Automatica, Vol. 2, pp. 211-234. Pergamon Press, 1965. Printed in Great Britain.

A TRANSFORM TECHNIQUE FOR MULTIVARIABLE, TIME-VARYING, DISCRETE-TIME, LINEAR SYSTEMS*

A. W. NAYLOR

Department of Electrical Engineering and the Cooley Electronics Laboratory,
University of Michigan, Ann Arbor, Michigan

Summary—It is shown that a transform technique developed for single-input, single-ouput time-varying, discrete-time, linear systems can be meaningfully extended to multivariable systems.

1. INTRODUCTION

IN THE last several years the use of matrices to characterize time-varying, discrete-time, linear systems has been growing. For example, FRIEDLAND [1] has done much important work in this area. CRUZ [2] has shown how this idea can be employed in the design of control systems. Recently the author showed [3] that many of the frequency response concepts of time-invariant systems can be generalized so that they are meaningful for time-varying systems. So far attention has been centered on single-input, single-output systems. It is the purpose of this article to show that the previously developed methods can be applied, after certain changes and re-interpretations, to multivariable systems.

The systems considered here are those with m input channels and n output channels which can be characterized by the following equation:

$$y^j(t_l) = \sum_{k=1}^{m} \sum_{s=1}^{N} g_{jk}[t_l, t_s]\, x^k(t_s) \begin{Bmatrix} j=1, \ldots, n \\ l=1, \ldots, N \end{Bmatrix} \tag{1.1}$$

where j, l, m, N are positive integers and

(i) $y^j(t_l)$ is the output on j-th output channel at the l-th sampling time,

(ii) $g_{jk}[t_l, t_s]$ is the output on the j-th output channel at the l-th sampling time caused by a unit input on the k-th input channel at the s-th sampling time,

(iii) $x^k(t_s)$ is the input on the k-th input channel at the s-th sampling time.

* The research reported in this article was supported through the Cooley Electronics Laboratory, task 12, The University of Michigan by the U.S. Army Electronics Research and Development Laboratory, Fort Monmouth, under Contract No. DA-36-039 sc-89227.

[1] B. FRIEDLAND, A Technique for the Analysis of Time-Varying Sampled-Data Systems, *Trans. Amer. Inst. Elect. Engrs.* **75**, 407–412 (1957).

[2] J. B. CRUZ, Sensitivity Considerations for Time-Varying Sampled-Data Feedback Systems, *IRE Transactions on Automatic Control* 228–236 (1961).

[3] A. W. NAYLOR, Generalized Frequency Response Concepts for Time-Varying, Discrete-Time Linear Systems, *IEEE Transactions on Circuit Theory*, September (1963).

It is assumed for convenience and not out of necessity that all input and output channels are sampled simultaneously, i.e., at

$$t_1, t_2, \ldots, t_l, \ldots, t_N .$$

It can be seen from (1.1) that it has also been assumed that the state of the system before t_1 is such that zero input yields zero output. In case we can not neglect the initial state, it can be incorporated into the input and a development similar to the one presented here can be carried out.

We can also write (1.1) as a partitioned matrix equation as follows:

$$\begin{bmatrix} \mathbf{y}(t_1) \\ \mathbf{y}(t_2) \\ \cdots \\ \mathbf{y}(t_N) \end{bmatrix} = \begin{bmatrix} G[t_1, t_1] & \cdots & G[t_1, t_N] \\ G[t_2, t_1] & \cdots & G[t_2, t_N] \\ \cdot & \cdot & \cdot \\ G[t_N, t_1] & \cdots & G[t_N, t_N] \end{bmatrix} \begin{bmatrix} \mathbf{x}(t_1) \\ \mathbf{x}(t_2) \\ \cdots \\ \mathbf{x}(t_N) \end{bmatrix} \tag{1.2}$$

where $\mathbf{y}(t_l)$ is a vector made up of the outputs in the n output channels at the l-th sampling time, $\mathbf{x}(t_s)$ is a vector made up of the inputs on the m input channels at the s-th sampling time, and $G[t_l, t_s]$ is an nxm matrix which relates $\mathbf{y}(t_l)$ to $\mathbf{x}(t_s)$. The notation can be further simplified by writing

$$\bar{\mathbf{y}} = G\bar{\mathbf{x}}, \text{ where } \bar{\mathbf{y}} = \begin{bmatrix} \mathbf{y}(t_1) \\ \mathbf{y}(t_2) \\ \cdots \\ \mathbf{y}(t_N) \end{bmatrix}, \quad \bar{\mathbf{x}} = \begin{bmatrix} \mathbf{x}(t_1) \\ \mathbf{x}(t_2) \\ \cdots \\ \mathbf{x}(t_N) \end{bmatrix}, \tag{1.3}$$

and G, an $nN \times mN$ matrix, is the partitional matrix in (1.2).

Since a physically realizable system is one whose present output is independent of future inputs, it is clear that (1.2) represents a physically realizable system if and only if $G[t_l, t_s]=0$ for all $s>l$, that is, if the matrix G is "lower staircase".

The matrix G represents a time-invariant system in (1.2) if $G[t_l, t_s]=G[t_l-t_s]$ for all l and s.

It is, of course, possible to consider a situation where G is an infinite matrix, that is, where an infinite number of sampling times, t_l, spread over the infinite (or, perhaps, semi-infinite) time interval are considered. Although the methods presented here can, with a few restrictions, be extended to cover infinite matrices, such an extension is of limited practical value for two reasons. First, it is usually very difficult if not impossible to obtain the solutions of the equations which arise. In fact, often the only way to handle an infinite G is to replace it by a finite G. Secondly, infinite time intervals are rarely of practical importance for the simple reason that systems are not operated that long. The main reason infinite time

intervals are sometimes considered instead of finite ones (e.g. when using the Laplace transform with time-invariant, continuous-time systems) is that the mathematical analysis is simplified. Here the mathematical analysis is simplified by maintaining the finite time interval.

2. BACKGROUND TO TRANSFORM METHODS

At this point it is worthwhile to reconsider the basic goal behind transform techniques. Simply stated, the goal is to transform, where possible, a given operation into the operation of multiplication by a function (e.g., the transfer function). Let L denote some given linear operation such that

$$y = Lx, \tag{2.1}$$

where x is an element of the input space* or domain and y is an element of the output space or range. L is transformed to a multiplication in the desired sense if an invertible transformation T and a function $\phi(\lambda)$ can be found such that

$$L = T^{-1}\phi(\lambda)T. \tag{2.2}$$

Thus, if L is a linear transformation of a space H_1 into itself, T is a linear transformation of H_1 onto a space H_2, and multiplication by $\phi(\lambda)$ is a linear transformation of H_2 into itself, we have the following situation:

$$\begin{array}{ccc} H_1 & \xrightarrow{L} & H_1 \\ \Big\downarrow T & & \Big\uparrow T^{-1} \\ H_2 & \xrightarrow{\phi(\lambda)} & H_2 \end{array}$$

A classic example of such a transformation to a multiplication is offered by the well-known use of the Laplace transform with linear, time-invariant systems. There T is the direct Laplace transform; $\phi(\lambda)$ is the transfer function with the complex numbers, λ, taken along the Wagner-Bromwich contour; and T^{-1} is the inverse Laplace transform. Another well known example in the same spirit is the diagonalization of a square matrix by means of a similarity transformation. There L is the matrix to be diagonalized; T is a non-singular (i.e., invertible) matrix; and $\phi(\lambda)$ is a diagonal matrix. This diagonal matrix can be viewed as equivalent to a function defined on the integers from 1 through N (L being an $N \times N$ matrix). Thus, $\phi(1) = \phi_1$, the first entry on the diagonal; $\phi(2) = \phi_2$, the second entry; and so on through $\phi(N) = \phi_N$. If, in similar manner, an arbitrary vector upon which the diagonal matrix operates is viewed as a function—say $x(\lambda)$, defined on the integers 1 to N—then operation with the diagonal matrix can be viewed as a multiplication of the function $x(\lambda)$ by the function $\phi(\lambda)$. This latter example leads one to refer to the general process as a "diagonalization of L" whether L is a matrix or not. Moreover, owing to (2.2) this diagonalization is said to be carried out on the basis of a similarity transformation. A detailed discussion of the philosophy behind such diagonalization applied to continuous-time systems is given by ZADEH [4]. Finally, it must be emphasized that such a diagonalization is not always possible.

* The element x might be a function or a sequence, for example; the corresponding spaces would be function or sequence spaces. In the case of the systems considered in this article the element x is a sequence and the space is a sequence space.

[4] L. ZADEH, A General Theory of Linear Signal Transmission Systems, *J. Franklin Inst.* 293–312 (1952).

