Controller Reduction: Concepts and Approaches

BRIAN D. O. ANDERSON, FELLOW, IEEE, AND YI LIU, STUDENT MEMBER, IEEE

Abstract—**This paper considers the problem of passing from a linear time-invariant high-order controller designed for a linear time-invariant plant (of presumably high order) to a low-order approximation of the controller. The approximation problem is often best posed as a frequency-weighted L_∞ approximation problem. Many different controller representations are possible, giving different performance of the various reduction algorithms.**

I. Introduction

SIMPLE linear controllers are normally preferred over complex linear controllers for linear time-invariant plants: there are fewer things to go wrong in the hardware or bugs to fix in the software; they are easier to understand; and the computational requirements are less. For this reason, there is a desire to have methods available for designing low-order controllers for high-order plants. Such methods can broadly be divided into two classes: *direct*, in which the parameters defining a low-order controller are computed by some optimization or other procedure; and *indirect*, in which a high-order controller is first found, and then a procedure used to simplify it.

Examples of direct methods include the work of Gangsaas *et al.* [1] (see the third case study in [1]) which draws on [2], and of Bernstein and Hyland [3]–[5]. Although this paper does not have such methods as its concern, they nevertheless demand comment, at least of a brief character.

The common philosophy in these methods is to seek to minimize a quadratic performance index subject to the constraint that the controller be of fixed degree (as well as stabilizing and time-invariant). Two key technical ideas underpin [2]. First, given a stabilizing controller, it is straightforward to compute the sensitivity of the performance index to a variation in a controller parameter; second, when a stabilizing controller is not initially known, one can often be found by working with a finite time performance index (which is not infinite-valued, obviously), and using its sensitivity to the controller parameters (quantities which are readily computable) to iterate the controller parameters so that the index is minimized. If the time interval over which the index is computed is made sufficiently long, the resulting controller should be stabilizing, provided a stabilizing controller of the assumed order exists.

The strategy in [3]–[5] is to determine a set of algebraic equations which constitute necessary (but not sufficient) conditions on the controller parameters to achieve a minimum value for the performance index. The equations have a structure which displays a parallel with those applicable in the full-order case, and the number of (scalar) equations equals the number of scalar controller parameters. The solution of these equations is not straightforward—and the multiplicity of solutions may be very great. Nevertheless, substantial progress has been made using homotopic methods; see, e.g., [6]. This more recent work gives grounds for believing that standard control design packages in the medium term could contain software suitable for a nonspecialist in numerical analysis.

There is at least one overriding theoretical issue in relation to these direct methods which deserves more attention. Often, conventional LQG design uses the performance index as a tool for achieving certain closed-loop goals relating to sensitivity, bandwidth, gain margin, etc. The LQG design process (before any controller reduction), may proceed iteratively, using results or empirical rules for adjusting weighting matrices in light of achieved closed-loop parameters such as those noted above. The extent to which this could be possible in direct reduced-order design is far from clear; is it legitimate to mimic the concept of loop transfer recovery, for example, in a direct reduced-order design? No one really knows at this point, and until such points are better understood, there may be some reluctance in embracing the direct reduced-order design approach for problems where minimization of a quadratic index is not the ultimate goal, but a device to achieve some other end.

As for indirect methods, there are at least two sophisticated approaches to the design of high-order controllers, LQG and H^∞, and at least for the former, a great deal of qualitative/conceptual knowledge exists which is vital in the applications of the design algorithms to practical problems. Less well developed are the procedures for reducing high-order controllers to low-order controllers; such procedures are the subject of this paper.

A somewhat crude approach (that nevertheless can often be successful) to controller reduction is modal reduction. In frequency domain terms, modal reduction amounts to representing a stable transfer function matrix in partial fraction form, and approximating by throwing away the summand with the smallest value of maximum $j\omega$-axis magnitude. As the poles approach the $j\omega$-axis, this method becomes equivalent asymptotically to a scheme known as approximation based on modal cost analysis (the underlying idea being to approximate so as to affect the cost to the least degree); it also becomes equivalent asymptotically to balanced realization truncation. See [7]–[9].

Yet another approach is to approximate the plant rather than the controller. Then a low-order controller is designed using a low order of approximation of the plant, with the low-order controller then used on the correct plant. There is both a general and a specific criticism of this idea. The general criticism is that in the overall design process leading to the controller, the approximation is carried out at an earlier step in the process than if the controller is approximated; each subsequent step in a design after an approximation propagates the effects of that approximation, and the ultimate effect at the end can be unclear, the more so the greater the number of design steps subsequent to the approximation. The specific criticism is that, as argued in [9], *satisfactory* approximation of the plant requires some knowledge in advance of the controller. So the designer is caught in an awkward logical loop. Iteration may be one way out. Also, the advance knowledge of the controller may in practice not have to be precise, but equivalent to or deducible from the knowledge embedded in closed-loop specifications supplied in advance.

It is crucial to accept that the problem of controller reduction is distinct from the problem of (open-loop) model reduction because of the presence of the plant. It is after all closed-loop performance (in the broadest sense of the word, i.e., stability, bandwidth,

Manuscript received December 12, 1987; revised February 7, 1989. Paper recommended by Associate Editor at Large, T. Johnson.
The authors are with the Department of Systems Engineering, Research School of Physical Sciences, Australian National University, Canberra, Australia.
IEEE Log Number 8928774.

0018-9286/89/0800-0802$01.00 © 1989 IEEE

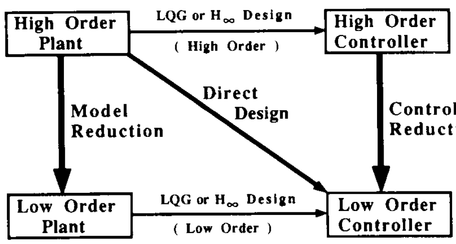


Fig. 1. Basic principles of low-order controller design.

robustness as well as performance index value) which should be well approximated, and obviously the plant as well as the controller is a determinant. This is argued in Section II, which makes the case that controller reduction can be regarded as a *frequency-weighted* L_∞ approximation problem, with no simple procedure best for defining the weight. Unfortunately, there are no nice algorithms available for solving such problems. There are, however, algorithms which appear to come close in practice, and can probably be proven to come close (closeness being measured using singular values of the error) in some cases. We describe such algorithms in Section III. Such algorithms often offer less appealing performance when used on unstable controllers—this motivates a study in Section IV of controller reduction using what are termed "fractional representations" of the controller. Section V shows how to trade-off stability and performance considerations in some of the reduction procedures, and Section VI contains concluding remarks.

Fig. 1 displays the alternative approaches we have described above to the problem of finding a low-order controller for a high-order plant.

Our original motivation for undertaking this work came from conversations with Boeing engineers, especially D. Gangsaas. Specific motivating examples have had plant orders between 8 and 55. Some of these are discussed or referenced later in the paper.

II. Controller Reduction and Frequency Weighting

There is a fundamental difference between model reduction and controller reduction. Model reduction is, at least normally, based on open-loop considerations. On the other hand, any controller reduction procedures, if rationally based, ought to take into account the existence of the plant. Controller reduction should after all preserve *closed-loop* stability, and (as far as possible) the *closed-loop* performance and *closed-loop* transfer function.

Making these arguments more precise turns out to generate frequency-weighted approximation problems, as we shall now show. The choice of frequency weight is influenced by the choice of criterion thought most important in the approximation process, viz. maintenance of stability margin, performance (the term being used in a loose sense), or closed-loop transfer function.

A. Stability Margin Considerations for Frequency Weighting

Let $G(s)$ be the transfer function matrix of a given linear time-invariant plant (with l inputs and m outputs), and let $K(s)$ be a stabilizing high-order compensator (obtained by some standard procedure). Let $K_r(s)$ be a reduced-order compensator, which we are seeking. Regard the closed-loop system with $K_r(s)$ replacing $K(s)$ as being equivalent to that of Fig. 2. It can then be concluded using this redrawing [10] (and it is now well known) that if:

i) $K(s)$ and $K_r(s)$ have the same number of poles in Re $(s) > 0$ and no poles on the imaginary axis; [1] and

[1] This restriction to no $j\omega$-axis poles can be circumvented by requiring K and K_r to have identical $j\omega$-axis poles and residues.

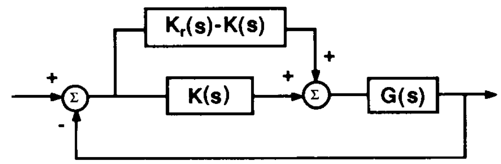


Fig. 2. Rearrangement of feedback system with reduced-order compensator.

