:=(1,0,...,0), k:=k+1,

and go to Step 2.

Step 14 Call SIG $(\tilde{A}_{11}^T, \tilde{C}_1^T, \tilde{B}_1^T)$

Return $\tilde{A}_{11}^T$, $\tilde{C}_1^T$, $\tilde{B}_1^T$

Go to step 1.

Step 15 Return $\tilde{A}$, $\tilde{B}$, $\tilde{C}$.

Remarks 1) In step 0 usually we can choose c = (1,0,...,0). Since A_{11} is upper Hessenberg, the existence of a row vector c in step 0 is obvious.

2) Clearly $\gamma_j \neq 0$ in step 3 both in the initialization (k=1) and subsequent iteration (k>1).

3) The elimination process is considered separately in step 6 and step 10 because we want to preserve the upper block triangular structure of the matrix $\tilde{A}$.

4) The multipliers, i.e., $\left(\frac{\gamma_j}{\gamma_1}\right)$, and $\left(\frac{a_{ij}}{a_{i,i+1}}\right)$, may have magnitude larger than one, hence there exists potential numerical instability.

5) For $k \geq 2$, step 2 means simply to take the first row of $(\tilde{A}-\lambda_{n_1+k}I)$ as c. Since $\tilde{A}$ now is in lower Hessenberg form, c has only the first two components zero. The following steps 4-12 can be simplified, similar to what we did in the algorithm ITE.

Theorem 5 Given an block upper triangular matrix

$$\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}$$

where eigenvalues of A are distinct, $\tilde{A}_{11}$ is $n_1 x n_1$ and upper Hessenberg, and $\{\lambda_{n_1+1},\ldots,\lambda_{n_1+n_2}\}$ are the eigenvalues of $\tilde{A}_{22}$. The algorithm BDF $(\tilde{A},\tilde{B},\tilde{C})$ transforms $\tilde{A}$ into a block diagonal matrix.

Remark If we call the algorithm BDF prior to step 5 of the modal reduction algorithm MR, we may then apply the Corollary of Theorem 1 and readily obtain the mode-reduced output $\hat{y}$. Thus the algorithm MR may be modified from step 4 on as follows.

Step 4' Call BDF (H_k,B_k,C_k)

$$\text{Return } \tilde{A} = \begin{bmatrix} \tilde{A}_{11} & 0 \\ 0 & \tilde{A}_{22} \end{bmatrix}, \tilde{B} = \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix} [\tilde{C} = \tilde{C}_1 \quad \tilde{C}_2]$$

Step 5' Solve the reduced-order system

$$\dot{\eta}_2 = \tilde{A}_{22}\eta_2 + \tilde{B}_2 u$$

and obtain the mode-reduced output

$$\hat{y} = \tilde{C}_2\eta_2$$

VI. SYMMETRICAL A MATRIX

For the special case that the A matrix in Eq. (1) is symmetric, orthogonal transformations should be used to preserve symmetry. Therefore,

we apply Householder transformation [11,pp.290-299] instead of the elementary stabilized transformations in the algorithm Ht to transform a matrix into Hessenberg form. The algorithm is described below.

Algorithm SHt (A,b)

Step 1 Find $\beta_{\underline{i}}$, $\underline{i} \in \{1,2,\ldots,n\}$, such that

$$|\beta_{\underline{i}}| \geq |\beta_i| \quad \text{for all } i \in \{1,2,\ldots,n\}$$

If $\beta_{\underline{i}} = 0$ go to step 9;

else go to step 2.

Step 2 Interchange rows 1 and $\underline{i}$ of b, A, and B, and interchange columns 1 and $\underline{i}$ of A and C.

Step 3 For i = 2,...,n, do the following

Subtract $(\frac{\beta_i}{\beta_1})$ x row 1 from row i of A and B.

Add $(\frac{\beta_i}{\beta_1})$x column i to column 1 of A and C.

Step 4 For k = 1,2,...,n-2, do steps 5-7

Step 5 $S = (\sum_{j=k+1}^{n} a_{ki}^2)^{1/2}$

$2K^2 = S^2 + (\text{sgn } a_{k,k+1})a_{k,k+1}S$

$u_{k+1} = a_{k,k+1} + (\text{sgn } a_{k,k+1})S.$

If s=0 go to step 9; else continue

Step 6 For i = 1,...,k, set $u_i = 0$;

for i = k+2,...,n, set $u_i = a_{ki}$.