In addition to the obvious fact that such a "diagonalization" simplifies the representation of the operator L (assuming L is "diagonalizable"), several specific aspects should be noted. Perhaps the most important of these relates to the combination of two operators, say L_1 and L_2, which can be diagonalized by the same transform T. In this case,

$$L_1 = T^{-1}\phi_1(\lambda)T$$

and

$$L_2 = T^{-1}\phi_2(\lambda)T \tag{2.3}$$

It immediately follows that

$$L_1L_2 = T^{-1}\phi_1(\lambda)\phi_2(\lambda)T$$

and

$$L_1 + L_2 = T^{-1}[\phi_1(\lambda) + \phi_2(\lambda)]T \tag{2.4}$$

Thus, T diagonalizes L_1L_2 and L_1+L_2, and the resulting "transfer functions" are simply $\phi_1(\lambda)\,\phi_2(\lambda)$ and $\phi_1(\lambda)+\phi_2(\lambda)$. Similarly, T also diagonalizes L_2L_1. Carrying (2.4) further, if $L_1, L_2, \ldots, L_n$ are a set of operators which can be diagonalized by T, then any polynomial function*, say $P[L_1, L_1^{-1}, \ldots, L_n, L_n^{-1}]$, of these operators and their inverses [where they exist] can be diagonalized by T and the "transfer function" is given by $P[\phi_1(\lambda), 1/\phi_1(\lambda), \ldots, \phi_n(\lambda), 1/\phi_n(\lambda)]$. Moreover, if P is invertible, T diagonalizes it and the transfer function is $1/P[\phi_1(\lambda), 1/\phi_1(\lambda), \ldots, \phi_n(\lambda), 1/\phi_n(\lambda)]$. Simply stated an operational calculus is obtained. An example is, of course, given by polynomials of the derivative operator, $\mathrm{d}/\mathrm{d}t$, and the Laplace transform as T.

It should be noted in (2.3) and (2.4) that a necessary condition for L_1 and L_2 to be diagonalized by the same T is $L_1L_2 = L_2L_1$. Since not all linear operators commute, one can not expect to find one T which will diagonalize all operators. Thus, an all-purpose transform, in the sense of (2.2), is not possible. On the other hand, it is possible to find T's which diagonalize all members of large classes of linear operators. One such class is made up of time-invariant linear differential operators.

In any event, it is true that given an operator L which can be diagonalized by a transform T, the operator $I+L$, where I is the identity operator, can always be diagonalized by T. That is, if $L = T^{-1}\phi(\lambda)\,T$ then $I+L = T^{-1}[1+\phi(\lambda)]T$ and the "transfer function" is $1+\phi(\lambda)$. If $I+L$ has an inverse, then T diagonalizes the inverse and $1/[1+\phi(\lambda)]$ is the transfer function of the inverse. Needless to say, operators of the form $I+L$ are of great interest in control theory. In fact, the key equation in classical control theory is

$$(I+L_1)y = L_2x \tag{2.5}$$

If L_1 and L_2 can both be diagonalized by T and $(I+L_1)$ has an inverse, (2.5) can be replaced by

$$Ty = \frac{\phi_2(\lambda)}{1+\phi_1(\lambda)}Tx \tag{2.6}$$

Obviously, this simple operational calculus is very useful. Unfortunately, it can easily occur that L_1 and L_2 in (2.5) can not be diagonalized by the same T; moreover, it may not be possible to diagonalize them at all.

Diagonalization, then, simplifies the representation of the operator and leads to operational calculi.

* This statement is valid for a larger class of functions than polynomials.

In the case of linear time-invariant systems it does more. There the form of the transfer function, $\phi(\lambda)$, yields insight into the nature of the system under consideration. In particular, the magnitude $/\phi(\lambda)/$ has significance and yields insight. Roughly speaking, where $/\phi(\lambda)/$ is large there is large transmission through the system and conversely. The underlying mathematical explanation is Parseval's (Plancherel's) theorem. Unfortunately, in the case of operators L which represent time-varying systems, the existence of a transform T which diagonalize L does not imply that an analogy to Parseval's (Plancherel's) theorem exists. In fact, there are many important situations where one does not exist.

Let us consider generalizations of Parseval's theorem a little further. Returning to (2.1) and (2.2)

$$y = Lx = T^{-1}\phi(\lambda)Tx$$

Assume that L is a bounded linear transformation from a Hilbert space H_1 into itself and T is an invertible transformation of H_1 onto a Hilbert space H_2. Let $Tx = X(\lambda)$, $Ty = Y(\lambda)$, and denote the inner products on H_1 and H_2 by $(y,x)_{H_1}$ and $(Y,X)_{H_2}$, respectively. In the case of square-integrable functions these become

$$(y,x)_{H_1} = \int y(t)x'(t)\mathrm{d}t$$

and (2.7)

$$(Y,X)_{H_2} = \int Y(\lambda)X'(\lambda)\mathrm{d}\lambda$$

where the $x'(t)$ denotes the complex conjugate of $x(t)$. The goal is to relate $(y,y)_{H_1}$, the "energy" of the output, to $Y(\lambda)$, the "transform domain" representation of the output, and, because $Y(\lambda) = \phi(\lambda)\ X(\lambda)$, also relate it to $\phi(\lambda)$ and $X(\lambda)$. It follows from (2.1) and (2.2) that

$$\begin{aligned}(y,y)_{H_1} &= (Lx,\ Lx)_{H_1} \\ &= (T^{-1}\phi(\lambda)Tx,\ T^{-1}\phi(\lambda)Tx)_{H_1} \\ &= (T^{-1}Y(\lambda),\ T^{-1}Y(\lambda))_{H_1}\end{aligned} \tag{2.8}$$

If the adjoint* of T^{-1} is designated by $(T^{-1})^{**}$ and $(T^{-1})^{**}\ T^{-1}$ is designated by Q, then (2.8) becomes

$$(y,y)_{H_1} = (Y(\lambda),\ QY(\lambda))_{H_2} \tag{2.9}$$

The usefulness of the above expression depends on the nature of the transformation Q.

The desirable situation is that $QY(\lambda)$ be easily expressible in terms of $Y(\lambda)$. For example, if T is a unitary transformation†, then $T^{-1} = T^{**}$; therefore, $Q = I$, the identity transformation, and (2.9) becomes

$$\int y(t)y'(t)\mathrm{d}t = \int Y(\lambda)Y'(\lambda)\mathrm{d}\lambda \tag{2.10}$$

which is a generalization of Parseval's theorem.

* Recall that the adjoint of an operator A which maps H_1 into H_2 is that operator A^{**}, mapping H_2 into H_1, for which $(Au,\ v)H_1 = (u, A^{**}v)\ H_2$ for all u in H_1 and all v in H_2.

† Recall that L can be "diagonalized" by a unitary transformation if and only if it is normal, i.e. it commutes with its adjoint, $L^{**}L = LL^{**}$. It is not true that all "diagonalizable" operators are normal.

Substituting $Y(\lambda)=\phi(\lambda)X(\lambda)$ into (2.10) yields

$$(y,y)_{H_1}=\int \phi(\lambda)\phi'(\lambda)X(\lambda)X'(\lambda)\mathrm{d}\lambda \tag{2.11}$$

For example, in the case of the Fourier transform pair*

$$X(\lambda)=\int_{-\infty}^{\infty} x(t)\varepsilon^{-j2\pi\lambda t}\mathrm{d}t$$

and

$$x(t)=\int_{-\infty}^{\infty} X(\lambda)\varepsilon^{j2\pi\lambda t}\mathrm{d}\lambda$$

where the implied T is unitary and, thus, $Q=\mathrm{I}$.
Assuming that the operator L represents the appropriate type of time-invariant system, it can be diagonalized by this T and (2.11) becomes

$$\int_{-\infty}^{\infty} y^2(t)\mathrm{d}t=\int_{-\infty}^{\infty} \phi(\lambda)\,\phi'(\lambda)\,X(\lambda)\,X'(\lambda)\,\mathrm{d}\lambda \tag{2.12}$$

Thus, at least when T is unitary and $Q=\mathrm{I}$, $|\phi(\lambda)|^2$ directly and simply characterizes the energy transfer capabilities of the system.

On the other hand, in general the transformation Q in (2.9) need not lead to a simple characterization of transfer properties such as (2.11) or (2.12). In fact, if an operator can be diagonalized by a transform T, there is no reason to expect (without further restriction) a simple correlation between the "transfer function" $\phi(\lambda)$ and the energy-transfer capabilities of the system. Thus, much of the insight and many of the analytic techniques associated with the use of transfer functions based on Fourier (or Laplace) transforms may not carry over to the general case. Unfortunately, this is usually the case for G-matrices corresponding to physically realizable systems.

In summary, then, the diagonalization discussed above has certain advantages and certain disadvantages.

