Multivariable Frequency-Weighted Order Reduction

Barton J. Bacon*
Purdue University, West Lafayette, Indiana
and
David K. Schmidt†
Arizona State University, Tempe, Arizona

Model or controller order reduction for closed-loop analysis and synthesis is considered. An algebraic framework utilizing projection operators is discussed, two reduction methods originally motivated in the time domain are examined, and their ties to the frequency domain explored. Causes for failure of these techniques to provide stable reduced-order compensators are established. Frequency-weighted techniques are introduced, and guidelines for weighting selection discussed. Properties of the frequency-weighted results are noted, and an example clearly illustrates the effects of frequency weighting.

I. Introduction

IN this paper, order reduction specifically tailored to dynamic components within the closed-loop system is considered. The search for lower-order representations of these components concerns both the control system designer and the dynamic system analyst. Whether the objective is to simplify implementation with a lower-order representation of the controller, or to simplify closed-loop system analysis with more tractable lower-order forms of the components,[1] it is hypothesized that both problems may be approached in the same manner.

In closed-loop system applications, the order-reduction technique chosen should have the following properties. Stability margins should be maintained by the reduced-order model, the technique should be reliable or it should exhibit no extensive numerical searches and involve no questions of existence, and it should provide some a priori clue to the component's effective order. The technique should also approximate component characteristics known to be critical, based on well-established synthesis and analysis tools. For example, a multivariable (multi-input/multi-output) (MIMO) order-reduction technique should certainly work in simple single-input/single-output (SISO) cases, for which the fundamentals of feedback were established in the frequency domain some 40 years ago by Bode.[2,3]

Fortunately, these classical design guidelines are not restricted to SISO systems, but have been extended to MIMO systems by the singular-value methods,[4] and multivariable loop shaping techniques.[5,6] To the designer, these tools provide rather specific guidelines on what type of approximation is admissible in closed-loop system applications. Namely, a good approximation to the frequency-response of the full-order system is required. Furthermore, the required accuracy of the approximation varies with frequency as a function of the *nominal* closed-loop system design. Hence, the design and the reduction steps are inexorably intertwined.

In this paper, these guidelines are discussed. In response to these guidelines, two model reduction techniques, originally formulated in the time domain, are re-examined here in the frequency domain. The two model-reduction techniques considered are the internally balanced approach[7] and the optimal aggregate approach.[8] These techniques are of particular interest, since each may be associated with a metric (norm) in the frequency domain. Due to the inherit properties of the resulting approximations, however, these techniques do not necessarily lead to desirable results in closed-loop system applications.

Frequency weightings in the above procedures will then be introduced, leading to better approximations in critical frequency ranges. The effect of these weightings on the solutions, the algorithms, and the properties of the resulting reduced-order model for both techniques shall be discussed and compared.

II. Important Closed-Loop System Characteristics

The standard feedback configuration to be studied is illustrated in Fig. 1. It consists of the interconnected plant matrix $G(s)$ and controller matrix $K(s)$, with excitations from commands y_c, measurement noise n, and disturbances d. All disturbances are assumed to be reflected to the measured outputs y, all signals are multivariable, in general, and both nominal mathematical models of $G(s)$ and $K(s)$ are represented by finite-dimensional, linear, time-invariant, transfer-function matrices. The configuration, if it is stable, has the following major properties:

1) Input-output behavior:

$$y = GK(I+GK)^{-1}(y_c - n) + (I+GK)^{-1}d \tag{1}$$

$$e = y_c - y$$

$$= (I+GK)^{-1}(y_c - d) + GK(I+GK)^{-1}n \tag{2}$$

2) Closed-loop transfer-matrix sensitivity to perturbations in the loop:

$$\Delta H_{cl} = [I + (GK)']^{-1}\Delta H_{ol} \tag{3}$$

In Eq. (3), ΔH_{cl} and ΔH_{ol} denote perturbations in the closed-loop system and in the open-loop system, respectively, caused by changes in the forward loop GK, where $H_{cl} = GK(I+GK)^{-1}$, $GK' = GK + \Delta(GK)$, and $\Delta H_{ol} = \Delta(GK)$.

The transfer functions described in Eqs. (1) and (2) determine how closely y tracks y_c, how well the system rejects disturbances, and how the system responds to sensor noise. Changes in performance due to changes in the loop-transfer function GK may be determined from the frequency response of the perturbed (matrix) inverse return difference $\{I + [GK(j\omega)]'\}^{-1}$, where it is

Received Dec. 30, 1986; revision received July 20, 1987.

*Graduate Research Assistant, School of Aeronautics and Astronautics; currently at NASA Langley Research Center. Student Member AIAA.

†Professor, Department of Mechanical and Aerospace Engineering. Associate Fellow AIAA.

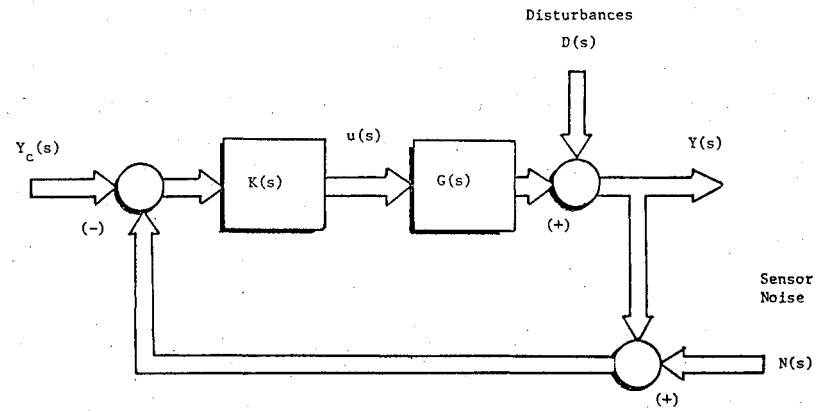

Fig. 1 The closed-loop system.

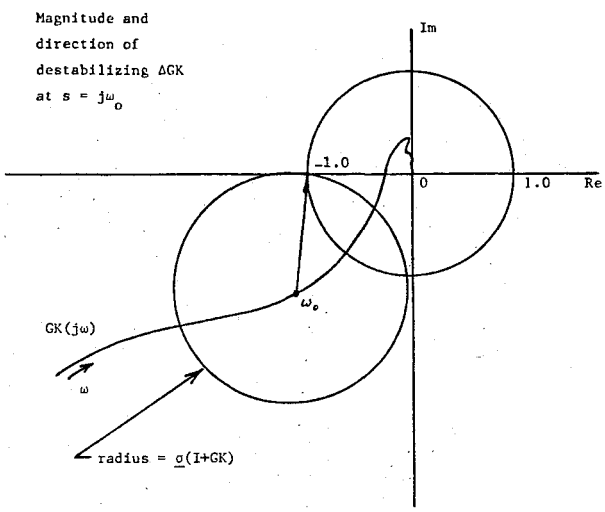

Fig. 2 Nyquist diagram.

useful to note that

$$(GK)'[I + GK']^{-1} \approx GK(I + GK)^{-1} + \Delta H_{cl} \tag{4}$$

for incremental changes in $\Delta(GK)$. Clearly, the frequency response of both G and K play a significant role in characterizing the closed-loop system.

Of paramount importance to both controller reduction and closed-loop system analysis is the requirement that the order-reduction method preserve the existing nominal stabilty robustness of the closed-looped system, where stability robustness here refers to the system's ability to retain stability in the face of loop uncertainties $\Delta(GK)$. The genesis of this uncertainty could be due to parameter variations, or unmodeled dynamics, in either G and K or both. Of course, it could arise from (lower-order) approximations to $G(s)$ or $K(s)$. It should be clear that merely obtaining closed-loop stability is insufficient.

Stability of the perturbed system is determined by the roots of the characteristic equation, the zeros of

$$\det[I + GK + \Delta(GK)] = 0 \tag{5}$$

Assuming the nominal closed-loop system and $\Delta(GK)$ are stable, the perturbed system is stable if

$$\det[I + GK + \varepsilon\Delta(GK)]|_{s=j\omega} \neq 0 \tag{6}$$

for all ω and all ε in the closed interval [0,1] (Ref. 4). Another way to express this is to require that $[I + GK(j\omega)] + \varepsilon\Delta[GK(j\omega)]$ remain nonsingular for all ω and all ε, $0 \leq \varepsilon \leq 1$.

The perturbed system is guaranteed to be stable[9] whenever

$$\underline{\sigma}[I + GK(j\omega)] > \bar{\sigma}[\Delta GK(j\omega)] \tag{7}$$

is satisfied for all ω; $\underline{\sigma}$ and $\bar{\sigma}$ denote the minimum and maximum singular values, respectively. Therefore, the key frequency ranges where stability robustness is potentially a problem is where

$$\underline{\sigma}[I + GK(j\omega)] \approx \bar{\sigma}\{\Delta[GK(j\omega)]\} \tag{8}$$

over the (physically) possible $\Delta(GK)$.