Step 7 $P = I - \frac{1}{2K^2} uu^T$

A:=PAP

B:=PB

C:=CP

Step 8 Output H:=A and $t = (\beta_1, 0, \ldots, 0)^T$

Step 9 Stop.

Remark Perhaps it is worthwhile to elaborate on Step 7 of the above algorithm SHt. Note that it is important to take full advantage of symmetry. The derivation of the following procedure is straightforward and can be found in [11,p.292].

Step 7-1 For i = k+1,...,n

$$p_i = \sum_{j=k+1}^{n} a_{ij} u_j$$

Step 7-2 $$\alpha = \frac{1}{4K^2} \sum_{j=k+1}^{n} u_j p_j$$

Step 7-3 For i = k+1,...,n

$$q_i = p_i - \alpha u_i$$

Step 7-4 $a_{k,k+1} := -(\text{sgn } a_{k,k+1})S$

For i = k+2,...,n, $a_{ki} := 0$

Step 7-5 For i = k+1,...,n; u = i,...,n

$$a_{ij} := a_{ij} - u_i q_j - q_i u_j.$$

Step 7-6 For i = k+1,...,n; j = 1,2,...,m

$$b_{ij} := b_{ij} - \frac{1}{2K^2} u_i \sum_{\ell=k+1}^{n} u_\ell b_{\ell j}$$

Step 7-7 For $i = 1,2,\ldots,p$; $j = k+1,\ldots,n$

$$c_{ij} := c_{ij} - \frac{1}{2K^2}\, u_j \sum_{\ell=k+1}^{n} c_{i\ell} u_\ell$$

Remark The total number of multiplications in reducing a symmetric A to a tridiagonal form by Householder's Method is essentially $\frac{2}{3} n^3$ compared with $\frac{4}{3} n^3$ in Givens' reduction, and n square roots compared with $\frac{1}{2} n^2$ in the Givens' Method [11,p.293]. Hence Householder's Method is employed here.

At each iteration, again by taking advantage of the special structures of H_k and $H_k b_k$, the algorithm can be simplified as below.

Algorithm SITE (H_k, B_k, C_k)

Step 1 Let $\beta_1 = h_{11}$ and $\beta_2 = h_{21}$

If $|\beta_2| > |\beta_1|$ go to step 2; else go to step 3.

Step 2 Interchange β_1 and β_2

Interchange rows 1 and 2 of H_k and B_k

and interchange columns 1 and 2 of H_k and C_k.

Step 3 Subtract $(\frac{\beta_2}{\beta_1})$ x row 1 from row 2 of H_k and B_k

Add $(\frac{\beta_2}{\beta_1})$ x column 2 to column 1 of H_k and C_k

Step 4 For $r = 1,2,\ldots,n-2$, do steps 5-11.

Step 5 $s = (a^2_{r,r+1} + a^2_{r,r+2})^{1/2}$

$2K^2 = S^2 + (\text{sgn } a_{r,r+1}) a_{r,r+1} S$

$u_{r+1} = a_{r,r+1} + (\text{sgn } a_{r,r+1}) S$

$u_{r+2} = a_{r,r+2}$

Step 6 For i = r+1, r+2, r+3, set

$$p_i = a_{i,r+1}u_{r+1} + a_{i,r+2}u_{r+2};$$

$$\alpha = \frac{1}{4K^2}\,(p_{r+1}u_{r+1} + p_{r+2}u_{r+2})$$

Step 7 For i = r+1, r+2, r+3, set

$$q_i = p_i - \alpha u_i$$

Step 8 $a_{r,r+1} := -(\text{sgn } a_{r,r+1})S$

$$a_{r,r+2} := 0$$

Step 9 For i = r+1, r+2, r+3; j = r+1, r+2, r+3

$$a_{ij} := a_{ij} - u_i q_j - q_i u_j$$

Step 10 For i = r+1, r+2; j = 1,2,...,m

$$b_{ij} := b_{ij} - \frac{1}{2K^2}\,u_i(u_{r+1}b_{r+1,j} + u_{r+2}b_{r+2},j)$$

Step 11 For i = 1,2,...,p; j = r+1, r+2

$$c_{ij} := c_{ij} - \frac{1}{2K^2}\,u_j\,(c_{i,r+1}u_{r+1} + c_{i,r+2}u_{r+2})$$

Step 12 Return $H_{k+1} = H_k$, $B_{k+1} = B_k$, $C_{k+1} = C_k$.