The advantages are as follows:

A_1. If the operators $L_1,\ldots,L_n$ can be diagonalized by a transformation T, then any polynomial function of these operators and their inverses (where they exist) can be diagonalized by T.
Moreover, if the polynomial function itself is invertible, its inverse can be diagonalized by T. These properties of diagonalization allow operational calculi to be devloped for classes of time-varying linear systems. In any event, if the transform T diagonalizes L and the operator $(I+L)$ is invertible, then T also diagonalizes $(I+L)$ and $(I+L)^{-1}$. The resulting operational calculus is often very useful.

A_2. In certain cases, for example when L is normal, L can be diagonalized in a way that leads to a meaningful generalization of Parseval's (Plancherel's) theorem.

The disadvantages are as follows:

D_1. Not all linear operators can be diagonalized in the above way. For example, not all matrices can be so diagonalized.

D_2. Given any transformation T, only a relatively small class of linear operators with the appropriate domain and range will be diagonalized by T.

D_3. In many important cases Parseval's (Plancherel's) theorem can not be generalized.

* Here $x(t)$ is restricted to the intersection of square-integrable and absolutely integrable functions. In order to consider all square-integrable functions the Fourier-Plancherel transform must be used.

In the next section, a transform is introduced which overcomes some of the above difficulties at the cost of sacrificing advantages. Moreover, it adds an advantage which not even the Laplace transform as usually applied to time-invariant systems has: for a multivariable system, it yields *one transfer function* instead of a matrix of transfer functions.

3. TRANSFORM TECHNIQUE

LANZCOS [5] has shown that *all* system matrices, G, can be decomposed as follows:

$$G=(Y\Delta f)\Lambda X^T, \tag{3.1}$$

where

(i) $\Delta f=1/N$ and is referred to as an increment of generalized frequency.

(ii) Y is a matrix whose columns are pairwise orthogonal to one another and each is of norm $\sqrt{N}$. At this point the norm employed is the familiar Euclidean norm. (This decomposition can be generalized so that norms based on arbitrary inner products can be employed). If n, the number of output channels, is greater than or equal to m, the number of input channels, then Y is a $(nN\times mN)$-matrix. If $n<m$, then Y is an $(nN\times nN)$-matrix.

(iii) Λ is a diagonal matrix with all non-negative entries. If $n\geqslant m$, Λ is an $(mN\times mN)$-matrix. If $n<m$ Λ is an $(nN\times nN)$-matrix.

(iv) X is a matrix (X^T is the transpose of X) whose columns are pairwise orthogonal to one another and each is of norm $\sqrt{N}$. If $n>m$, X^T is an $(mN\times mN)$-matrix. If $n<m$, is an $(nN\times mN)$-matrix. This decomposition (see Appendix) is essentially the same as the one used with single-input, single-output systems [3]. The fact that it is valid for nonsquare matrices allows the present extension to multivariable systems.

As in the case of single-input, single-output systems, it is shown below that X^T acts as a direct transform, Λ acts as a transfer function, and $(Y\Delta f)$ acts as an inverse transform.

Case 1: $n>m$

Assume for the moment that there are more output than input channels, that is, $n>m$. The matrices in the decomposition are then of the following structure:

$$Y=\begin{bmatrix} y^1{}_1(t_1) & y^1{}_2(t_1) & \cdots & y^1{}_{mN}(t_1) \\ y^2{}_1(t_1) & y^2{}_2(t_1) & \cdots & y^2{}_{mN}(t_1) \\ \cdots & \cdots & \cdots & \cdots \\ y^n{}_1(t_1) & y^n{}_2(t_1) & \cdots & y^n{}_{mN}(t_1) \\ y^1{}_1(t_2) & y^1{}_2(t_2) & \cdots & y^1{}_{mN}(t_2) \\ y^2{}_1(t_2) & y^2{}_2(t_2) & \cdots & y^2{}_{mN}(t_2) \\ \cdots & \cdots & \cdots & \cdots \\ y^n{}_1(t_2) & y^n{}_2(t_2) & \cdots & y^n{}_{mN}(t_2) \\ \cdots & \cdots & \cdots & \cdots \\ y^1{}_1(t_N) & y^1{}_2(t_N) & \cdots & y^1{}_{mN}(t_N) \\ y^2{}_1(t_N) & y^2{}_2(t_N) & \cdots & y^2{}_{mN}(t_N) \\ \cdots & \cdots & \cdots & \cdots \\ y^n{}_1(t_N) & y^n{}_2(t_N) & \cdots & y^n{}_{mN}(t_N) \end{bmatrix} \tag{3.2}$$

[5] C. LANCZOS, Linear Systems in Self-Adjoint Form, *Amer. Math. Monthly* **65**, 665–679 (1958).

An equivalent, simplified form of this matrix is

$$Y=\begin{bmatrix} \mathbf{y}_1(t_1) & \mathbf{y}_2(t_1) & \dots & \mathbf{y}_{mN}(t_1) \\ \mathbf{y}_1(t_2) & \mathbf{y}_2(t_2) & \dots & \mathbf{y}_{mN}(t_2) \\ \dots & \dots & \dots & \dots \\ \mathbf{y}_1(t_N) & \mathbf{y}_2(t_N) & \dots & \mathbf{y}_{mN}(t_N) \end{bmatrix} \tag{3.3}$$

where

$$\mathbf{y}_j(t_k)=\begin{bmatrix} y^1{}_j(t_k) \\ y^2{}_j(t_k) \\ \dots \\ y^n{}_j(t_k) \end{bmatrix} \qquad \begin{Bmatrix} j=1, \dots, mN \\ k=1, \dots, N \end{Bmatrix} \tag{3.4}$$

An even more concise notation is

$$Y=[\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \dots, \bar{\mathbf{y}}_{mN}], \tag{3.5}$$

where

$$\bar{\mathbf{y}}_j=\begin{bmatrix} \mathbf{y}_j(t_1) \\ \mathbf{y}_j(t_2) \\ \dots \\ \mathbf{y}_j(t_N) \end{bmatrix} \tag{3.6}$$

Thus, the vectors $\bar{\mathbf{y}}_j$, $j=1, 2, \dots, mN$, are mutally orthogonal and of norm $\sqrt{N}$, that is, $(\bar{\mathbf{y}}_j, \bar{\mathbf{y}}_k)=\delta_{jk}\ 1/\Delta f=\delta_{jk}N$, where $\delta_{jk}=$Kronecker delta symbol.
H re

$$(\bar{\mathbf{y}}_j, \bar{\mathbf{y}}_k)=y^1{}_j(t_1)y^1{}_k(t_1)+y^2{}_j(t_1)y^2{}_k(t_1)+ \dots +y^n{}_j(t_N)y^n{}_k(t_N) \tag{3.7}$$

The matrix is an $(mN \times mN)$ diagonal matrix with non-negative main diagonal entries, that is,

$$\Lambda=\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & 0 \\ & & \lambda_3 & \\ 0 & & \dots & \\ & & & \lambda_{mN} \end{bmatrix}, \tag{3.8}$$

where $\lambda_i \geq 0$, $i=1, 2, \dots, mN$.

The matrix X^T is an $(mN \times mN)$-matrix with mutually orthogonal rows of the form

$$X^T = \begin{bmatrix} x^1{}_1(t_1), x^2{}_1(t_1), \ldots, x^m{}_1(t_1), x^1{}_1(t_2), x^2{}_1(t_2), \ldots, x^m{}_1(t_2), \ldots \\ \qquad \ldots, x^1{}_1(t_N), \ldots, x^m{}_1(t_N) \\ x^1{}_2(t_1), x^2{}_2(t_1), \ldots, x^m{}_2(t_1), x^1{}_2(t_2), x^2{}_2(t_2), \ldots, x^m{}_2(t_2), \ldots, \\ \qquad \ldots, x^1{}_2(t_N). \ldots, x^m{}_2(t_N) \\ \qquad \ldots \\ x^1{}_{mN}(t_1), x^2{}_{mN}(t_1), \ldots, x^m{}_{mN}(t_1), x^1{}_{mN}(t_2), \ldots, x^m{}_{mN}(t_2), \ldots \\ \qquad \ldots, x^1{}_{mN}(t_N), \ldots, x^m{}_{mN}(t_N) \end{bmatrix} \tag{3.9}$$

This matrix can be simplified to*

$$X = \begin{bmatrix} \mathbf{x}_1(t_1) & \mathbf{x}_2(t_1) & \ldots & \mathbf{x}_{mN}(t_1) \\ \mathbf{x}_1(t_2) & \mathbf{x}_2(t_2) & \ldots & \mathbf{x}_{mN}(t_2) \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{x}_1(t_N) & \mathbf{x}_2(t_N) & \ldots & \mathbf{x}_{mN}(t_N) \end{bmatrix} \tag{3.10}$$

or

$$X = [\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \ldots, \bar{\mathbf{x}}_{mN}], \tag{3.11}$$

where substitutions analagous to those used with the Y-matrix are employed. Moreover the vectors $\bar{\mathbf{x}}_j$, $j=1, 2, \ldots, mN$, are mutually orthogonal and $(\bar{\mathbf{x}}_j, \bar{\mathbf{x}}_k)=\delta_{jk}/\Delta f=\delta_{jk}N$. It is important to note that $G\bar{\mathbf{x}}_i=\lambda_i\bar{\mathbf{y}}_i$, $i=1, 2, \ldots, mN$. This relation and the orthogonality of the $\bar{\mathbf{x}}_i$-vectors and the $\bar{\mathbf{y}}_i$-vectors is the key to what follows.