For SISO systems, Eq. (8) is easily interpreted on a Nyquist diagram (see Fig. 2). In this case, the left-hand side of Eq. (8) corresponds to the distance between a point on the Nyquist locus $GK(s)$ at $s = j\omega_o$, and the critical point $s = -1$. From Fig. 2 the following statements are apparent:

1) For any $\omega = \omega_o$, there exists a $\Delta GK(j\omega_o)$ satisfying Eq. (8) that drives $GK'(j\omega_o)$ to -1, i.e., closed-loop instability.

2) Not all $\Delta GK(j\omega_o)$ satisfying Eq. (8) necessarily make $GK'(j\omega) = -1$.

Analogous statements can be made regarding MIMO systems. Consequently, Eq. (8) defines frequency ranges where stability robustness is potentially a problem, and certain frequency ranges are more critical than others.

The frequency ranges determined by Eq. (8) include the crossover frequency [i.e., the frequency range where $\underline{\sigma}(GK) \approx 1$, $\bar{\sigma}(GK) \approx 1$] where relatively small variations in the loop GK can sometimes be destabilizing. Moreover, it includes frequencies where small changes in G or K can create large $\Delta(GK)$ satisfying Eq. (8) (e.g., systems with near pole-zero cancellations within GK in the vicinity of the $j\omega$ axis).

Although stability robustness is the primary factor that emphasizes a certain frequency range, recall the classical notion of loop shaping in control synthesis. It is desired that the plant-plus-controller open-loop transfer function gain be large at low frequency, to attenuate disturbances and desensitize the closed-loop system to open-loop parameter variations. Small-loop gains are desired, however, at high frequency, to limit the effect of sensor noise and unmodeled high-frequency dynamics. Moreover, the transition between the high and low gain must be done in a stable, robust manner [i.e., such that Eq. (7) is satisfied].[4]

Introduction of any reduced-order (i.e., approximate) compensator (or plant) into the loop alters the loop shape from the "true" loop shape. *If* the loop shape described above is still achieved, however, deviations from the true loop may occur in the high- and low-frequency ranges, and not significantly affect the results of the design. At low frequencies, the compensator must simply provide adequate loop gain, since this desensitized the closed-loop system to model (and controller) errors occurring there. Also, in the high-frequency range the loop gain must be adequately attenuated to allow for modeling uncertainty. *Consequently, an extremely accurate approximation of the full-order system is not required in either the low- or high-frequency*

range as long as the two above criteria are met. This leaves the midfrequency range or crossover region as the critical region that should be *accurately* modeled by any reduced-order approximation.

III. Projections and Model Reduction

Therefore, given a higher-order, linear system (model) $G(s)$, the task is to find a lower-order model $G_r(s)$ that well approximates $G(s)$ for $s = j\omega$ over a prespecified frequency range $(\omega_1 < \omega < \omega_2)$. An algebraic framework for this problem, in terms of the state-space of the higher-order system, is now discussed. Although somewhat of a digression, this state-space interpretation shall be utilized in the later discussion of the frequency-weighting techniques.

Consider the following linear, time invariant system:

$$\dot{x} = Ax + Bu \qquad x \in R^n,\ u \in R^m \tag{9a}$$

$$y = Cx \qquad y \in R^p \tag{9b}$$

$$G(s) = C(sI - A)^{-1}B \tag{9c}$$

A common underlying structure of most state-space model-reduction procedures that shall be invoked here involves the decomposition of the original state-space into two components. It is desirable that one component dominates the important characteristics of the system to be modeled, whereas the other is less significant. (Hence, some realization of what is important is necessary at the outset.)

Decomposition is uniquely determined by the choice of two disjoint subspaces $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ of dimension r and n-r, respectively, whose direct sum spans the state-space $R^n = \boldsymbol{T}_1 \oplus \boldsymbol{T}_2$, i.e., for every $x(s) \in R^n$, $x(s)$ can be uniquely decomposed as $x(s) = x_1(s) + x_2(s)$, where $x_1(s) \in \boldsymbol{T}_1$ and $x_2(s) \in \boldsymbol{T}_2$. Under the problem stated, $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ must be chosen such that $x_1(s)$ more significantly influences the frequency response of the system to be approximated, $G(j\omega)$, over (ω_1,ω_2). An rth-order approximation of $x_1(s)$ then defines the state-space of the rth order model.

The decomposition of the state-space, as well as the construction of the reduced-order model, is more easily visualized by performing a state transformation $x(s) = Tz(s)$ on the given realization of $G(s)$. If $T = [T_1 T_2]$ and $T^{-T} = [U_1 U_2]$ where the columns of T_1 and T_2 are any arbitrary span of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$, respectively, then the transformed system

$$z(s) = [sI - \hat{A}]^{-1}\hat{B}u(s) \tag{10a}$$

$$y(s) = \hat{C}z(s) \tag{10b}$$

reflects such a decomposition of the state-space in that $x_1(s) = T_1 z_1(s)$ and $x_2(s) = T_2 z_2(s)$ where

$$\hat{A} = \begin{bmatrix} U_1^T A T_1 & U_1^T A T_2 \\ U_2^T A T_2 & U_2^T A T_2 \end{bmatrix} \qquad \hat{B} = \begin{bmatrix} U_1^T B \\ U_2^T B \end{bmatrix}$$

$$\hat{C} = [CT_1 | CT_2]$$

and $z(s) = [z_1^T(s) z_2^T(s)]^T$. The reduced-order model

$$z_r(s) = (sI_r - A_r)^{-1}B_r u(s) \tag{11a}$$

$$\hat{y}(s) = C_r z_r(s) \tag{11b}$$

where $A_r = U_1^T A T_1$, $B_r = U_1^T B$, and $C_r = CT_1$ provides an approximation of $x_1(s)$, $T_1 z_r(s)$. Given a state-space realization of the original system $G(s)$, the reduced-order model (transfer function) and the decomposition $[x_1(s)$ and $x_2(s)]$ are both uniquely determined by choice of subspaces $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$.

Unfortunately, the columns of T_1 and T_2 spanning $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$, respectively, are not uniquely determined. Any T_1 and T_2 with column spaces equal to $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$, respectively, produce the same reduced-order model (transfer function). To overcome that arbitrariness in representing $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$, and in defining the reduced-order model, a square matrix uniquely defined by both $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ is used to define the reduced-order model, given the state-space realization of $G(s)$. This square matrix is the projection matrix P_{12} (Ref. 11) taken on $\boldsymbol{T}_1$ along $\boldsymbol{T}_2$ and equals $P_{12} = T_1 U_1^T$ (note $U_1^T T_1 = I_r$). Assuming $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are defined as above, the projection on $\boldsymbol{T}_1$ along $\boldsymbol{T}_2$ is the linear transformation P_{12} defined by $x_1(s) = P_{12}x(s)$.

Note that P_{12} uniquely determines $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ [i.e., $R(P_{12}) = \boldsymbol{T}_1$ and $\eta(P_{12}) = \boldsymbol{T}_2$, where $R(.)$ and $\eta(.)$ denote the range and null space of $(.)$, respectively], and $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ uniquely determine P_{12} (Refs. 12 and 13); P_{12} also uniquely determines $G_r(s)$, given the state-space of the original system. Any other factorization of P_{12}, say $V_1 S_1^T$, leads to a model $(S_1^T A V_1,\ S_1^T B,\ CV_1)$ that is related to the above model by an $r \times r$ similarity transformation.

Various criteria have been used to define $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ yielding, of course, to the various reduction approaches. Two such approaches will now be highlighted in this context.

IV. Two Model-Reduction Methods

Two model-reduction techniques originally formulated in the time domain are re-examined here in the frequency domain. These reduction procedures are the internally balanced approach and the optimal aggregate approach, with similar algebraic structures, but motivated by two different metrics (i.e., norms) on the frequency-response error matrix $E(j\omega) = G(j\omega) - G_r(j\omega)$.

The internally balanced approach was shown by Enns[10] to be associated with the ∞ norm

$$\|E(j\omega)\|_\infty \triangleq \sup_\omega \bar{\sigma}[E(j\omega)] \tag{12}$$

that bounds the magnitude of the i-j elements of $E(j\omega)$ for all ω. The optimal aggregate approach, however, involves the 2-norm

$$\|E(j\omega)\|_2^2 \triangleq \mathrm{Tr}\,\frac{1}{2\pi}\int_{-\infty}^{\infty} E(j\omega)E^T(-j\omega)\,d\omega \tag{13}$$

the square integral sum over all i-j elements of $E(j\omega)$.

Internally Balanced Approach

The principal component analysis that originally motivated the internally balanced realization[7] is to be re-examined. This approach focuses on the identification of n principal state directions t_i, $= 1,...,n$ used to define the subspaces $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ cited previously, which in turn define the internally balanced reduced-order model. The frequency-domain properties and implications of these directions are considered first.