For a symmetrical A matrix, if we replace Ht by SHt and ITE by SITE in our modal reduction algorithm, the iterations will converge to a block diagonal matrix. Therefore, we may apply the Corollary of Theorem 1 and readily obtain the mode-reduced output $\hat{y}$. The modified modal reduction algorithm for symmetrical A matrix is summarized below.

Algorithm SMR (A,B,C)

Step 1 Call INI (A,B,C), replacing Ht by SHt

Return H_0, B_0, C_0,

Set k=0

Step 2 If $h_{n_1+1,n_1} \simeq 0$ for some n_1, go to step 4;
else continue

Step 3 Call SITE (H_k, B_k, C_k)

Return H_{k+1}, B_{k+1}, C_{k+1}

Set k:=k+1, go to step 2

Step 4 Set $\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & 0 \\ 0 & \tilde{A}_{22} \end{bmatrix}$ $\tilde{B} = \tilde{B}_k = \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix}$, $\tilde{C} = C_k = [\tilde{C}_1 \quad \tilde{C}_2]$

where $\tilde{A}_{11}$ is $n_1 x n_1$, $\tilde{A}_{22}$ is $n_2 x n_2$ and tridiagonal.

Step 5 Solve the reduced-order system

$$\dot{\eta}_2 = \tilde{A}_{22}\eta_2 + \tilde{B}_2 u$$

and obtain the mode-reduced output

$$\hat{y} = \tilde{C}_2 \eta_2$$

VII. ANOTHER MODE-RETENTION CRITERION

Consider a linear time-invariant system

$$\dot{x} = Ax + Bu \tag{41}$$

$$y = Cx \tag{42}$$

Suppose that the eigenvalues of A are distinct. Let $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ be a given sub-set of the eigenvalues of A. Suppose we want a reduced order model which retains only the contributions of the modes associated with $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$. The following algorithm will accomplish this in finite steps, the proof of which can be easily constructed by applying Theorem 3, Lemma 1, and similar arguments in the proof of Theorem 5. Let $t = (1, 0, \ldots, 0)^T$.

Algorithm AMR (A,B,C)

Step 1 Call INI (A,B,C)

Return H_0, B_0, C_0

Set $H_1 = H_0$, k=1.

Step 2 If $h_{n-r+1,n-r} = 0$, go step 4;

else go to step 3.

Step 3 $c_k = (H_k - \alpha_k I)t$

Call Ht (H_k, c_k)

Return H_{k+1}, t_{k+1}

Set k:=k+1, go to step 2

Step 4 Set $\tilde{A} = H_k = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}$, $\tilde{B} = \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix}$, $\tilde{C} = [\tilde{C}_1 \quad \tilde{C}_2]$

Step 5 Find the eigenvectors q^j, j = 1,2,...,r, of $\tilde{A}_{22}$.

Step 6 For j = 1,2,...,r, solve for p^j

$$(\tilde{A}_{11} - \alpha_j I)p^j = -\tilde{A}_{12}q^j$$

Let $\tilde{Q}_{12} = [p^1 p^2, \ldots, p^r]$ and $\tilde{Q}_{22} = [q^1 q^2, \ldots, q^r]$

Step 7 Solve the reduced-roder system

$$\dot{\eta}_2 = \tilde{A}_{22}\eta_2 + \tilde{B}_2 u$$

and obtain the mode-reduced output

$$\hat{y} = [\tilde{C}_1 \tilde{Q}_{12} \tilde{Q}_{22}^{-1} + \tilde{C}_2]\eta_2$$

REFERENCES

1. L. O. Chua and P. M. Lin, Computer-Aided Analysis of Electronic Circuits, Prentice-Hall, 1975.

2. J. M. Undrill, "Dynamic stability calculations for an arbitrary number of interconnected synchronous machines," IEEE Trans. Power App. & Syst., vol. PAS-87, March 1968, pp. 835-844.

3. M. Lal and M. E. Van Valkenburg, "Reduced-order modeling of large-scale linear systems," in Large-Scale Dynamical Systems ed. by R. Saeks, Point Lobos Press, 1976, pp. 127-166.

4. E. J. Davidson, "A method for simplifying linear dynamic systems," IEEE Trans. Aut. Control., vol. AC-11, Jan. 1966, pp. 93-101.

