Design of Multi-Input Multi-Output Systems with Minimum Sensitivity

WILLIAM J. LUTZ AND S. LOUIS HAKIMI, FELLOW, IEEE

Abstract —A state-space approach is used to find a state-space realization of a transfer function $\hat{h}(s)$ which minimizes a sensitivity measure. The sensitivity measure to be minimized is the continuous-time equivalent of a sensitivity measure defined by Tavsanoğlu and Thiele [6] for digital filters. A class of realizations that minimizes this sensitivity measure is discussed and a method for finding the sparsest realization in this class is presented. The theory is extended to include the case of multi-input multi-output (MIMO) systems realizing a transfer function matrix $\hat{H}(s)$.

I. Introduction

THE STATE-SPACE design of a continuous-time filter consists of finding a suitable set of state-space equations that realize a desired transfer function $\hat{h}(s)$. Then, an active realization that simulates these equations yields the desired circuit. The state-space approach to the design of RC-active filters was first introduced by Kerwin *et al.* [1], who used this approach mainly to design high-quality second-order biquadratic filters. The state-space equations corresponding to a transfer function $\hat{h}(s)$ are not unique, thus one may select among the realizations of $\hat{h}(s)$ one (or a subclass) that minimizes a suitable sensitivity measure. Motivated by this observation, MacKay and Sedra [2] used a numerical optimization technique to find low-sensitivity state-space filters.

There has also been much research on low-sensitivity state-space digital filters, e.g., see [3]–[5]. In particular, Mullis and Roberts [5] have produced closed-form results concerning minimum round off noise digital filters. Recently, Tavsanoğlu and Thiele [6] and Thiele [7] have considered the optimal design of state-space digital filters which simultaneously minimize the coefficient sensitivity and roundoff noise. Of particular interest here is the sensitivity measure used by Tavsanoğlu and Thiele and the fact that Thiele has derived the conditions to minimize the sensitivity measure. This means that an optimal design can be found directly.

Here we use a scalar sensitivity measure for continuous-time state-space filters that is equivalent to the sensitivity measure for digital filters introduced by Tavsanoğlu and Thiele [6]. Thiele's conditions to minimize the sensitivity measure also apply to the continuous-time case. A class of realizations that minimize the sensitivity measure is discussed and it is shown how orthogonal similarity transformations can be used to produce the sparsest realization (i.e., the realization with the least number of components) in the class.

The major result of this paper is to extend this theory to include the case of multi-input multi-output (MIMO) systems realizing a transfer function matrix $\hat{H}(s)$.

Manuscript received February 25, 1986; revised October 30, 1987. This work was supported in part by the National Science Foundation under Grant DMC-8406854 and under Grant ECS-8511211. This paper was recommended by Associate Editor R. W. Liu.

W. J. Lutz is with Nicolet Instrument Corporation, Madison, WI 53711.

S. L. Hakimi is with the Department of Electrical Engineering and Computer Science, University of California, Davis, CA 95616.

IEEE Log Number 8822396.

II. Preliminary Observations and Notation

A. State-Space Formulation

To design a filter one normally begins with its specified transfer function $\hat{h}(s)$. The most general form for a scalar, single-input single-output (SISO), nth-order transfer function is

$$\hat{h}(s)=\hat{h}(\infty)+\frac{b_{n-1}s^{n-1}+\cdots+b_1s+b_0}{s^n+a_{n-1}s^{n-1}+\cdots+a_1s+a_0}$$

where the coefficients a_i and b_i are real constants. It is assumed that $\hat{h}(s)$ has no poles in the closed right half plane. It is also assumed that $\hat{h}(\infty)=0$, and thus $\hat{h}(s)$ is a strictly proper stable rational function. It can be seen that this assumption does not lead to a loss of generality. It is well known that such a function $\hat{h}(s)$ can be represented by a set of state equations in the form

$$\dot{x}(t)=Ax(t)+bu(t)$$
$$y(t)=c^Tx(t)$$

where $x(t)$ is an n-dimensional state vector, $u(t)$ is the scalar input, $y(t)$ is the scalar output, $A\in\mathbb{R}^{n\times n}$, and $b,c\in\mathbb{R}^n$. An active realization, using op-amps and R and C elements, of the above state-space description of the filter provides us with a realization of $\hat{h}(s)$. See MacKay and Sedra for details [2].

For convenience, we refer to the state-variable equations as the set $\{A,b,c^T\}$. It is easily shown that the transfer function associated with $\{A,b,c^T\}$ is

$$\hat{h}(s)=c^T(sI-A)^{-1}b.$$

The state-space equations $\{A,b,c^T\}$ corresponding to a transfer function $\hat{h}(s)$ are not unique, since for any arbitrary nonsingular matrix T,

$$\{\tilde{A},\tilde{b},\tilde{c}^T\}=\{T^{-1}AT,T^{-1}b,c^TT\}$$

will also have transfer function $\hat{h}(s)$. In what follows, we find a realization $\{\tilde{A}, \tilde{b}, \tilde{c}^T\}$ that has $\hat{h}(s)$ as its transfer function and also minimizes a scalar sensitivity measure.

Two column n-vectors of transfer functions associated with $\{A, b, c^T\}$ are defined as

$$\hat{f}(s) = [\hat{f}_1(s) \cdots \hat{f}_n(s)]^T \triangleq (sI - A)^{-1}b$$

$$\hat{g}(s) = [\hat{g}_1(s) \cdots \hat{g}_n(s)]^T \triangleq (sI - A^T)^{-1}c.$$

These two vectors play a central role in the theory that follows. It is easily shown that each entry of $\hat{f}(s)$ or $\hat{g}(s)$ is a strictly proper stable rational function.

It can be seen that $f(t) \triangleq \mathscr{L}^{-1}\{\hat{f}(s)\} = e^{At}b$ and $g(t) \triangleq \mathscr{L}^{-1}\{\hat{g}(s)\} = e^{A^T t}c$, where $\mathscr{L}^{-1}$ denotes the inverse Laplace transform. Associated with $f(t)$ and $g(t)$ are the controllability Gramian matrix W_c and the observability Gramian matrix W_o, respectively. These matrices are defined as [8]

$$W_c \triangleq \int_0^\infty f(t) f^T(t)\, dt = \int_0^\infty e^{At} b b^T e^{A^T t}\, dt$$

and

$$W_o \triangleq \int_0^\infty g(t) g^T(t)\, dt = \int_0^\infty e^{A^T t} c c^T e^{At}\, dt.$$

We assume $\hat{h}(s)$ has no pole–zero cancellations, and thus $\{A, b, c^T\}$ is completely controllable and completely observable. Then, the controllability Gramian W_c and the observability Gramian W_o are both nonsingular symmetric, positive definite $n \times n$ constant matrices [9].

B. Sensitivity Functions

In practice it is not possible to realize the coefficients of $\{A, b, c^T\}$ exactly. Thus for the realization $\{A, b, c^T\}$, with transfer function $\hat{h}(s)$, we define the sensitivity of $\hat{h}(s)$ with respect to the elements of A, b, and c to be the (partial) derivative of $\hat{h}(s)$ with respect to these elements. The derivative of $\hat{h}(s)$ with respect to the matrix A, denoted $\partial\hat{h}(s)/\partial A$, is an $n \times n$ matrix whose ijth entry is $\partial\hat{h}(s)/\partial a_{ij}$. The derivative of $\hat{h}(s)$ with respect to the vector b, denoted $\partial\hat{h}(s)/\partial b$, is an $n \times 1$ column vector, whose ith entry is $\partial\hat{h}(s)/\partial b_i$. Similarly, $\partial\hat{h}(s)/\partial c$ is an $n \times 1$ column vector, whose ith entry is $\partial\hat{h}(s)/\partial c_i$.