In general, "state directions" refer to any set of basis vectors $(t_i, i = 1,...,n)$ that span R^n, where any state vector $x \in R^n$ can be uniquely expressed as

$$x = \sum_{i=1}^{n} t_i(u_i^T x) \tag{14}$$

the sum of n components (vectors) each directed along t_i. Here, $u_i^T, i = 1,...,n$, denotes the corresponding reciprocal basis.[13] Principal directions are then a particular set of basis vectors. The system eigenvectors, for example, would be another.

What is required is the means to ascertain the relative importance of individual state directions with respect to both the system's input and output behavior. This task is clearly relevant in the selection of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$, because it is the system's input-output behavior (input and output considered together) that determines a system's transfer function and frequency response.

Two frequency-dependent matrices of interest are

$$\bar{X}(j\omega) = (j\omega I - A)^{-1}B \tag{15a}$$

$$\bar{Y}(j\omega) = C(j\omega I - A)^{-1} \tag{15b}$$

$\bar{X}(j\omega)$ reflects the system's input behavior, since each column is the state vector frequency response associated with the related input, whereas $\bar{Y}(j\omega)$ reflects the system's output behavior. The controllability and observability grammians

$$X = \frac{1}{2\pi}\int_{-\infty}^{\infty} \bar{X}(j\omega)\bar{X}^T(-j\omega)\,d\omega \tag{16a}$$

$$Y = \frac{1}{2\pi}\int_{-\infty}^{\infty} \bar{Y}^T(-j\omega)\bar{Y}(j\omega)\,d\omega \tag{16b}$$

are related to $\bar{X}(j\omega)$ and $\bar{Y}(j\omega)$ as shown. Note finally that $\bar{X}(j\omega)$ and $\bar{Y}(j\omega)$ are ultimately related to the system's frequency response $G(j\omega) = C\bar{X}(j\omega) = \bar{Y}(j\omega)B$.

Consider the decomposition of $\bar{X}(j\omega)$ and $\bar{Y}(j\omega)$, written with respect to first an *arbitrary* basis (state directions) t_i. This decomposition may be expressed in terms of the projection $P_i = t_i u_i^T$ where

$$I_n = \sum_{i=1}^{n} P_i = \sum_{i=1}^{n} t_i u_i^T \tag{17}$$

For $\bar{X}(j\omega)$, this implies

$$\bar{X}(j\omega) = (j\omega I - A)^{-1}B = \sum_{i=1}^{n} P_i(j\omega I - A)^{-1}B \tag{18a}$$

$$\bar{X}(j\omega) = t_1 x_1(j\omega) + \cdots + t_n x_n(j\omega) \tag{18b}$$

where $x_i(j\omega) = u_i^T\bar{X}(j\omega)$ for all $i = 1,n$. For $\bar{Y}(j\omega)$

$$\bar{Y}(j\omega) = C(j\omega I - A)^{-1} = \sum_{i=1}^{n} C(j\omega I - A)^{-1}P_i \tag{19a}$$

$$\bar{Y}(j\omega) = y_1(j\omega)u_1^T + \cdots + y_n(j\omega)u_n^T \tag{19b}$$

where $y_i(j\omega) = \bar{Y}(j\omega)t_i$ for all $i = 1,n$. Clearly, state directions t_i more significant to the system's input behavior generally yield larger components of $\bar{X}(j\omega)$, or $x_i(j\omega)$ and $t_i u_i^T\bar{X}(j\omega)$ are large. Conversely, state directions more significant to the system's output behavior yield larger components of $\bar{Y}(j\omega)$, or $y_i(j\omega)$ and $\bar{Y}(j\omega)t_i u_i^T$ are large. Consider now the *principal* state directions.

By definition, the principal directions decompose X and Y, the grammians related to $\bar{X}(j\omega)$ and $\bar{Y}(j\omega)$, respectively, into the following outer product sums:

$$X = \sum_{i=1}^{n} t_i v_{c_i} t_i^T \tag{20a}$$

$$Y = \sum_{i=1}^{n} u_i v_{o_i} u_i^T \tag{20b}$$

where v_{o_i} and v_{c_i} are nonnegative scalars, and where $u_i^T t_i = 1$ and $u_i^T t_j = 0$ for all $i \neq j$. For t_i to be a principal direction, the corresponding $x_i(j\omega)$ and $y_i(j\omega)$ defined in Eqs. (18) and (19) must be orthogonal, or

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} x_i(j\omega)x_j^T(-j\omega)\,d\omega = 0 = \frac{1}{2\pi}\int_{-\infty}^{\infty} y_i^T(-j\omega)y_j(j\omega)\,d\omega \tag{21}$$

for all $i \neq j$. The scalars v_{c_i} and v_{o_i} are related to $x_i(j\omega)$ and $y_i(j\omega)$ by

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} x_i(j\omega)x_i^T(-j\omega)\,d\omega = v_{c_i} \tag{22a}$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} y_i^T(-j\omega)y_i(j\omega)\,d\omega = v_{o_i} \tag{22b}$$

for $i = 1,n$.

Therefore, the "principal state" input component $x_i(j\omega)$ is that portion of $\bar{X}(j\omega)$ directed along t_i, whereas the principal state component $y_i(j\omega)$ is that portion of $\bar{Y}(j\omega)$ due to t_i, i.e., $\bar{Y}(j\omega)t_i$. Moore[7] postulated that the importance of t_i to the *input-output behavior* (e.g., frequency response) of the system to be approximated is reflected by the relative magnitude of the product $v_{c_i}v_{o_i}$, where

$$\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} |x_i(j\omega)|^2\,d\omega\right)\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} |y_i(j\omega)|^2\,d\omega\right) = v_{c_i}v_{o_i} = h_i^2 \tag{23}$$

The scalar h_i is the ith Hankel singular value of $G(s)$,[14] all of which are real, nonnegative, and invariant to state-space transformation. The matrix product

$$XY = \sum_{i=1}^{n} t_i v_{c_i} v_{o_i} u_i^T \tag{24}$$

obtained from Eqs. (20) shows that the corresponding principal directions t_i are eigenvectors of XY. These eigenvectors are uniquely defined (within scaling) if the eigenvalues of XY are distinct. The h_i^2 are these eigenvalues. The internally balanced algorithm is summarized in Table 1.

The reduced-order model may also be interpreted in terms of the subspaces $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$: $\boldsymbol{T}_1$ is spanned by the t_i corresponding to the r *largest* h_i, whereas $\boldsymbol{T}_2$ is spanned by the remaining n-r principal directions. If the h_i are assumed ordered such that $h_{i+1} \leq h_i$, then as long as $h_r > h_{r+1}$, the internally balanced model is uniquely defined.[7] The internally balanced method's projection matrix is

$$P_{12} = T_1 U_1^T = \sum_{i=1}^{r} t_i u_i^T = \sum_{i=1}^{r} P_i$$

where the P_i correspond to the directions associated with the r largest Hankel singular values.

Table 1 Internally balanced algorithm

Given: HOS state-space description A,B,C
Find: rth-order LOS

Step 1: Solve for X and Y

$$AX + XA^T + BB^T = 0$$

$$A^TY + YA + C^TC = 0$$

Step 2: Find T and Σ where $XY = T\Sigma^2 T^{-1}$, $T = [T_r, T_{n-r}], T^{-T} = [U_r, U_{n-r}]$

$$\overset{2}{\Sigma} = \begin{bmatrix} \overset{2}{\sum_r} & 0 \\ 0 & \overset{2}{\sum_{n-r}} \end{bmatrix} \quad \text{where } \sum_r = \text{diag}(h_i) \qquad i = 1,r$$

$$\sum_{n-r} = \text{diag}(h_i) \qquad i = r+1,n \text{ and}$$

$$h_1 \geq h_2 \geq \cdots \geq h_r > h_{r+1} \geq \cdots h_n \geq 0$$

Step 3: LOS is defined by

$$A_r = U_r^T A T_r, \qquad B_r = U_r^T B, \qquad \text{and } C_r = CT_r$$

To understand the implications of the principal state directions and the Hankel singular values in the frequency domain, consider what affects $\bar{X}(j\omega)$ and $\bar{Y}(j\omega)$, Eqs. (16). Each of the defining integrals are over *all frequencies*. Furthermore, the relative magnitudes of the Hankel singular values, which determine the t_i that spans T_1 and T_2, are determined by the magnitudes of $|x_i(j\omega)|^2$ and $|y_i(j\omega)|^2$ ($|y_i(j\omega)|^2 = y_i^T(-j\omega)y_i(j\omega)$), which can be large *at any frequencies*, and principal directions are important if the associated h_i is large. Intuitively, then, the internally balanced approach will tend to lead to good lower-order approximations in frequency ranges where $|x_i(j\omega)|$ or $|y_i(j\omega)|$ and, hence, the magnitude of $G(j\omega)$ is large.

