OPTIMUM SYSTEMS

IN

MULTI-DIMENSIONAL RANDOM PROCESSES

by

MICHAEL CHASE DAVIS

Lieutenant, U.S. Navy

B.S., U.S. Naval Academy
(1953)

SUBMITTED IN PARTIAL FULFILLMENT

OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 1961

Signature of Author - - - - - - - - - - - - -
Department of Electrical Engineering, May 13, 1961

Certified by - - - - - - - - - - - - - - - - -
Thesis Supervisor

Accepted by - - - - - - - - - - - - - - - - - -
Chairman, Departmental Committee on Graduate Students

OPTIMUM SYSTEMS IN MULTI-DIMENSIONAL RANDOM PROCESSES

by

MICHAEL CHASE DAVIS

Lieutenant, U.S. Navy

Submitted to the Department of Electrical Engineering on May 13, 1961, in partial fulfillment of the requirements for the degree of Doctor of Science.

ABSTRACT

This thesis deals with random processes which are stationary, ergodic, and described by correlation functions or power density spectra. An attempt has been made to develop a new approach to the study and control of random processes which is simple, stresses physical rather than mathematical interpretation, and is valid when a number of statistically related processes are to be processed simultaneously. Among the origianl and fundamental results of this investigation are:

(1) A closed-form solution is presented for the optimum multi-dimensional system in the Wiener sense. This system operates on n correlated random input signals and produces m desired outputs, each of which has minimum mean-square error. The solution is dependent upon the factorization of a matrix $\Phi(s)$ of the cross-power density spectra of the input signals into two matrices, such that $\Phi(s) = G(-s) \cdot G^{T}(s)$. The nxn physical system G(s) determined from this procedure must be realizable and inverse realizable, and is the system which would reproduce the measured statistics when excited by n uncorrelated white noise sources.

(2) A general solution is derived for the above matrix factorization problem, valid without regard to order, providing the spectra satisfy a realizability requirement. The method employs a series of simple matrix transformations which manipulate the original matrix into desired forms. The key to this solution is a general procedure for reducing a matrix with polynomial elements to impotent form, having a constant determinant. This latter step is also an original contribution to the theory of matrices with algebraic elements. With this solution to the matrix factorization problem, essentially no conceptual difference remains between single and multi-dimensional random processes.

(3) The optimum single or multi-dimensional prediction operation is

shown to result from a continuous measurement of the current state variables of the hypothetical model G(s) which can create the random process from white noise excitation. These state variables are then weighted according to their decay as initial conditions in the desired prediction time and the decayed output or outputs are the desired prediction. Thus, expected behavior of the random process over all future time is compactly summarized in the current values of these state variables.

(4) It is proved that correlation functions measured between two variables in a linear system can be viewed as an initial condition response of this system. Also, the well-known Wiener-Hopf equation is shown merely to require that every error be uncorrelated with past values of every input signal.

(5) If one or more noisy signals have a power density spectra matrix $\Phi(s)$, which can be factored into $G(-s)\, G^{T}(s)$, and if G(s) is separated such that $G(s) = S(s) + N(s)$, where S(s) and N(s) have signal and noise poles, respectively, then it is shown that the optimum filter is a unity feedback system with forward transference $S(s)\, N^{-1}(s)$. This very general result is valid for single or multi-dimensional optimum filtering problems.

(6) A quantitative substitute for the Nyquist sampling theorem is presented which is concerned with a measure of the irrecoverable error inherent in representing a continuous random process by its samples. Also, the new results in continuous random process theory derived herein are extended to the discrete case.

(7) The concept of "state" of a random process is advanced as fundamental information for control use. Two new design principles are discussed for the bang-bang control of a linear system subject to a random input. In one, suitable for multi-dimensional full throw control, the determinate Second Method of Lyapunov is extended to include random processes.

The basic contributions of this thesis are (1) a complete theory of multi-dimensional random processes, (2) a simple physical explanation for the optimum linear filter and predictor using white-noise generating models, and (3) a new approach to stochastic control problems, especially those involving saturation, using the concept of the "state" of a random process.

Thesis Supervisor: Ronald A. Howard

Title: Assistant Professor of Electrical Engineering

ACKNOWLEDGMENT

The author is very grateful to Professor Ronald A. Howard for his support and encouragement throughout this research, and to LCDR John R. Baylis, USN and Professor Amar G. Bose for their valuable assistance.

To Captain Edward S. Arentzen USN and Professor Murray F. Gardner, thanks are due for their active encouragement during the entire doctoral program. The author is indebted to the Bureau of Ships, U. S. Navy, for providing the financial support which made this investigation possible.

