Efficient Model Reduction for the Control of Large-Scale Systems

Richard Colgren

1 Introduction

When analyzing and controlling large-scale systems, it is extremely important to develop efficient modeling processes. The key dynamic elements must be identified and spurious dynamic elements eliminated. This allows the controls engineer to implement the optimal control strategy for the problem at hand. Model reduction techniques provide an extremely effective way to address this requirement.

In this chapter the evolution of model reduction techniques for designing control systems for large-scale systems is summarized. These start with simple approaches such as spectral decomposition and simultaneous gradient error reduction and then progresses through a variety of balanced and related model reduction approaches. Motivations for the use of these methods are given. Emphasis is given to approaches which are easily applied to the large, generalized models which are created by Computer Aided Design (CAD) tools. A U-2S example application is provided and described.

The provided example of a large-scale system of very high order is a model of an air vehicle with significant aeroservoelastic coupling. Future unstable vehicles, or those employing features such as relaxed static stability, can display aeroservoelastic coupling. Evaluation of the interactions between the dynamic modes must be accomplished using an efficient, integrated modeling approach. The extensive use of composite materials has also resulted in greater aeroservoelastic coupling. Finite element models of complex systems are sparse and are also intrinsically of very high order. Such large-scale systems must be subjected to comprehensive analysis.

The methods discussed in this chapter preserve the frequency response characteristics of the system model being examined while reducing its size to one practical for direct controls design. They support a variety of optimal and robust control system

R. Colgren
Vice President and Chief Scientist, Viking Aerospace LLC, 100 Riverfront Road, Suite B, Lawrence, KS 66044, USA
e-mail: rcolgren@gmail.com

J. Mohammadpour and K.M. Grigoriadis, *Efficient Modeling and Control, of Large-Scale Systems*, DOI 10.1007/978-1-4419-5757-3_2,

design approaches. Some control system design methods require the use of all of the system's states. Without the use of model reduction techniques the full state control system design would be too large to practically implement. States that would be included within the controller might also be outside the bandwidth of the servos or actuators. This might generate an ill-conditioned problem. The use of model reduction methods provides a way to generate full state controller solutions of reduced size, and to further simplify these full state feedback solutions once they have been generated.

2 Spectral Decomposition

This method is an efficient way to generate a reduced order model of a large-scale system when all of its subsystems are decoupled or are at most weakly coupled. It was initially developed by Chiodi and Davis at the Lockheed Corporation [1]. It depends on systems having distinct eigenvalues within well separated frequency responses as shown in Fig. 1. These modal groupings are then decoupled using the eigenvector matrix.

The starting point for this method is the standard *n*th order state space representation given in (1).

$$dx/dt = Ax + Bu; y = Cx \tag{1}$$

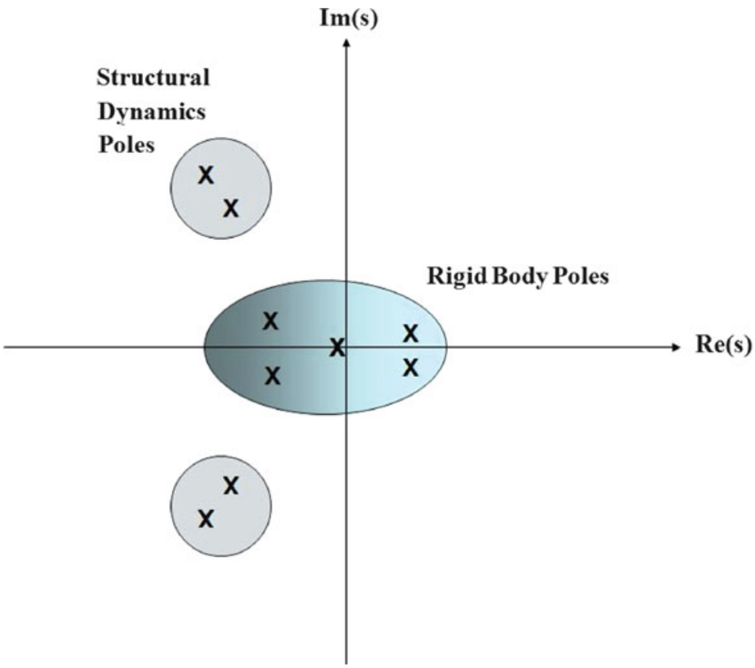

Fig. 1 Distinct frequency range groupings of a model's eigenvalues

The system matrix A is then represented using the system's eigenvalue and eigenvector matrices as shown in (2).