It can be shown that (e.g., see [10])

$$\frac{\partial\hat{h}(s)}{\partial a_{ij}} = \hat{g}_i(s)\hat{f}_j(s), \quad \frac{\partial\hat{h}(s)}{\partial b_i} = \hat{g}_i(s), \quad \text{and}$$

$$\frac{\partial\hat{h}(s)}{\partial c_i} = \hat{f}_i(s).$$

Then, it is obvious that $\partial\hat{h}(s)/\partial b = \hat{g}(s)$ and $\partial\hat{h}(s)/\partial c = \hat{f}(s)$. Using $\hat{g}_i(s)\hat{f}_j(s)$ as the ijth element in the definition of $\partial\hat{h}(s)/\partial A$ gives the result $\partial\hat{h}(s)/\partial A = \hat{g}(s)\hat{f}^T(s)$.

C. Norms

The coefficient sensitivity functions $\partial\hat{h}(s)/\partial A = \hat{g}(s)\hat{f}^T(s)$, $\partial\hat{h}(s)/\partial b = \hat{g}(s)$, and $\partial\hat{h}(s)/\partial c = \hat{f}(s)$ are not attractive for comparing the sensitivity of various realizations, $\{\tilde{A}, \tilde{b}, \tilde{c}^T\}$, of a transfer function $\hat{h}(s)$ because they are matrix-valued and vector-valued functions of s. What we need for comparison purposes is a suitable norm. To generate a norm for such matrix-valued (vector-valued) functions, we can combine those for constant matrices (vectors) and scalar-valued functions as follows.

Let $\hat{A}(s)$ be an $m \times n$ matrix with elements that are strictly proper stable rational functions. Then $\hat{A}(j\omega) \in L_p^{m\times n}(-\infty, \infty)$. Let $\|\circ\|_F$ denote the Frobenius norm of a matrix in $L_p^{m\times n}$, i.e.,

$$\|\hat{A}(j\omega)\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |\hat{a}_{ij}(j\omega)|^2\right)^{1/2} = \left(\operatorname{tr}\left(\hat{A}^*(j\omega)\hat{A}(j\omega)\right)\right)^{1/2}.$$

Then, we define the "matrix L_p norm" to be

$$\|\hat{A}\|_{F,L_p} \triangleq \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \|\hat{A}(j\omega)\|_F^p\, d\omega\right)^{1/p}, \qquad \text{for } 1 \leqslant p < \infty.$$

This definition is suitable even for an $m \times 1$ matrix, i.e., a vector. Note that the Frobenius matrix norm becomes the familiar Euclidean vector norm for this case. Taking the matrix L_p norm of a coefficient sensitivity function is equivalent to applying the well-known Schoeffler criterion [11] to a coefficient sensitivity function, and then taking an L_p norm of the resulting single real-valued function of ω.

We will only use the L_1 and L_2 norms. We note that the L_2 norm can often be given a physical interpretation as a measure of the energy in a signal set. Even more important, we can express the L_2 norm in the time domain by using Parseval's theorem. One can easily show that the time domain equivalent of $\|\hat{A}\|_{F,L_2}$ becomes (apply Parseval's theorem term by term)

$$\|A\|_{F,L_2} = \left(\int_0^\infty \|A(t)\|_F^2\, dt\right)^{1/2}.$$

Noting that $\|A\|_F^2 = \operatorname{tr}(A(t)A^T(t))$, and interchanging trace and integration, we can write $\|A\|_{F,L_2}^2 = \operatorname{tr}\int_0^\infty A(t)A^T(t)\, dt$. We also note that, in general, the Gramian of the matrix $A(t)$ is $W \triangleq \int_0^\infty A(t)A^T(t)\, dt$. Thus the matrix L_2 norm squared of an $m \times n$ matrix $A(t)$, whose elements are real functions of time, is simply the trace of the Gramian of the matrix $A(t)$. This important result is due to Moore [8].

D. Scalar Sensitivity Measure

We will use the scalar sensitivity measure M given by

$$M = \left\|\frac{\partial\hat{h}}{\partial A}\right\|_{F,L_1}^2 + \left\|\frac{\partial\hat{h}}{\partial b}\right\|_{E,L_2}^2 + \left\|\frac{\partial\hat{h}}{\partial c}\right\|_{E,L_2}^2$$

$$= \left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\left\|\frac{\partial\hat{h}}{\partial A}\right\|_F dw\right)^2 + \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\|\frac{\partial\hat{h}}{\partial b}\right\|_E^2 dw + \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\|\frac{\partial\hat{h}}{\partial c}\right\|_E^2 dw.$$

This is the sensitivity measure used by Tavsanoğlu and Thiele for digital filters [6].

The reader can easily verify that

$$\left\| \frac{\partial \hat{h}(j\omega)}{\partial A} \right\|_F = \|\hat{g}(j\omega)\|_E \|\hat{f}(j\omega)\|_E.$$

Thus we can write

$$M = \frac{1}{2\pi}\int_{-\infty}^{\infty} \|\hat{g}\|_E \|\hat{f}\|_E \, d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty} \|\hat{g}\|_E^2 \, d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty} \|\hat{f}\|_E^2 \, d\omega.$$

For convenience, we let M_1 denote the first term and M_2 denote the sum of the last two terms in the expression for M. Using Parseval's theorem, M_2 can be expressed in the time domain as $M_2 = \int_0^\infty \|g(t)\|_E^2 \, dt + \int_0^\infty \|f(t)\|_E^2 \, dt$. Using the result of Moore [8], we see that $M_2 = \operatorname{tr} W_o + \operatorname{tr} W_c$.

We would like to be able to express M_1 as a product of a norm involving $\hat{g}(s)$ and a norm involving $\hat{f}(s)$. Using the Cauchy–Schwarz inequality, we can write

$$\begin{aligned} M_1 &= \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \left\| \frac{\partial \hat{h}(j\omega)}{\partial A} \right\|_F d\omega \right)^2 \\ &= \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \|\hat{g}\|_E \|\hat{f}\|_E \, d\omega \right)^2 \\ &\leqslant \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \|\hat{g}\|_E^2 \, d\omega \right) \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \|\hat{f}\|_E^2 \, d\omega \right) \\ &= \left(\int_0^{\infty} \|g(t)\|_E^2 \, dt \right) \left(\int_0^{\infty} \|f(t)\|_E^2 \, dt \right) \\ &= \operatorname{tr} W_o \operatorname{tr} W_c. \end{aligned}$$

This result is the reason for taking the L_1 norm of $\|\partial \hat{h}(j\omega)/\partial A\|_F$, instead of using the more desirable L_2 norm.

We have just shown that $M = M_1 + M_2 \leqslant \operatorname{tr} W_o \operatorname{tr} W_c + \operatorname{tr} W_o + \operatorname{tr} W_c$. The upper bound for the scalar sensitivity measure M will be denoted by θ; thus $\theta = \operatorname{tr} W_o \operatorname{tr} W_c + \operatorname{tr} W_o + \operatorname{tr} W_c$. This result was first obtained by Thiele for digital filters [7].

E. Change of Coordinates

A state-space realization $\{A, b, c^T\}$ of a transfer function $\hat{h}(s)$ is not unique, since the change of coordinates $x(t) = T\tilde{x}(t)$, where $T \in \mathbb{R}^{n\times n}$ is any nonsingular matrix, results in another realization $\{\tilde{A}, \tilde{b}, \tilde{c}^T\} = \{T^{-1}AT, T^{-1}b, c^T T\}$ that also has the transfer function $\hat{h}(s)$. The two matrices $\tilde{A}$ and A are said to be similar and the transformation defined by $\tilde{A} = T^{-1}AT$ is called a similarity function.