It can be seen from (3.1) that the first step in the operation of a decomposed G on an arbitrary input, $\bar{\mathbf{x}}$, is $X^T\bar{\mathbf{x}}$. It will now be shown that this first step can be interpreted as taking the direct transform of $\bar{\mathbf{x}}$ to obtain its generalized frequency domain representation. First note that the orthogonal set of vectors, $\bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_{mN}$, spans the linear space of all possible inputs.

Thus, an arbitrary input, $\bar{\mathbf{x}}$, can be uniquely represented in the form

$$\bar{\mathbf{x}} = \sum_{j=1}^{mN} r_j\bar{\mathbf{x}}_j\Delta f \tag{3.12}$$

where the r_j's are constants. Taking the inner product of both sides of (3.12) with $\bar{\mathbf{x}}_i$ yields

$$r_i=(\bar{\mathbf{x}}, \bar{\mathbf{x}}_i) \qquad i=1, 2, \ldots, mN \tag{3.13}$$

for the determination of the r_i's. In a manner similar to that employed in the single-input, single-output case [3], the sequence made up of the r_i's may be considered to be a "frequency domain representation" or transform of $\bar{\mathbf{x}}$. Moreover, a piecewise constant function

* Note that here X and not XT is written.

$R_x(f)$, of generalized frequency can be introduced as the frequency domain representation of $\bar{\mathbf{x}}$. This function is obtained by associating an increment Δf of generalized frequency with each $\bar{\mathbf{x}}_i$-vector. Thus, $R_x(f)$ is defined as follows

$$R_X(f)=\begin{cases} r_i \text{ for } (i-1)\Delta f \leq f < i\Delta f, & i=1, 2, \ldots, mN \\ 0 \text{ for } m<f<n \end{cases} \tag{3.14}$$

where, again, $\Delta f = 1/N$. An illustration is shown in Fig. 1. The frequency domain is defined to include the interval $m<f<n$ and $R_x(f)$ is defined to be equal to zero on this interval so that the input and output frequency domain representations can be compatible. In general,

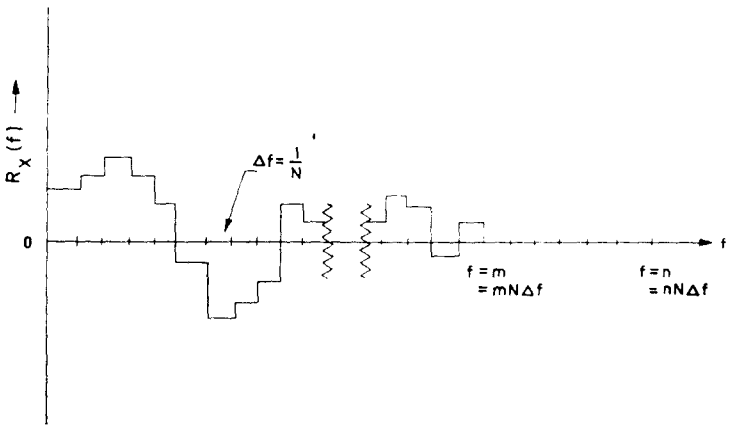

FIG. 1. Typical generalized frequency domain representation of an input.

the generalized frequency domain is defined to be $0 \leq f < \max[m, n]$. Since it can be seen from (3.11) and (3.13) that

$$X^T\bar{\mathbf{x}} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \cdots \\ r_{mN} \end{bmatrix} \tag{3.15}$$

it follows that $X^T\bar{\mathbf{x}}$ can indeed be viewed as the direct transform, that is, $X^T\bar{\mathbf{x}}$ yields $R_x(f)$.

Next, it is easily shown that this generalized frequency domain representation leads to a meaningful generalization of Parseval's (Plancherel's) theorem. In fact, trivial calculations show that

$$(\bar{\mathbf{x}}, \bar{\mathbf{x}}) = \sum_{j=1}^{mN} r_j^2 \Delta f = \int_0^n R_X^2(f)\, df, \tag{3.16}$$

that is, $R_x^2(f)$ can be viewed as "energy" per unit bandwidth.

In a similar manner, the generalized frequency domain representation of an arbitrary output, $\bar{\mathbf{y}}$, is given by

$$\bar{\mathbf{y}} = \sum_{j=1}^{mN} c_j \bar{\mathbf{y}}_j \Delta f, \tag{3.17}$$

and it follows that a function $C_Y(f)$, analogous to $R_X(f)$, can be defined in the generalized frequency domain, $0 \leq f < n$, by

$$C_Y(f) = \begin{cases} c_i \text{ for } (i-1)\Delta f \leq f < i\Delta f, & i = 1, 2, \ldots, mN \\ 0 \text{ for } m \leq f < n \end{cases} \tag{3.18}$$

so that

$$(\bar{\mathbf{y}}, \bar{\mathbf{y}}) = \int_0^n C_Y^2(f)\, df \tag{3.19}$$

and $C_Y(f)$ can be viewed as the transform of $\bar{\mathbf{y}}$. It should be noted that an arbitrary vector in the nN-dimensional vector space which contains the output vectors, $\bar{\mathbf{y}}$, need not be a linear combination of the $\bar{\mathbf{y}}_j$-vectors as in (3.17). The $\bar{\mathbf{y}}_j$-vectors span only the mN-dimensional subspace which is the range of the G-matrix under consideration. Therefore, this transform is not, as presented here, applicable to the whole nN-dimensional vector space. $C_Y(f)$ has been defined equal to zero for $m \leq f < n$ in order to emphasize the fact that the range of G is a proper subspace (recall that here we are considering the case $m < n$).

The functions $R_X(f)$ and $C_Y(f)$, then, are the frequency domain representations of the input and output, respectively. The subscripts X and Y indicate that the transforms are with respect to the $\bar{\mathbf{x}}_i$- and $\bar{\mathbf{y}}_i$-vectors, respectively. It is now easy to introduce a generalized transfer function relating $R_X(f)$ and $C_Y(f)$. It is clear from the decomposition (3.1) that the output corresponding to an input $\bar{\mathbf{x}}_i$ is $\lambda_i \bar{\mathbf{y}}_i$. Thus, in the spirit of the definitions of $R_X(f)$ and $C_Y(f)$, the transfer function for the system can be defined by

$$\Lambda(f) = \begin{cases} \lambda_i \text{ for } (i-1)\Delta f \leq f < i\Delta f, & i = 1, 2, \ldots, mN \\ 0 \text{ for } m \leq f < n. \end{cases} \tag{3.20}$$

It follows that the generalized frequency domain representation for the system operation is given by

$$C_Y(f) = \Lambda(f) R_X(f) \qquad 0 \leq f < n \tag{3.21}$$

and from (3.19) that

$$(\bar{\mathbf{y}}, \bar{\mathbf{y}}) = \int_0^n \Lambda^2(f)\, R_X^2(f)\, df \tag{3.22}$$

This equation along with (3.16) and (3.19) is, of course, one of the primary justifications employing the present decomposition of G. It can be seen by comparing (3.22) with (2.11) or (2.12) that it is the desired generalization of Parseval's (Plancherel's) theorem and $\Lambda(f)$ simply and meaningfully characterizes the "energy" transfer properties of the system under consideration.

Finally, it follows from (3.5) and (3.17) that the matrix multiplication of $\Lambda X^T \bar{\mathbf{x}}$ by $(Y\Delta f)$ is equivalent to transforming $C_Y(f)$ to the time domain by the "inverse transform" $(Y\Delta f)$.

Case II: $m \geq n$

So far it has been assumed that $n > m$; if $m > n$ it is merely necessary to interchange the roles of m and n. The frequency domain becomes $0 < f < m$. The matrix Y becomes an $(nN \times nN)$-instead of an $(nN \times mN)$-matrix, Λ becomes an $(nN \times nN)$-instead of an $(mN \times mN)$-matrix, and X^T becomes an $(nN \times mN)$-matrix instead of an $(mN \times mN)$-matrix.