This is consistent with the results of Enns,[10] that the rth-order internally balanced approximation $G_r(s)$ satisfies

$$\|G(j\omega) - G_r(j\omega)\|_\infty \le 2 \sum_{i=r+1}^{n} h_i \tag{25}$$

Here, the norm is defined by Eq. (14), and the h_i are the Hankel singular values truncated in the reduction. Minimizing this unweighted norm tends to lead to approximations that are biased toward frequency ranges where the magnitude of $G(j\omega)$ is large. These concepts will be demonstrated by example in the sequel.

Optimal Aggregation

The second method to be considered produces the model that tends to minimize the norm $\|G(j\omega) - G_r(j\omega)\|_2^2$, given the model is one of a class of reduced-order models called aggregates. The definition of an aggregate model, as well as the determination of the aggregate that minimizes $\|G(j\omega) - G_r(j\omega)\|_2^2$, are the two topics of this section.

An aggregate model is any reduced-order model obtained via the state transformation structure shown in Eqs. (10) and (11) such that $U_i^T A T_2^T = 0$. Therefore, the state trajectory of the aggregate model is a linear combination, an "aggregation" of the state trajectory[15] of the original, or

$$z_r(j\omega) = U_i^T x(j\omega) \tag{26}$$

for all ω. Premultiplying this by T_1 implies

$$T_1(j\omega I - A_r)^{-1}B_r = P_{12}(j\omega I - A)^{-1}B \tag{27}$$

for all ω. Here, A_r and B_r are defined by Eqs. (11), and $P_{12} = T_1 U_1^T$ denotes the "aggregate" projection. Note that the aggregate model's state-space *equals* a component of the original system's state-space $x_1(j\omega) = T_1 z_r(j\omega)$.

The projection P_{12} defines the reduced-order model, given a state-space realization of the system to be approximated. The "aggregate" projection is any projection matrix taken *along* an n-r dimensional subspace (T_2) invariant with respect to A, i.e., T_2 can be spanned by a subset of the eigenvectors of A. No restriction, however, is placed on T_1, other than $T_1 \oplus T_2 = R^n$. Consequently, the property of aggregation, Eq. (26), is tied only to the condition imposed on T_2.

Under Eq. (27) a closed-form expression for the frequency response error is readily available.

$$E(j\omega) = [G(j\omega) - G_r(j\omega)] = C(I - P_{12})(j\omega I - A)^{-1}B$$
$$= CP_{21}(j\omega I - A)^{-1}B \tag{28}$$

where P_{21} is the projection matrix on T_2 along T_1, i.e., $P_{21} = T_2 U_2^T$. This expression is the key to obtaining the optimal aggregate model.

The optimal rth-order aggregate model is the rth-order aggregate model $G_r^*(s)$ that minimizes the cost functional

$$J = \mathrm{Tr}\frac{1}{2\pi}\int_{-\infty}^{\infty} E(j\omega)E^T(-j\omega)\,d\omega \tag{29}$$

Table 2 Optimal aggregate algorithm (limited search)

Given: HOS state-space description A,B,C
Find: rth-order LOS
Step 1: Solve for X

$$AX + XA^T + BB^T = 0$$

Step 2: For each real eigenvector of A, v_i, define

$$w_i = v_i$$

For each complex conjugate pair of eigenvectors of A

$$v_i = v_{R_i} \pm j v_{I_i} \text{ define}$$

$$w_i = [v_{R_i} v_{I_i}]$$

Step 3: Over all combinations of w_i, let $T_{2_j} = [w_i,..w_j..w_k]$ such that dim $(T_{2_j}) = n$-r, and evaluate

$$J_j = \mathrm{Tr}[CT_{2_j}(T_{2_j}^T X^{-1} T_{2_j})^{-1} T_{2_j}^T C^T]$$

Step 4: For T_{2_j} that produces the minimum J_j define

$$P_{12} = I_n - T_2(T_2^T X^{-1} T_2)^{-1} T_2^T X^{-1}, \text{ where } T_2 = T_{2_j}$$

Step 5: Find T and Λ where $P_{12} = T\Lambda T^{-1}$, $T = [T_r, T_{n-r}]$, $T^{-T} = [U_r, U_{n-r}]$ and

$$\Lambda = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

Step 6: LOS is defined by

$$A_r = U_r^T A T_r, \qquad B_r = U_r^T B, \qquad \text{and } C_r = C T_r$$

Using Eq. (28), J may be expressed as

$$J = \mathrm{Tr}(CP_{21}XP_{21}^T C^T) \tag{30}$$

with X given by Eqs. (15) and (16). As a result, the problem may be restated as identifying the P_{21}, subject to the condition of aggregation, that minimizes J. Details of the solution[16] for the optimal aggregate are described in Table 2 and in the Appendix.

In contrast to the internally balanced approach, one set of state directions cannot be used to characterize the optimal nth-order aggregate for all r, $1 \le r \le n-1$. Aggregation results whenever the relation $U_1^T A T_2 = 0$ is satisfied. This means that the eigenvalues of A_r are a subset of those in A *not* corresponding to the eigenspace defined by T_2. Consequently, $G(s)$ and $G_r(s)$ share the same set of r poles. The zeros in $G_r(s)$ are different, however, to compensate for the missing poles such that the cost functional is minimized. Note also that models obtained by modal truncation[17] (i.e., $U_1^T A T_2 = 0$ and $U_2^T A T_1 = 0$) are also in this set. For any set of modes truncated, however, a better model (with respect to the cost functional) is obtained by determining the reduced-order model as in Table 2.

As in the internally balanced approach, the expression $\bar{X}(j\omega) = (j\omega I - A)^{-1}B$ is a determining factor regarding the system's input behavior. However, by restricting T_2 to be invariant with respect to A, the decomposition of $\bar{X}(j\omega)$ here has direct implications regarding the system's input/output behavior as well.

$$G(j\omega) = C\bar{X}(j\omega) = CT_1 U_1^T \bar{X}(j\omega) + CT_2 U_2^T \bar{X}(j\omega)$$
$$= G_r(j\omega) + E(j\omega) \tag{31}$$

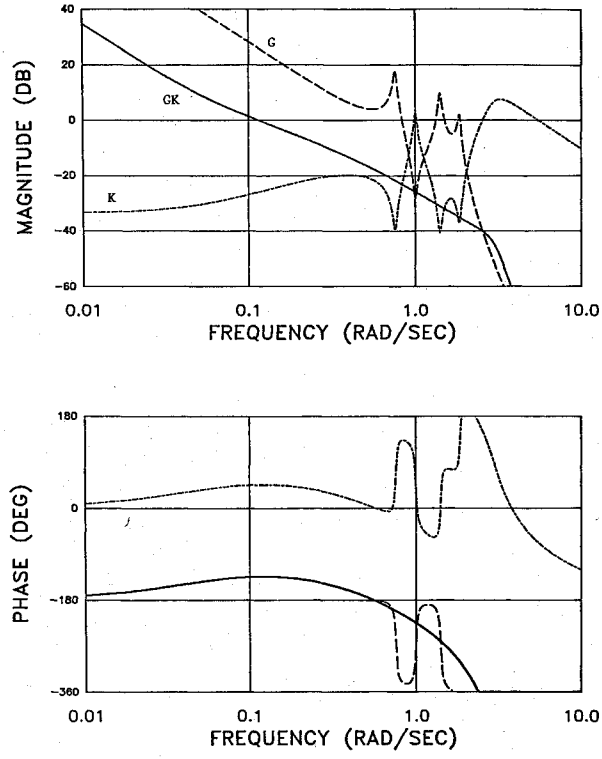

Fig. 3 Nominal loop GK, plant G, and controller K.

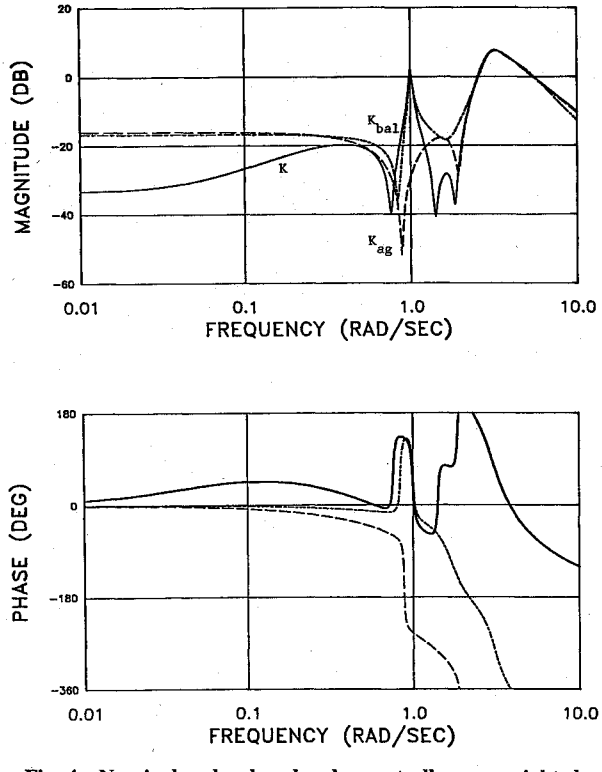

Fig. 4 Nominal and reduced-order controllers: unweighted.