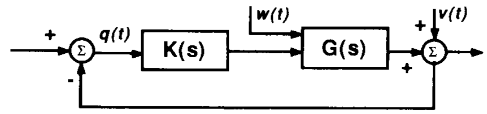


Fig. 3. Original feedback scheme showing noise excitation.

ii) either

$$\|[K(s)-K_r(s)]G(s)[I+K(s)G(s)]^{-1}\|_\infty<1 \tag{2.1}$$

or

$$\|[I+G(s)K(s)]^{-1}G(s)[K(s)-K_r(s)]\|_\infty<1 \tag{2.2}$$

then $K_r(s)$ is a stabilizing compensator. (The notation $\|A(s)\|_\infty$ means $\sup_\omega \max_i \lambda_i^{1/2}[A^*(j\omega)A(j\omega)]$. Here, $\lambda_i[X]$ denotes the ith eigenvalue of the matrix X.)

This suggests a minimization problem: find a $K_r(s)$ satisfying i) which at the same time minimizes the left side of (2.1) or (2.2), and has prescribed degree. The matrix $G(I + KG)^{-1} = (I + GK)^{-1}G$ acts as a *weighting matrix* in this case.

Remarks 2.1: It is easy to see that $\bar{\sigma}\{G(j\omega)[I + K(j\omega)G(j\omega)]^{-1}\}$ is small when either $\bar{\sigma}\{G(j\omega)\}$ is small or $\underline{\sigma}\{K(j\omega)\}$ is large and so often the frequency weighting obtained from the above stability margin argument will be greatest near the unity gain crossover frequencies of the loop gain $G(j\omega)K(j\omega)$. Here $\bar{\sigma}$ and $\underline{\sigma}$ stand for the largest and the smallest singular value of the matrix, respectively. This means that it is more important to have accurate approximation in this band, an idea familiar from classical control.

Remark 2.2: If a $K(s)$ of nth degree is designed by an LQG optimal procedure, and we then find the lower order $K_r(s)$ which minimizes $\|[K(j\omega) - K_r(j\omega)]G(j\omega)[I + K(j\omega)G(j\omega)]^{-1}\|_\infty$ (or $\|[I + GK]^{-1}G(K - K_r)\|_\infty$, in the other case), there is no implication that $K_r(s)$ is in any sense LQG optimal.

B. Performance Considerations for Frequency Weighting

Consider the original closed-loop system in the presence of process and measurement noise, $w(t)$ and $v(t)$, as depicted in Fig. 3. It is possible to compute the spectrum $\Phi_{qq}(j\omega)$ of the noise process of $q(t)$. (We assume stationarity of the excitation noises and closed-loop stability, so that the spectrum exists.)

In order that a low-order approximation $K_r(s)$ to $K(s)$ be a good approximation, it is important that it be most accurate in those frequency bands encountered in actual operation on performance. Thus, if $q(t)$ has little spectral energy in one band, $K(j\omega)$ need not be closely approximated there by $K_r(j\omega)$, while if the spectral energy in another band is high, approximation needs to be accurate. Let $V(j\omega)$ be a stable, minimum-phase spectral factor of $\Phi_{qq}(j\omega)$. (Thus, $VV^* = \Phi_{qq}$.) Then the approximation problem becomes: find $K_r(j\omega)$ of nominated degree such that:

i) $K(s)$ and $K_r(s)$ have the same number of poles in Re $(s) > 0$ and no poles on the imaginary axis;

ii) $\|[K(s) - K_r(s)]\ V(s)\|_\infty$ is minimized.

Note that replacement of $K(s)$ by $K_r(s)$ will change the spectrum of $q(\cdot)$ and to this extent, the choice of error is heuristic.

C. Closed-Loop Transfer Function Considerations for Frequency Weighting

The closed-loop transfer function matrices with $K(s)$, $K_r(s)$ are

$$W(s)=G(s)K(s)[I+G(s)K(s)]^{-1}=I-[I+G(s)K(s)]^{-1}$$

$$W_r(s)=G(s)K_r(s)[I+G(s)K_r(s)]^{-1}=I-[I+G(s)K_r(s)]^{-1}.$$

Approximately, i.e., neglecting terms of second order in $K_r - K$, there holds

$$W_r(s)-W(s)=[I+G(s)K(s)]^{-1}G(s)\cdot[K_r(s)-K(s)][I+G(s)K(s)]^{-1}$$

and this suggests the following approximation problem: find $K_r(j\omega)$ of nominated degree so that:

i) $K(s)$ and $K_r(s)$ have the same number of poles in Re $(s) > 0$ and no $j\omega$-axis poles;

ii) $\| V_1(s)[K_r(s) - K(s)] V_2(s)\|_\infty$ is minimized, where $V_1 = (I + GK)^{-1}G$, $V_2 = (I + GK)^{-1}$.

Comparing ii) to (2.2) shows that there is reduced weighting placed on frequencies in the high loop-gain region in this third approach as compared to the first approach.

D. Further Issues

- Other weights may be appropriate on occasion. For example, if the spectrum of external inputs were known, a minimum-phase spectral factor could appear as an additional weight $V_3(s)$, multiplying $V_2(s)$ on the right.
- As will be later seen, other representations of the controller lead to different frequency-weighted problems, formulated, however, with the same conceptual basis (e.g., stability) as above.
- One would have to expect that concentration on stability could lead to poorly performing controllers, while concentration on other performance measures could lead to instability.
- The weighted approximation problem cannot in general be easily solved. Related problems can, however, be comparatively easily solved, as described in the sequel.
- The approximation problems posed are not fully appropriate for controllers with unstable or $j\omega$-axis poles. Consider a controller containing a pure integrator. The approximation problem posed demands that any approximation also contain a pure integrator with *precisely* the same residue. This shows that the approximation problem is in some way unnecessarily restrictive, since a good approximation in practice need not meet this requirement.
- Quite apart from the suitability of the approximation problems posed, frequency-weighted (or for that matter unweighted) approximation when unstable poles occur in the object being approximated can cause further headaches in the actual approximation process. One approach is to copy the unstable part (under additive decomposition) of $K(s)$ into $K_r(s)$ and then just to approximate the stable part of $K(s)$ with the (lower order) stable part of $K_r(s)$.
- None of the formulated frequency-weighted approximation problems above requires the high-order controller $K(s)$ to have been designed by LQG methods; the method based on performance considerations does, however, require an underlying assumption concerning noise statistics.

III. Frequency-Weighted Model Reduction Techniques

In this section, we discuss the problem of determining the solution to one of the frequency-weighted reduction problems posed in the previous section. Our starting point is that the problems as posed are not analytically solvable, and in many cases "brute force" optimization may be impractical because of the dimensionality of the parameter vector defining the reduced-order controller. Therefore, we must consider reduction methods which do not necessarily achieve the minimum of the frequency-weighted L_∞ index, but in some way effect a good frequency-weighted approximation. Note that, at least in principle, any reduction method can be coupled with any frequency-weighted problem.

There are now at least three important and also rather popular state-space based model reduction techniques, namely, truncation of the internally balanced realization [11], [12], Hankel norm optimal approximation [13]-[16], and q-covariance equivalent realization (q-COVER) [17]-[20]. In their usual form, they replace one stable high-order model by a second stable low-order model that usually is *not* an optimal L_∞ approximation; further, usually no frequency weighting is employed. There are available for the first two methods bounds on the L_∞ error of approximation. Frequency-weighted versions of the first two methods are also available [9], [21], [22], but an L_∞ error bound is available only for frequency-weighted Hankel norm reduction [22]. There is not to this point a complete frequency-weighted q-COVER approximation procedure, although some initial work has been done [23].

Through lack of a better alternative, these reduction techniques have been or can be used for controller reduction. In the remainder of this section we shall highlight some features of these techniques. It is important to remember that all these techniques are faute de mieux procedures that can only come close to solving the L_∞ approximation problems defined in the last section.

A. Balancing Approximation

Given an nth-order, linear time-invariant, and asymptotically stable system with transfer function matrix $G(s)$, a minimal realization of $G(s) = C(sI - A)^{-1}B$ is internally balanced if $\{A, B, C\}$ satisfy

$$A\Sigma+\Sigma A'+BB'=0 \tag{3.1}$$

$$A'\Sigma+\Sigma A+C'C=0 \tag{3.2}$$

$$\Sigma=\text{diag}\,\{\sigma_1,\sigma_2,\cdots,\sigma_n\},\ \sigma_i\geqslant\sigma_{i+1}>0,\ i=1,\cdots,n-1. \tag{3.3}$$

The matrix Σ is both the controllability and observability Gramian. Partition the system $\{A, B, C\}$ as

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}\}r,\quad B=\begin{bmatrix}B_1\\ B_2\end{bmatrix}\}r$$

$$C=[C_1\ \ C_2],\ \Sigma=\begin{bmatrix}\Sigma_1 & 0\\ 0 & \Sigma_2\end{bmatrix}\}r. \tag{3.4}$$

Set $A_r = A_{11}$, $B_r = B_1$, $C_r = C_1$. Then the reduced-order system $\{A_r, B_r, C_r\}$ is a good approximation of system $\{A, B, C\}$ if $\sigma_r \gg \sigma_{r+1}$. In fact, we have the following two properties.

i) Subsystems $\{A_{ii}, B_i, C_i\}$, $i = 1, 2$, are asymptotically stable if $\sigma_r > \sigma_{r+1}$ [12].

ii) There exists a frequency domain error bound for the balancing approximation [9], [16], [24]:

$$\|[C(j\omega I-A)^{-1}B-C_r(j\omega I-A_r)^{-1}B_r]\|_\infty \leqslant 2(\sigma_{r+1}+\cdots+\sigma_n)=2\ \text{tr}\ (\Sigma_2). \tag{3.5}$$

It also turns out [16] that *any* approximation of $C(j\omega I - A)^{-1}B$ of degree r necessarily has an L_∞ error of at least σ_{r+1}. Sometimes the actual L_∞ error achieved on the left of (3.5) can be compared against this generally unachievable lower bound.