$$dx/dt = V\Lambda\ V^{-1}x + Bu; y = Cx \tag{2}$$

This system description can then be expanded into the following series representation as provided in (3). Note that the sum is initially computed using the full number of states within the model, but that for control system design it is computed using the eigenvalues to be obtained. Also, please note that $V_iV_i^{-1}$ is the residue of the eigenvalue λ_i in $(sI - A)^{-1}$.

$$dx/dt = [\sum V_iV_i^{-1}\lambda_i]x + Bu; y = C_x \tag{3}$$

From this representation the system can be reduced into its individual elements as is done in (4)

$$dx_i/dt = \underline{A}_ix + \underline{B}_iu \tag{4}$$

The spectral decomposition process is described in Fig. 2. The system eigenvalues and eigenvectors are next calculated. For each eigenvalue to be retained within the reduced order model the eigenvector and its inverse are multiplied to generate $V_iV_i^{-1}$. Each matrix element is multiplied by its corresponding eigenvalue λ_i. The transformed state matrix is then constructed by solving for the sum as given in (5) for all the retained eigenvalues.

$$\underline{A} = \sum V_iV_i^{-1}\lambda_i \tag{5}$$

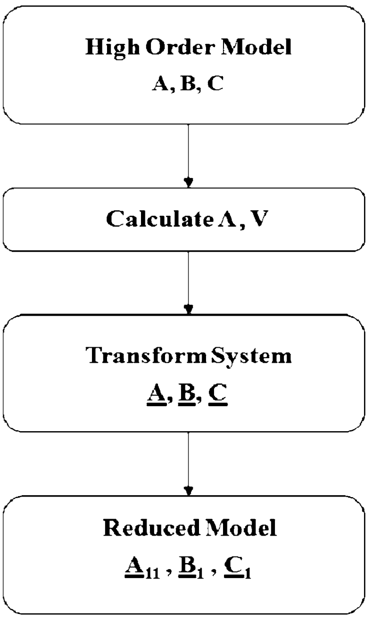

Fig. 2 Spectral decomposition process

To calculate $\underline{B} = TB$ and $\underline{C} = CT$ the transformation matrix T must also be calculated. This transformation is provided as in (6).

$$T = \sum V_i V_i^{-1} \tag{6}$$

After spectral decomposition the system is transformed into that provided as in (7)–(9).

$$d\underline{x}_1/dt = \underline{A}_{11}\underline{x}_1 + \underline{A}_{12}\underline{x}_2 + \underline{B}_1 u \tag{7}$$

$$d\underline{x}_2/dt = \underline{A}_{21}\underline{x}_1 + \underline{A}_{22}\underline{x}_2 + \underline{B}_2 u \tag{8}$$

$$y = \underline{C}_1\underline{x}_1 + \underline{C}_2\underline{x}_2 \tag{9}$$

The reduced system model as described by the states to be used for the design of the control system is given as in (10) and (11).

$$d\underline{x}_1/dt = \underline{A}_{11}x_1 + \underline{B}_1 u \tag{10}$$

$$y = \underline{C}_1\underline{x}_1 \tag{11}$$

3 Simultaneous Gradient Error Reduction

This method provides an efficient way to generate a reduced order model of a large-scale system whether all of its subsystems and their outputs are decoupled or not. It does require that all of its inputs are no more than weakly coupled to each other. For flying qualities prediction and for aeroservodynamic modeling [2,3] the author originally developed this method for simultaneously fitting several frequency responses depicting high order systems to low order transfer functions including time delays [4, 5]. This method has been applied to the design the control systems for several large-scale aerospace systems.

Simultaneous gradient error reduction is applied over a finite bandwidth. Controls analysis and design are always accomplished over a finite bandwidth. This allows modes outside of the control bandwidth to be neglected without impacting the responsiveness or the robustness of the closed loop system. As an example, for representative handling qualities it is believed that the low and the high order system models must have sufficiently similar frequency responses between approximately 1 to 10 rps. This means that, in the case of such a fit, that the pilot will judge the dynamics to be equivalent between the high order and the reduced order system models.