ADVANTAGES AND DISADVANTAGES OF DECOMPOSITION

In the foregoing section the original operation has been transformed to a multiplication by a function $\Lambda(f)$ defined on the interval $0 \leqslant f < n$ (or $0 \leqslant f < m$ depending on Case I or II). However, this has not been accomplished by means of the similarity transformation discussed in Section 3, for the inverse transform implied by $(Y\Delta f)$ is not necessarily the inverse of the direct transform implied by X^T. The question is how do the advantages and disadvantages of this transform method relate to those listed at the end of Section 2.

What about an operational calculus? Consider the following salient points. First, consider tandem operation of two systems G_1 and G_2. Given that

$$G_1 = (Y_1 \Delta f) \Lambda_1 X_1{}^T$$

and

$$G_2 = (Y_2 \Delta f) \Lambda_2 X_2{}^T$$

for operational calculus purposes the desirable situation is to have*

$$G_1 G_2 = (Y_1 \Delta f) \Lambda_1 \Lambda_2 X_2{}^T$$

A sufficient condition for this reduction to take place is $Y_2 = X_1$, for then $X_1{}^T (Y_2 \Delta f) = I$. When G_1 and G_2 are invertible this condition is necessary, for then

$$(Y_1 \Delta f) \Lambda_1 \Lambda_2 X_2{}^T = (Y_1 \Delta f) \Lambda_1 X_1{}^T (Y_2 \Delta f) \Lambda_2 X_2{}^T$$

which yields

$$\Lambda_1 \Lambda_2 = \Lambda_1 X_1{}^T (Y_2 \Delta f) \Lambda_2$$

and

$$X_1{}^T (Y_2 \Delta f) = I$$

which implies that $Y_2 = X_1$. In case G_1 or G_2 or both are not invertible, it again follows that

$$\Lambda_1 \Lambda_2 = \Lambda_1 X_1{}^T (Y \Delta f) \Lambda_2$$

When Λ_1 and Λ_2 are both invertible (this can happen in spite of G_1 and/or G_2 not being invertible), $Y_2 = X_1$ is still a necessary and sufficient condition.

If either Λ_1 or Λ_2 or both are not invertible, that is, have zero entries on the main diagonal, then $Y_2 = X_1$ is no longer a necessary condition for some of the rows in $X_1{}^T$ and some of the columns of Y_2 may be multiplied by the possible zero entries in Λ_1 and Λ_2, respectively. It follows that the necessary condition becomes that $Y_2 = X_1$ except for the columns multiplied by zero. However, since these latter columns are not uniquely determined and can be chosen so that the corresponding ones in Y_2 and X_1 are equal to one another, it can be said that the desired reduction takes place if and only if Y_2 and X_1 can be selected so that they equal one another. Finally, note that this is a general statement that applies in all cases where the decomposition of G_1 or G_2 is not unique.

It follows from the foregoing remarks that given $G_1 = (Y_1 \Delta f) \Lambda_1 X_1{}^T$, one representation for the set of all G_2's for which $G_1 G_2 = (Y_1 \Delta f) \Lambda_1 \Lambda_2 X_2{}^T$ is

$$G_2 = (X_1 \Delta f) \Lambda_2 X_2{}^T,$$

where Λ_2 and X_2 are arbitrary or constrained by the requirements of physical realizability.

* Here it is assumed that the indicated matrix multiplication makes sense.

In any event, this class of G_2's is not empty; therefore, the concept of the product of two transfer functions representing tandem operation does carry over in a certain sense. On the other hand, it is true, for example, that the transfer function of G^2 is not necessarily $\Lambda^2(f)$, and this would be true if a similarity transformation were used.

Next consider the question of the decomposition of $I+G$.* If the decomposition of G is given by

$$G=(Y\Delta f)\Lambda X^T$$

and $Y\neq X$, it follows that $(Y\Delta f)\, IX^T\neq I$. Then

$$I+G\neq(Y\Delta f)(I+\Lambda)X^T,$$

which means that except for the special case $Y=X$, decompositions of $I+G$ and G cannot be carried out by the same Y and X transformations. This is not the case, of course, when similarity transformations are used.

Still on the subject of an operational calculus, the parallel connection of systems is also of interest. Let G represent the parallel connection of G_1 and G_2; that is, $G=G_1+G_2$. It can be seen that if $G_1=(Y\Delta f)\Lambda_1 X^T$ and $G_2=(Y\Delta f)\Lambda_2 X^T$, that is, if G_1 and G_2 have the same Y and X-matrices, than $G=(Y\Delta f)(\Lambda_1+\Lambda_2)X^T$. Thus, the same transforms (Y and X) are associated with G_1, G_2, and G, and the transfer function Λ for the sum G_1+G_2 is the sum $\Lambda_1+\Lambda_2$.

In summary, when the decomposition presented in this article is employed a multiplication and addition of transfer functions results for certain classes of tandemly and parallel connected systems—just as multiplication and addition of transfer functions results for certain, presumably other, classes of tandemly and parallel connected systems when a decomposition based on a similarily transformation (Section 2) is used.

In regard to a generalization of Parseval's (Plancherel's) theorem, it is clear from the foregoing discussion that a meaningful generalization is always possible. This is an extremely important property and one which is not present when similarity transformations are employed.

Another extremely important property of the present decomposition, and one not present with a similarity transformation, is that it can be applied to an arbitrary system matrix, G.

One last point: The nature of the transform implied by the matrix X^T and its constituent $\bar{\mathbf{x}}_i$-vectors should be carefully appreciated. Since arbitrary inputs, $\bar{\mathbf{x}}$, are represented as linear combinations of the $\bar{\mathbf{x}}_i$-vectors, the $\bar{\mathbf{x}}_i$'s can be viewed as basic or fundamental inputs, and the response to these fundamental inputs completely characterizes the system. A key point is that these fundamental inputs can and probably do involve simultaneous inputs on more than one channel, which is in contrast, for example, to the approach to time-invariant multivariable systems implied when the final system characterization is a matrix made up of transfer functions. There each column of the transfer function matrix is the Laplace transform of the output when a unit impulse is applied to one input channel *while the other input channels have zero input.*

Thus, the present approach might be characterized as treating all input channels simultaneously rather than one at a time. Related remarks can be made regarding the $(Y\Delta f)$ matrix and its constituent $\bar{\mathbf{y}}_i$-vectors.

* Here G is assumed to be square.

5. EXTENSION OF SINGLE-VARIABLE RESULTS TO MULTIVARIABLE SYSTEMS

It has been shown [3] that many important results follow from the matrix decomposition discussed in the foregoing section when it is applied to matrices representing two-port systems (single-input, single-output channel systems). Since the same decomposition has been applied here to matrices representing multivariable systems, it is not too surprising that many of the results pertaining to two-port systems also pertain to multivariable systems. Some of these extensions are outlined below:

Gain: The entry λ_i on the main diagonal of the Λ-matrix is referred to as the *gain* over the frequency interval $(i-1)\Delta f \leq f < i\Delta f$.

Gain-squared bandwidth product:

The gain-squared bandwidth product, $\Phi(G)$, is defined by

$$\Phi(G) = \int_0^{\max\ [m,n]} \Lambda^2(f)\mathrm{d}f = \sum_{i=1}^{\min\ [m,n]} \lambda_i^2 \Delta f. \tag{5.1}$$

Moreover, it should be noted that

$$\Phi(G) = \sum_{i,j,k,l} g_{ij}^2[t_k, t_l]\Delta f$$

where the $g_{ij}[t_k, t_l]$'s are the elements of G.

Norm of G:

The norm of G, $\|G\|$, is defined by

$$\|G\| = \max_{\|\bar{\mathbf{x}}\|=1} \|G\bar{\mathbf{x}}\|.$$

It can be shown that

$$\|G\| = \max_i \lambda_i \tag{5.2}$$

Assuming for the moment that the λ_i's are distinct and that $\lambda_1 > \lambda_2 > \ldots > \lambda_{nN}$ (or λ_{mN}), it follows that for *all* $\bar{\mathbf{x}}$-vectors of a fixed norm, say $\sqrt{N}$, the one causing an output vector with a maximum norm is $\bar{\mathbf{x}}_1$ and the corresponding output vector is $\lambda_1\bar{\mathbf{y}}_1$. Considering all input vectors of norm $\sqrt{N}$ which are orthogonal to $\bar{\mathbf{x}}_1$, the one which causes an output vector with a maximum norm is $\bar{\mathbf{x}}_2$ and the corresponding output vector is $\lambda_2\bar{\mathbf{y}}_2$. This pattern continues through $\bar{\mathbf{x}}_{nN}$ (or $\bar{\mathbf{x}}_{mN}$) and $\lambda_{nN}\bar{\mathbf{y}}_{nN}$ (or $\lambda_{mN}\bar{\mathbf{y}}_{mN}$). This is an important property of the decomposition presented and shows that the $\bar{\mathbf{x}}_i$ and $\bar{\mathbf{y}}_i$-vectors characterize the "extremal" inputs and outputs of the system.