Enns [9], [24] introduced frequency weighting into the balancing technique. Consider an asymptotically stable frequency-weighting function $W_i(s) = E_i + C_i(sI - A_i)^{-1}B_i$ as an input weighting to the asymptotically stable system $G(s)$ in Fig. 4. The

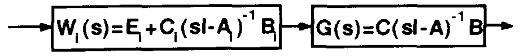


Fig. 4. Introduction of input weighting.

basic idea is to change the controllability Gramian to reflect the introduction of the frequency weighting. Thus, we find a "*frequency-weighted*" controllability Gramian which equals the observability Gramian and then both are diagonalized. In outline, this is done as follows.

Let system matrices of the cascade system be defined as

$$\tilde{A}=\begin{bmatrix} A & BC_i \\ 0 & A_i \end{bmatrix}, \tilde{B}=\begin{bmatrix} BE_i \\ B_i \end{bmatrix}.$$

Assume

$$\tilde{U}=\begin{bmatrix} U & U_{21}' \\ U_{21} & U_{22} \end{bmatrix}$$

is the solution of the following Lyapunov equation:

$$\tilde{A}\tilde{U}+\tilde{U}\tilde{A}'+\tilde{B}\tilde{B}'=0. \tag{3.6}$$

Now, U can be regarded as the frequency-weighted controllability Gramian for the original system. Let Y be the observability Gramian, i.e., Y satisfies

$$YA+A'Y+C'C=0. \tag{3.7}$$

Consider a coordinate basis change to (A, B, C) which makes

$$\begin{aligned} U_{\text{new}} &= Y_{\text{new}} \\ &= \text{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_n\}, \lambda_i \geqslant \lambda_{i+1}, \\ & \quad i=1, 2, \cdots, n-1. \end{aligned} \tag{3.8}$$

There is no change to (A_i, B_i, C_i). We call this new realization a frequency-weighted balanced realization.

Now, as before, the frequency-weighted approximation is achieved by eliminating rows and columns of A, B, C corresponding to smallest $\{\lambda_n, \lambda_{n-1}, \cdots, \lambda_{r+1}\}$.

Remark 3.1: The dual procedure for output frequency weighting is similar. In this case, one uses a frequency-weighted observability Gramian instead of a frequency-weighted controllability Gramian. One can also carry out a two-sided weighted balancing approximation by diagonalizing and equalizing the frequency-weighted controllability Gramian, and the frequency-weighted observability Gramian.

Remark 3.2: For input frequency-weighted or output frequency-weighted balancing approximation, the reduced-order system is generically asymptotically stable. But a proof of the stability of the reduced-order system is lacking for the two-sided frequency-weighted balancing reduction procedure.

Remark 3.3: No error bound formula analogous to (3.5) is available.

Remark 3.4: We now can apply this technique to the controller reduction problem. As we discussed before, we set $V(s) = G(s)[I + K(s)G(s)]^{-1} = C_i(sI - A_i)^{-1} B_i$ to be the input frequency weighting or the output frequency weighting for the reduction of controller $K(s)$.

Remark 3.5: If $K(s)$ is determined by combining an estimator and state feedback law, and is stable, then (3.6) generically has order $2n$. Otherwise, if $K(s)$ is stable, of order n, generically (3.6) has order $3n$. If $K(s)$ has unstable poles, these do not contribute to the order of (3.6), as we approximate only the stable part of $K(s)$.

Remark 3.6: An alternative method of introducing frequency weighting is suggested in the work of [21] on frequency-weighted Hankel norm reduction. First note that $\|(K - K_r)V\|_\infty = \|(K - K_r)\bar{V}\|_\infty$ where $\bar{V}(-s)$ and $\bar{V}^{-1}(-s)$ are stable and $VV^* = \bar{V}\bar{V}^*$ (given a nonsingular V with no $j\omega$-axis poles and zeros, such a $\bar{V}$ can always be found). Let $[\ ]_+$ denote the operation of taking the stable part. Let $L = [K\bar{V}]_+$. Let L_r approximate L (no frequency weighting) and set $K_r = [L_r\bar{V}^{-1}]_+$. Note that deg K = deg L, deg K_r = deg L_r, if K is stable.

Remark 3.7: There have been other approaches employing balancing to achieve controller reduction. Jonckheere and Silverman [25] suggested balancing of the two Riccati equations in the LQG design procedure, followed by truncation. As pointed out in [26], this scheme does not eliminate any uncontrollable or unobservable modes in the controller (which can arise in LQG design), and so it appears unattractive. Also no frequency domain error bound is available for it. The same idea was advanced also by Verriest; see [27]. Yousuff and Skelton [26] proposed use of an unweighted balancing approximation directly on the controller if it is stable. This scheme has a frequency domain error bound and has no restriction that the controller be obtained from an LQG design. Further, it certainly eliminates uncontrollable or unobservable controller modes. However, as we have argued, for controller reduction it is better to have frequency weighting to best maintain the stability margin or some aspect of the closed-loop performance. If the controller is unstable, a modification of the scheme is available [26], but the underlying rationale for the modification (apart from its practicality) is hard to see. A variant of the modification with more rationale was suggested by Davis and Skelton [28]. However, the same objection applies to this variant as applies to Jonckheere and Silverman's scheme.

B. Hankel Norm Approximation

Another very important model reduction approach is Hankel norm optimal approximation, which has been first considered in [13], then in [14] and [15]. Glover [16] characterized all stable approximations of a linear time-variant stable system $G(s)$ of McMillan degree n by $G_r(s)$ of McMillan degree r $(r < n)$ which minimize the "Hankel norm" error $\|G(s) - G_r(s)\|_H$. Note that an exact solution of an approximation problem is achieved, but the approximation problem solved is different from that for which a solution is desired, involving the Hankel norm rather than L_∞ norm. There is, however, a connection. If $G_r(s)$ is an optimal Hankel norm approximant of order r to $G(s)$, then

$$\|G-G_r\|_\infty \leqslant (\sigma_{r+1}+\cdots+\sigma_n) \tag{3.9}$$

while, as earlier noted, no rth-order approximant can ever achieve $\|G - G_r\|_\infty < \sigma_{r+1}$. Here the σ_i are the singular values appearing in the balancing realization theory, ordered with $\sigma_i \geqslant \sigma_{i+1}$. (They are also known as the Hankel singular values.) Glover's calculations actually involve manipulation of a balanced realization of G to get G_r, the manipulations being somewhat more complicated than mere truncation.

Note that (3.9) gives an error bound half that for balanced truncation, but in (3.9), $G_r(s)$ is allowed to have $G_r(\infty)$ nonzero, while in (3.5), $G_r(\infty)$ must be zero. Latham and Anderson [21] proposed a frequency-weighted version of the Hankel norm approximation to find a stable $G_r(s)$ of McMillan degree r, which minimized the Hankel norm error $\|[G(s) - G_r(s)]V(s)\|_H$ with $V(s)$ and $V^{-1}(s)$ completely unstable. (The approach of Remark 3.6 combined with Glover's method was used.) A formula bounding $\|[G(s) - G_r(s)]V(s)\|_\infty$ has been derived [22], although it is not as easy to evaluate the bound as in the unweighted case.

Reference [15] reports use of the unweighted procedure to simplify controllers. (The bulk of this reference suggests that a reduced-order controller can be constructed by reducing the plant using unweighted Hankel norm reduction, and designing a controller. However, Section III-D suggests Hankel norm reduction of the controller. In Example 1 presented below, it is this latter approach which is followed.)

C. q-COVER Approximation

The basic idea of q-covariance equivalent approximations applied to controller reduction is to approximate a high-order controller by a low-order one with two properties [17]–[20]:

i) the first q Markov parameters of the high- and low-order controllers are the same;

ii) the output covariances and their first q derivatives evaluated at time $t = 0$ of both high- and low-order controllers are equal. (This presupposes noise excitations to the closed loop.)