It is easily shown (e.g., see Moore [8]) that

$$\tilde{W}_c = T^{-1} W_c T^{-T}$$

and

$$\tilde{W}_o = T^T W_o T.$$

Thus the controllability Gramian W_c and the observability Gramian W_o, and hence the upper bound θ of the scalar sensitivity measure M, depend on the choice of coordinates (state-variables). The Gramians $\tilde{W}_c$ and W_c are congruent, as are $\tilde{W}_o$ and W_o. Since a congruence transformation preserves the properties of symmetry and positive definiteness, $\tilde{W}_c$ and $\tilde{W}_o$ are symmetric, positive definite matrices. However, a congruence transformation preserves only the signs of the eigenvalues of a matrix, and not the eigenvalues themselves. Therefore, unless T is orthogonal $(T^T = T^{-1})$, the eigenvalues of $\tilde{W}_c$ will not equal those of W_c and same for $\tilde{W}_o$ and W_o. As a result, $\operatorname{tr} \tilde{W}_c$ is different, in general, than $\operatorname{tr} W_c$. Similarly, $\operatorname{tr} \tilde{W}_o$ is different, in general, then $\operatorname{tr} W_o$. Thus, we conclude that the upper bound θ of the sensitivity measure M is altered by a general similarity transformation (change of coordinates). However, we also note the important result that the upper bound θ is invariant if we perform an orthogonal similarity transformation (i.e., T is orthogonal).

The eigenvalues of $\tilde{W}_c\tilde{W}_o$ (or $\tilde{W}_o\tilde{W}_c$) are invariant under any similarity transformation of the system, since $\tilde{W}_c\tilde{W}_o = T^{-1}W_cW_oT$. Also, the eigenvalues of W_cW_o are always real and positive [5]. The Hankel singular values of $\hat{h}(s)$ [12] are defined by

$$\sigma_i \triangleq \left[\lambda_i(W_cW_o)\right]^{1/2}, \qquad i = 1, \cdots, n$$

where $\lambda_i(W_cW_o)$ denotes the eigenvalues of W_cW_o, $i = 1, \cdots, n$. Note that the Hankel singular values are invariant and depend only on the choice of transfer function $\hat{h}(s)$. The Hankel singular values were called "second-order modes" by Mullis and Roberts [5].

F. Conditions to Minimize the Sensitivity Measure M

We now present a theorem that gives the conditions under which the upper bound θ of the scalar sensitivity measure M is a minimum. Our presentation closely follows that of Thiele [7].

Theorem: Let $\tilde{W}_c$ be the controllability Gramian and $\tilde{W}_o$ be the observability Gramian of a minimal realization $\{\tilde{A}, \tilde{b}, \tilde{c}^T\}$ of the transfer function $\hat{h}(s)$. Both $\tilde{W}_c$ and $\tilde{W}_o$ are $n \times n$ real, symmetric and positive definite matrices. Then

$$\begin{aligned} \theta &= \operatorname{tr} \tilde{W}_o \operatorname{tr} \tilde{W}_c + \operatorname{tr} \tilde{W}_o + \operatorname{tr} \tilde{W}_c \\ &\geqslant \left(\sum_{i=1}^{n} \sigma_i \right)^2 + 2\left(\sum_{i=1}^{n} \sigma_i \right) \end{aligned}$$

where $\{\sigma_i\}$ are the Hankel singular values of $\hat{h}(s)$, $[\sigma_i = (\lambda_i(W_cW_o))^{1/2}]$. The lower bound is achieved if and only if $\tilde{W}_c = \tilde{W}_o$.

Proof: See Thiele [7]. An alternate proof has been given by Lutz [10].

We know that $\tilde{W}_c = T^{-1}W_cT^{-T}$ and $\tilde{W}_o = T^TW_oT$. Then the upper bound θ will be minimized if we can find a transformation T such that $\tilde{W}_c = \tilde{W}_o$.

G. Balanced Realizations

A realization $\{\tilde{A}, \tilde{b}, \tilde{c}^T\}$ is called a balanced realization [8], if the controllability Gramian $\tilde{W}_c$ and the observability Gramian $\tilde{W}_o$ are diagonal and equal. Moore has shown that there exists a nonsingular $T \in \mathbb{R}^{n \times n}$ such that $\tilde{W}_c = T^{-1} W_c T^{-T} = \tilde{W}_o = T^T W_o T = \Sigma$, where

$$\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_n) \quad \text{and} \quad \sigma_i = (\lambda_i(W_c W_o))^{1/2}.$$

Since a balanced realization has $\tilde{W}_c = \tilde{W}_o = \Sigma$, the balanced realization of a transfer function $\hat{h}(s)$ satisfies the conditions for minimizing the upper bound θ. A balanced realization can be obtained from any other minimal realization $\{A, b, c^T\}$ by a unique similarity transformation T. Procedures for finding this balancing transformation have been given by Moore [8] and Laub [13]. Because the upper bound θ is invariant if we perform an orthogonal similarity transformation (i.e., T is orthogonal), any realization orthogonally similar to a balanced realization minimizes the upper bound θ.

When the system is balanced, $\|\hat{f}\|_E = \|\hat{g}\|_E$ and thus $\|\hat{f}\|_{E, L_2} = \|\hat{g}\|_{E, L_2}$. A proof of the first equality is given in Appendix 2. The second equality is evident from the fact that $\mathrm{tr}\, \tilde{W}_c = \mathrm{tr}\, \tilde{W}_o$ for a balanced realization. As a result, when the system is balanced:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \left\| \frac{\partial \hat{h}}{\partial A} \right\|_F d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \|\hat{f}\|_E^2 \, d\omega$$
$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \|\hat{g}\|_E^2 \, d\omega.$$

In other words, the L_1 norm of $\|\partial \hat{h}(j\omega)/\partial A\|_F$ becomes equal to the L_2 norm of either $\|\hat{f}(j\omega)\|_E$ or $\|\hat{g}(j\omega)\|_E$ when the realization is balanced. Also, $\|\partial \hat{h}/\partial A\|_{F, L_1}^2$ becomes equal to its upper bound of $\mathrm{tr}\, \tilde{W}_o \,\mathrm{tr}\, \tilde{W}_c$. This occurs for the balanced realization, when the upper bound $\mathrm{tr}\, \tilde{W}_o \,\mathrm{tr}\, \tilde{W}_c$ achieves its *minimum*.

One practical disadvantage of the balanced realization is that no elements in the vectors b and c are zero and few, if any, elements in the A matrix are zero. This means that when constructing a filter as a balanced realization, close to the maximum possible number of amplifiers and passive components will be required. No one, to our knowledge, has proven that every element of A, b, and c will be nonzero in a balanced realization. However, Verriest and Kailath have proven a "less pessimistic" theorem, which states that no element of the b and c vectors, nor an element on the principal diagonal of A will be zero in a balanced realization [14]. It is interesting to note that several published examples have A matrices with *no* zero elements [8], [15], [16]. In any event, it is clear that a balanced realization has "dense" A, b, and c and is not an efficient realization in terms of component count.

H. Householder Transformations

We now show that orthogonal similarity transformations can be used to introduce zeros in the elements of the c (or b) vector and the A matrix. This allows us to reduce the system to a simpler form, while leaving the upper bound θ of the sensitivity measure M unaltered.

We consider a special class of orthogonal matrices known as Householder transformations. Householder transformations are matrices of the form $P = I - 2(uu^T/u^T u)$, where u is an arbitrary $n \times 1$ column vector. Householder transformations are symmetric ($P^T = P$), orthogonal ($P^T P = I$), and involutory ($P^2 = I$) [17]. Thus a Householder transformation has $P = P^T = P^{-1}$. It is well known that for a suitable choice of u, the corresponding Householder transformation P can be used to zero all but a single entry of the column vector c when the vector c is premultiplied by P. (See Golub and Van Loan [18, algorithm 3.3-1].) It is also known that Householder transformations can be used as similarity transformations which will reduce an arbitrary matrix $A \in \mathbb{R}^{n \times n}$ to the upper Hessenberg matrix H in $n-2$ steps. That is, $H = P_{n-2} \cdots P_1 A P_1 \cdots P_{n-2}$, where H is upper Hessenberg ($h_{ij} = 0$ whenever $i \geqslant j+1$), and P_i, $i = 1, \cdots, n-2$, are suitably chosen Householder transformations. (See [18, Golub and Van Loan, Algorithm 7.4-2].) Combining these two facts, it can be shown that there exists an orthogonal matrix Q such $Q^{-1} A Q$ is upper Hessenberg and $Q^T c$ is equal to $(\|c\|, 0, \cdots, 0)^T$. The orthogonal matrix Q will be the product of $n-1$ Householder transformations P. Further details are available in Lutz [10]. For $(n-1)$ steps, a maximum of $\sum_{i=1}^{n-1}(n-i)$ zeros can be introduced in A and c by this technique.