This procedure uses a conjugate gradient search routine to adjust the parameters of transfer functions including pure time delays with the desired form until their frequency responses are as close as possible to the frequency responses from the high order dynamics model. This is a multi-variable approach whereby several output frequency responses to the same input can be analyzed simultaneously. The

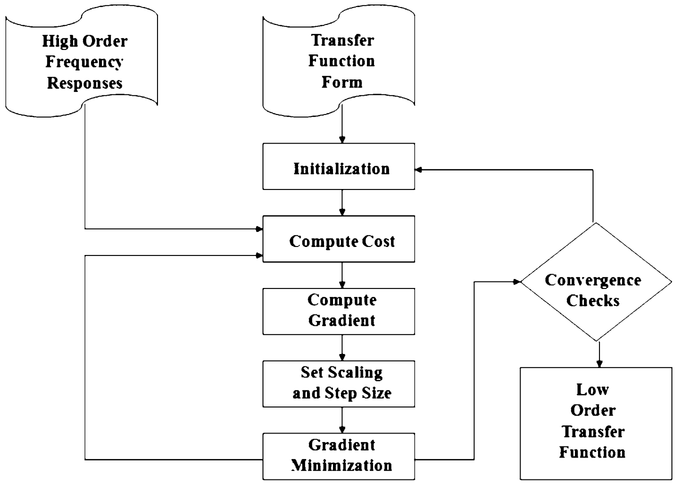

Fig. 3 Simultaneous gradient error reduction

user is allowed to fix, free, or simultaneously fit the various modes, including time delays, after selecting the order of the transfer function [5] This procedure is outlined in Fig. 3.

The desired result is to match a transfer function to each of the given plant's magnitude $A(\omega)$ and phase frequency response $\Theta\omega$ data over the control bandwidth. Expressed in the form of a complex number $A(\omega)e^{j\Theta(\omega)}$, these plant transfer functions should respond in an equivalent manner to the full order plant model over the desired frequency range for the given pair of inputs and outputs. The reduced order transfer function representation is given in factored form and incorporates a pure time delay. All of the output responses to a single large-scale system input can be simultaneously fit using this approach. The output response models to the other large-scale system's inputs are generated in subsequent analysis.

The cost function J represents the fit error between the plant data and the reduced order model. It is formulated as the integral of the absolute value of the error squared between the data and the reduced order model as shown in (12)

$$J = \int_0^\infty |G(j\omega) - A(\omega)e^{j\Theta(\omega)}|^2 d\omega \tag{12}$$

The fit error is weighted over each infinitesimal frequency interval. This fit error can be assigned a relative importance as a function of frequency. A conjugate gradient routine is used for error reduction because of its simplicity. Parameters are constrained to maintain the same sign during minimization. This is done to maintain stability characteristics according to the Nyquist Stability Criteria. The gradient ΔJ can be expressed analytically by differentiating the cost function given in (13).

$$\Delta J = Re\sum[G(j\omega_i) - A(\omega_i))e^{j\Theta\omega_i}]\Delta G(j\omega)d\omega \tag{13}$$

$G(j\omega)$ is defined in (14). In this equation $Z(j\omega)$ is composed of the partial derivatives of $G(j\omega)$ divided by their corresponding factors.

$$\Delta G(j\omega) = G(j\omega)Z(j\omega) \tag{14}$$

Gradient search techniques generally require parameter scaling to efficiently converge. An ideal scaling uses a cost function equally sensitive to each parameter. The law to accomplish this scales each parameter by its own magnitude as in (15). Re-scaling is done after completing each separate iteration.

$$P_i = b_i/|b_i| \tag{15}$$

The elements in the scaled gradient J are formed using (16).

$$\delta\, J/\delta\, p_i = \delta\, b_i/\delta\, p_i * \delta\, J/\delta\, b_i = |b_i| * \delta\, J/\delta\, b_i \tag{16}$$

Simultaneously fit parameters are constrained within bounds derived by initially allowing each transfer function to converge to an independent solution. An unweighted average is then computed, and is recomputed after each search. This is continued until all reduced order transfer functions have converged, the simultaneously fit parameters have converged, or the maximum number of iterations have been exceeded. In studies it was shown that the use of weighted averages did not improve the cost function of the final result. Its use also created convergence stability problems. If the problem as formulated does not have a true joint minimum, the gradient search routine will give the best estimated joint minimum.