In case the λ_i's are not all distinct, as assumed above, any linear combination of $\bar{\mathbf{x}}_i$-vectors associated with equal λ_i's with a fixed norm (say norm equal $\sqrt{N}$) yields an output vector whose norm is (1) independent of the linear combination used, and (2) the maximum output norm possible over the appropriate subspace of inputs. For example, if $\lambda_1 > \lambda_2 > \ldots > \lambda_j = \lambda_{j+1} > \lambda_{j+2} > \ldots > \lambda_{nN}$ (or λ_{mN}) and the input vectors of norm $\sqrt{N}$ which are orthogonal to $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \ldots, \bar{\mathbf{x}}_{j-1}$ are considered, the $\bar{\mathbf{x}}$'s associated with the maximum output norm are all linear combinations of the form $a\bar{\mathbf{x}}_j + b\bar{\mathbf{x}}_{j+1}$, where $a^2 + b^2 = 1$, and the corresponding outputs are $a\lambda_j\bar{\mathbf{y}}_j + b\lambda_{j+1}\bar{\mathbf{y}}_{+1} = \lambda_j(a\bar{\mathbf{y}}_j + b\bar{\mathbf{y}}_{j+1})$.

It is interesting to note that the foregoing extremal properties of the $\bar{\mathbf{x}}_i$- and $\bar{\mathbf{y}}_i$-vectors show that the decomposition of a G-matrix corresponding to a physically realizable system is equivalent to a sequence of optimalization problems. Assume that the usual state variable description of a physically realizable system is given, that is,

$$\mathbf{z}(t_{l+1}) = A(t_l)z(t_l) + B(t_l)\mathbf{x}(t_l)$$
$$\mathbf{y}(t_l) = C(t_l)\mathbf{z}(t_l) + D(t_l)\mathbf{x}(t_l)$$

where $\mathbf{z}(t_l)$ is the state of the system at the l-th sampling time, $\mathbf{y}(t_l)$ and $\mathbf{x}(t_l)$ are as previously defined, and A, B, C, D are time dependent matrices. The initial state is assumed equal to zero, that is, $z(t_1)=0$. The system matrix, G, is, of course, implicitly determined by this equation. In order to simplify the discussion, assume the λ_i's of G are distinct and $\lambda_1 > \lambda_2 > \ldots$.

The output $\bar{\mathbf{y}}_1$ and the input $\bar{\mathbf{x}}_1$ are determined by the following optimalization problem:

Select that input $\bar{\mathbf{x}}$ subject to the constraint $\|\bar{\mathbf{x}}\| < \sqrt{N}$ which maximizes the norm of the output $\|\bar{\mathbf{y}}\| = \|G\bar{\mathbf{x}}\|$. The selected input is $\bar{\mathbf{x}}_1$ and the corresponding output is $\lambda_1\bar{\mathbf{y}}_1$.

The vectors $\bar{\mathbf{y}}_2$ and $\bar{\mathbf{x}}_2$ are determined by the following problem:

Select that input $\bar{\mathbf{x}}$ subject to the constraints $\|\bar{\mathbf{x}}\| < \sqrt{N}$ and $(\bar{\mathbf{x}}, \bar{\mathbf{x}}_1)=0$ which maximizes the norm of the output $\|\bar{\mathbf{y}}\|$.

This sequence of optimalization problems can be continued until all $\bar{\mathbf{y}}_i$- and $\bar{\mathbf{x}}_i$-vectors are determined. It seems likely that computational algorithms based on these optimalization problems will be of great value.

Bandwidth:

As in the case of two-port systems, a meaningful generalization of the concept of bandwidth is given by

$$Bw = \frac{\Phi(G)}{\|G\|^2}. \tag{5.3}$$

Physical realizability:

Let* the columns of the matrix X^T be designated by $\bar{\mathbf{u}}_j(t_k)$, where

$$\mathbf{u}_j(t_k) = \begin{bmatrix} x^j{}_1(t_k) \\ x^j{}_2(t_k) \\ x^j{}_3(t_k) \\ \cdots \\ x^j{}_{mN}(t_k) \end{bmatrix} \qquad \begin{cases} k=1, 2, \ldots, N \\ j=1, \ldots, m \end{cases} \tag{5.4}$$

The $\bar{\mathbf{u}}_j(t_k)$'s are referred to here as the input ensemble vectors. Let the rows of the matrix $(Y\Delta f)\Lambda$ be designated by $\bar{\mathbf{v}}_j(t_k)$:,

$$\bar{\mathbf{v}}_j(t_k) = \begin{bmatrix} \lambda_1 y^j{}_1(t_k) \\ \lambda_2 y^j{}_2(t_k) \\ \cdots \\ \lambda_{mN} y^j{}_{mN}(t_k) \end{bmatrix} \qquad \begin{cases} k=1, 2, \ldots, N \\ j=1, 2, \ldots, n \end{cases} \tag{5.5}$$

*It is assumed here that $n \geq m$. If $m > n$ the subscripts must be altered accordingly.

The $\bar{\mathbf{v}}_j(t_k)$'s are referred to as the output ensemble vectors. As mentioned before, the matrix G corresponds to a physically realizable system if

$$G[t_k, t_r]=0 \text{ for } r>k. \tag{5.6}$$

This is easily shown to be the case if and only if

$$(\bar{\mathbf{v}}_j(t_k), \mathbf{u}_l(t_r))=0 \text{ for } r>k. \tag{5.7}$$

That is, the output ensemble vector at time t_k must be orthogonal to all input ensemble vectors occuring later in time. Roughly speaking, the output at any time is "independent" of all future inputs.

6. EXAMPLE

As mentioned before, CRUZ [2] has discussed the design of control systems from the system matrix viewpoint. A point in his approach is the selection of the desired system matrix, G_D (CRUZ uses the symbol "F"), and the desired sensitivity matrix, S, (a concept which is not discussed here). In simple cases, once G_D and S are given the appropriate compensation system can be determined in a straightforward manner. However, it is not often clear what the initial selection for G_D (and S) should be. The purpose of the following example is to indicate the manner in which the transform technique presented here yields insight into this latter problem. In order to simplify the example, questions of feedback are left aside and only the filter aspects of the problem are considered.

Suppose that the plant under consideration is a two-input, two-output characterized by the following difference equation:

$$\begin{bmatrix} y^1(k) \\ y^2(k) \end{bmatrix} = \begin{bmatrix} \frac{k-1}{10} & 1-\frac{k-1}{10} \\ 1-\frac{k-1}{10} & \frac{k-1}{10} \end{bmatrix} \begin{bmatrix} y^1(k-1) \\ y^2(k-1) \end{bmatrix} + \begin{bmatrix} x^1(k) \\ x^2(k) \end{bmatrix} \quad (k=1, 2, \ldots, 8)$$

$$\begin{bmatrix} y^1(0) \\ y^2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{6.1}$$

Denote the system matrix implied by this difference equation by G_P. Let the problem be to select a compensation system G_C so that G_PG_C is equal to or some approximation to a desired total system behavior characterized by a system matrix G_D.

The desired system G_D is selected on the basis of (1) wanting to decouple output channel number one (two) from input channel number two (one), and (2) wanting the two uncoupled channels to act as ideal filters (see [3]).

In fact, it is assumed that in the decomposition of G_D

$$\lambda_1 = 1,\ \bar{\mathbf{x}}_1 = \bar{\mathbf{y}}_1 = [1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0]$$

$$\lambda_2 = 1,\ \bar{\mathbf{x}}_2 = \bar{\mathbf{y}}_2 = [0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1]$$

$$\lambda_3 = 1,\ \bar{\mathbf{x}}_3 = \bar{\mathbf{y}}_3 = [1,0,1,0,1,0,1,0,-1,0,-1,0,-1,0,-1,0]$$

$$\lambda_4 = 1,\ \bar{\mathbf{x}}_4 = \bar{\mathbf{y}}_4 = [0,1,0,1,0,1,0,1,0,-1,0,-1,0,-1,0,-1] \tag{6.2}$$

$$\lambda_5 = 1,\ \bar{\mathbf{x}}_5 = \bar{\mathbf{y}}_5 = [1,0,1,0,-1,0,-1,0,-1,0,-1,0,1,0,1,0]$$

$$\lambda_6 = 1,\ \bar{\mathbf{x}}_6 = \bar{\mathbf{y}}_6 = [0,1,0,1,0,-1,0,-1,0,-1,0,-1,0,1,0,1]$$

$$\lambda_7 = \lambda_8 = \ldots = \lambda_{16} = 0$$

Since the remaining $\bar{\mathbf{x}}_i$'s are multiplied by the zero λ_i's, we need not specify them. Given the Y, Λ, X matrices, the implied system matrix (6.3) is shown on the next page.
Since the entries above the dotted line in (6.3) are not all zero, it follows that the resulting G_D does not represent a physically realizable system. Thus, it is not possible to obtain the originally desired system. At this point there are many ways to proceed. In order, again, to keep this example simple, the easiest path is chosen. Replace G_D by the matrix G_{DR} corresponding to the physically realizable system which is the best approximation to G_D in the gain-squared bandwidth product sense, that is, G_{DR} minimizes $\Phi(G_D - G_{DR})$.