Finally, we recall that for a general similarity transformation, $\tilde{W}_c = T^{-1} W_c T^{-T}$ and $\tilde{W}_o = T^T W_o T$. When T is a Householder transformation P, $\tilde{W}_c = P W_c P$ and $\tilde{W}_o = P W_o P$, so that if W_c and W_o were equal before a Householder transformation, $\tilde{W}_c$ and $\tilde{W}_o$ are equal afterwards. Also, since $\tilde{W}_c$ is similar to W_c and $\tilde{W}_o$ is similar to W_o, $\mathrm{tr}\, \tilde{W}_c$ and $\mathrm{tr}\, \tilde{W}_o$ are invariant. Thus the upper bound θ of the sensitivity measure M remains unchanged as we use orthogonal Householder transformations to change our initial balanced realization (which we can find by direct means) into a simpler (i.e., more zero elements) realization.

III. MIMO Case

A. State-Space Formulation

We would now like to extend the previous theory to include the case of multi-input multi-output (MIMO) systems realizing a matrix of transfer functions $\hat{H}(s)$. The state variable description for such a system is given by the state equations

$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t)$$

were $x(t)$ is an $n \times 1$ state vector, $u(t)$ is an $l \times 1$ input vector, and $y(t)$ is an $m \times 1$ output vector; A, B, and C are real constant matrices, with $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times l}$, and $C \in \mathbb{R}^{m \times n}$. If we are given a matrix of transfer functions $\hat{H}(s)$, a set of state equations that realize $\hat{H}(s)$ can always be found. However, the process is rather complicated and the reader is referred to Chen [9, chap. 6] for details

[9]. We now have l inputs and m outputs and so $\hat{\boldsymbol{H}}(s) = \boldsymbol{C}(s\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{B}$ is an $m \times l$ matrix of transfer functions. The ijth element of $\hat{\boldsymbol{H}}(s)$, denoted by $\hat{h}_{ij}(s)$, is the transfer function between the ith output and the jth input, with all other inputs set equal to zero.

The transfer function vectors $\hat{\boldsymbol{f}}(s)$ and $\hat{\boldsymbol{g}}(s)$ of the SISO case become the transfer function matrices $\hat{\boldsymbol{F}}(s)$ and $\hat{\boldsymbol{G}}(s)$ in the MIMO case. The transfer function matrix $\hat{\boldsymbol{F}}(s)$ is the $n \times l$ matrix defined as $\hat{\boldsymbol{F}}(s) = (s\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{B}$. Similarly, the transfer function matrix $\hat{\boldsymbol{G}}(s)$ is the $n \times m$ matrix defined as $\hat{\boldsymbol{G}}(s) = (s\boldsymbol{I}-\boldsymbol{A}^T)^{-1}\boldsymbol{C}^T$. As in the SISO case, each entry of $\hat{\boldsymbol{H}}(s)$, $\hat{\boldsymbol{F}}(s)$, and $\hat{\boldsymbol{G}}(s)$ is a strictly proper stable rational function. We also note that $\boldsymbol{F}(t) = \mathscr{L}^{-1}\{\hat{\boldsymbol{F}}(s)\} = e^{\boldsymbol{A}t}\boldsymbol{B}$ and that $\boldsymbol{G}(t) = \mathscr{L}^{-1}\{\hat{\boldsymbol{G}}(s)\} = e^{\boldsymbol{A}^Tt}\boldsymbol{C}^T$.

The controllability Gramian matrix $\boldsymbol{W}_c$ and the observability Gramian matrix $\boldsymbol{W}_o$ are defined as

$$\boldsymbol{W}_c \triangleq \int_0^\infty \boldsymbol{F}(t)\boldsymbol{F}^T(t)\,dt = \int_0^\infty e^{\boldsymbol{A}t}\boldsymbol{B}\boldsymbol{B}^T e^{\boldsymbol{A}^Tt}\,dt$$

and

$$\boldsymbol{W}_o \triangleq \int_0^\infty \boldsymbol{G}(t)\boldsymbol{G}^T(t)\,dt = \int_0^\infty e^{\boldsymbol{A}^Tt}\boldsymbol{C}^T\boldsymbol{C}e^{\boldsymbol{A}t}\,dt$$

for the MIMO case [8]. The matrices $\boldsymbol{W}_c$ and $\boldsymbol{W}_o$ are again $n \times n$ symmetric positive definite constant matrices in the MIMO case.

B. The Derivative of a Matrix with Respect to a Matrix

When we attempt to define the sensitivity of $\hat{\boldsymbol{H}}(s)$ with respect to the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, or $\boldsymbol{C}$, we realize that we must take the derivative of a matrix with respect to a matrix. There are at least three definitions for the derivative of a matrix with respect to a matrix in current use [19], and the one we shall use is the following. Let $\boldsymbol{Y} = [y_{kl}]$ be an $p \times q$ matrix and $\boldsymbol{X} = [x_{ij}]$ be an $m \times n$ matrix, with the elements of $\boldsymbol{Y}$ being functions of the elements of $\boldsymbol{X}$. Then $\partial \boldsymbol{Y}/\partial \boldsymbol{X}$ is an $mp \times nq$ matrix, which can be partitioned into $m \times n$ submatrices with the klth partition equal to $\partial y_{kl}/\partial \boldsymbol{X}$. In other words,

$$\frac{\partial \boldsymbol{Y}}{\partial \boldsymbol{X}} = \left[\begin{array}{c|c|c|c} \dfrac{\partial y_{11}}{\partial \boldsymbol{X}} & \dfrac{\partial y_{12}}{\partial \boldsymbol{X}} & \cdots & \dfrac{\partial y_{1q}}{\partial \boldsymbol{X}} \\ \hline \dfrac{\partial y_{21}}{\partial \boldsymbol{X}} & \dfrac{\partial y_{22}}{\partial \boldsymbol{X}} & \cdots & \dfrac{\partial y_{2q}}{\partial \boldsymbol{X}} \\ \hline \vdots & \vdots & & \vdots \\ \hline \dfrac{\partial y_{p1}}{\partial \boldsymbol{X}} & \dfrac{\partial y_{p2}}{\partial \boldsymbol{X}} & \cdots & \dfrac{\partial y_{pq}}{\partial \boldsymbol{X}} \end{array}\right].$$

Here, the derivative of the scalar y_{kl} with respect to the matrix $\boldsymbol{X}$, denoted $\partial y_{kl}/\partial \boldsymbol{X}$, is a matrix of the same order as $\boldsymbol{X}$, whose ijth entry is $\partial y_{kl}/\partial x_{ij}$. For this definition, the $[(k-1)m+i,\ (l-1)n+j]$th entry of $\partial \boldsymbol{Y}/\partial \boldsymbol{X}$ is $\partial y_{kl}/\partial x_{ij}$.