In aggregation, the first component $T_1 U_i^T \bar{X}(j\omega)$ determines the reduced-order model, whereas the second component $T_2 U_2^T \bar{X}(j\omega)$ is the reduction error. Because the 2-norm is minimized, one can expect $G_r(j\omega)$ to approximate $G(j\omega)$ over regions (ω_1,ω_2) where $|\bar{G}(j\omega)|(\omega_2-\omega_1)$ is relatively large, where $|\bar{G}(j\omega)|$ is some mean value of $|G(j\omega)|$ over (ω_1,ω_2). In other words, the approximation occurs in regions of potentially large *integrated* errors. In contrast to the ∞ norm, magnitude alone does not dictate which characteristics will tend to be well approximated.

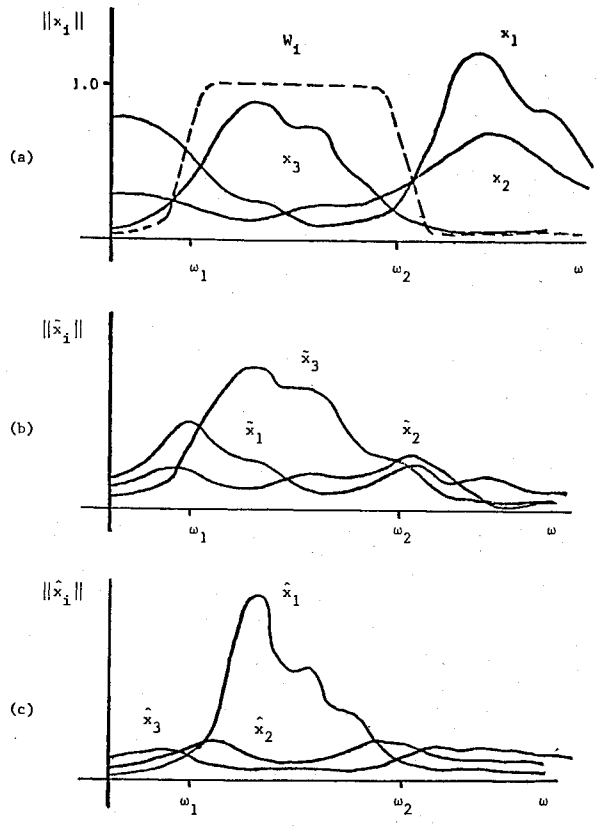

Fig. 5 Principal components.

Example

To illustrate some of the key features of both procedures in the context of a controller reduction problem, consider an example originally used by Enns[10] and revisited by Liu and Anderson.[18] The closed-loop system to be examined consists of a plant G that is linear, time-invariant, SISO, and of eighth order. The plant consists of four disks (unity inertia) axially connected by a flexible shaft (unity spring constant). Torque is applied to this system at one disk and angular measurements of another disk are obtained. The system is nonminimum phase, and the transfer function‡ is

$$G(s) = \frac{0.0064(4.85)[0.02,1.00][-.40,5.66]}{(0)(0)[0.02,0.76][0.02,1.41][0.02,1.85]} \tag{32}$$

A ninth-order compensator K is to be approximated that is stable and minimum phase.

$$K(s) = \frac{31.1(0.05)[0.02,0.76][0.02,1.41][0.02,1.85]}{(0.52)[0.02,1.0][0.95,2.74][0.60,2.87][0.20,2.99]} \tag{33}$$

The nominal loop $GK(j\omega)$, the plant $G(j\omega)$, and the compensator $K(j\omega)$ are shown in Fig. 3. $GK(j\omega)$, of course, determines the critical crossover frequency range. In this particular example, crossover ($|GK(j\omega_c)| = 1$) occurs around $\omega_c = 0.1$ rad/s, and the corresponding phase margin is approximately 45 deg. Moreover, the gain margin for the nominal system is

‡(ζ,ω_n) corresponds to $(s^2 + 2\zeta\omega_n s + \omega_n^2)$ and (p) corresponds to $(s+p)$.

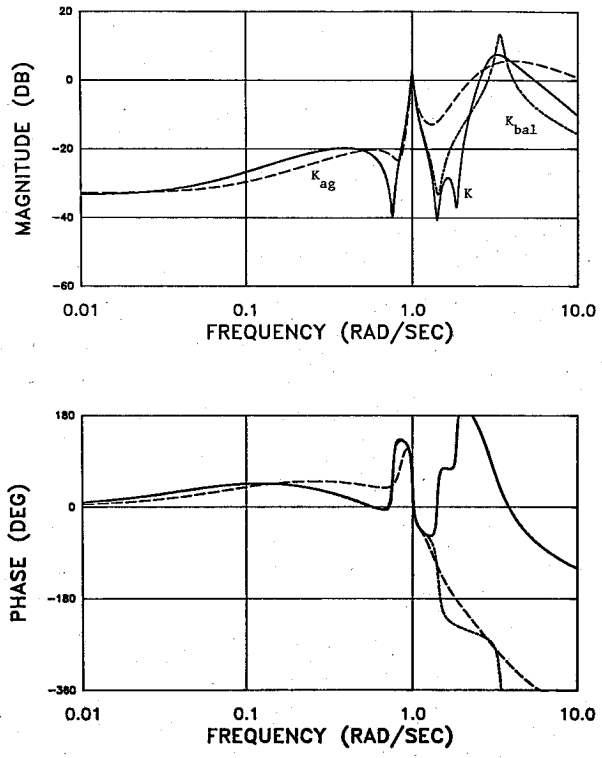

Fig. 6 Full- and reduced-order controllers: Butterworth filter weighting.

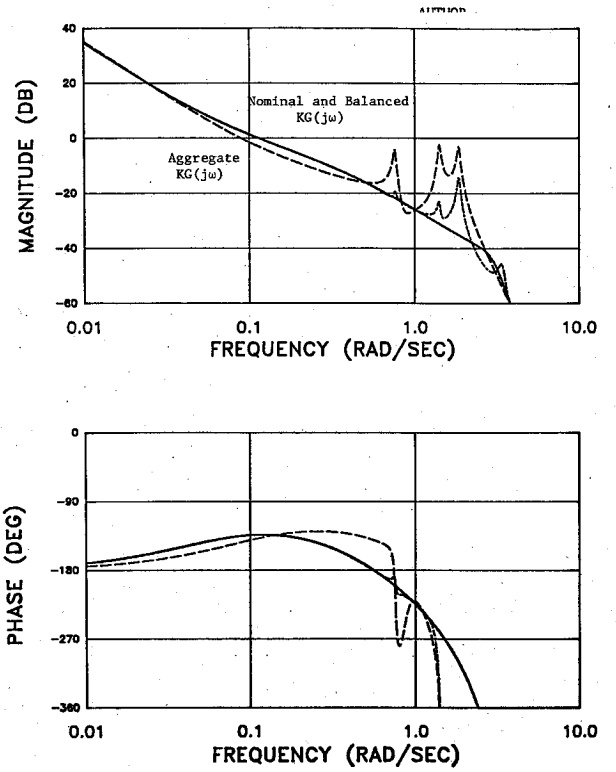

Fig. 7 Nominal and reduced-order loop shapes: Butterworth filter weighting.

almost 20 dB at 0.55 rad/s, with the gain rolling off at about −40 dB/dec. above 1.0 rad/s. Since the reduced-order compensator is expected to preserve this nominal stability robustness, quantified by these margins, a good frequency response approximation of K is sought, say, between 0.01 and 1 rad/s. Let the desired order of the reduced-order controller be sixth.

Application of the internally balanced approach to this problem produces the following sixth-order approximation for K:

$$K_B(s) = \frac{0.9145(17.3)[0.02,0.84][-0.17,1.70]}{[0.02,1.0][0.14,3.06][0.37,4.87]} \tag{34}$$

An optimal sixth-order aggregate approximation is:

$$K_A(s) = \frac{0.1254(237)[-0.01,0.88][-0.03,1.94]}{[0.95,2.74][0.60,2.87][0.20,2.99]} \tag{35}$$

The corresponding frequency responses are shown in Fig. 4, along with that of the full-order K. These reduced-order compensators would not stabilize the loop. Both approaches are seen to yield poor approximations near crossover, tending instead to well model $K(j\omega)$ in the frequency range where $K(j\omega)$ has large magnitude, above the critical frequency range. This tendency, as discussed previously, is characteristic of both approaches, thus demonstrating here by example why some form of frequency weighting is necessary.