Since

$$\Phi(G) = \sum_{i,j,k,l} g_{ij}^2[t_k, t_l]\Delta f,$$

it follows that G_{DR} is simply (6.3) with each of the entries above the dotted line set equal to zero.

At this point it is important to obtain more insight into the nature of the new system G_{DR}. This insight is obtained from the decomposition of G_{DR} which is given in (6.4), (6.5) and (6.6).
The $\bar{\mathbf{x}}_i$-vectors in the X-matrix have been arranged according to the number of sign changes occuring in the vector. Thus, the k-th $\bar{\mathbf{x}}$-vector has $(k-1)$ sign changes. It happens for the system under consideration that this arrangement of the $\bar{\mathbf{x}}_i$'s causes the $\bar{\mathbf{y}}_i$'s to be arranged in the same manner. Roughly speaking, a vector with a few sign changes can be viewed as a "low frequency" vector and one with many sign changes can be viewed as a "high frequency" vector.

We see that approximating G_D with G_{DR} has not destroyed the decoupling between the two channels. However, each channel, as was to be expected [3], is no longer an ideal filter. On the other hand, each still has roughly the desired characteristics. High gain is associated with "low frequencies", that is, $\lambda_1, \lambda_2, \ldots, \lambda_6$ are large, and low gain is associated with "high frequencies", that is, $\lambda_7, \lambda_8, \ldots, \lambda_{16}$ are small. Moreover, the $\bar{\mathbf{x}}_i$-and $\bar{\mathbf{y}}_i$-vectors obtained are roughly comparable with the desired ones.

Finally, the system matrix, G_C, for the compensation is obtained from the equation $G_C = G_P^{-1} G_{DR}$, where the matrix G_P^{-1} is simply a restatement of the difference equation (6.1). The resulting G_C is given in (6.7).

(6.3) $$G_D=(Y\Delta f)\Lambda X^T=\tfrac{1}{8}$$

$$\begin{bmatrix}
3 & 0 & 3 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 \\
0 & 3 & 0 & 3 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 \\
3 & 0 & 3 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 \\
0 & 3 & 0 & 3 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & 3 & 0 & 3 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\
0 & 1 & 0 & 1 & 0 & 3 & 0 & 3 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\
1 & 0 & 1 & 0 & 3 & 0 & -3 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\
0 & 1 & 0 & 1 & 0 & 3 & 0 & 3 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\
-1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 3 & 0 & 3 & 0 & 1 & 0 & 1 & 0 \\
0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 3 & 0 & 3 & 0 & 1 & 0 & 1 \\
-1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 3 & 0 & 3 & 0 & 1 & 0 & 1 & 0 \\
0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 3 & 0 & 3 & 0 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 3 & 0 & 3 & 0 \\
0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 3 & 0 & 3 \\
1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 3 & 0 & 3 & 0 \\
0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 3 & 0 & 3
\end{bmatrix}$$

where $\Delta f=\dfrac{1}{N}=\dfrac{1}{8}$

(6.4) $Y=$

$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$	$\bar{y}_5$	$\bar{y}_6$	$\bar{y}_7$	$\bar{y}_8$	$\bar{y}_9$	$\bar{y}_{10}$	$\bar{y}_{11}$	$\bar{y}_{12}$	$\bar{y}_{13}$	$\bar{y}_{14}$	$\bar{y}_{15}$	$\bar{y}_{16}$
0.33	0.00	0.60	0.00	0.74	0.00	−1.43	0.00	1.54	0.00	1.18	0.00	−0.96	0.00	−0.52	0.00
0.00	0.33	0.00	0.60	0.00	0.74	0.00	−1.43	0.00	1.54	0.00	1.18	0.00	−0.96	0.00	−0.52
0.57	0.00	1.04	0.00	1.32	0.00	−1.78	0.00	−0.43	0.00	−1.59	0.00	0.76	0.00	0.40	0.00
0.00	0.57	0.00	1.04	0.00	1.32	0.00	−1.78	0.00	−0.43	0.00	−1.59	0.00	0.76	0.00	0.40
1.00	0.00	0.94	0.00	0.07	0.00	0.76	0.00	−1.07	0.00	−0.28	0.00	−1.36	0.00	−1.57	0.00
0.00	1.00	0.00	0.94	0.00	0.07	0.00	0.76	0.00	−1.07	0.00	−0.28	0.00	−1.36	0.00	−1.57
1.54	0.00	1.29	0.00	−0.32	0.00	1.02	0.00	0.44	0.00	0.91	0.00	0.82	0.00	1.06	0.00
0.00	1.54	0.00	1.29	0.00	−0.32	0.00	1.02	0.00	0.44	0.00	0.91	0.00	0.82	0.00	1.06
1.06	0.00	−0.69	0.00	−0.85	0.00	−0.24	0.00	1.00	0.00	−0.65	0.00	1.30	0.00	−1.59	0.00
0.00	1.06	0.00	−0.69	0.00	−0.85	0.00	−0.24	0.00	1.00	0.00	−0.65	0.00	1.30	0.00	−1.59
1.57	0.00	−1.17	0.00	−0.82	0.00	−1.06	0.00	−0.62	0.00	−0.14	0.00	−0.96	0.00	1.03	0.00
00.0	1.57	0.00	−1.17	0.00	−0.82	0.00	−1.06	0.00	−0.62	0.00	−0.14	0.00	−0.96	0.00	1.03
0.47	0.00	−0.88	0.00	1.27	0.00	−0.23	0.00	−1.30	0.00	1.52	0.00	1.03	0.00	−0.55	0.00
0.00	0.47	0.00	−0.88	0.00	1.27	0.00	−0.23	0.00	−1.30	0.00	1.52	0.00	1.03	0.00	−0.55
0.63	0.00	−1.18	0.00	1.61	0.00	1.32	0.00	1.02	0.00	−0.65	0.00	−0.57	0.00	0.31	0.00
0.00	0.63	0.00	−1.18	0.00	1.61	0.00	1.32	0.00	1.02	0.00	−0.65	0.00	−0.57	0.00	0.31

$$\Lambda = \begin{bmatrix} \lambda_1 & & & & & & & & & & & & & & & \\ 0.73 & & & & & & & & & & & & & & & \\ & \lambda_2 & & & & & & & & & & & & & & \\ & 0.73 & & & & & & & & & & & & & & \\ & & \lambda_3 & & & & & & & & & & & & & \\ & & 0.74 & & & & & & & & & & & & & \\ & & & \lambda_4 & & & & & & & & & & & & \\ & & & 0.74 & & & & & & & & & & & & \\ & & & & \lambda_5 & & & & & & & & & & & \\ & & & & 0.81 & & & & & & & & & & & \\ & & & & & \lambda_6 & & & & & & & & & & \\ & & & & & 0.81 & & & & & & & & & & \\ & & & & & & \lambda_7 & & & & & & & & & \\ & & & & & & 0.35 & & & & & & & & & \\ & & & & & & & \lambda_8 & & & & & & & & \\ & & & & & & & 0.35 & & & & & & & & \\ & & & & & & & & \lambda_9 & & & & & & & \\ & & & & & & & & 0.25 & & & & & & & \\ & & & & & & & & & \lambda_{10} & & & & & & \\ & & & & & & & & & 0.25 & & & & & & \\ & & & & & & & & & & \lambda_{11} & & & & & \\ & & & & & & & & & & 0.21 & & & & & \\ & & & & & & & & & & & \lambda_{12} & & & & \\ & & & & & & & & & & & 0.21 & & & & \\ & & & & & & & & & & & & \lambda_{13} & & & \\ & & & & & & & & & & & & 0.22 & & & \\ & & & & & & & & & & & & & \lambda_{14} & & \\ & & & & & & & & & & & & & 0.22 & & \\ & & & & & & & & & & & & & & \lambda_{15} & \\ & & & & & & & & & & & & & & 0.23 & \\ & & & & & & & & & & & & & & & \lambda_{16} \\ & & & & & & & & & & & & & & & 0.23 \end{bmatrix} \tag{6.5}$$