In the scalar case, q is the actual dimension of the reduced-order controller. The mechanics of the reduction procedure involve the use of Hessenberg form representations. The basic idea behind q-COVER approximation is to match transient behavior *and* steady-state behavior; the transient behavior is reflected in the Markov coefficients and the steady-state behavior in the covariance data. Of course, to the extent that frequency weighting cannot be accommodated, the technique is deficient; one must assume white noise excitation of the controllers. But note that equality of the first q-Markov parameters of two controllers implies (nontrivially) the corresponding equality for the first q-Markov parameters of the two associated *closed-loop* systems.

So the failure to use frequency weighting is more of concern in securing the right steady-state behavior. Preservation of closed-loop stability is not directly sought or secured in q-COVER approximation. Examples of reduced-order controller design using q-COVERS can be found in [20].

D. Additional Remarks

There are important, partly concealed, limitations of the methods, even frequency-weighted methods, espoused above.

- No approximation is actually optimal from an L_∞ point of view, although this is what we want.
- The approximation procedures do not cope well with nonstable controllers. It is *not* the case that copying of the unstable part of the high-order controller into the low-order controller is likely to be optimal, see, e.g., [29], and as noted in the last section, one can even query the appropriateness of the weighted L_∞ approximation problem when the high-order controller is unstable.
- Frequency-weighted Hankel-norm reduction can only be carried out for square plants (admittedly the major case of interest).
- Frequency-weighted q-COVER approximation has hardly been developed; see [23].
- In comparing frequency-weighted balanced truncation and frequency-weighted Hankel norm reduction on examples, the former has almost always proved to yield better approximations, despite the fact that in the unweighted case, Hankel norm reduction is superior.

Some of these issues are addressed in the succeeding section.

IV. Controller Reduction Using Fractional Representations

To this point, we have indicated at least three ways in which a frequency-weighted approximation problem can be posed to reflect the goal of controller reduction. We have also indicated three approaches for approximately solving these problems (albeit with various restrictions on two of the approaches). In this section, we shall indicate a further choice—that of controller representation. To this point, we have essentially considered a controller representation via a rational transfer function matrix $K(s)$, or as the sum of a stable rational $K_+(s)$ and an entirely unstable rational $K_-(s)$; in approximation, $K_+(s)$ is approximated as $K_{+r}(s)$, $K_-(s)$ is unaltered and copied into the reduced-order controller $K_r(s) = K_{+r}(s) + K_-(s)$. In this section, we shall switch from this additive view of a controller to a multiplicative view.

For the most part, in this section we shall consider controllers formed from a combination of a stabilizing feedback law and an estimator. At one point, we shall even suppose an LQG-based design. Thus, for a given linear time-invariant system $G(s) = C(sI_n - A)^{-1}B$, and an LQ index

$$J=\int_0^\infty (x'Qx+u'Ru)\,dt, \qquad Q'=Q\geqslant 0,\ R'=R>0 \quad (4.1)$$

we design the optimal feedback gain F which minimizes (4.1). Let L be the Kalman gain for the estimator. Then the compensator is

$$K(s)=F(sI_n-A+BF+LC)^{-1}L. \quad (4.2)$$

If matrices F, L are found by any method whatsoever such that $A - BF$ and $A - LC$ have all eigenvalues in Re $(s) < 0$, then F, L can be regarded as defining an estimator-based controller, and the formula (4.2) remains valid.

Now, irrespective of how F, L are determined, define

$$\begin{bmatrix} Y(s) & -N(s) \\ X(s) & D(s) \end{bmatrix} \triangleq \begin{bmatrix} I_l & 0 \\ 0 & I_m \end{bmatrix} + \begin{bmatrix} -F \\ C \end{bmatrix} (sI_n-A+BF)^{-1}[B\ \ L] \quad (4.3)$$

and

$$\begin{bmatrix} \tilde{D}(s) & \tilde{N}(s) \\ -\tilde{X}(s) & \tilde{Y}(s) \end{bmatrix} \triangleq \begin{bmatrix} I_l & 0 \\ 0 & I_m \end{bmatrix} + \begin{bmatrix} F \\ -C \end{bmatrix} (sI_n-A+LC)^{-1}[B\ \ L]. \quad (4.4)$$

Then it has been proven (see [30], [31]) that

$$G(s)=X(s)Y^{-1}(s)=\tilde{Y}^{-1}(s)\tilde{X}(s) \quad (4.5a)$$

$$K(s)=N(s)D^{-1}(s)=\tilde{D}^{-1}(s)\tilde{N}(s) \quad (4.5b)$$

and the Bezout identity holds

$$\begin{bmatrix} \tilde{D}(s) & \tilde{N}(s) \\ -\tilde{X}(s) & \tilde{Y}(s) \end{bmatrix}\begin{bmatrix} Y(s) & -N(s) \\ X(s) & D(s) \end{bmatrix} = \begin{bmatrix} I_l & 0 \\ 0 & I_m \end{bmatrix} = \begin{bmatrix} Y(s) & -N(s) \\ X(s) & D(s) \end{bmatrix} \cdot \begin{bmatrix} \tilde{D}(s) & \tilde{N}(s) \\ -\tilde{X}(s) & \tilde{Y}(s) \end{bmatrix}. \quad (4.6)$$

The transfer function matrices $N(s)$, $D(s)$, $\tilde{N}(s)$, $\tilde{D}(s)$, $X(s)$, $Y(s)$, $\tilde{X}(s)$, and $\tilde{Y}(s)$ are all stable. Hence, the Bezout identity (4.5) means that $N(s)D^{-1}(s)$ $(\tilde{D}^{-1}(s)\tilde{N}(s))$ is a stable right (left) coprime factorization of $K(s)$. At the same time, $X(s)Y^{-1}(s)$ $(\tilde{Y}^{-1}(s)\tilde{X}(s))$ is a stable right (left) coprime factorization of $G(s)$.

A. Coprime Factorization Reduction Without Frequency Weighting (Performance Considerations)

With the controller defined as in (4.2), we can draw the closed loop as shown in Fig. 5.

Now think of the controller as being defined by *an interconnection rule together with a stable transfer function matrix*

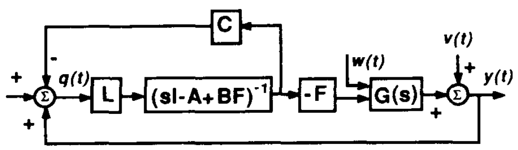


Fig. 5. State feedback/estimation based controller with plant.

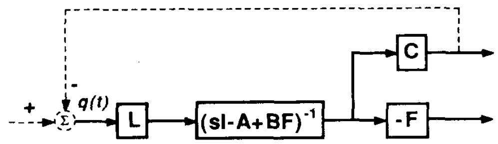


Fig. 6. "Controller" (heavy line) and interconnection rule (dashed line).

viz.

$$\begin{pmatrix} C \\ F \end{pmatrix} (sI_n - A + BF)^{-1} L = \begin{bmatrix} D(s) - I_m \\ N(s) \end{bmatrix} \tag{4.7}$$

(see Fig. 6—the minus sign attached to F is inessential). This suggests that we might approximate

$$\begin{bmatrix} D(s) - I_m \\ N(s) \end{bmatrix}$$

by some

$$\begin{bmatrix} D_1(s) - I_m \\ N_1(s) \end{bmatrix}$$

or, approximate

$$\begin{bmatrix} D(s) \\ N(s) \end{bmatrix} \text{ by } \begin{bmatrix} D_1(s) \\ N_1(s) \end{bmatrix}$$

with the second matrix having order r ($r < n$), and then recover a reduced-order controller by replicating the interconnection rule. Of course, this is equivalent to approximating the parts of a fractional representation of the controller [31], where such representations are over stable rational transfer functions. However, the utility of such an approximation must be verified. *How* we reduce the controller, i.e., balancing truncation, Hankel norm, or q-COVER is a separate issue. Note that there are no difficulties with instabilities in the object being approximated, as $D(s)$ and $N(s)$ are guaranteed stable.

Should we introduce frequency weighting? If we take the spectral (performance oriented) viewpoint, as opposed to the stability or closed-loop transfer function viewpoint, the answer is no. For the process $q(t)$ depicted in Fig. 5 is the *innovations* process, and as such is white, with covariance $V\delta(t - \tau)$, the covariance of the measurement noise process $v(t)$. The only weighting which we should then use is a *constant* matrix, $V^{1/2}$. Fortuitously, the redescription of the controller eliminates the frequency weighting; the idea of avoiding frequency weighting in this way comes indirectly from [20], and it allows application of q-COVER approximation without the concern that the need for frequency weighting is being overlooked.

As shown by examples in [32], this method of controller reduction can be very efficacious.