V. Frequency Weighting

It will be shown that if the original error $E(j\omega)$ is replaced with some type of frequency-weighted error $E(j\omega)W_i(j\omega)$, then the order-reduction results may be enchanced. In this section, two topics shall be discussed: the incorporation and the selection of frequency weighting. An example will illustrate key results.

Incorporation

In both the internally balanced approach and the optimal aggregate approach, the expression $\bar{X}(j\omega) = (j\omega I - A)^{-1}B$ determines the system's input behavior. The decomposition of $\bar{X}(j\omega)$ shown in Eqs. (18), or

$$\bar{X}(j\omega) = T_1 U_1^T \bar{X}(j\omega) + T_2 U_2^T \bar{X}(j\omega) \tag{36}$$

appears in the respective procedures. The problems arise because $\bar{X}(j\omega)$ describes the system's input behavior over all frequencies. In other words, $\bar{X}(j\omega)$ determines X [Eqs. (16)], which is the state covariance of the system driven by white noise with unity intensity. The dominant components in the above decomposition may not, in fact, dominate $G(j\omega)$ in the frequency range of interest.

Filtering the input to the system using some filter $W_i(j\omega)$ changes the system's input behavior to $\hat{\bar{X}} = \bar{X}(j\omega)W_i(j\omega)$. The effect of this filtering on the resulting reduced-order approximation will first be interpreted assuming that the state directions used in the reduction remain unchanged from those in Eq. (36).

In Fig. 5a, the magnitudes of three unweighted components $|x_i(j\omega)|$ are displayed. Assume the $x_i(j\omega)$ are ordered here in decreasing importance to the system's input, as determined by Eqs. (22). Note that the dominant components (x_1 and x_2) do not dominate $\bar{X}(j\omega)$ [Eq. (36)] over the prespecified frequency range of interest (ω_1,ω_2).

Now, let the input be filtered via a bandpass filter $W_i(j\omega)$ over (ω_1,ω_2), the magnitude of which is shown as a dashed line in Fig. 5a. As noted above, using the same t_i, the decomposition of the filtered state $\hat{\bar{X}}(j\omega)$ is now

$$\hat{\bar{X}}(j\omega) = t_1\tilde{x}_1(j\omega) + \cdots + t_n\tilde{x}_n(j\omega) \tag{37a}$$

$$\tilde{x}_k(j\omega) = u_k^T\hat{\bar{X}}(j\omega) = x_k(j\omega)W_i(j\omega) \tag{37b}$$

for all $k = 1,n$. From the magnitude of $\tilde{x}_i(j\omega)$, displayed in Fig. 5b, it is apparent that the filtering can result in a reordering of the importance of $\tilde{x}_i$ (or t_i) with respect to the system's input behavior. This reordering can be determined by the relative

magnitudes of

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}|\tilde{x}_i(j\omega)|^2\,d\omega=\frac{1}{2\pi}\int_{-\infty}^{\infty}|x_i(j\omega)W_i(j\omega)|^2\,d\omega \quad (38)$$

for $i=1,n$.

Of course, for an actual principal component analysis using $\bar{X}(j\omega)W_i(j\omega)$, the principal directions $\hat{t}_i$ also change as a result of the filtering, so then

$$\hat{\bar{X}}(j\omega)=\hat{t}_1\hat{x}_1(j\omega)+\cdots+\hat{t}_n\hat{x}_n(j\omega) \quad (39)$$

where $\hat{x}_i(j\omega)=\hat{u}_1^T\hat{\bar{X}}(j\omega)$ for $i=1,n$. The resulting magnitudes of $\hat{x}_i(j\omega)$ might be as shown in Fig. 5c. The directions dominating $\hat{\bar{X}}(j\omega)$ are now the most important to the system's input behavior over (ω_1,ω_2).

The relative importance of $\hat{t}_i$ with respect to the system's input/output behavior $[G(j\omega)]$ over (ω_1,ω_2), of course, is still dependent on the decomposition of $Y=C(j\omega I-A)^{-1}$ as well. But, if $W_i(s)$ is well-attenuated outside the frequency band of interest (ω_1,ω_2), integrals analogous to Eqs. (16) and those defining the Hankel singular values, Eq. (23), tend to be band-limited over (ω_1,ω_2). As a result, the principal directions/Hankel singular values would also reflect the weighting on the system's frequency response over (ω_1,ω_2). The largest h_i would correspond to principal directions most critical to $G(j\omega)$ over (ω_1,ω_2).

Unfortunately, neither the internally balanced nor the optimal aggregate approach directly leads to an approximation of the form $G_r(j\omega)W_i(j\omega)$ with the filter intact after reduction. But, under the assumption that both the system to be approximated and the weighting are asymptotically stable, the addition of frequency weighting involves only one minor change to the respective algorithms.[16] The state covariance matrix X, used in both the routines, is replaced by the state covariance matrix of the system driven by white noise filtered by $W_i(s)=C_i(sI-A_i)^{-1}B_i$. This matrix $\hat{X}$ is obtained by simply solving the Lyapunov equation:

$$\begin{bmatrix}A & BC_i\\ 0 & A_i\end{bmatrix}\begin{bmatrix}\hat{X} & X_{12}\\ X_{12}^T & X_{22}\end{bmatrix}+\begin{bmatrix}\hat{X} & X_{12}\\ X_{12}^T & X_{22}\end{bmatrix}\begin{bmatrix}A^T & 0\\ C_i^TB^T & A_i^T\end{bmatrix}+\begin{bmatrix}0 & 0\\ 0 & B_iB_i^T\end{bmatrix}=0 \quad (40)$$

$\hat{X}$ also corresponds to the following integral:

$$\hat{X}=\frac{1}{2\pi}\int_{-\infty}^{\infty}(j\omega I-A)^{-1}BW_i(j\omega)W_i^T(-j\omega)B^T\times(-j\omega I-A^T)^{-1}\,d\omega \quad (41)$$

The matrix Y remains unchanged, as given in Eqs. (18) and (20). Now one proceeds in a manner directly analogous to the unweighted approaches, except that the decompositions of $\hat{X},Y,\hat{\bar{X}}(j\omega)$, and $\bar{Y}(j\omega)$ are used instead of their unweighted counterpart.

For reducing unstable systems, the unstable part of $G(s)$ may be extracted via partial function expansion, and the remaining stable part of $G(s)$ reduced. Then the unstable subsystem is added back to that reduced-order (sub) system. Unstable weightings should be avoided, or only the stable factors[10] used.

Properties of the Weighted Results

One major difference between the frequency-weighted internally balanced approach and its unweighted counterpart is that a norm bound like Eq. (25) has not been established. Also, the (unweighted) internally balanced projection defines a reduced-order model with grammians [Eqs. (16) using the reduced-order model] related to those of the original system by

$$X_r=U_1^TXU_1 \quad (42a)$$

$$Y_r=T_1^TYT_1 \quad (42b)$$

However, with frequency weighting $\hat{X}_r\neq U_1^T\hat{X}U_1$, although $Y_r=T_1^TYT_1$. Both Eqs. (50) and (51) are significant with respect to the stability of the reduced-order model. Without weighting, the guarantee that either $X_r>0$ or $Y_r>0$ is enough to insure that no eigenvalues of A_r have positive real parts, and those that have zero real parts are uncontrollable or unobservable, respectively.[19] With frequency weighting, stability is still guaranteed but only under the condition $Y_r>0$.

With regard to the frequency-weighted optimal aggregate, desirable properties of the reduced-order model transfer directly with frequency weighting because the solution was obtained by minimizing $\|[G(j\omega)-G_r(j\omega)]W_i(j\omega)\|_2^2$ over all aggregate $G_r(j\omega)$. It can be shown[16] that the properties

$$T_1(j\omega I-A_r)^{-1}B_rW_i(j\omega)=P_{12}(j\omega I-A)^{-1}BW_i(j\omega)\ [\text{Eq. (27)}] \quad (43)$$

$$J=\text{Tr}(CP_{21}\hat{X}P_{21}^TC^T)\ [\text{Eq. (30)}] \quad (44)$$

$$J_{\min}=\min_{G_r\in\tilde{A}_r}\|[G(j\omega)-G_r(j\omega)]W_i(j\omega)\|_2^2\ [\text{Eq. (29)}] \quad (45)$$

are preserved. Here $\tilde{A}_r$ denotes the set of rth-order aggregate models, P_{12} denotes an aggregate projection, and $P_{21}=I-P_{12}$. (T_2 is still an n-r dimensioned subspace invariant with respect to A.) The key expression here is Eq. (43), which results by postmultiplying Eq. (27) by $W_i(j\omega)$. The solution for the frequency-weighted optimal aggregate, therefore, proceeds directly in the same manner as the unweighted version. The only difference is that an integral corresponding to $\hat{X}$ emerges in place of the one corresponding to X in Eqs. (16).