(6.6) $X =$

$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_4$	$\bar{x}_5$	$\bar{x}_6$	$\bar{x}_7$	$\bar{x}_8$	$\bar{x}_9$	$\bar{x}_{10}$	$\bar{x}_{11}$	$\bar{x}_{12}$	$\bar{x}_{13}$	$\bar{x}_{14}$	$\bar{x}_{15}$	$\bar{x}_{16}$
0.63	0.00	1.18	0.00	1.61	0.00	−1.32	0.00	1.02	0.00	0.65	0.00	−0.57	0.00	−0.31	0.00
0.00	0.63	0.00	1.18	0.00	1.61	0.00	−1.32	0.00	1.02	0.00	0.65	0.00	−0.57	0.00	−0.31
0.47	0.00	0.88	0.00	1.27	0.00	0.23	0.00	−1.30	0.00	−1.52	0.00	1.03	0.00	0.00	0.55
1.57	0.00	1.17	0.00	−0.82	0.00	1.06	0.00	−0.62	0.00	0.14	0.00	−0.96	0.00	−1.03	0.00
0.00	1.57	0.00	1.17	0.00	−0.82	0.00	1.06	0.00	−0.62	0.00	0.14	0.00	−0.96	0.00	−1.03
1.06	0.00	0.69	0.00	−0.85	0.00	0.24	0.00	1.00	0.00	0.65	0.00	1.30	0.00	1.59	0.00
0.00	1.06	0.00	0.69	0.00	−0.85	0.00	0.24	0.00	1.00	0.00	0.65	0.00	1.30	0.00	1.59
1.54	0.00	−1.29	0.00	−0.32	0.00	−1.02	0.00	0.44	0.00	−0.91	0.00	0.82	0.00	−1.06	0.00
0.00	1.54	0.00	−1.29	0.00	−0.32	0.00	−1.02	0.00	0.44	0.00	−0.91	0.00	0.82	0.00	−1.06
1.00	0.00	−0.94	0.00	0.07	0.00	−0.76	0.00	−1.07	0.00	0.28	0.00	−1.36	0.00	1.57	0.00
0.57	0.00	−1.04	0.00	1.33	0.00	1.18	0.00	−0.43	0.00	1.59	0.00	0.76	0.00	−0.40	0.00
0.00	0.57	0.00	−1.04	0.00	1.33	0.00	1.18	0.00	−0.43	0.00	1.59	0.00	0.76	0.00	−0.40
0.33	0.00	−0.60	0.00	0.74	0.00	1.43	0.00	1.54	0.00	−1.18	0.00	−0.96	0.00	0.52	0.00
0.00	0.33	0.00	−0.60	0.00	0.74	0.00	1.43	0.00	1.54	0.00	−1.18	0.00	−0.96	0.00	0.52

(6.7) $G_C = \frac{1}{8}$

$$\left[\begin{array}{cc|cc|cc|cc|cc|cc|cc|cc}
3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2.7 & -2.7 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-2.7 & 2.7 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.4 & -2.4 & 0.4 & -2.4 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-2.4 & 0.4 & -2.4 & 0.4 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.7 & -0.7 & 0.7 & -0.7 & 2.1 & -2.0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.7 & 0.7 & -0.7 & 0.7 & -2.1 & 2.1 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1.4 & -0.6 & -1.4 & -0.6 & -0.2 & -1.8 & -0.2 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.6 & -1.4 & -0.6 & -1.4 & -1.8 & -0.2 & -1.8 & -0.2 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.5 & 0.5 & -0.5 & 0.5 & 0.5 & -0.5 & 0.5 & -0.5 & 1.5 & -1.5 & 3 & 0 & 0 & 0 & 0 & 0 \\
0.5 & -0.5 & 0.5 & -0.5 & -0.5 & 0.5 & -0.5 & 0.5 & -1.5 & 1.5 & 0 & 3 & 0 & 0 & 0 & 0 \\
1.6 & 0.4 & 1.6 & 0.4 & -1.6 & -0.4 & -1.6 & -0.4 & -0.8 & -1.2 & -0.8 & -1.2 & 3 & 0 & 0 & 0 \\
0.4 & 1.6 & 0.4 & 1.6 & -0.4 & -1.6 & -0.4 & -1.6 & -1.2 & -0.8 & -1.2 & -0.8 & 0 & 3 & 0 & 0 \\
0.3 & -0.3 & 0.3 & -0.3 & -0.3 & 0.3 & -0.3 & 0.3 & 0.3 & -0.3 & 0.3 & -0.3 & 0.9 & -0.9 & 3 & 0 \\
-0.3 & 0.3 & -0.3 & 0.3 & 0.3 & -0.3 & 0.3 & -0.3 & -0.3 & 0.3 & -0.3 & 0.3 & -0.9 & 0.9 & 0 & 3
\end{array}\right]$$

7. CONCLUSIONS

It has been shown that multivariable, time-varying, discrete-time, linear systems can be handled in a manner equivalent to single-input, single-output systems. A transform technique for the latter systems is extended to multivariable systems, and frequency response concepts are shown to carry over in a straightforward way to multivariable systems. The key point of the present development has been a de-emphasis of the channelized character of the input or output and the treatment of an arbitrary input or output as a single vector in a linear vector space. Thus, much of the insight associated with single-input, single-output systems has validity for multivariable systems.

Although a connection has been made between the transform technique introduced here and certain optimalization problems, it is important to note the philosophical difference between the "optimalization" approach to time-varying linear systems and the approach presented here. At the risk of gross oversimplification, it can be said that the "optimalization" approach is based on ignoring all but optimum system behavior. Whereas, the approach presented here is an attempt to obtain insight into all aspects of system behavior, suboptimum as well as optimum.

Acknowledgments—The author wishes to thank the following colleagues for their helpful comments: Prof. F. N. BAILEY, Mr. J. A. COLLIGAN, Mr. R. FISCHL, Prof. K. IRANI, Prof. E. L. MCMAHON, Prof. A. J. PENNINGTON, Prof. G. R. SELL, Mr. F. M. WALTZ.

APPENDIX

It is worthwhile to make a few remarks regarding the details of this decompositon.

(a) Given a system matrix G the decomposition indicated in (3.1) can be carried out in the following manner. Since

$$G^T=(X\Delta f)\Lambda Y^T$$

and

$$Y^TY=I\frac{1}{\Delta f}$$

even for rectangular Y's, it follows that

$$G^TG=(X\Delta f)\Lambda^2X^T$$

Thus, the columns of the matrix X are eigenvectors of G^TG and the main diagonal entries in Λ are positive square roots of eigenvalues of G^TG. Given X and Λ, the Y-matrix is determined from the relation $Y\Lambda=GX$.

(b) If the matrix G has at least as many rows as columns, the square matrix G^TG is the same size as the square matrix Λ^2; therefore, all the eigenvalues and eigenvectors of G^TG enter into the formation of Λ and X.

(c) If G has more columns than rows, the square matrix G^TG is larger than the square matrix Λ^2. However, the ranks of G^TG and Λ^2 are the same. Thus, all the nonzero eigenvalues and associated eigenvectors of G^TG enter into the formulation of Λ and X. If Λ is not filled by the nonzero eigenvalues, the remaining entries on the main diagonal of Λ are equal to zero.

Note that there are two ways for a G with more columns than rows to have a non-trivial null space: (1) it automatically has one of dimension equal to the number of columns minus the number of rows, and (2) it has one of larger dimension if the rows of G are not linearly independent.

(d) The matrix Λ is uniquely determined. The matrices X and Y are not. If the entries in Λ are distinct this lack of uniqueness is trivial and arises from the fact that (-1) times an eigenvector is also an eigenvector. Thus, the matrix X is determined except for a sign ambiguity on each of its columns. If the entries in Λ are not distinct, the lack of uniqueness is slightly more complicated. The geometric multiplicity of certain eigenvalues of G^TG is greater than one, which means, of course, that each subspace spanned by the eigenvectors associated with each of these eigenvalues has dimension greater than one. Therefore, *any* orthogonal basis for one of these subspace can be used for columns in X. An extreme example is given by the decomposition of the identity matrix I; in that case any orthogonal matrix can be used for the matrix X.

Assuming that Λ and X are chosen, Y is uniquely determined if Λ has no zero main diagonal entry. Otherwise the columns of Y which are orthogonal to the range of G are not uniquely determined but this is very trivial.

Résumé—Il est montrè qu'une technique de transformation, developpée pour des systèmes linéaires, à temps discret, variant dans le temps et à entrèe et sortie uniques, peut etre étendue d'une manière significative, à des systèmes à variables multiples.

Zusammenfassung—Gezeigt wird, wie ein für einschleifige lineare zeitvariante Abtastsysteme entwickeltes Transformationsverfahren auf entsprechende vermaschte Systeme ausgedehnt werden kann.

Абстракт—Показано, что техника преобразованний, разработанная для линейных систем с дискретным временем, переменных по времени и обладающих лишь одной входной и одной выходной координатой, может быть распространена надлежащим образом на многокоординатные системы.