5. J. M. Undrill and A. E. Turner, "Construction of power system electromechanical equivalent by modal analysis," IEEE Trans. Power App. & Systems, vol. PAS-90, Sept. 1971, pp. 2049-2059.

6. J. E. Van Ness, H. Zimmer, and M. Cutter, "Reduction of dynamic models of power systems," Proc. 1973 PICA Conference, pp. 105-112.

7. H. Y. Altalib and P. C. Krause, "Dynamic equivalents by combination of reduced order models of system components," IEEE Trans. Power App. & Systems, vol. PAS-95, Sept. 1976, pp. 1535-1544.

8. W. W. Price, "NPCC/GE study of model analysis equivalents," GE Report, 1974.

9. P. V. Kokotovic, "A Ricatti equation for block-diagonization of ill-conditioned systems," IEEE Trans. Aut. Control, vol. AC-20, Dec. 1975, pp. 812-814.

10. C. A. Desoer, Notes for a Second Course on Linear Systems, Van Nostrand Reinhold, 1970.

11. J. H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1966.

APPENDIX

A. Proof of Theorem 1.

Let the k-th column of the $n \times n$ matrix $\tilde{Q}$ (resp. Q) be an eigenvector of $\tilde{A}$ (resp. A) corresponding to the eigenvalue λ_k. From the definition of Q we have $AQ = Q\Lambda$ where $\Lambda = \text{diag}\,(\lambda_1, \lambda_2, \ldots \lambda_{n_1+n_2})$. Hence $TAT^{-1}\,TQ = TQ\Lambda$ or $\tilde{A}(TQ) = (TQ)\Lambda$. We conclude that

$$\tilde{Q} = TQ \tag{A1}$$

We may partition both $\tilde{Q}$ and Q as follows

$$\begin{bmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ \tilde{Q}_{21} & \tilde{Q}_{22} \end{bmatrix}, \quad \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$$

where $\tilde{Q}_{11}$ and Q_{11} are $n_1 \times n_1$.

Combining the transformations (8) and (15), we have

$$\eta = TQ\xi = \tilde{Q}\xi \tag{A2}$$

in view of (A1).

By definition we have

$$\begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix} \begin{bmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ \tilde{Q}_{21} & \tilde{Q}_{22} \end{bmatrix} = \begin{bmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ \tilde{Q}_{21} & \tilde{Q}_{22} \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \tag{A3}$$

where $\Lambda_1 = \text{diag}\,(\lambda_1, \ldots \lambda_{n_1})$ and $\Lambda_2 = \text{diag}\,(\lambda_{n_1+1}, \ldots \lambda_{n_1+n_2})$, or

$$\tilde{A}_{11}\,\tilde{Q}_{11} + \tilde{A}_{12}\,\tilde{Q}_{21} = \tilde{Q}_{11}\,\Lambda_1 \tag{A4}$$

$$\tilde{A}_{22}\,\tilde{Q}_{21} = \tilde{Q}_{21}\,\Lambda_1 \tag{A5}$$

$$\tilde{A}_{11}\,\tilde{Q}_{12} + \tilde{A}_{12}\,\tilde{Q}_{22} = \tilde{Q}_{12}\,\Lambda_2 \tag{A6}$$

$$\tilde{A}_{22}\,\tilde{Q}_{22} = \tilde{Q}_{22}\,\Lambda_2 \tag{A7}$$

Since none of $\{\lambda_1,\ldots,\lambda_{n_1}\}$ is an eigenvalue of $\tilde{A}_{22}$, eq. (A5) implies that $\tilde{Q}_{21} = 0$. Eq. (A7) implies that the columns of $\tilde{Q}_{22}$ are eigenvectors of $\tilde{A}_{22}$. Since the eigenvalues of $\tilde{A}_{22}$ are distinct, $\tilde{Q}_{22}$ is nonsingular.

We may rewrite (A2) as

$$\begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix} = \begin{bmatrix}\tilde{Q}_{11} & \tilde{Q}_{12}\\ 0 & \tilde{Q}_{22}\end{bmatrix}\begin{bmatrix}\xi_1\\ \xi_2\end{bmatrix} \tag{A8}$$

hence $\eta_2 = \tilde{Q}_{22}\xi_2$, where $\tilde{Q}_{22}$ is nonsingular.