4 Balancing

This method offers an efficient way to generate a reduced order model of a large-scale system whether all of its subsystems and their inputs and outputs are decoupled or not. This method for reducing the order of large-scale system models was originally developed by Enns while he was at Stanford [6]. An internally balanced system representation has input and output grammians that are equal and diagonal [7]. The magnitude of each diagonal element provides a measure of the controllability and observability of the corresponding state [8]. This is used as a guide in selecting those states which are the least controllable or observable for elimination [9]. The sum of the diagonal elements of the grammian corresponding to the states eliminated provides an error bound between the high order representation and the reduced order model.

Balancing relies on the notion of a system as a mapping from the inputs to the servos or actuators and then to the sensor outputs [10]. This mapping is viewed as a combination of the reachability mapping from the actuator input signals to the state vector and an observability mapping from the state to the sensor output signals. It follows from this mapping concept that the state is an intermediate quantity between the inputs in the past and the sensor outputs in the future. The minimum amount of control energy required to reach the desired state vector is inversely proportional to the reachability grammian. Similarly, the amount of sensor output energy generated by the state vector is proportional to the observability grammian. Balancing examines the amount each state component participates in the mapping from the input to the output. The model reduction problem is now reduced into one of truncating small terms from the partial fraction decomposition.

The following section pertains to a system again formulated as in (1), where A and B are controllable and A and C are observable. Given these assumptions, the controllability grammian U and the observability grammian Y are the solutions to the Lyapunov equations given as in (17) and (18)

$$d\underline{x}/dt = \underline{A}x + \underline{B}u = TAT^{-1}x + T^{-1}Bu \tag{17}$$

$$y = \underline{Cx} = CTx \tag{18}$$

The transformation matrix T also relates the grammians through equations (19) and (20) using an algorithm first developed by Laub [11, 12]. These contragedient transformations (where both U and Y are diagonal) can be calculated either as the best conditioned contragedient transformation or as an internally balanced transformation.

$$\underline{U} = T^{-1}UT^{-1} \tag{19}$$

$$\underline{Y} = T'YT \tag{20}$$

For an internally balanced transformation in (21) applies. In this representation the columns of T are the eigenvectors of the product UY.

$$UY = T^{-1}\Lambda\ T \tag{21}$$

To compute the balanced representation of the original large-scale system model, it first must be decoupled into stable and unstable subsystem models. Balancing is used for this step to generate a transformation based on the system's eigenvalues. For simplicity, neutrally stable modes can be slightly perturbed to make them marginally stable. If the conditionality of the resulting stable projection is a concern, all neutrally stable modes can also be preserved as a part of the unstable system. This can lead to a higher order final reduced system model. The resulting system is given in (22). In this equation $G_+(s)$ is the stable subsystem and $G_-(s)$ is the unstable subsystem

$$G(s) = C(sI - A)^{-1}B = G_+(s) + G_-(s) \tag{22}$$

Without input or output weighting the stable subsystem is transformed into a real Schur form. Next the Cholesky factors L_u and L_y of U and Y are computed. A singular value decomposition of the product L_yL_u is then performed. The singular values and the corresponding vectors are arranged in order of decreasing singular values. The transformation T is then computed in a final Schur transformation.

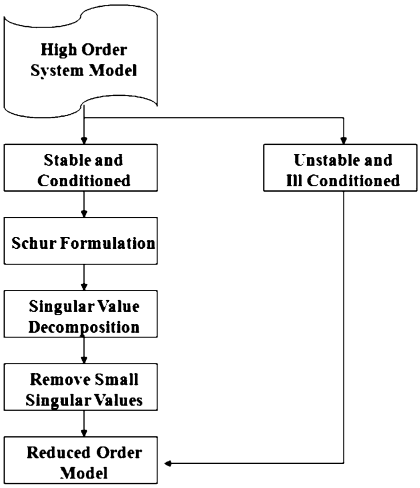

Fig. 4 Balancing process

States which have relatively small singular values and thus have a relatively small effect on the response of the large-scale system are then truncated from the model of the stable subsystem. The unstable and ill conditioned subsystem is completely preserved throughout this process. It is next recombined with the reduced order representation of the stable subsystem. This process is shown in Fig. 4.