We would now like to find $\partial \hat{\boldsymbol{H}}(s)/\partial \boldsymbol{A}$. This requires us to find the ikth partition $\partial \hat{h}_{ik}(s)/\partial \boldsymbol{A}$, which seems difficult to compute. However, it is easy to compute $\partial \hat{\boldsymbol{H}}(s)/\partial a_{rs}$, and so we take a roundabout approach of finding $\partial \hat{\boldsymbol{H}}(s)/\partial a_{rs}$ and then using Graham's "First Transformation Principle" [20] to compute $\partial \hat{h}_{ik}(s)/\partial \boldsymbol{A}$. We briefly discuss Graham's "First Transformation Principle." Consider a matrix product of the form $\boldsymbol{Y} = \boldsymbol{M}\boldsymbol{X}\boldsymbol{N}$ where $\boldsymbol{X} = [x_{rs}]$ is of order $m \times n$, $\boldsymbol{Y} = [y_{ij}]$ is of order $l \times q$, and $\boldsymbol{M}$ and $\boldsymbol{N}$ are matrices compatible with $\boldsymbol{X}$ and which are independent of $\boldsymbol{X}$. It can be shown that under these conditions [20]

$$\frac{\partial y_{ij}}{\partial \boldsymbol{X}} = \boldsymbol{M}^T\boldsymbol{E}_{ij}\boldsymbol{N}^T$$

where $\boldsymbol{E}_{ij}$ is an elementary matrix of the same order as the matrix $\boldsymbol{Y}$. It can also be shown that [20]

$$\frac{\partial \boldsymbol{Y}}{\partial x_{rs}} = \frac{\partial(\boldsymbol{M}\boldsymbol{X}\boldsymbol{N})}{\partial x_{rs}} = \boldsymbol{M}\boldsymbol{E}_{rs}\boldsymbol{N}$$

where $\boldsymbol{E}_{rs}$ is an elementary matrix of the same order as the matrix $\boldsymbol{X}$. Graham's "First Transformation Principle" states that $\partial y_{ij}/\partial \boldsymbol{X}$ is a transformation of $\partial \boldsymbol{Y}/\partial x_{rs}$ and vice versa. For example, to obtain $\partial y_{ij}/\partial \boldsymbol{X}$ from $\partial \boldsymbol{Y}/\partial x_{rs}$ we replace $\boldsymbol{M}$ by $\boldsymbol{M}^T$, $\boldsymbol{N}$ by $\boldsymbol{N}^T$, and $\boldsymbol{E}_{rs}$ by $\boldsymbol{E}_{ij}$ (changing the order of the elementary matrix as appropriate).

C. Coefficient Sensitivity Functions

We now find $\partial \hat{h}_{ik}(s)/\partial \boldsymbol{A}$ and then $\partial \hat{\boldsymbol{H}}(s)/\partial \boldsymbol{A}$. Since $\hat{\boldsymbol{H}}(s) = \boldsymbol{C}(s\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{B}$ and, in general, $\partial \boldsymbol{P}^{-1}/\partial z = -\boldsymbol{P}^{-1}\partial \boldsymbol{P}/\partial z \boldsymbol{P}^{-1}$, it can be shown that

$$\frac{\partial \hat{\boldsymbol{H}}(s)}{\partial a_{rs}} = -\boldsymbol{C}(s\boldsymbol{I}-\boldsymbol{A})^{-1}\frac{\partial(s\boldsymbol{I}-\boldsymbol{A})}{\partial a_{rs}}(s\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{B}$$

or

$$\frac{\partial \hat{\boldsymbol{H}}(s)}{\partial a_{rs}} = \hat{\boldsymbol{G}}^T(s)\boldsymbol{E}_{rs}\hat{\boldsymbol{F}}(s) \qquad (m \times l \text{ result})$$

where $\boldsymbol{E}_{rs}$ is an $n \times n$ elementary matrix of the same order as $\boldsymbol{A}$. Applying Graham's "First Transformation Principle" gives $\partial \hat{h}_{ik}(s)/\partial \boldsymbol{A} = \hat{\boldsymbol{G}}(s)\boldsymbol{E}_{ik}\hat{\boldsymbol{F}}^T(s)$ ($n \times n$ result) where $\boldsymbol{E}_{ik}$ is an $m \times l$ elementary matrix (same order as $\hat{\boldsymbol{H}}(s)$). Since $\boldsymbol{E}_{ik} = \boldsymbol{e}_i\boldsymbol{e}_k^T$, we can write

$$\frac{\partial \hat{h}_{ik}(s)}{\partial \boldsymbol{A}} = \hat{\boldsymbol{G}}(s)\boldsymbol{e}_i\boldsymbol{e}_k^T\hat{\boldsymbol{F}}^T(s) = \left(\hat{\boldsymbol{G}}(s)_{\cdot i}\right)\left(\hat{\boldsymbol{F}}(s)_{\cdot k}\right)^T$$

where, in general, $\boldsymbol{M}_{\cdot i}$ denotes the ith column of the matrix $\boldsymbol{M}$. Thus we can write

$$\frac{\partial \hat{\boldsymbol{H}}(s)}{\partial \boldsymbol{A}} = \left[\begin{array}{c|c|c|c} \hat{\boldsymbol{G}}_{\cdot 1}(\hat{\boldsymbol{F}}_{\cdot 1})^T & \hat{\boldsymbol{G}}_{\cdot 1}(\hat{\boldsymbol{F}}_{\cdot 2})^T & \cdots & \hat{\boldsymbol{G}}_{\cdot 1}(\hat{\boldsymbol{F}}_{\cdot l})^T \\ \hline \hat{\boldsymbol{G}}_{\cdot 2}(\hat{\boldsymbol{F}}_{\cdot 1})^T & \hat{\boldsymbol{G}}_{\cdot 2}(\hat{\boldsymbol{F}}_{\cdot 2})^T & \cdots & \hat{\boldsymbol{G}}_{\cdot 2}(\hat{\boldsymbol{F}}_{\cdot l})^T \\ \hline \vdots & \vdots & & \vdots \\ \hline \hat{\boldsymbol{G}}_{\cdot m}(\hat{\boldsymbol{F}}_{\cdot 1})^T & \hat{\boldsymbol{G}}_{\cdot m}(\hat{\boldsymbol{F}}_{\cdot 2})^T & \cdots & \hat{\boldsymbol{G}}_{\cdot m}(\hat{\boldsymbol{F}}_{\cdot l})^T \end{array}\right]$$

or

$$\frac{\partial \hat{H}(s)}{\partial A} = \begin{bmatrix} \hat{G}_{\cdot 1} \\ \hline \hat{G}_{\cdot 2} \\ \hline \vdots \\ \hline \hat{G}_{\cdot m} \end{bmatrix} \left[(\hat{F}_{\cdot 1})^T \,\vdots\, (\hat{F}_{\cdot 2})^T \,\vdots\, \cdots \,\vdots\, (\hat{F}_{\cdot l})^T \right].$$

To simplify the notation, we introduce the vec operator [20]. If A is an $m \times n$ matrix, then we define

$$\operatorname{vec} A = \begin{bmatrix} A_{\cdot 1} \\ \hline A_{\cdot 2} \\ \hline \vdots \\ \hline A_{\cdot n} \end{bmatrix}$$

where $A_{\cdot i}$ denotes the ith column of A. Note that vec A is an $mn \times 1$ column vector. Using the vec operator, we can write

$$\frac{\partial \hat{H}(s)}{\partial A} = \underset{(mn \times 1)}{(\operatorname{vec} \hat{G}(s))} \underset{(1 \times nl)}{(\operatorname{vec} \hat{F}(s))^T} \qquad mn \times nl \text{ result.}$$

We now find the $nm \times l^2$ matrix $\partial \hat{H}(s)/\partial B$, with the ikth partition given by $\partial \hat{h}_{ik}(s)/\partial B$. The ikth partition is

$$\frac{\partial \hat{h}_{ik}(s)}{\partial B} = \hat{G}(s) E_{ik}(I_l) = \hat{G}(s) e_i e_k^T = (\hat{G}(s)_{\cdot i}) e_k^T$$

where E_{ik} is an $m \times l$ elementary matrix. Forming the matrix $\partial \hat{H}(s)/\partial B$ by using the above expression as the ikth partition and simplifying leads to

$$\frac{\partial \hat{H}(s)}{\partial B} = \left[(\operatorname{vec} \hat{G}) \mathbf{0} \cdots \mathbf{0} \,\vdots\, \mathbf{0} (\operatorname{vec} \hat{G}) \cdots \mathbf{0} \,\vdots\, \cdots \,\vdots\, \mathbf{0}\mathbf{0} \cdots (\operatorname{vec} \hat{G}) \right].$$