The mode-reduced output as defined in (12) can now be expressed as

$$\begin{aligned}\hat{y} &= CT^{-1}{}_T\begin{bmatrix}Q_{12}\\ Q_{22}\end{bmatrix}\xi_2\\ &= CT^{-1}\begin{bmatrix}\tilde{Q}_{12}\\ \tilde{Q}_{22}\end{bmatrix}\tilde{Q}_{22}^{-1}\eta_2\end{aligned} \tag{A9}$$

where we have used eq. (A1) and $\xi_2 = \tilde{Q}_{22}^{-1}\eta_2$. Eq. (A9) is the same as eq. (18). Note also (19) is in fact (A6). Q.E.D.

B. Proof of Theorem 2.

Given a pair (A,b) the algorithm Ht(A,b) will generate a pair (H,t) such that H is Hessenberg and $t = (\tau,0,\ldots,0)^T$ provided it does not exit from step 9.

We are going to show that if $\{b,Ab,\ldots,A^{n-1}b\}$ spans $\mathbb{R}^n$, the algorithm Ht(A,b) will never exit from step 9. First notice that it can never come to step 9 via step 1 because b can not be the zero vector. Now suppose the algorithm comes to step 9 via step 5, then let the current A

matrix and b vector be denoted by $\underline{A}$ and $\underline{b}$, respectively, we have

$$\underline{A} = \begin{bmatrix} \underline{A}_{11} & \underline{A}_{12} \\ 0 & \underline{A}_{22} \end{bmatrix}$$

where $\underline{A}_{11}$ is $j' \times j'$ and Hessenberg. Let the transformations so far be represented by $\underline{\mathcal{J}}$, which is of course nonsingular. Hence

$$\underline{b} = \underline{\mathcal{J}} b = (\tau, 0, \ldots, 0)^T$$

$$\underline{A} = \underline{\mathcal{J}} A \underline{\mathcal{J}}^{-1}$$

Note that $\underline{A}^i \underline{b} \in$ span $\{\underline{b}, \underline{A}\underline{b}, \ldots, \underline{A}^{j'-1}\underline{b}\}$ for $i \geq j'$, i.e., $\{\underline{b}, \underline{A}\underline{b}, \ldots, \underline{A}^{n-1}\underline{b}\}$ does not span $\mathbb{R}^n$, which contradicts the assumption that $\{b, Ab, \ldots, A^{n-1}b\}$ spans $\mathbb{R}^n$. Q.E.D.

C. Lemma 1. Let $H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}$, where H_{11} is a $k \times k$ upper Hessenberg matrix with nonzero subdiagonal elements $(h_2, h_3, \ldots, h_k)$; the ij-th element of H_{21}, $(H_{21})_{ij}$, is given below:

$$(H_{21})_{ij} = \begin{cases} h_{k+1} & \text{for } i = 1,\ j = k \\ 0 & \text{otherwise} \end{cases} \qquad \text{(A10)}$$

Suppose that the eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ of H are distinct. Let us denote the corresponding eigenvectors by $(v_1, v_2, \ldots, v_n)$. Let $b = (1, 0, \ldots, 0)^T$. Then $h_{k+1} = 0$ if and only if

$$b = \sum_{i=1}^{k} \theta_{\sigma i} v_{\sigma i} \quad \text{with} \quad \theta_{\sigma i} \neq 0 \quad \text{for} \quad i = 1, 2, \ldots, k \ . \qquad \text{(A11)}$$

Moroever $(v_{\sigma 1}, v_{\sigma 2}, \ldots, v_{\sigma k})$ are the eigenvectors of H associated with the eigenvalues of H_{11}.

Proof: ($\Leftarrow$). The set of k vectors $\{b, Hb, \ldots, H^{k-1}b\}$ are linearly independent because $h_i \neq 0$ for $i = 2, \ldots, k$. Hence the span of $\{b, Hb, \ldots, H^{k-1}b\}$ is the same as the span of the k linearly independent vectors $\{v_{\sigma 1}, v_{\sigma 2}, \ldots, v_{\sigma k}\}$. Clearly any vector in the span of $\{b, Hb, \ldots, H^{k-1}b\}$ has the p-th component, $p \geq k+1$ zero. Now the vector $H^k b = \sum_{i=1}^{k} \theta_{\sigma i} \lambda_{\sigma i}^k v_{\sigma i}$ is in the span of $\{v_{\sigma 1}, v_{\sigma 2}, \ldots, v_{\sigma k}\}$, hence in the span of $\{b, Hb, \ldots, H^{k-1}b\}$, and its (k+1)-th component is $h_{k+1} h_k \cdots k_2$. Therefore $h_{k+1} = 0$.