Remark 4.1: The dual approach to the above is to start with the left stable coprime factorization of $K(s) = \tilde{D}^{-1}(s)\tilde{N}(s)$, then to approximate $[\tilde{D}(s) - I_l \ \tilde{N}(s)] = F(sI_n - A + LC)^{-1}[B \ L]$ by some low order $[\tilde{D}_1(s) - I_l \ \tilde{N}_1(s)]$ and implement the reduced-order controller as $K_r(s) = \tilde{D}_1^{-1}(s)\tilde{N}_1(s)$. (The significance of the transfer function matrix being approximated is evident from Fig. 7.) In this case, however, we are unable to justify constant frequency weighting (at the output now, by duality, see [32, Remark (C)]). As it turns out, for some examples this procedure works well and for some it works badly. The Hankel singular values of a $[\tilde{D}(s) \ \tilde{N}(s)]$ can be very different to those of $[D'(s) \ N'(s)]'$, and in particular the number of significant singular values may be very different. It is the number of significant Hankel singular values which determines a rough lower limit to the order of a reduced-order controller achieving a good approximation, and having the procedure work badly is often equivalent to having a large proportion of the singular values being significant.

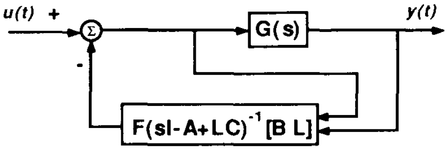


Fig. 7. State feedback law/estimator design of controller.

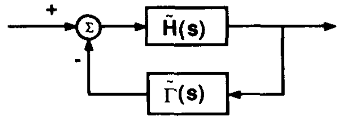


Fig. 8. Redrawing of the scheme of Fig. 7.

The fact that the efficacy of the procedure above and its dual cannot always be guaranteed (with primal and dual often having quite different quality of result in terms of stability) leads us to focus on the use of a stability-oriented method in conjunction with the above style of controller description. The key is to exploit the Bezout identity. This can be done when we start with either a left coprime or right coprime realization; we shall illustrate the former in the next subsection, having considered the latter in this subsection.

B. Coprime Factorization Reduction with Bezout Identity Induced Frequency Weighting (Stability Margin Consideration)

(The ideas of this subsection are developed at length in [33].) Refer again to Fig. 7, and regard the compensator as a two (vector) input, single (vector) output system with transfer function

$$\tilde{\Gamma}(s) = [\tilde{D}(s) - I_l \ \tilde{N}(s)]. \tag{4.8}$$

Regard the "plant" as defined by

$$\tilde{H}(s) = \begin{bmatrix} I_l \\ G(s) \end{bmatrix} \tag{4.9}$$

so that Fig. 7 is equivalent to Fig. 8. Let us seek to approximate $\tilde{\Gamma}(s)$ by a lower order stable $\tilde{\Gamma}_r(s)$. Taking the stability point of view, we should seek to minimize (see Section II)

$$\rho = \|[\tilde{\Gamma}(s) - \tilde{\Gamma}_r(s)]\tilde{H}(s)[I + \tilde{\Gamma}(s)\tilde{H}(s)]^{-1}\|_\infty. \tag{4.10}$$

Defining $\tilde{V}(s) = \tilde{H}(s)[I + \tilde{\Gamma}(s)\tilde{H}(s)]^{-1}$, it is easy, using Bezout identity, to prove that

$$\begin{aligned} \tilde{V}(s) &= \tilde{H}(s)[I + \tilde{\Gamma}(s)\tilde{H}(s)]^{-1} \\ &= \begin{bmatrix} Y(s) \\ X(s) \end{bmatrix} = \begin{bmatrix} I_l \\ 0 \end{bmatrix} + \begin{bmatrix} -F \\ C \end{bmatrix} (sI_n - A + BF)^{-1}B \end{aligned} \tag{4.11}$$

where $Y(s)$, $X(s)$ were as defined in (4.3). Hence, our goal becomes one of minimizing or finding a procedure that will approximately minimize

$$\tilde{\rho}=\left\| [\tilde{D}(s)-\tilde{D}_r(s)\ \ \tilde{N}(s)-\tilde{N}_r(s)] \begin{bmatrix} Y(s) \\ X(s) \end{bmatrix} \right\|_\infty \tag{4.12}$$

over stable $[\tilde{D}_r(s)\ \tilde{N}_r(s)]$ of prescribed degree.

It is not hard to verify that the weighting matrix in (4.12) is given by

$$\begin{bmatrix} Y(s) \\ X(s) \end{bmatrix} = \begin{bmatrix} (I+K(s)G(s))^{-1}\tilde{D}^{-1}(s) \\ G(s)(I+K(s)G(s))^{-1}\tilde{D}^{-1}(s) \end{bmatrix}$$

which shows that the reduction error is being weighted by a weighted version of two standard closed-loop transfer functions. The transfer function $G(s)(I + K(s)G(s))^{-1}$ appeared in Section II.

Provided $[\tilde{D}_r\ \tilde{N}_r]$ causes $\tilde{\rho} < 1$, we know from Section II-A that the controller based on $[\tilde{D}_r\ \tilde{N}_r]$ will necessarily be stabilizing. Obviously though, the smaller $\tilde{\rho}$ is, the better off we are likely to be.

Remark 4.2: A dual result is available. We think of the controller as being defined by

$$\Gamma(s)=\begin{bmatrix} D(s)^{-1} \\ N(s) \end{bmatrix}. \tag{4.13}$$

$$A=\begin{bmatrix}
-0.161 & -6.004 & -0.58215 & -9.9835 & -0.40727 & -3.982 & 0.0 & 0.0 \\
1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0
\end{bmatrix}$$

(See Fig. 6 again.) The actual controller transfer function matrix is $N(s)D^{-1}(s)$, and can be obtained by implementing $\Gamma(s)$ and an interconnection rule. Carrying through reasoning analogous to the above leads to the conclusion that we should use $[\tilde{Y}(s)\ \tilde{X}(s)]$ as the output frequency weighting, seeking D_r, N_r such that $[D_r'\ N_r']'$ has a prescribed degree, and

$$\rho=\left\| [\tilde{Y}(s)\ \tilde{X}(s)] \begin{bmatrix} D(s)-D_r(s) \\ N(s)-N_r(s) \end{bmatrix} \right\|_\infty \tag{4.14}$$

is minimum. We remark that $\rho < 1$ (which guarantees stability with the reduced-order controller) is itself guaranteed if

$$\left\| \begin{bmatrix} D-D_r \\ N-N_r \end{bmatrix} \right\|_\infty < \frac{1}{\|[\tilde{Y}\ \tilde{X}]\|_\infty}. \tag{4.15}$$

In [31], see Lemma 3.2, a more restrictive condition guaranteeing closed-loop stability was given on the approximating controller.

Remark 4.3: The frequency weightings are all stable, as are the components in the fractional representations of the controller. So there are no problems associated with open-loop instability (or $j\omega$-axis poles) of $K(s)$. Moreover, the methods do not require an underlying LQG assumption (in order, for example, to guarantee whiteness of some signal), but only the existence of F, L for which $A - BF$ and $A - LC$ have all left-half plane eigenvalues.

Remark 4.4: q-COVER approximation methods have not yet been completely extended to cope with weighting, and Hankel norm approximations require invertible weights. Hence, without further development, we necessarily use balanced truncation. The dimension of the key Lyapunov equation is $2n \times 2n$; however, simple algebraic transformations readily break this down, so that only one $n \times n$ equation has to be solved in obtaining the frequency-weighted controllability Gramian. Compare this to the usual frequency-weighted balanced truncation procedure where the dimension is $3n \times 3n$ (for an arbitrary nth-order controller) and $2n \times 2n$ (for an nth-order controller obtained from feedback law and estimator).

Remark 4.5: In the above description, we have used particular factional representations of plant and controller which are the subject of the special Bezout identity of [30]. A particularly attractive property is that the McMillan degree of ND^{-1} is generally that of $[N'\ D']'$ and similarly for $N_rD_r^{-1}$ and $[N_r'\ D_r']'$. So reduction of the McMillan degree of $[N'\ D']'$ through approximations by $[N_r'\ D_r']'$ is equivalent to reduction of the McMillan degree of ND^{-1} by $N_rD_r^{-1}$, which is our real objective. While in principle we could use any fractional representations of plant and controller for much of the above analysis, we would be defeated in terms of getting a useful result unless this crucial property concerning the McMillan degree was retained.

Example 1: We used the example in [32] to compare the effects of different controller reduction procedures including Bezout identity type of frequency weighting. The plant to be controlled is a four-disk system, represented as a linear, time-invariant, single input and single output, unstable, nonminimum-phase plant of eighth order. It was studied by Enns [7]. The plant $G(s) = C(sI - A)^{-1}B$ with transfer function described in [32] has minimal realization:

$$B=\begin{bmatrix} 1.0 \\ 0.0 \\ 0.0 \\ 0.0 \\ 0.0 \\ 0.0 \\ 0.0 \\ 0.0 \end{bmatrix} \quad C'=\begin{bmatrix} 0.0 \\ 0.0 \\ 6.4432\times10^{-3} \\ 2.3196\times10^{-3} \\ 7.1252\times10^{-2} \\ 1.0002 \\ 0.10455 \\ 0.99551 \end{bmatrix}. \tag{4.16}$$

(In [32] this is incorrectly termed an observable canonical form realization.) The weightings for an LQG design are given by $Q = q_1H'H$, $R = 1$, with $H = [0, 0, 0, 0, 0.55, 11, 1.32, 18]$ and $q_1 = 10^{-6}$. And the filter covariance matrices are $W = q_2BB'$, $V = 1$, where q_2 is a design parameter.