This expression can be seen to be equal to

$$\frac{\partial \hat{H}(s)}{\partial B} = (\operatorname{vec} \hat{G}(s)) (\operatorname{vec} I_l)^T.$$

Finally, we must find the $m^2 \times nl$ matrix $\partial \hat{H}(s)/\partial C$, with the ikth partition given by $\partial \hat{h}_{ik}(s)/\partial C$. The ikth partition is

$$\frac{\partial \hat{h}_{ik}(s)}{\partial C} = (I_m) E_{ik} \hat{F}(s)^T = e_i e_k^T \hat{F}(s)^T = e_i (\hat{F}(s) e_k)^T$$
$$= e_i (\hat{F}(s)_{\cdot k})^T$$

where E_{ik} is an $m \times l$ elementary matrix. Forming the matrix $\partial \hat{H}(s)/\partial C$ by using the above expression as the ikth partition and simplifying leads to

$$\frac{\partial \hat{H}(s)}{\partial C} = (\operatorname{vec} I_m) (\operatorname{vec} \hat{F}(s))^T.$$

To summarize, our sensitivity functions for the MIMO case are

$$\frac{\partial \hat{H}(s)}{\partial A} = (\operatorname{vec} \hat{G}(s)) (\operatorname{vec} \hat{F}(s))^T \qquad mn \times nl$$

$$\frac{\partial \hat{H}(s)}{\partial B} = (\operatorname{vec} \hat{G}(s)) (\operatorname{vec} I_l)^T \qquad mn \times l^2$$

$$\frac{\partial \hat{H}(s)}{\partial C} = (\operatorname{vec} I_m) (\operatorname{vec} \hat{F}(s))^T \qquad m^2 \times nl.$$

In general, if y is an $n \times 1$ column vector, then $\operatorname{vec} y = \operatorname{vec} y^T = y$. As a result, the above expressions automatically include the SISO case.

D. Sensitivity Measure

We would like to obtain a scalar sensitivity measure from these sensitivity functions. Before doing so, however, we introduce two facts that will simplify the process.

Fact 1: If w and v are column vectors and $X = wv^T$, then $\operatorname{vec} X = \operatorname{vec}(wv^T) = v \otimes w$, where $\otimes$ denotes the Kronecker product (see Appendix I) [19].

Fact 2: The Frobenius norm (squared) of a matrix A is given by $\|A\|_F^2 = (\operatorname{vec} A)^* (\operatorname{vec} A)$.

Since the matrix L_p norm is given by $\|\hat{A}\|_{F, L_p} = (1/2\pi \int_{-\infty}^{\infty} \|\hat{A}\|_F^p \, d\omega)^{1/p}$, we proceed to first find $\|\partial \hat{H}/\partial A\|_F$, $\|\partial \hat{H}/\partial B\|_F$, and $\|\partial \hat{H}/\partial C\|_F$.

We start with $\partial \hat{H}(s)/\partial A$, which is given by $\partial \hat{H}(s)/\partial A = (\operatorname{vec} \hat{G}(s))(\operatorname{vec} \hat{F}(s))^T$. We can use Facts 1 and 2 to write $\operatorname{vec}(\partial \hat{H}/\partial A) = (\operatorname{vec} \hat{F}) \otimes (\operatorname{vec} \hat{G})$, and

$$\left\| \frac{\partial \hat{H}}{\partial A} \right\|_F^2 = \left[\operatorname{vec}\left(\frac{\partial \hat{H}}{\partial A} \right) \right]^* \left[\operatorname{vec}\left(\frac{\partial \hat{H}}{\partial A} \right) \right]$$
$$= \left[(\operatorname{vec} \hat{F})^* \otimes (\operatorname{vec} \hat{G})^* \right] \left[(\operatorname{vec} \hat{F}) \otimes (\operatorname{vec} \hat{G}) \right]$$
$$= (\operatorname{vec} \hat{F})^* (\operatorname{vec} \hat{F}) \otimes (\operatorname{vec} \hat{G})^* (\operatorname{vec} \hat{G})$$
$$= \|\hat{F}\|_F^2 \cdot \|\hat{G}\|_F^2.$$

Thus $\|\partial \hat{H}/\partial A\|_F = \|\hat{F}\|_F \cdot \|\hat{G}\|_F$.

We have used several properties of Kronecker products listed in Appendix I.

For $\partial \hat{H}(s)/\partial B$ we can use Facts 1 and 2 to write $\operatorname{vec}(\partial \hat{H}/\partial B) = (\operatorname{vec} I_l) \otimes (\operatorname{vec} \hat{G})$, and

$$\left\| \frac{\partial \hat{H}}{\partial B} \right\|_F^2 = \left[\operatorname{vec}\left(\frac{\partial \hat{H}}{\partial B} \right) \right]^* \left[\operatorname{vec}\left(\frac{\partial \hat{H}}{\partial B} \right) \right]$$
$$= \left[(\operatorname{vec} I_l)^* \otimes (\operatorname{vec} \hat{G})^* \right] \left[(\operatorname{vec} I_l) \otimes (\operatorname{vec} \hat{G}) \right]$$
$$= (\operatorname{vec} I_l)^T (\operatorname{vec} I_l) \otimes (\operatorname{vec} \hat{G})^* (\operatorname{vec} \hat{G})$$
$$= l \otimes (\operatorname{vec} \hat{G})^* (\operatorname{vec} \hat{G})$$
$$= l \|\hat{G}\|_F^2.$$

In a similar fashion, it is easily shown that $\operatorname{vec}(\partial \hat{H}/\partial C) = (\operatorname{vec}\hat{F}) \otimes (\operatorname{vec} I_m)$, and thus $\|\partial \hat{H}/\partial C\|_F^2 = m\|\hat{F}\|_F^2$.

With the above results, it is now a simple matter to obtain a scalar sensitivity measure. We still take our scalar sensitivity measure M to be

$$M \triangleq \left\|\frac{\partial \hat{H}}{\partial A}\right\|_{F,L_1}^2 + \left\|\frac{\partial \hat{H}}{\partial B}\right\|_{F,L_2}^2 + \left\|\frac{\partial \hat{H}}{\partial C}\right\|_{L_2}^2$$

$$= \left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\left\|\frac{\partial \hat{H}}{\partial A}\right\|_F d\omega\right)^2 + \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\|\frac{\partial \hat{H}}{\partial B}\right\|_F^2 d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty}\left\|\frac{\partial \hat{H}}{\partial C}\right\|_F^2 d\omega$$

$$= \left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\|\hat{F}\|_F \cdot \|\hat{G}\|_F d\omega\right)^2 + \frac{l}{2\pi}\int_{-\infty}^{\infty}\|\hat{G}\|_F^2 d\omega + \frac{m}{2\pi}\int_{-\infty}^{\infty}\|\hat{F}\|_F^2 d\omega.$$

Using the Cauchy–Schwarz inequality on the first term gives

$$M \leqslant \left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\|\hat{F}\|_F^2 d\omega\right)\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\|\hat{G}\|_F^2 d\omega\right) + \frac{l}{2\pi}\int_{-\infty}^{\infty}\|\hat{G}\|_F^2 d\omega + \frac{m}{2\pi}\int_{\infty}^{\infty}\|\hat{F}\|_F^2 d\omega.$$

The expression

$$\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\|\hat{F}\|_F^2 d\omega\right)\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\|\hat{G}\|_F^2 d\omega\right) + \frac{l}{2\pi}\int_{-\infty}^{\infty}\|\hat{G}\|_F^2 d\omega + \frac{m}{2\pi}\int_{-\infty}^{\infty}\|\hat{F}\|_F^2 d\omega$$

is equal to $\operatorname{tr} W_c \operatorname{tr} W_o + l \operatorname{tr} W_o + m \operatorname{tr} W_c$, so we can write

$$M \leqslant \operatorname{tr} W_c \operatorname{tr} W_o + l \operatorname{tr} W_o + m \operatorname{tr} W_c$$

where W_c is the controllability Gramian matrix and W_o is the observability Gramian matrix of the MIMO system.