($\Rightarrow$). Let the eigenvectors of H correspond to the eigenvalues $(\lambda_{\sigma 1}, \ldots, \lambda_{\sigma k})$ of H_{11} be denoted by $(v_{\sigma 1}, \ldots, v_{\sigma k})$. Let $v_{\sigma i} = \begin{bmatrix} v_{\sigma i}^1 \\ v_{\sigma i}^2 \end{bmatrix}$, by definition

$$\begin{bmatrix} H_{11} & H_{12} \\ 0 & H_{22} \end{bmatrix} \begin{bmatrix} v_{\sigma i}^1 \\ v_{\sigma i}^2 \end{bmatrix} = \lambda_{\sigma i} \begin{bmatrix} v_{\sigma i}^1 \\ v_{\sigma i}^2 \end{bmatrix} \tag{A12}$$

Since $\lambda_{\sigma i}$ is not an eigenvalue of H_{22}, $v_{\sigma i}^2 = 0$. Thus the k linearly independent set of vectors $\{v_{\sigma 1}, \ldots, v_{\sigma k}\}$ span the subspace $\{x = (x_1, \ldots, x_n)^T | x_p = 0, \text{ for } p > k\}$. But $b = (1, 0, \ldots, 0)^T$, we may write

$$b = \sum_{i=1}^{k} \theta_{\sigma i} v_{\sigma i} \tag{A13}$$

On the other hand, since H_{11} is Hessenberg with nonzero subdiagonal elements, the set of k vectors $\{b, Hb, \ldots, H^{k-1}b\}$ is linearly independent. This implies that $\theta_{\sigma i} \neq 0$ for all $i = 1, 2, \ldots, k$.

Q.E.D.

D. Proof of Theorem 3.

Clearly if the algorithm SIG works, then the theorem is proved. We are going to show that the algorithm SIG works. Specifically, we claim that

(i) SIG may result in an A'_{11} in step 5 such that $k > j'$.

(ii) SIG will result in an A'_{11} in step 5 such that $k > j'$ in no more than j' iterations of the loop consisting of the steps 3-4-5-6-3.

Let the input to SIG be denoted by $\underline{A}$, $\underline{B}$, $\underline{C}$, where

$$\underline{A} = \begin{bmatrix} \underline{A}_{11} & \underline{A}_{12} \\ 0 & \underline{A}_{22} \end{bmatrix}$$

and $\underline{A}_{11}$ is $j' \times j'$ and Hessenberg with nonzero subdiagonal. Let $\{\lambda_{\alpha 1},\ldots,\lambda_{\alpha j'}\}$ and $\{\lambda_{\beta 1},\ldots,\lambda_{\beta(n-j')}\}$ be the eigenvalues of $\underline{A}_{11}$ and $\underline{A}_{22}$ respectively, and $\{e_{\alpha 1},\ldots,e_{\alpha j'}\}$ and $\{e_{\beta 1},\ldots,e_{\beta(n-j')}\}$ be the corresponding eigenvectors of $\underline{A}$. Let $\underline{b} = (1,0,\ldots,0)^T$. We may write

$$\underline{b} = \sum_{i=1}^{j'} \phi_{\sigma i} e_{\sigma i} + \sum_{i=1}^{n-j'} \phi_{\beta i} e_{\beta i} \qquad \text{(A14)}$$

The "only if" part of Lemma 1 implies that

$$\phi_{\beta i} = 0 \quad \text{for all} \quad i \in \{1,\ldots,n-j'\}$$
$$\phi_{\alpha i} \neq 0 \quad \text{for all} \quad i \in \{1,\ldots,j'\}$$

Now let us define

$$s \triangleq (0,\ldots,\sigma,0,\ldots,0)^T$$

with σ in the $(j'+1)$-th position.