The following table extends [32, Table 2]. It depicts the closed-loop stability of the system with reduced-order controllers of different orders obtained by different reduction methods, starting with different LQG designs. These designs are obtained by adjusting the input noise intensity. For a minimum-phase plant, choosing q_2 large would correspond to loop transfer recovery based designs. The unweighted and weighted right coprime factorization design referred to in the table corresponds to the schemes set out in [32] and Section IV-A (unweighted), and described in Remark 4.2 (weighted).

In terms of ensuring closed-loop stability with the reduced-order controller, the right coprime factorization frequency-weighted method is the best.

It is clear that the worst methods are those with no weighting in the approximation; this is to be expected, as the plant is not taken into account in the approximation process. Two of the five methods are tied directly to stability considerations, viz. the Enns scheme and the scheme based on Bezout induced weighted with a right coprime realization. There is no technical reason presented in

TABLE I
CLOSED-LOOP STABILITY OF THE REDUCED-ORDER CONTROLLERS
(S = STABLE, U = UNSTABLE)

Controller Order	Controller Reduction Method	$q_2 = 100$	$q_2 = 1000$	$q_2 = 2000$
7	Unweighted Balanced Truncation (Yousuff and Skelton) [*Y*]	*U*	*U*	*U*
	(Davis and Skelton) [*D*]	*S*	*S*	*S*
	Unweighted Hankel-Norm Reduction	*S*	*U*	*S*
	Weighted Balanced Truncation (Enns) [*E*]	*S*	*S*	*S*
	Unweighted Right Coprime Factorization (Liu and Anderson) [*RCFU*]	*S*	*S*	*U*
	Weighted Right Coprime Factorization [*RCFW*]	*S*	*S*	*S*
6	*Y*	*U*	*U*	*U*
	D	*S*	*S*	*S*
	C	*U*	*U*	*U*
	E	*S*	*S*	*S*
	RCFU	*S*	*S*	*U*
	RCFW	*S*	*S*	*S*
5	*Y*	*U*	*U*	*U*
	D	*S*	*U*	*U*
	C	*U*	*U*	*U*
	E	*S*	*S*	*S*
	RCFU	*S*	*S*	*S*
	RCFW	*S*	*S*	*S*
4	*Y*	*U*	*U*	*U*
	D	*S*	*U*	*U*
	C	*U*	*U*	*U*
	E	*S*	*S*	*U*
	RCFU	*S*	*S*	*S*
	RCFW	*S*	*S*	*S*
3	*Y*	*U*	*U*	*U*
	D	*U*	*U*	*U*
	C	*U*	*U*	*U*
	E	*S*	*S*	*S*
	RCFU	*S*	*U*	*U*
	RCFW	*S*	*S*	*S*
2	*Y*	*U*	*U*	*U*
	D	*U*	*U*	*U*
	C	*S*	*U*	*U*
	E	*U*	*U*	*U*
	RCFU	*S*	*S*	*S*
	RCFW	*S*	*S*	*S*

this paper which might suggest that one scheme should be better than the other, except perhaps if the controller is open-loop unstable; this is, however, not the case here. The use of unweighted right coprime factorization here is neither better nor worse than the Enns scheme; because of its simplicity, and greater concentration on performance issues, it therefore merits consideration as one scheme to try. There are of course a number of other aspects of closed-loop performance, once stability has been ensured, which may lead to one arrangement being preferable to another; see [32] for a discussion of some of these for this example.

The author's opinion is that no one method is likely to prove universally more attractive than any other; Table I in effect shows that three methods (at least) are *prima facie* attractive for this example.

We remark that an extension of this table can be found in [34], in conjunction with a discussion of another type of controller reduction scheme. This scheme performs very well on this example; the key idea behind it is to throw away those entries of the controller state vector which are least correlated with the plant state vector.

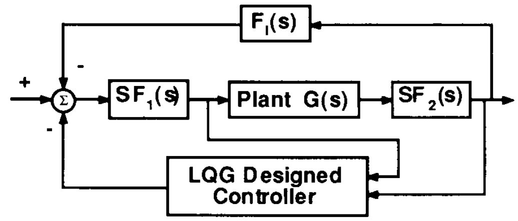


Fig. 9. General setup allowing integral control and/or shaping filter.

C. Coprime Factorization Reduction with More General Weighting

In many control system design problems, sometimes a single LQG designed controller is not enough. For instance, one may need an extra integral control loop to zero the steady-state error, or a shaping filter at input or even output to achieve some performance objective. (See Fig. 9 for an illustration that simultaneously encompasses several possibilities.) Controller reduction may be required with the constraint that shaping filters or integral feedback loops are maintained. We now seek to explain how this may easily be achieved, using the preceding ideas.

Let us consider a special case in detail.

Assume $F_I(s) = 0$, $SF_2(s) = I$, $SF_1(s) = D_F + C_F(sI - A_F)^{-1}B_F$, $G(s) = C(sI - A)^{-1}B$, and $\tilde{\Gamma}(s) = F(sI - A + LC)^{-1}[B, L]$ as the LQG controller. Then the closed-loop system is equivalent to that of Fig. 8 with

$$\tilde{H}(s) = \begin{bmatrix} I_l \\ G(s) \end{bmatrix} SF_1(s). \tag{4.17}$$

With $\tilde{\Gamma}_r(s)$ the reduced-order controller, we seek $\tilde{\Gamma}_r(s)$ to minimize

$$\|[\tilde{\Gamma}(s) - \tilde{\Gamma}_r(s)]\, V(s)\| \tag{4.18}$$

where

$$V(s) = \tilde{H}(s)[I + \tilde{\Gamma}(s)\tilde{H}(s)]^{-1}. \tag{4.19}$$

It is not too difficult to show that

$$V(s) = \begin{bmatrix} D_F \\ 0 \end{bmatrix} + \begin{bmatrix} C_F & -D_F F \\ 0 & C \end{bmatrix} \cdot \left[sI_{n+f} - \begin{bmatrix} A_F & -B_F F \\ BC_F & A - BD_F F \end{bmatrix} \right]^{-1} \begin{bmatrix} B_F \\ BD_F \end{bmatrix}. \tag{4.20}$$

It is very interesting to note that now the frequency weighting has order $n + f$. Were we to use the nonfactorization procedure of Enns' frequency weighting, the order of weighting would be of the order of $2n + f$ in this case; worse, the shaping filter structure is lost.

Note here that generally speaking if the filter $SF_1(s)$ is not a constant, we do not readily obtain a dual formulation of the same problem.

V. CONTROLLER REDUCTION WITH CONSIDERATION OF THE CLOSED-LOOP PERFORMANCE

It is obvious that the purpose of controller reduction is not only to maintain closed-loop stability but also to maintain the closed-loop performance as much as possible. In previous sections, we have concentrated on the stability problem and it is evident that the

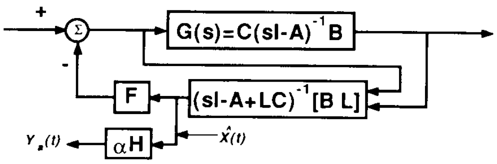


Fig. 10. Introduction of artificial output.

Bezout identity type of frequency weighting in controller reduction handles this issue well. Now, we shift to the performance problem.

We recall the LQG design procedure described in Section IV. For the given linear time-invariant system $G(s) = C(sI - A)^{-1} B$ and the LQ index

$$J=\int_0^{\infty} (x'Qx+u'Ru)\, dt \qquad Q'=Q \geqslant 0,\ R'=R>0 \quad (5.1)$$

one can design the feedback gain F which minimizes (5.1) by full state feedback law $u = -Fx$. If not all states are measurable or there is process noise and measurement noise, one has to design an estimator to obtain the estimate of the states $\hat{x}$ to form the feedback law as $u = -F\hat{x}$.

Let $Q = H'H$. The index (5.1) then suggests that we should be concerned about getting a good approximation of Hx, or the next best thing, $H\hat{x}$. Continuing this heuristic argument, this suggests that we add an extra output y_a to the controller, see Fig. 10, and reflect this into our statement of the approximation problem.