For MIMO systems, the scalar sensitivity measure M has the upper bound $\theta = \operatorname{tr} W_c \operatorname{tr} W_o + l \operatorname{tr} W_o + m \operatorname{tr} W_c$, which will be altered by a similarity transformation.

E. Minimum Sensitivity Realization of $\hat{H}(s)$

We now determine the conditions under which the upper bound θ is a minimum. The first term of θ is $\operatorname{tr} W_c \operatorname{tr} W_o$. Thus, there is no difference between the SISO case and the MIMO case as far as the first term is concerned. It has been shown [7], [10] that

$$\operatorname{tr} W_c \operatorname{tr} W_o \geqslant \left(\sum_{i=1}^{n} \sigma_i\right)^2$$

where equality is achieved if and only if $W_c = \alpha W_o$, for some scalar $\alpha \neq 0$.

We now derive an inequality for the second and third terms of the upper bound θ. Consider $[\sqrt{m \operatorname{tr} W_c} - \sqrt{l \operatorname{tr} W_o}]^2 \geqslant 0$. Expanding gives $m \operatorname{tr} W_c + l \operatorname{tr} W_o \geqslant 2\sqrt{ml}(\operatorname{tr} W_c \operatorname{tr} W_o)^{1/2}$, with equality if and only if $\sqrt{m \operatorname{tr} W_c} = \sqrt{l \operatorname{tr} W_o}$. Using the previous result for the first term, we can write

$$m \operatorname{tr} W_c + l \operatorname{tr} W_o \geqslant 2\sqrt{ml}\left(\sum_{i=1}^{n} \sigma_i\right)$$

with equality if and only if $W_c = (l/m)W_o$. Notice that the conditions that minimize the sum of the second and third terms of θ also achieve equality for the first term (with $\alpha = l/m$). Thus, the minimum possible value of θ is

$$\theta_{\min} = \left(\sum_{i=1}^{n} \sigma_i\right)^2 + 2\sqrt{ml}\left(\sum_{i=1}^{n} \sigma_i\right)$$

and this value is achieved if and only if $W_c = \alpha W_o$, where $\alpha = l/m$.

The easiest way to find a coordinate system such that $\tilde{W}_c = (l/m)\tilde{W}_o$ is as follows. First, find the balanced realization of $\hat{H}(s)$ using one of the available algorithms [8], [13]. For the balanced realization, $\tilde{W}_c = \tilde{W}_o = \Sigma$, where $\Sigma = \operatorname{diag}(\sigma_1, \cdots, \sigma_n)$ and $\sigma_i = (\lambda_i(W_c W_o))^{1/2}$. If we use a similarity transformation $T = \sqrt[4]{m/l}\, I$, then $\overline{W}_c = \sqrt[2]{l/m}\,\Sigma$ and $\overline{W}_o = \sqrt[2]{m/l}\,\Sigma$ and $\overline{W}_c = (l/m)\overline{W}_o$, as desired. Thus, by simply scaling the balanced realization by the transformation $T = \sqrt[4]{m/l}\, I$, the upper bound θ will be minimized. Note that $\overline{W}_c$ and $\overline{W}_o$ are still diagonal, but they are no longer equal.

Finally, we note that a realization such that $\overline{W}_c = (l/m)\overline{W}_o$ will have few, if any, zero elements in $\overline{A}$, $\overline{B}$, or $\overline{C}$, since it is simply a scaled balanced realization. As before, we can use Householder transformations to first introduce zeros into a column of $\overline{C}$ and to then reduce $\overline{A}$ to upper Hessenberg form. For a Householder transformation P, $\breve{W}_c = P\overline{W}_c P$ and $\breve{W}_o = P\overline{W}_o P$, so that if $\overline{W}_c = l/m\overline{W}_o$ before a Householder transformation, $\breve{W}_c = (l/m)\breve{W}_o$ afterwards. Also, since $\breve{W}_c$ is similar to $\overline{W}_c$ and $\breve{W}_o$ is similar to $\overline{W}_o$, then $\operatorname{tr} \breve{W}_c = \operatorname{tr} \overline{W}_c$ and $\operatorname{tr} \breve{W}_o = \operatorname{tr} \overline{W}_o$. Thus $m \operatorname{tr} \breve{W}_c = l \operatorname{tr} \breve{W}_o$ and the upper bound θ is unchanged by a Householder transformation.

IV. Conclusions

A scalar sensitivity measure M has been presented for both the SISO continuous-time state-space realization of a transfer function $\hat{h}(s)$ and the MIMO continuous-time state-space realization of a transfer function matrix $\hat{H}(s)$. For either case, this measure can be expressed in terms of the controllability Gramian matrix W_c and the observability Gramian matrix W_o. It was shown that the balanced realization (scaled balanced realization) minimizes the upper bound θ of the scalar sensitivity measure M in the SISO (MIMO) case. The advantage of the balanced realization is that it can be computed directly from any equivalent realization. The disadvantage is that balanced realizations tend to have few, if any, zero elements in the A, B, or C matrices and are, therefore, inefficient in terms of component count. This disadvantage was overcome by using (orthogonal) Householder transformations to first zero all but a single element of a column of $\boldsymbol{C}$ and to then reduce $\boldsymbol{A}$ to an upper Hessenberg form. These Householder transformations leave the upper bound θ of the scalar sensitivity measure M unchanged and result in a simple realization which has up to $\Sigma_{i=1}^{n-1}(n-i)$ more zero elements in $\boldsymbol{A}$ and $\boldsymbol{C}$ than the balanced realization.

Appendix I
Kronecker Product

$\boldsymbol{A}=[a_{ij}]$ is an $p\times q$ matrix and $\boldsymbol{B}=[b_{rs}]$ is an $m\times n$ matrix. The Kronecker product of $\boldsymbol{A}$ and $\boldsymbol{B}$, denoted by $\boldsymbol{A}\otimes\boldsymbol{B}$, is the $pm\times qn$ matrix [20]

$$\boldsymbol{A}\otimes\boldsymbol{B}=\left[\begin{array}{c:c:c} a_{11}\boldsymbol{B} & \cdots & a_1q\boldsymbol{B} \\ \hdashline \vdots & & \vdots \\ \hdashline a_{p1}\boldsymbol{B} & \cdots & a_{pq}\boldsymbol{B} \end{array}\right].$$

Notice that the elements of $\boldsymbol{A}\otimes\boldsymbol{B}$ consist of all possible products $a_{ij}b_{rs}$.

Some important properties of the Kronecker product are listed below without proof. (See Graham [20] for proofs and additional properties.)

(i) $(\boldsymbol{A}\otimes\boldsymbol{B})^T=\boldsymbol{A}^T\otimes\boldsymbol{B}^T$.

(ii) $(\boldsymbol{A}\otimes\boldsymbol{B})^{-1}=\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1}$.

(iii) For λ a scalar,

$$\lambda\otimes\boldsymbol{A}=\lambda\boldsymbol{A}=\boldsymbol{A}\otimes\lambda=\boldsymbol{A}\lambda.$$

(iv) $(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D})=\boldsymbol{AC}\otimes\boldsymbol{BD}$

provided the dimensions of the matrices are such that the matrix products $\boldsymbol{AC}$ and $\boldsymbol{BD}$ exist.