4.1 Techniques Not Requiring Balancing

Other model reduction methods with an equivalent range of applicability have been designed to avoid the balancing process. One such method uses the Hankel Minimum Degree Approximate algorithm as described in [12]. This method is similar to balanced additive model reduction routines but can produce a reduced order model more reliably when the desired reduced model has nearly controllable and/or observable states [13]. These conditions are equivalent to having Hankel singular values very close to the machine accuracy.

For a stable system the Hankel singular values indicate the relative energy of each state with respect to the state energy of the entire system [14]. The reduced order system is directly determined by examining the system's Hankel singular values. An optimal reduced order system is selected using this algorithm to satisfy the error bound criterion regardless of the order selected at the beginning of the process [12]. Given the state space representation of the system as shown in (23), along with the value selected for k (the desired reduced order), the reduced order model is generated.

$$dx/dt = Ax + Bu; y = Cx \tag{23}$$

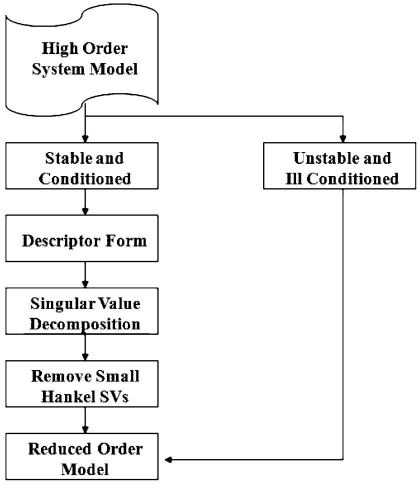

Fig. 5 Hankel minimum degree approximate process

The following steps, as shown in Fig. 5, produce a similarity transformation to truncate the original state space system into the k^{th} order reduced model. Weights on the original model's inputs and/or its outputs can make the model reduction algorithm focus on the specific frequency range of interest. These weights are required to be stable, minimum phase, and invertible. Note that as in the previous section the unstable subsystem must be combined with the reduced order stable subsystem to create the final reduced order model.

In more detail the steps to accomplish model reduction without balancing using the Hankel Minimum Degree Approximate algorithm are given as follows. First, the original large-scale system model must be decoupled into stable and unstable subsystem models. Next, the controllability and observability grammians P and Q must be generated. Using these grammians the descriptor given in (24) is generated

$$E = QP(j\omega) - \rho^{2} I \tag{24}$$

where $\sigma_k > \rho \geq \sigma_{k+1}$.

A singular value decomposition is then accomplished on the descriptor. The system is next transformed to generate the system representation given in (25) and (26).

$$dx\ dt = U'(\rho^2 A' + QAP)Vx + U'(QB - C')u \tag{25}$$

$$y = [CP - \rho B']Vu \tag{26}$$

The resulting system is then partitioned and truncated to become a kth order system. The final kth order Hankel Minimum Degree Approximate is the stable part of the state space realization. Its unstable part must be recombined with the reduced order model of the stable part.

The nature of the resulting error between the original system $G(j\omega)$ and the final reduced order model is discussed in [14]. This error is described by an all-pass function. A detailed description of the Hankel Minimum Degree Approximate algorithm can be found in [13].

4.2 Balancing Over A Disk

This procedure is a further development of the balancing technique. It was originally developed by Jonckheere at the University of Southern California [15, 16]. Its advantages are that it gives precise frequency response bounds over a desired bandwidth. Elimination of eigenvalues by inspection makes the balancing better conditioned and computationally efficient as well as further decreasing the final size of the reduced order system.

As was done using the previous two approaches, the unstable subsystem is removed and only the stable subsystem is processed and reduced. Note that with proper sign changes the unstable subsystem could be balanced over a disk. Because of the Nyquist Stability Criteria the number of open loop poles should not be reduced. In actual application work the number of unstable poles represents a very small percentage of the total number of poles. Thus reducing the number of unstable poles would have little effect on the dimension of the final reduced order system.