We may write

$$s = \sum_{i=1}^{j'} \sigma\theta_{\sigma i} e_{\sigma i} + \sum_{i=1}^{n-j'} \sigma\phi_{\beta i} e_{\beta i}$$

It can be shown, using similar arguments following eqs. (A4-A7), that the set of eigenvectors $\{e_{\alpha 1},\ldots,e_{\alpha j'}\}$ spans the first j' coordinates, hence

$$s \notin \text{span}\{e_{\alpha 1},\ldots,e_{\alpha j'}\}$$

or

$$\phi_{\beta i} \neq 0 \quad \text{for some} \quad i \in \{1,\ldots,n-j'\} \tag{A15}$$

Consider $b = \underline{b} + s$, or

$$b = \sum_{i=1}^{j'} (\phi_{\sigma i} + \sigma\theta_{\sigma i}) e_{\sigma i} + \sum_{i=1}^{n-j'} \sigma\phi_{\beta i} e_{\beta i} \tag{A16}$$

Comparing b with $\underline{b}$ we notice that because of (A15) new nonzero component $\sigma\phi_{\beta i}$ along some eigenvectors $e_{\beta i}$ will be introduced. On the other hand, some components along $e_{\alpha i}$ may vanish as a consequence of the possibility

$$\phi_{\sigma i} + \sigma\theta_{\sigma i} = 0 \quad \text{for some} \quad i \in \{1,\ldots,j'\} \tag{A17}$$

If the number of new components introduced is greater than the number of components vanished, then b has more nonzero components. The "if" part of Lemma 1 implies that the application of $\underline{\underline{H}}t$ to $\underline{A}$ and b will then result in a new upper left Hessenberg block with dimension strictly greater than j'.

On the other hand, consider $(\phi_{\sigma i} + \sigma\theta_{\sigma i})$, $i \in \{1,\ldots,j'\}$ as a function of σ. Clearly, if $\sigma \neq -\frac{\phi_{\sigma i}}{\theta_{\sigma i}}$, $i = 1,\ldots,j'$, then $(\phi_{\sigma i} + \sigma\theta_{\sigma i}) \neq 0$ for all $i \in \{1,\ldots,j'\}$. Since each iteration in the loop (steps 3-4-5-6-3) we change the value of σ, in no more than j' iterations we will come across a vector b which has more nonzero components of the eigenvectors than $\underline{b}$ has.

Consequently the outer loop consisting of the steps 1 through 7 will terminate eventually when k reaches n. Q.E.D.

E. Proof of Theorem 4.

We are going to apply the "if" part of Lemma 1 to the pair (H_k, b_k). First let us relate the eigenvectors of H_k to that of A, and relate the coordinates of b_k relative to the eigenvectors of H_k to the coordinates of b relative to the eigenvectors of A.

Referring to the notations we used in the outline of the modal reduction algorithm, we have

$$H_k = \mathcal{J}_k H_{k-1} \mathcal{J}_k^{-1}$$
$$= \mathcal{J}_k \mathcal{J}_{k-1} \cdots \mathcal{J}_0 A \mathcal{J}_0^{-1} \cdots \mathcal{J}_k^{-1} \qquad \text{(A18)}$$

$$b_k = \frac{1}{\alpha_k} \mathcal{J}_k b_{k-1}$$
$$= \frac{1}{\alpha_k \cdots \alpha_0} \mathcal{J}_k \cdots \mathcal{J}_0 A^k b \qquad \text{(A19)}$$

Let us define

$$T_k \triangleq \mathcal{J}_k \mathcal{J}_{k-1} \cdots \mathcal{J}_0 \qquad \text{(A20)}$$

$$\alpha^{(k)} \triangleq \alpha_k \alpha_{k-1} \cdots \alpha_0 \qquad \text{(A21)}$$

Hence

$$H_k = T_k A T_k^{-1} \tag{A22}$$

$$b_k = T_k \frac{A^k b}{\alpha(k)} \tag{A23}$$

The component of $b_k = T_k \frac{A^k b}{\alpha(k)}$ along the eigenvector $T_k e_i$ of $H_k = T_k A T_k^{-1}$ is $\frac{\phi_i \lambda_i^k}{\alpha(k)}$.

Now we claim that

$$\frac{\phi_j \lambda_j^k}{\alpha(k)} \to 0 \quad \text{as} \quad k \to \infty \quad \text{for all} \quad j \geq n_1+1 \,. \tag{A24}$$

Note that

$$b_k = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \sum_{i=1}^{n} \frac{\phi_i \lambda_i^k}{\alpha(k)} T_k e_i \tag{A25}$$

hence

$$\sum_{i=1}^{n} \left| \frac{\phi_i \lambda_i^k}{\alpha(k)} \right| \leq M \quad \text{for some} \quad M \in \mathbb{R} \tag{A26}$$

$$|\alpha(k)| \geq \frac{1}{M} \sum_{i=1}^{n} |\phi_i \lambda_i^k| \geq \frac{1}{M} \phi_{\bar{i}} \lambda_{\bar{i}}^k \quad \text{for any} \quad \bar{i} \in \{1,\ldots,n_1\} \tag{A27}$$