(v) For $\boldsymbol{x}$ and $\boldsymbol{y}$ arbitrary vectors,

$$\boldsymbol{x}^T\otimes\boldsymbol{y}=\boldsymbol{y}\boldsymbol{x}^T=\boldsymbol{y}\otimes\boldsymbol{x}^T.$$

Appendix II
Theorem

Let the realization $\{\tilde{\boldsymbol{A}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{c}}^T\}$ be a balanced realization. Let

$$\hat{\boldsymbol{f}}(s)\triangleq(s\boldsymbol{I}-\tilde{\boldsymbol{A}})^{-1}\tilde{\boldsymbol{b}}$$

and

$$\hat{\boldsymbol{g}}(s)\triangleq(s\boldsymbol{I}-\tilde{\boldsymbol{A}}^T)^{-1}\tilde{\boldsymbol{c}}.$$

Then,

$$\|\hat{\boldsymbol{f}}\|_E=\|\hat{\boldsymbol{g}}\|_E$$

where

$$\|\hat{\boldsymbol{f}}\|_E=\sqrt{\hat{\boldsymbol{f}}^*(j\omega)\hat{\boldsymbol{f}}(j\omega)}$$

and

$$\|\hat{\boldsymbol{g}}\|_E=\sqrt{\hat{\boldsymbol{g}}^*(j\omega)\hat{\boldsymbol{g}}(j\omega)}.$$

Proof: For any controllable realization $\{\tilde{\boldsymbol{A}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{c}}^T\}$, a unique invertable symmetric matrix $\boldsymbol{R}$ exists such that [21]

$$\boldsymbol{R}\tilde{\boldsymbol{A}}=\tilde{\boldsymbol{A}}^T\boldsymbol{R}\quad\text{and}\quad\boldsymbol{R}\tilde{\boldsymbol{b}}=\tilde{\boldsymbol{c}}.$$

As a result, $\hat{\boldsymbol{g}}(s)=\boldsymbol{R}\hat{\boldsymbol{f}}(s)$.

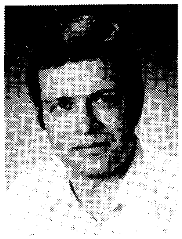
When the realization $\{\tilde{\boldsymbol{A}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{c}}^T\}$ is a balanced realization, $\boldsymbol{R}=\operatorname{diag}(\pm1)$, [22]. Thus for a balanced realization,

$$\hat{\boldsymbol{g}}^*(j\omega)\hat{\boldsymbol{g}}(j\omega)=\hat{\boldsymbol{f}}^*(j\omega)\boldsymbol{R}^T\boldsymbol{R}\hat{\boldsymbol{f}}(j\omega)=\hat{\boldsymbol{f}}^*(j\omega)\hat{\boldsymbol{f}}(j\omega)$$

and the result follows.

References

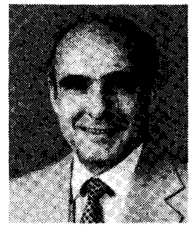
[1] W. J. Kerwin, L. P. Huelsman, and R. W. Newcomb, "State-variable synthesis for insensitive integrated circuit transfer functions," *IEEE J. Solid-State Circuits*, vol. SC-2, pp. 87–92, Sept. 1967.
[2] R. MacKay and A. Sedra, "Generation of low-sensitivity state-space active filters," *IEEE Trans. Circuits Syst.*, vol. CAS-27, pp. 863–870, Oct. 1980.
[3] L. B. Jackson, "Roundoff noise bounds derived from coefficient sensitivities for digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-23, pp. 481–485, Aug. 1976.
[4] S. Y. Hwang, "Minimum uncorrelated unit noise in state-space digital filtering," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-25, pp. 273–281, Aug. 1977.
[5] C. T. Mullis and R. A. Roberts, "Synthesis of minimum roundoff noise fixed point digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-23, pp. 551–562, Sept. 1976.
[6] V. Tavsanoğlu and L. Thiele, "Optimal design of state-space digital filters by simultaneous minimization of sensitivity and roundoff noise," *IEEE Trans. Circuits Syst.*, vol. CAS-31, pp. 884–888, Oct. 1984.
[7] L. Thiele, "Design of sensitivity and round-off noise optimal state-space discrete systems," *Int. J. Circuit Theory Appl.*, vol. 12, pp. 39–46, 1984.
[8] B. C. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 17–32, Feb. 1981.
[9] C-T. Chen, *Linear System Theory and Design*. New York: Holt, Rinehart and Winston, 1984.
[10] W. J. Lutz, "Design of multi-input multi-output systems with minimum sensitivity," Ph.D. dissertation, Northwestern University, Evanston, Illinois, June 1986.
[11] M. S. Ghausi and K. R. Laker, *Modern Filter Design: Active RC and Switched Capacitor*. Englewood Cliffs, NJ: Prentice-Hall, 1981.
[12] K. Glover, "All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds," *Int. J. Cont.*, vol. 39, no. 6, pp. 1115–1193, 1984.
[13] A. J. Laub, "Computation of 'balancing' transformations," in *Proc. 1980 JACC*, San Francisco, CA, 1980.
[14] E. I. Verriest and T. Kailath, "On generalized balanced realizations," *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 833–844, Aug. 1983.
[15] K. V. Fernando and H. Nicholson, "Singular perturbational model reduction of balanced systems," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 466–468 Apr. 1982.
[16] F. W. Fairman, S. S. Mahil, and J. A. DeAbreu, "Balanced realization algorithm for scalar continuous-time systems having simple poles," *Int. J. Syst. Sci.*, vol. 15, no. 6, pp. 685–694, 1984.
[17] G. W. Stewart, *Introduction to Matrix Computations*. New York: Academic, 1973.
[18] G. Golub and C. VanLoan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins Univ. Press, 1983.
[19] D. S. G. Pollock, "Tensor products and matrix differential calculus," *Linear Algebra and Its Appl.*, vol. 67, pp. 169–193, 1985.
[20] A. Graham, *Kronecker Products and Matrix Calculus With Applications*. New York: Halsted, (div. of Wiley), 1981.
[21] B. D. O. Anderson, "On the computation of the Cauchy index," *Q. Appl. Math.*, vol. 29, pp. 577–582, 1972.
[22] D. A. Wilson and A. Kumar, "Symmetry properties of balanced systems," *IEEE Trans. Automatic Contr.*, vol. AC-28, pp. 927–929, Sept. 1983.



William J. Lutz received the B.A. degree from DePauw University, Greencastle, Indiana, in 1969, and the B.S., M.S., and Ph.D. degrees in electrical engineering in 1976, 1977, and 1986, respectively, from Northwestern University, Evanston, Illinois.

From 1977 to 1983, he was with Auditory Research Laboratory, Northwestern University. During 1986, he was a MTS at Bell Communications Research, Red Bank, NJ. Since February 1987, he has been with Nicolet Instrument Corporation, Madison, WI as a Senior Design Engineer.

Dr. Lutz is a member of Tau Beta Pi and the Audio Engineering Society.



S. Louis Hakimi (S'56–M'59–SM'67–F'72) received the B.S. (Bronze Tablet), M.S. and Ph.D. degrees from the University of Illinois at Urbana, in electrical engineering in 1955, 1957, and 1959, respectively.

He was an Assistant Professor of Electrical Engineering at the University of Illinois (Urbana) during 1959 to 1961. He joined Northwestern University as an Associate Professor of Electrical Engineering in 1961 and was promoted to Full Professor in 1966. From 1972 to 1977, he served as Chairman of the Electrical Engineering Department. From September 1977 to August 1986 he was a professor in Electrical Engineering and Computer Science, Industrial Engineering/Management Science, and Applied Mathematics at Northwestern University. In September 1986, he joined University of California at Davis as Professor and Chairman of the Department of Electrical Engineering and Computer Science. His research interests lie in applications of graph theory and combinatorics to circuits, network theory, coding theory, operations research and computer science.