The author is appreciative of the competency of Mrs. Jutta Budek, who performed the final typing of this manuscript.

Finally, the author wishes to thank the "unsung heroine", his wife Beverly, for her gracious acceptance of the trying demands of the thesis research and presentation.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT.)

CHAPTER I.

INTRODUCTION

The word "random" is an adjective which mankind has come to use in apology for unwillingness or inability to measure fundamental causes for events observed in Nature. Of these events, the random process which goes on continuously and indefinitely has captured the interest of mathematicians and engineers. There is something compelling about attempting to describe that which is ever changing, and thus undescribable.

This thesis is concerned with random processes in their simplest form -- with statistics that do not change with time, and whose properties are adequately described by the well-known correlation functions. Many able researchers have cleared this path and it could well be asked, like an echo from the Second World War, "Is this trip necessary?"

To begin with, a research investigation is generally based on aggravation, either with what is not known or with what is known. In this work, the latter case is true. It is the opinion of the author that the classic and beautiful core theory of Wiener in this area, by its very mathematical eloquence, has tended to suppress a more fundamental understanding of what can be known in a random process and what cannot.

In essence, the original work of this thesis starts with the well-known fact that the random processes considered here act <u>as if</u> they came from a linear system which is excited by the most random of signals, "white" noise. This linear system specifies the particular random process, and focussing attention on its determinate structure is a more satisfying approach, at least to the engineer, than is accepting the manipulation of statistical properties of the ever-changing output of this system.

Some of the unsolved problems and prominent possibilities in

random process theory which come to mind for possible attack are:

(1) Conventionally, derivations in the Wiener theory are made for optimum systems in the time domain. A pure transform approach appears much more desirable.

(2) A general closed-form solution for the optimum multi-dimensional system has not yet been given in the literature.

(3) A means has not yet been found for determining a physical system capable of reproducing signals with the given statistics of multi-dimensional random processes.

(4) The fundamental results of Wiener theory are the optimum predictor and filter. It may be possible that these have a very simple interpretation in terms of the equivalent white-noise driven system.

(5) The correlation functions of many observed random processes have the appearance of an initial condition response of a linear system. If this is true, what linear system and what initial conditions?

(6) What effect would white noise have if suddenly applied to an otherwise quiescent linear system?

(7) There is no valid measure of the inherent error due to sampling of a random process to replace the "Go-No Go" nature of the Nyquist Sampling Theorem.

(8) If a linear theory produces all the knowable information about an input random process, is there some way of intelligently using this to control a physical system which has limitations such as saturation? No suitable approach to the on-off or bang-bang control problem with random excitation has been made which makes complete use of this information.

(9) If a random process is to be examined by means of investigation of an effective physical system, can some determinate approaches to systems analysis such as the "Second Method of Lyapunov" be extended to include random processes?

This thesis provides a quantitative answer to each of these questions or possibilities. The author believes that the results found in this

thesis investigation, because of their simplicity and generality, provide the most effective means for understanding the nature of stationary random processes.

CHAPTER II.

DERIVATION OF OPTIMUM SINGLE AND MULTIDIMENSIONAL SYSTEMS

2.1 Introduction

This chapter is concerned with linear systems which operate on stationary random processes so as to minimize a quadratic measure of error between the desired and actual outputs. In the case of a single random signal, perhaps corrupted by noise, the results of this theory have been known for over a decade. Why, then, is it necessary to retrace such well-worn steps?

There are two reasons for this apparent duplication. First of all, the author feels that the time-domain derivations found in many standard texts of the optimum Wiener filter are unnecessarily complicated and tend to obscure the basic simplicity of the ideas expressed. Secondly and more important, when the optimum system to process two or more signals simultaneously is derived, the conventional methods rapidly become enmeshed in their own symbology, whereas the steps of the single-signal frequency domain approach to be described in this chapter allow direct extension to the multi-dimensional case.

2.2 Historical perspective

In this country, the origin of the statistical theory of optimum linear systems was the wartime work of Wiener[1]. A parallel development in Russia at approximately the same time was made by Kolmogorov[2]. The structure of the basic theory was thus well-formed by 1950 for problems involving prediction and filtering of a single stationary random process in the presence of additive noise. Significant extensions and clarification of Wiener's work were made by Zadeh and Ragazzini[3], Bode and Shannon[4], Blum[5], Lee[6], Pike[7], and Newton[8]. The latter's work was of particular significance, since it introduced the concept of optimization with constraints in order to satisfy certain practical engineering

requirements of a system which the basic theory neglected. In the last decade, graduate-level control systems engineering texts have generally emphasized the statistical approach. These include books by Truxal[9], Newton[10], Smith[11], Seifert and Steeg[12], and Lanning and Battin[13].