Weighting Selection

Because of the issues raised in Sec. II regarding key frequency ranges of the nominal loop, the appropriate $W_i(s)$ should possess the following properties:

1) $W_i(j\omega)$ must be dependent on the nominal loop shape $KG(j\omega)$ or $GK(j\omega)$.

2) $W_i(j\omega)$ must emphasize frequency ranges where $\underline{\sigma}[I+GK(j\omega)]\approx\bar{\sigma}[\Delta GK(j\omega)]$ or $\underline{\sigma}[I+KG(j\omega)]\approx\bar{\sigma}[\Delta KG(j\omega)]$.

3) $W_i(j\omega)$ should normalize the error $E(j\omega)$ with respect to the magnitude of the system to be approximated over the frequency range of interest.

4) $W_i(j\omega)$ must be asymptotically stable.

The need for properties 1, 2, and 4 has been addressed previously. Property 3 is a consequence of the norms described in Eqs. (12) and (13).

With a knowledge of the intended crossover-frequency range, and the basic trend (e.g., lead or lag behavior) of the system to be approximated, a weighting might be obtainable by inspection. For example, a pass-band filter simply emphasizing the crossover range may be adequate. A later example will demonstrate this. In applications to model reduction, for example, a detailed description of the "nominal" high-order controller is not necessary with this approach, in contrast to the following technique.

A more elegant weighting technique was originally[10] motivated by the stability robustness theorem,[4,9] which states that over all stable perturbations ΔGK, the perturbed closed-loop system will remain stable if

$$\|(\Delta GK)(I+GK)^{-1}\|_\infty<1 \quad (46)$$

or equivalently, if

$$\|(\Delta KG)(I+KG)^{-1}\|_\infty < 1 \tag{47}$$

By considering the substitution of a lower-order approximation in the loop as a perturbation of the nominal loop shape, Enns[10] postulated that a reasonable choice for plant weighting is $W_{i_G} = K(I+GK)^{-1}$ for plant reduction, whereas a reasonable choice for the controller weighting is $W_{i_K} = G(I+KG)^{-1}$. Consider, now, the properties associated with this choice of frequency weighting.

First, the stability robustness theorem pinpoints frequency ranges consistent with Eq. (7). Considering Eq. (46) and recalling how the singular value is related to $\|.\|_\infty$, the theorem states that stability is guaranteed when

$$\bar{\sigma}(\Delta GK) < \underline{\sigma}(I+GK) \tag{48}$$

for all ω. Intuitively, then, frequencies (and directions) where $K(I+GK)^{-1}$ is "big" correspond to frequencies where small changes in G, ΔG, could make

$$\bar{\sigma}(\Delta GK) \approx \underline{\sigma}(I+GK) \tag{49}$$

[Note, for example, the scalar case. $K(1+KG)^{-1}$ is big where K is big, and/or when $KG \approx -1$!] As a result, this weighting automatically highlights frequencies where stability robustness is *potentially* a critical factor.

Assuming the nominal loop shapes have high gains at low frequencies, Enns' frequency weightings also satisfy property 3, because $GW_{i_G}(s)$ and $KW_{i_K}(s)$ are simply the closed-loop transfer function. Therefore,

$$GW_{i_G}(s) \approx I_p \qquad KW_{i_K}(s) \approx I_m \tag{50}$$

over the bandwidth of the closed-loop system. Moreover, since the frequency-weighted system is the *nominal* closed-loop (assumed to be stable), property 4 is also satisfied.

Example

The effects of frequency weighting are to be observed in the context of the controller reduction example previously presented. First, consider a weighting consisting simply of a third-order low-pass Butterworth filter, with break frequency of 1.0 rad/s, plus approximate integral action (since the compensator is generating lead). Here

$$W_i(s) = \frac{1}{(s+0.01)(s^3+2s^2+2s+1)} \tag{51}$$

The reduced-order compensators obtained using this weighting are shown in Fig. 6. Comparing Figs. 6 and 4 clearly demonstrates the improvement in the approximation over the critical 0.01–1.0 rad/s frequency range. The loop shapes $GK(j\omega)$ with these compensators are shown in Fig. 7. The weighted internally balanced compensator K_B yields almost precisely the same gain and phase margin as the full-order compensator. The compensators obtained using the above weighting are

$$K_B(s) = \frac{1.496(0.051)[0.03,0.758][-0.04,1.43]}{[0.89,0.61][0.02,1.0][0.04,3.38]} \tag{52a}$$

$$K_A(s) = \frac{11.7(0.098)[0.08,0.856][-0.22,1.33]}{[0.02,1.0][0.95,2.74][0.60,2.87]} \tag{52b}$$

Finally, consider the weighting suggested by Enns, or

$$W_i(s) = K(I+GK)^{-1} \tag{53}$$

The resulting reduced-order compensators in this case are shown in Fig. 8, while the associated loops GK are shown in Fig. 9. Even better results are demonstrated here, especially for the weighted-aggregate approach. The reduced-order compensators in this case are

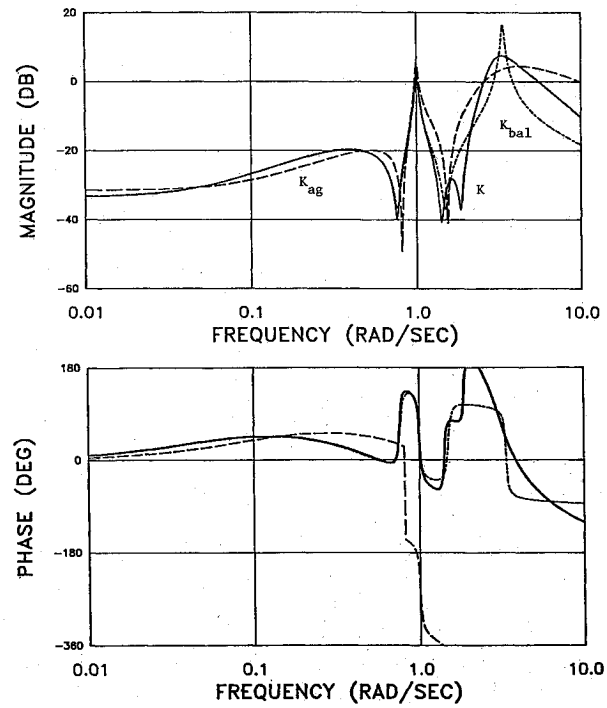

Fig. 8 Nominal and reduced-order controllers: Enns weighting.

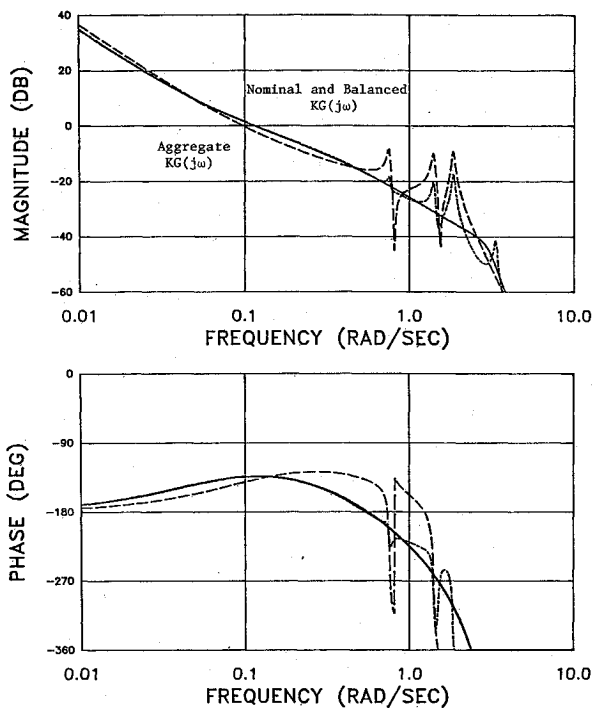

Fig. 9 Nominal and reduced-order loop shapes: Enns weighting.

$$K_B(s) = \frac{1.11(0.052)[0.03,0.766][0.03,1.49]}{[0.84,0.56][0.02,1.0][0.02,3.32]} \tag{54a}$$

$$K_A(s) = \frac{10.6(0.098)[4.0\times10^{-5},0.816][-0.004,1.54]}{[0.02,1.0][0.95,2.74][0.60,2.87]} \tag{54b}$$

VI. Conclusion

A major point of this paper is that regardless of the model-reduction technique selected, for closed-loop system applications it is imperative that the frequency domain character of the approximation method be understood. As shown, this is especially true for techniques not originally formulated in the frequency domain. If the technique actually approximates the key response over the critical frequency ranges, acceptable results can be expected. However, if not, certain measures may have to be instigated. Such a measure is frequency weighting.