Thus, we might seek to approximate

$$\begin{bmatrix} F \\ H \end{bmatrix} (sI_n - A + LC)^{-1}[B\ L]$$

rather than just $F(sI_n - A + LC)^{-1}[B\, L]$, as in the right coprime factorization methods. To develop the idea further, note that the performance index focuses our attention on $u = -F\hat{x}$, or more accurately $R^{1/2}u = -R^{1/2}F\hat{x}$. This suggests that it might be worthwhile to replace F by $R^{1/2}F$ (the negative sign is inessential). Recall that without the presence of H we are "stability-based" in our thinking, whereas introduction of H allows performance-based thinking. By introducing a scalar parameter $\alpha > 0$, we can adjust our relative weighting of the two. This leads us then to focus on

$$\begin{bmatrix} R^{1/2}F \\ \alpha H \end{bmatrix} (sI_n - A + LC)^{-1}[B\ L]$$

as the object to be reduced. Write the reduced-order object as

$$\begin{bmatrix} R^{1/2}F_1 \\ \alpha H_1 \end{bmatrix} (sI_r - A_1)^{-1}[B_1\ L_1].$$

Then the reduced-order controller is defined by $F_1(sI_r - A_1)^{-1}[B_1\, L_1]$. Both the unweighted left coprime factorization scheme, see Remark 4.1 [32, Remark (C)], and the frequency-weighted scheme, see Section IV-B, can be used for reduction, with the latter to be preferred, addressing as it does the stability problem directly.

Note also that the scaling $R^{1/2}$ and α play a very important role in improving the closed-loop performance. Although $R^{1/2}$ is fixed, it is very important to use it as a scaling for multiinput systems. A properly chosen α can make much difference in performance; in many cases, if we choose large α, we can (not unexpectedly) run into instability problems. Note also that since H has m rows, we can in principle use a diagonal scaling matrix $\Lambda = \text{diag}\,\{\alpha_1, \alpha_2, \cdots, \alpha_m\}$ instead of the scalar α in reduction.

Example 2: Now we use an example to illustrate how the artificial output involvement in reduction can affect the closed-loop performances in the coprime factorization reduction scheme, with or without frequency weighting. The system and LQG controller design setup are the same as for Example 1. The index is

$$J=\int_0^{\infty} (x'Qx+u'Ru)\, dt=\int_0^{\infty} (x'H'q_1Hx+u^2)\, dt$$

where q_1 and H are defined as in Example 1 and $R = 1$. In the first case, with the design parameter $q_2 = 1.0$, we reduce the eighth-order controller to a seventh-order one without use of frequency weighting, via the left coprime factorization scheme. As illustrated in Fig. 11, if we do not involve the artificial output in the reduction (i.e., $\alpha = 0$), the closed-loop performance as indicated by the step response is a very poor approximation. But if we use $\alpha = 9$ in balancing $[F'\ \alpha H']'\,(sI_n - A + LC)^{-1}\,[B\ L]$, then reducing the controller, the closed-loop performance improves dramatically. The gain margins for the three cases are 14.38 dB (full order), −3.55 dB and 14.26 dB ($\alpha = 0$), and 14.65 dB ($\alpha = 9$). The phase margins are (in degrees) 48.86 (full order), 22.34 ($\alpha = 0$), and 47.9 ($\alpha = 9$). So there is apparently little damage to robustness in using $\alpha = 9$. In the second case, with the design parameter $q_2 = 100$ and with $\alpha = 0$, the unweighted left coprime factorization balancing reduction scheme yielded a stabilizing sixth-order controller (with lower order controllers not stabilizing). If we use the Bezout identity type frequency weighting for reduction (frequency-weighted balancing), still with $\alpha = 0$, we obtain all stabilizing controllers from third- to seventh-order (another indication of the effectiveness of the Bezout identity type of frequency weighting in preserving stability). Then we introduce the artificial output to the reduction scheme. A properly chosen scaling factor α can make a significant difference in the closed-loop performance, as is illustrated in Fig. 12, where we reduced the controller to fifth-order with $\alpha = 35$ (via frequency-weighted balancing). The gain margins are 16.95 dB (full order), 17.76 dB ($\alpha = 0$), −15.92 dB, and 14.45 dB ($\alpha = 35$). The phase margins are (in degrees) 53.12, 12.53, and 41.06. Again, there is little damage to robustness.

We have also used this reduction scheme on a complicated example, the pitch control system of an F-111 airplane. The plant is a multiinput and multioutput system with twenty third-order LQG controller, fifth-order filter, and integral control loop, and the controller is reduced to fourth order. Performance appears very similar to that obtained with a preexisting fourth-order controller obtained via modal reduction. The results for this example will be reported elsewhere.

VI. Conclusions

Throughout this paper, we have exhibited many choices facing those seeking to reduce the order of a controller. A number can be listed as follows:

- frequency-weighted/unweighted approximation;
- choice of weight based on stability margin, spectrum, or closed-loop transfer functions;
- reduction method based on balanced truncation, Hankel norm, or q-COVER;
- conventional representation or fractional representation;
- exclusion/inclusion of augmented output to improve performance.

It is crucial that the plant be taken into account in the choice of method for controller simplification, and this means use of frequency weighting in all cases save when a right coprime factorization of the plant is being used, and performance or spectral considerations are used to define the frequency domain approximation problem.

Several other rules of thumb can be suggested. Balanced truncation appears superior to Hankel norm approximation and fractional representations should certainly be considered for open-loop unstable controllers. Computational burdens tend to be lower for the fractional representation approaches (provided that the

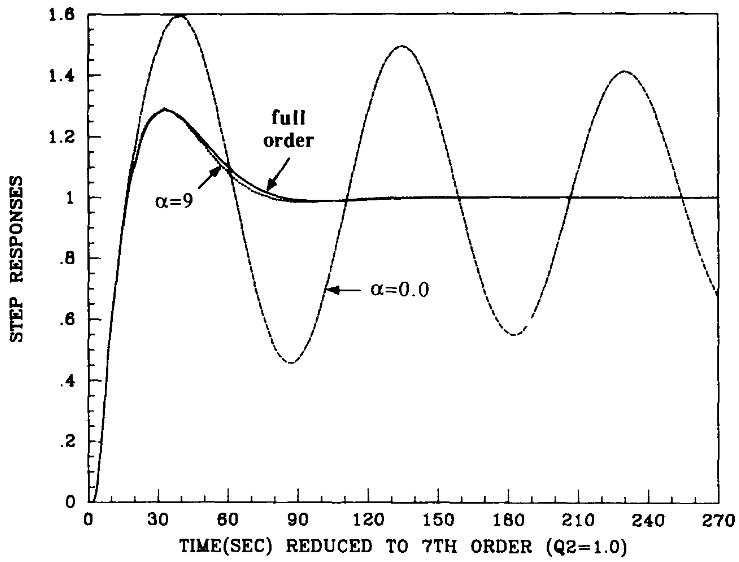


Fig. 11. Effect of considering performance in reduction.

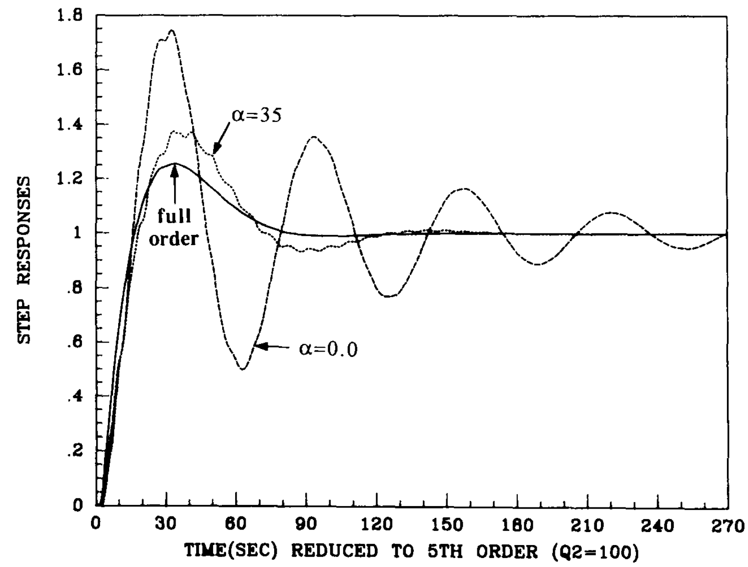


Fig. 12. Effect of considering performance in reduction.

controller is a combination of state estimation with feedback law). However, the main rule of thumb we would advance is, in relation to any one problem to check a multiplicity of approaches. Just as in classical control, it is recognized that different control objectives can conflict, and different controller designs can all be legitimate, so is this true with more complicated control problems.

We have also thrown up a number of questions or open problems, for example:

- frequency-weighted q-COVER reduction;
- preservation of stability in balanced truncation with input *and* output weighting;
- nonsquare frequency-weighted Hankel norm reduction;
- reduction of unstable systems;
- advance prediction that a particular reduction method will out-perform another.

In connection with this last point, let us note that examples and recent unpublished work [35] suggest that left fractional controller representations are to be preferred to right representations if the filter has been designed using loop recovery ideas, and right representations preferred to left representations if the LQ law has been designed using loop recovery ideas. Moreover, in using the preferred representation there will often be little difference between the effect of frequency-weighted or unweighted reduction.