Standard balancing has infinite bandwidth. It thus does not necessarily provide the smallest possible H_∞ error. What is desired is the smallest frequency response error over the selected bandwidth. An infinite bandwidth is not required. From the Hankel Singular Values for each pole the error bound is generated as in (27).

$$sup|G(s) - \underline{G}(s)| = \Sigma\delta \tag{27}$$

To develop a reduced order model optimized over a finite bandwidth and to improve the conditionality of the stable subsystem to be balanced, the subsystem is balanced over a disk. This is accomplished over a region as shown in Fig. 6. The disk is placed in the complex plane based on two considerations. First, the full order model $G(s)$ must be analytic in the disk [10, 17]. For the targeted error bound the disk must be placed to exclude any eigenvalues of $G(j\omega)$. Second, the disk should cover the interval of real frequencies $(-\Omega < \omega < \Omega)$. This bandwidth limit Ω is usually based on the response limitations imposed by the control servos or actuators. This guarantees that the error bound only includes this critical frequency range. A good rule is to use $\alpha = \frac{1}{2}$ (the largest real part of the poles of $G(j\omega)$). This is done using a bilinear mapping as shown in Fig. 7. The error bound becomes that given in (28)

$$sup|G(s) - \underline{G}_{DISK(s)}| = \Sigma\delta_{DISK} \tag{28}$$

Note that $\delta_{DISK} \leq \delta$. Those eigenvalues that have small singular values are deleted from the stable subsystem. The reduced order stable subsystem is recombined with the unstable subsystem. The overall process for reducing the order of a large-scale system over a disk is presented in Fig. 8.

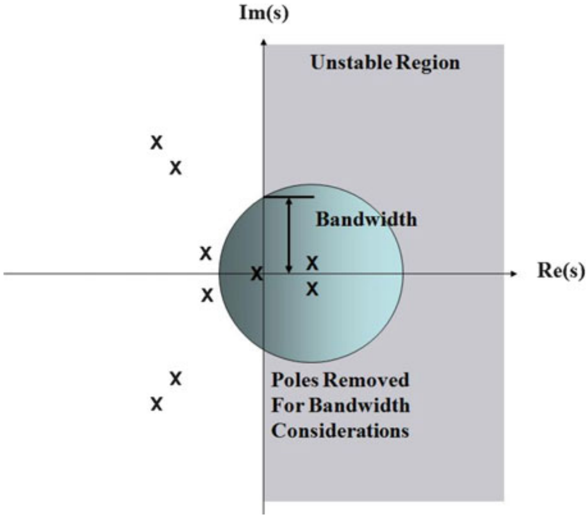

Fig. 6 Balancing over a Disk

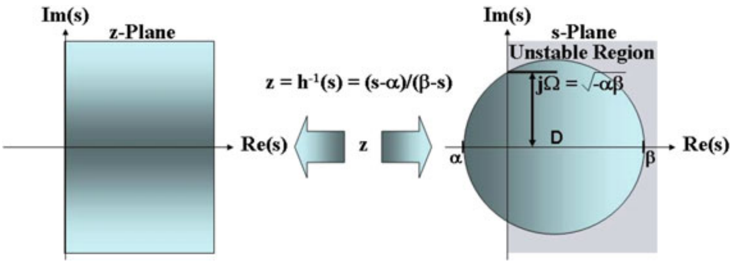

Fig. 7 Bilinear mapping

5 Example: Large-Scale System Application

The model reduction techniques discussed in the previous sections are based on minimizing the error between the high order and the low order models as measured within the frequency domain. Using these approaches, a large-scale 140th order multiple input, multiple output full dynamics model of the U-2S including rigid body and structural dynamics modes was reduced in order [4,18]. Within this section 90th, 80th and 40th order reduced models are documented. The response of the 90th order model is, by visual inspection, identical to the 140th order large-scale system's response. The response of the 80th order model begins to show differences from that of the large-scale system. The response shown in Fig. 9 is that of the roll rate due to a rudder input. This difference results from the migration of two extremely high frequency, lightly damped zeros from the left half into the right half of the complex plane. The large-scale system has ten non-minimum phase zeros while the 80th order model has 12 non-minimum phase zeros. Although the frequency response