The importance of a good frequency-domain approximation in closed-loop system applications was underscored. The use of model reduction in closed-loop systems is in fact feasible because some portions of the frequency response are more critical than others. The determining factors are the nominal control system design and some well-known properties of feedback.

An algebraic framework for studying state-space model reduction was discussed, two reduction methods originally considered in the time domain were examined, and the ties to the frequency domain were explored. Although direct application of these techniques failed, the leading causes were established.

Frequency weighting was then logically incorporated into the procedures, and guidelines for selecting frequency weightings were suggested. Properties and limitations of the frequency-weighted results were compared to those from the corresponding unweighted techniques. An example was presented that clearly illustrated the improvements due to the frequency weighting.

Appendix: Optimal Aggregate Solution

Using Eq. (28), the task of identifying the optimal aggregate model becomes the equivalent task of identifying the projection P_{21} that minimizes

$$J = \mathrm{Tr}(CP_{21}XP_{21}^T C^T) \tag{A1}$$

subject to the condition that the subspace *on* which the projection P_{12} is taken is an n-r subspace invariant with respect to A, where $P_{21} = I - P_{12}$ is the projection matrix on $\boldsymbol{T}_2$ along $\boldsymbol{T}_1$, and the matrix X is the controllability grammian defined in Eqs. (16).

The solution for the optimal aggregate is obtained in two steps. First, for an arbitrary $\boldsymbol{T}_2$ subject to the condition of aggregation, the $\boldsymbol{T}_1$ is identified that minimizes J; then, the optimal $\boldsymbol{T}_2$ is identified.

Let $\boldsymbol{T}_2$ be any n-r-dimensional subspace invariant with respect to A. Furthermore, let $\boldsymbol{T}_2$ be an arbitrary span of $\boldsymbol{T}_2$ and define $P_{21} = T_2U_2^T$ (i.e., $U_2^T T_2 = I_{n-r}$). The optimal $\boldsymbol{T}_1$ subject to this $\boldsymbol{T}_2$ is solely dependent on the choice of U_2^T (i.e., $\boldsymbol{T}_1$ equals the null space of U_2^T). The task then is to find U_2^T that minimizes J.

$$J = \mathrm{Tr}(CT_2U_2^T XU_2T_2^T C^T) \tag{A2}$$

subject to the constraint that $T_2^T U_2 = I_{n-r}$. Employing the Lagrange multiplier Q, the Lagrangian may be written as

$$L(U_2,Q) = \mathrm{Tr}(CT_2U_2^T XU_2T_2^T C^T) - 2\mathrm{Tr}[Q(T_2^T U_2 - I_{n-r})] \tag{A3}$$

The necessary conditions are:

$$0 = \frac{\partial L}{\partial U_2} \Rightarrow XU_2T_2^T C^T CT_2 - T_2Q^T = 0$$

$$0 = \frac{\partial L}{\partial Q} \Rightarrow T_2^T U_2 = I_{n-r}$$

yielding

$$Q^T = (T_2^T X^{-1} T_2)^{-1}(T_2^T C^T CT_2)$$

$$U_2 = X^{-1}T_2(T_2^T X^{-1}T_2)^{-1}$$

producing

$$P_{21} = T_2(T_2^T X^{-1}T_2)^{-1}T_2^T X^{-1} \tag{A4}$$

Assuming the eigenvalues of A are distinct, the search for the desired $\boldsymbol{T}_2$ is conducted over a finite set of n-r-dimensioned subspaces invariant with respect to A. Let v_i denote the eigenvector of A. For each real eigenvector of A, say v_i, define

$$w_i = v_i \tag{A5}$$

Whereas for each complex conjugate pair of eigenvectors of A, say $v_i = v_{R_i} + jv_{I_i}$ and $v_{i+1} = v_i^*$, define

$$w_i = (v_{R_i}, v_{I_i}) \tag{A6}$$

The finite set of n-r-dimensioned subspaces invariant with respect to A can be characterized by the finite set $\{T_{2_1},...,T_{2_j},...,T_{2_k}\}$. Each T_{2_j} consists of *combinations* of w_i, $T_{2_j} = \{w_1...w_i,...w_m\}$, such that T_{2_j} has n-r columns.

Each T_{2_j} actually corresponds to a suboptimal aggregate model of the original system. To identify the minimizing T_{2_j}, first note that if T_{2_j} is inserted into Eq. (A4) and the resulting P_{21} into Eq. (A1), the cost associated with that aggregate model is

$$J_j = \mathrm{Tr}[CT_{2_j}(T_{2_j}^T X^{-1}T_{2_j})^{-1}T_{2_j}^T C^T] \tag{A7}$$

The optimal aggregate is obtained from the projection

$$P_{12} = I_n - T_2(T_2^T X^{-1}T_2)^{-1}T_2^T X^{-1} \tag{A8}$$

or the projection on $\boldsymbol{T}_1$ along $\boldsymbol{T}_2$. The reduced-order model is obtained from a *spectral decomposition of* P_{12}. Let the columns of T_r and U_r be, respectively, the eigenvectors and reciprocal eigenvectors of P_{12} corresponding to the eigenvalues of P_{12} equal to unity. (The rest of the eigenvalues are zero.) P_{12} equals $T_rU_r^T$ and the optimal aggregate model is defined as

$$A_r = U_r^T AT_r \qquad B_r = U_r^T B \qquad C_r = CT_r \tag{A9}$$

Acknowledgment

This work was sponsored by the Naval Air Development Center under Contract N62269-83-C-0220. Mr. Marvin Walters was the technical monitor.

References

[1]Hodgkinson, J., Lamanna, W. J., and Heyde, J. L., "Handling Qualities of Aircraft with Stability and Control Augmentation – A Fundamental Approach," *Aeronautical Journal*, Feb. 1976, pp. 75–81.

[2]McRuer, D., Ashkenas, I., and Graham, D., *Aircraft Dynamics and Automatica Control*, Princeton University Press, Princeton, NJ, 1973.

[3]Bode, H. W., *Network Analysis and Feedback Amplifier Design*, Van Nostrand, Princeton, NJ, 1945.

[4]Doyle, J. C. and Stein, G., "Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis," *IEEE Transactions on Automatic Control*, Vol. AC-26, Feb. 1981.

[5]Kwakernaak, H. and Sivan, R., *Linear Optimal Control Systems*, Wiley-Interscience, New York, 1972.

[6]Doyle, J. C. and Stein, G., "Robustness with Observers," *IEEE Transactions on Automatic Control*, Vol. AC-24, No. 4, Aug. 1979.

[7]Moore, B. C., "Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction," *IEEE Transactions on Automatic Control*, Vol. AC-26, Feb. 1981.

[8]Mitra, D., "The Reduction of Complexity of Linear, Time-Invariant Dynamic Systems," *Proceedings of the 4th IFAC Congress*, 1969, pp. 19–33.

[9]Doyle, J. C., Wall, J. E., and Stein, G., "Performance and Robustness Analysis for Structured Uncertainties," *Proceedings of the 21st IEEE Conference on Decision and Control*, IEEE, New York, Vol. 2, Dec. 1982.

[10]Enns, D. F., "Model Reduction for Control System Design," Ph.D. Dissertation, Dept. of Aeronautics and Astronautics, Stanford Univ., Stanford, CA, June 1984.

[11]Takenchi, K., Yanai, H., and Mukherjec, B. N., *The Foundations of Multivariate Analysis: A Unified Approach by Means of Projection onto Linear Subspaces*, Wiley Eastern Ltd., New Delhi, India, 1982.

[12]Halmos, P. R., *Finite-Dimensional Vector Spaces*, Van Nostrand, Princeton, NJ, 1958.

[13]Brogan, W. L., *Modern Control Theory*, Quantum, New York, 1974.

[14]Glover, K., "All Optimal Hankel-Norm Approximations of Linear Multivariable Systems and Their L^{∞}-Error Bounds," *International Journal of Control*, Vol. 39, No. 6, pp. 1115–1193.

[15]Mahmond, M. S. and Singh, M. G., *Large Scale Systems Modelling*, Pergamon, Oxford, 1981.

[16]Bacon, B. J., "Order Reduction in Closed-Loop Systems," Ph.D. Dissertation, School of Aeronautics and Astronautics, Purdue Univ., West Lafayette, IN, 1988.

[17]Davidson, E. J., "A Method for Simplifying Linear Dynamic Systems," *IEEE Transactions on Automatic Control*, Vol. AC-11, 1966, pp. 93–101.

[18]Liu, Y. and Anderson, B. D. O., "Controller Reduction Via Stable Factorization and Balancing," *International Journal of Control*, Vol. 44, No. 2, 1986, pp. 507–531.

[19]Pernebo, L. and Silverman, L. M., "Model Reduction in Balanced State-Space Representation," *IEEE Transactions on Automatic Control*, AC-27, 1982, pp. 382–387.