Thus, for any $j \geq n_1 + 1$,

$$\left| \frac{\phi_j \lambda_j^k}{\alpha(k)} \right| \leq M \left| \frac{\phi_j}{\phi_{\bar{i}}} \right| \cdot \left| \frac{\lambda_j}{\lambda_{\bar{i}}} \right|^k \tag{A28}$$

But

$$\left(\frac{\lambda_j}{\lambda_i}\right)^k \to 0 \quad \text{as} \quad k \to \infty \quad \text{for any} \quad j \geq n_1+1 \ , \tag{A29}$$

so $\left|\frac{\phi_j \lambda_j^k}{\alpha^{(k)}}\right| \to 0$ as $k \to \infty$. Consequently,

$$b_k \to \sum_{i=1}^{n_1} \frac{\phi_i \lambda_i^k}{\alpha^{(k)}} T_k e_i \tag{A30}$$

From Lemma 1, we conclude that $h_{n_1+1}^{(k)} \to 0$ as $k \to \infty$.

Furthermore, when H_k converges to an upper triangular matrix $\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}$, those components for which the coordinates $\left(\frac{\phi_i \lambda_i^k}{\alpha^{(k)}}\right)$ do not vanish, i.e., $i = 1,2,\ldots,n$, will correspond to the modes associated with the eigenvalues of $\tilde{A}_{11}$, as implied by Lemma 1. Therefore the eigenvalues of $\tilde{A}_{22}$ are $\{\lambda_{n_1+1},\ldots,\lambda_{n_1+n_2}\}$. Q.E.D.

F. Proof of Theorem 5.

Essentially the algorithm BDF does the following.[7] It starts with a vector c_0 and the matrix $\tilde{A}^T$. At each k, the algorithm transforms the pair (H_k, c_k), where $c_k = (\tilde{A}^T - \lambda_{n_1+k} I)c_{k-1}$ and $H_1 = \tilde{A}^T$, into a pair (H_{k+1}, t_{k+1}) such that H_{k+1} is upper Hessenberg and $t_{k+1} = (\tau,0,\ldots,0)^T$. The transformation at each k is basically the same as in the algorithm Ht except the first n_1 rows (columns) and the last n_2 rows (columns) are

[7] In order to apply directly our previous results, we will consider the transpose of $\tilde{A}$ and the column vector c_0, etc., in the proof. However, in the algorithm we work directly with $\tilde{A}$ and the row vector c, etc. The transformation there is to bring $\tilde{A}$ to a lower Hessenberg form.

not allowed to interchange in order to preserve the upper triangular structure of $\tilde{A}$. Remarks 1) and 2) following the presentation of the algorithm BDF indicate that BDF will not break down. We are going to complete the proof of Theorem 5 in two steps.

(1) Let $\{f_1,\ldots,f_{n_1}\}$ and $\{f_{n_1+1},\ldots,f_n\}$ be the eigenvectors of $\tilde{A}^T$ corresponding to the eigenvalues of $\tilde{A}_{11}^T$ and $\tilde{A}_{22}^T$ respectively. We claim that if $c_0 = \sum_{i=1}^{n} \phi_i f_i$ such that $\phi_i \neq 0$ for all $i = 1,2,\ldots,n$, then the (n_1+1,n_1) element of H_{n_2+1} is zero.

Consider $c_1 = (\tilde{A}^T-\lambda_{n_1+1})c_0 = \sum_{i=1}^{n} \lambda_i\phi_i f_i - \lambda_{n_1+1} \sum_{i=1}^{n} \phi_i f_i$. So c_1 does not have a component along f_{n_1+1}. Because the way we define c_k, $c_{n_2} = \sum_{i=1}^{n_1} \lambda_i^{n_2}\phi_i f_i$. By applying the "if" part of Lemma 1, we prove the claim.

(2) We claim that the algorithm will generate a vector $c_0 = \sum_{i=1}^{n} \phi_i f_i$ such that $\phi_i \neq 0$ for all $i = 1,2,\ldots,n$.

This becomes obvious once one notices that step 14 in BDF does exactly the same as step 3 in INI. Therefore the claim can be easily proved following the proof of Theorem 3. Q.E.D.