Two other theoretical issues, not raised in the paper to this point, should be noted:

- the possibility of using multiplicative rather than additive error criteria; in this connection, see [36] for results on (unweighted) multiplicative model reduction;
- the need to use correct scaling of inputs and outputs before approximation; see [37] for a number of valuable insights.

Acknowledgment

Many helpful discussions with U. L. Ly and J. B. Moore are gratefully acknowledged.

References

[1] D. Gangsaas, K. R. Bruce, J. D. Blight, and U. -L. Ly, "Application of modern synthesis to aircraft control: Three case studies," *IEEE Trans. Automat. Contr.*, vol. AC-31, pp. 995-1104, Nov. 1986.

[2] U. -L. Ly, "A design algorithm for robust low order controller," Ph.D. dissertation, Dep. Aeronaut. Astronaut., Stanford Univ., Stanford, CA, 1982.

[3] D. S. Bernstein and D. C. Hyland, "The optimal projection equations for fixed-order dynamic compensation," *IEEE Trans. Automat. Contr.*, vol. AC-29, pp. 1034-1037, Nov. 1985.

[4] D. C. Hyland and D. S. Bernstein, "The optimal projection equations for model reduction and the relationships among the methods of Wilson, Skelton, and Moore," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 1201-1211, Dec. 1985.

[5] D. S. Bernstein and D. C. Hyland, "Optimal projection/maximum entropy stochastic modeling and reduced-order design synthesis," in *Proc. IFAC Workshop on Model Errors and Compensation,* Boston, MA, 1985.

[6] S. W. Greeley and D. C. Hyland, "Reduced order compensation: LQG reduction versus optimal projection using a homotopic continuation method," in *Proc. 26th IEEE Conf. Decision Contr.*, Los Angeles, CA, 1987, pp. 742-747.

[7] R. E. Skelton, *Dynamic Systems Control.* New York: Wiley, 1988.

[8] E. A. Jonckheere and L. M. Silverman, "Singular value analysis of deformable systems," in *Proc. 20th IEEE Conf. Decision Contr.*, Los Angeles, CA, 1981.

[9] D. Enns, "Model reduction for control systems design," Ph.D. dissertation, Dep. Aeronaut. Astronaut., Stanford Univ., Stanford, CA, 1984.

[10] J. C. Doyle and G. Stein, "Multivariable feedback design: Concepts for a classical/modern synthesis," *IEEE Trans. Automat. Contr.*, vol. AC-26, no. 1, pp. 4-16, 1981.

[11] B. C. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Automat. Contr.*, vol. AC-26, no. 1, pp. 17-32, 1982.

[12] L. Pernebo and L. M. Silverman, "Modern reduction via balanced state space representations," *IEEE Trans. Automat. Contr.*, vol. AC-27, no. 2, pp. 382-387, 1982.

[13] V. M. Adamjan, D. Z. Arov, and M. G. Krein, "Analytic properties of Schmidt pairs for Hankel operator and the generalized Schur-Takagi problem," *Math. USSR Sbornik*, vol. 15, no. 1, pp. 31-73, 1971.

[14] S. Y. Kung and D. W. Lin, "Optimal Hankel-norm model reductions: Multivariable systems," *IEEE Trans. Automat. Contr.*, vol. AC-26, no. 4, pp. 832-852, 1981.

[15] K. Glover and D. J. N. Limebeer, "Robust multivariable control system design using optimal reduced order plant methods," in *Proc. ACC*, San Francisco, CA, 1983, p. 664.

[16] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds," *Int. J. Contr.*, vol. 39, no. 6, pp. 1115-1193, 1984.

[17] A. Yousuff, D. A. Wagie, and R. E. Skelton, "Linear system approximation via covariance equivalent realizations," *J. Math. Anal. Appl.*, vol. 106, no. 1, pp. 91-115, 1985.

[18] A. Yousuff and R. E. Skelton, "Covariance equivalent realizations with application to model reduction of large-scale systems," *Contr. Dynam. Syst.*, vol. 22, pp. 273-348, 1985.

[19] B. D. O. Anderson and R. E. Skelton, "The generation of all q-Markov covers," *IEEE Trans. Circuits Syst.*, vol. CAS-35, no. 4, pp. 375-384, 1988.

[20] A. Yousuff and R. E. Skelton, "Controller reduction by component cost analysis," *IEEE Trans. Automat. Contr.*, vol. AC-29, no. 6, pp. 520-530, 1984.

[21] G. A. Latham and B. D. O. Anderson, "Frequency-weighted optimal Hankel norm approximation of stable transfer functions," *Syst. Contr. Lett.*, vol. 5, no. 4, pp. 229-236, 1985.

[22] B. D. O. Anderson, "Weighted Hankel norm approximation: Calculation of bounds," *Syst. Contr. Lett.*, vol. 7, no. 4, pp. 247-255, 1986.

[23] R. E. Skelton and B. D. O. Anderson, "Weighted q-Markov covers," submitted for publication.

[24] D. F. Enns, "Model reduction with balanced realizations: An error bound and a frequency weighted generalization," in *Proc. 23rd Conf. Decision Contr.*, Las Vegas, NV, 1984, pp. 127-132.

[25] E. A. Jonckheere and L. M. Silverman, "A new set of invariants for linear systems—Application and approximation," presented at the Int. Symp. Theory Networks & Syst., Santa Monica, CA, 1981.

[26] A. Yousuff and R. E. Skelton, "A note on balanced controller reduction," *IEEE Trans. Automat. Contr.*, vol. AC-29, no. 3, pp. 254-257, 1984.

[27] E. J. Verriest, "Suboptimal LQG design via balanced realizations," in *Proc. 20th IEEE Conf. Decision Contr.*, San Diego, CA, 1981, pp. 686-687.

[28] J. A. Davis and R. E. Skelton, "Another balanced controller reduction algorithm," *Syst. Contr. Lett.*, vol. 4, no. 2, pp. 79-83, 1984.

[29] P. J. Parker and B. D. O. Anderson, "Unstable rational function approximation," *Int. J. Contr.*, vol. 46, no. 5, pp. 1783-1801, 1987.

[30] C. N. Nett, C. A. Jacobson, and M. J. Balas, "A connection between state-space and doubly coprime fractional representations," *IEEE Trans. Automat. Contr.*, vol. AC-29, no. 9, pp. 831-832, 1984.

[31] M. Vidyasagar, *Control System Synthesis: A Factorization Approach.* Cambridge, MA: M.I.T. Press, 1985.

[32] Y. Liu and B. D. O. Anderson, "Controller reduction via stable factorization and balancing," *Int. J. Contr.*, vol. 44, no. 2, pp. 507-531, 1986.

[33] Y. Liu, B. D. O. Anderson, and U. -L. Ly, "Coprime factorization controller reductions with Bezout identity induced frequency weighting," submitted for publication.

[34] C. de Villemagne and R. E. Skelton, "Controller reduction using canonical interactions," *IEEE Trans. Automat. Contr.*, vol. 33, pp. 740-750, Aug. 1988.

[35] Y. Liu and B. D. O. Anderson, "Frequency weighted controller reduction methods and loop transfer recovery," submitted for publication.

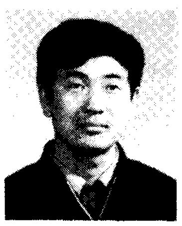
[36] K. Glover, "Multiplicative approximation of linear multivariable systems with L^∞ error bounds," in *Proc. ACC*, Seattle, WA, 1986, pp. 1705-1709.

[37] J. P. Keller and D. Bonvin, "Selection of inputs for the purpose of model reduction and control design," in *Proc. 10th IFAC World Congress*, Munich, 1987, vol. 2, pp. 226-231.

Brian D. O. Anderson (S'62-M'66-SM'74-F'75) was born in Sydney, Australia, in 1941. He received the B.S. degree in pure mathematics and electrical engineering from the University of Sydney, Sydney, Australia, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1966.

He is currently Professor and Head of the Department of Systems Engineering at the Australian National University, Canberra, Australia. From 1967 through 1981, he was Professor of Electrical Engineering at the University of Newcastle, New South Wales, Australia. He has also held appointments as Visiting Professor at a number of universities in the United States, Australia, and Europe. He is coauthor of several books and his research interests are in control and signal processing.

Dr. Anderson is a Fellow of the Australian Academy of Science, the Australian Academy of Technological Sciences, and the Royal Society, and an Honorary Fellow of the Institute of Engineers, Australia. Previously, he was Editor of *Automatica* and is currently President-Elect of IFAC.



Yi Liu (S'87) was born in Kunming, China, in 1958. He received the B.Sc. degree in mathematics from Zhongshan (Sun Yatsen) University, Guangzhou, China, in 1982, and the Ph.D. degree in systems engineering from the Australian National University, Canberra, Australia, in 1989.

From 1982 to 1985, he was a Research Assistant in the Control Theory Laboratory, Institute of System Sciences, Academia Sinica, Beijing, China. His research interests are in multivariable control, theoretical and computational aspects of model and controller reduction, and H^∞-norm optimal design.