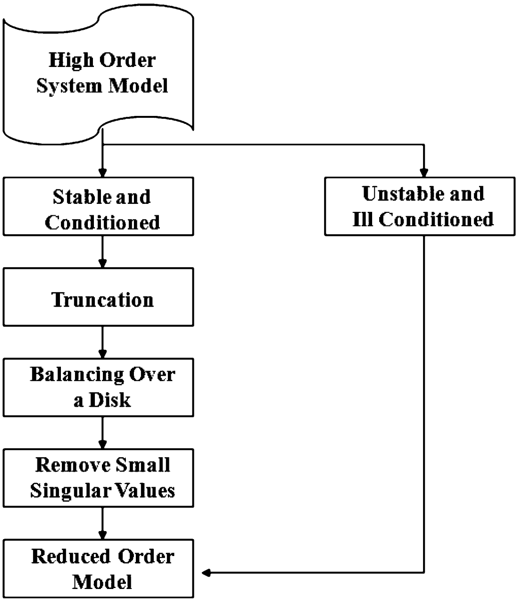

Fig. 8 Asymptotic Balancing Process

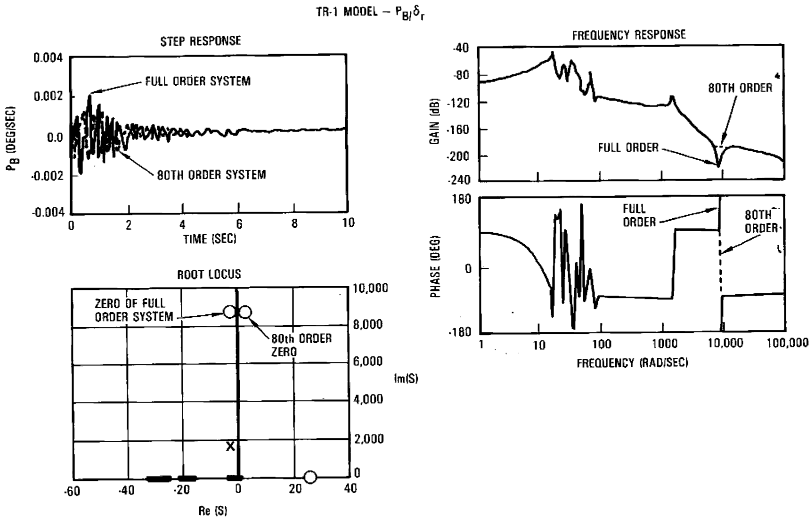

Fig. 9 Effect of non-minimum phase zeros

match shows very little variation between these two model's responses up to the frequency of this complex zero, the difference between the transient responses must be noted. In practice this difference is so far outside the control bandwidth of the servos that it will not have an impact on the design of the control laws.

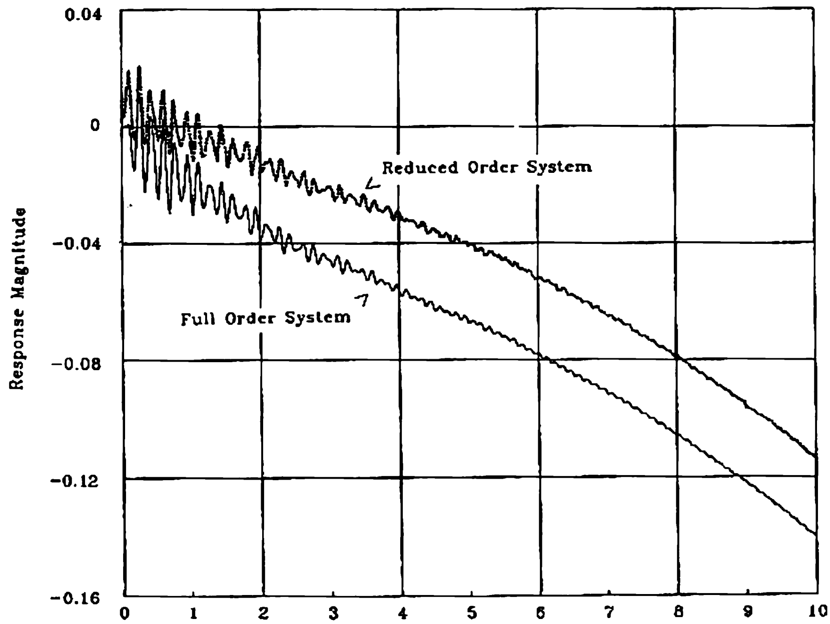

Fig. 10 Effect of Residualization

When the model is further reduced to 40th order a steady state offset error appears in the pitch rate response due to an elevator input. This offset is seen in Fig. 10. Residualization is easily used to shift the low order response to nearly overlay the 140^{th} order large-scale system's equivalent response. Only minor differences in the peaks of the first 5 oscillations are seen after residualization. Besides this minor error the residualized 40th order system's transient response overlays that of the 140th order large-scale system. This residualized system's time response is omitted from Fig. 10 for clarity.

References

1. Davis W. J. and Chiodi O. A. A method for the decoupling of equations by spectral decomposition. *Lockheed Report LR 27479*, Dec 1975
2. Socolinsky D. A. Wolff L. B., Neuheisel J. D., and Eveland C. K. Background information and user guide for mil-f-8785c, military specification, flying qualities of piloted airplanes. *AFWAL-TR-81-3109*, Sept 1981
3. Seidel R. C. Transfer-function-parameter estimation from frequency response data - a fortran program. *NASA TM X-3286*, Sept 1975
4. Colgren R. D. Methods for model reduction. *AIAA-88-4144*, pages 15–17, Aug 1998
5. Colgren R. D. *Simultaneous Fit Equivalent Systems Program – User's Guide*. Lockheed California Company, Burbank, CA, 1985
6. Enns D. *Model Reduction for Control System Design*. PhD thesis, Department of Aeronautics and Astronautics Stanford University, 1984

7. Laub A. J., Heath M. T., Paige C. C., and Ward R. C. Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms. *IEEE Transactions on Automatic Control*, Feb 1987
8. Kailath T. *Linear Systems*. Prentice-Hall, Upper Saddle River, 1980
9. Safonov M. G. and Chiang R. Y. A schur method for balanced model reduction. *IEEE Trans. on Automat. Contr.*, 34(7):729–733, july 1989
10. Colgren R. D. and Jonckheere E. A. Balanced model reduction for large aircraft models. *MTNS91*, June 1991
11. Anderson E. Bai Z., Bischof C., Blackford S., Demmel J., Dongarra J., Du Croz J., Greenbaum A., Hammarling S., McKenney A., and Sorensen D. Lapack User's Guide, 1999
12. Balas G., Chiang R., Packard A., and Safonov M. Robust control toolbox 3 user's guide. *Revision for Version 3.3.3 (Release 2009a), Natick, MA, March 2009.*, March 2009
13. Safonov M. G., Chiang R. Y., and Limebeer D. J. N. Optimal hankel model reduction for nonminimal systems. *IEEE Transactions on Automatic and Control*, 35(4):496–502, April 1990
14. Glover K. All optimal hankel norm approximation of linear multivariable systems, and their l infinity – error bounds. *International Journal of Control*, 39(6):1145–1193, 1984
15. Jonckheere, E. A. and Silverman, L. M. A new set of invariants for linear systems – application to reduced order controller design. *IEEE Transactions on Automatic Control*, AC-28:953–964, 1983
16. Jonckheere, E. A. and Silverman, L. M. A new set of invariants for linear systems – application and approximation. In *International Symposium on Theory, Networks and Systems*, Santa Monica, CA, 1981
17. Colgren and R. D. Finite bandwidth model reduction applied to the advanced supersonic transport. *COMCON3, Victoria, BC, Canada*, Oct 1991
18. Dudginski R. J., and Colgren R. D. Time domain effects of model order reduction. In *3rd IEEE International Symposium on Intelligent Control*, pp. 24–26, Arlington, VA, Aug 1988

http://www.springer.com/978-1-4419-5756-6

Efficient Modeling and Control of Large-Scale Systems
Mohammadpour, J.; Grigoriadis, K.M. (Eds.)
2010, XX, 335 p., Hardcover
ISBN: 978-1-4419-5756-6