In the multi-dimensional case, the theory is not as well-developed. Westcott[14] derived an optimum configuration for the two-dimensional case. Amara[15] used a partial matrix approach and successfully derived the optimum unrealizable configuration, but his realizable solution was only applicable upon very restricted signal conditions. Hsieh and Leondes[16] presented a method for solving for the optimum system involving undetermined coefficients, but the meaning of their solution was obscured by the formidable notation employed and no proof of the adequacy of their method was offered.

2.3 Summary of linear statistical theory

Figure 2.1 shows a typical time record of a random process involving two variables, x and y. The signals to be considered under this theory are stationary; that is, they have statistical properties which do not change with time. Also, these statistical properties can be approximated by measurements made on a single long but finite time-recording of the particular continuous signal -- that is, the processes satisfy the ergodic hypothesis.

Figure 2.1 Typical random processes

The objective of statistical analysis of a random process is to detect cause-effect relationships between events -- or signal levels -- separated in time. The basic tools in this analysis are the auto-correlation and the cross-correlation functions. The auto-correlation function, $\varphi_{xx}(\tau)$, is defined as the average value of the product of the instantane-

ous signal and the signal level τ seconds later.

$$\varphi_{xx}(\tau) \triangleq E\left\{x(t) \cdot x(t+\tau)\right\} \tag{2.1}$$

where the symbol $\triangleq$ is a defining equality and the operator $E\{\cdot\}$ means "the expected value of". Expressed in integral form for the class of signals considered,

$$\varphi_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} dt \quad x(t) \cdot x(t+\tau) \tag{2.2}$$

Figure 2.2 shows a typical auto-correlation function. Note that it is even about the $\tau = 0$ axis, $\varphi_{xx}(\tau) = \varphi_{xx}(-\tau)$, since replacing t by $t - \tau$ in Equation 2.2 does not affect its value. The maximum value of $\varphi_{xx}(\tau)$ is at $\tau = 0$ for any stationary signal observed in the real world (a proof is given by Truxal[9].)

The cross-correlation function, $\varphi_{xy}(\tau)$, is defined as the average value of the product of the instantaneous signal level of one variable, x, and that of another signal, y, τ seconds later.

$$\varphi_{xy}(\tau) \triangleq E\left\{x(t) \cdot y(t+\tau)\right\} \tag{2.3}$$

$$\varphi_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} dt \quad x(t) \cdot y(t+\tau) \tag{2.4}$$

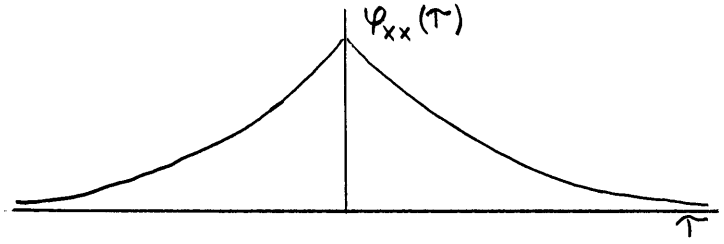


Figure 2.2 A typical auto-correlation function

In this case, replacing t by $t - \tau$ in the integral form yields the definition of $\varphi_{yx}(-\tau)$, and the peak value of $\varphi_{xy}(\tau)$ does not necessarily occur at the origin. Summarizing,

$$\varphi_{xx}(-\tau) = \varphi_{xx}(\tau) \tag{2.5}$$

$$\varphi_{xy}(-\tau) = \varphi_{yx}(\tau) \tag{2.6}$$

The auto-correlation functions and all possible cross-correlation functions among members of a set of random signals completely describe the particular process for the purposes of a linear theory.

One significant use of the auto-correlation function is that $\varphi_{xx}(0)$ is, by definition, the mean square value of x. For example, this makes it a useful measure of the accuracy of a system when the signal concerned is the error.

Since the correlation functions (for $\tau > 0$) have the same appearance as transient signals observed in linear systems it is logical to define the Laplace transforms of these functions and inquire as to their potential use. As the functions are defined for both positive and negative τ, the bilateral or "two-sided" Laplace transform is selected for use. The bilateral Laplace transform evaluates the positive-time part of a signal just as the one-sided Laplace transform does, but the negative-time portion has the sign of t changed (i e: "flipped over" the t = 0 axis), evaluated as a positive-time signal, and the sign of s, the transform variable, is changed to -s.

In order to ensure a one-to-one correspondence between the transform and the time-domain expression, it is necessary to specify that all poles in the right half plane (or "negative" poles) correspond to functions in negative time and not unstable functions in positive time.

In this work, the bilateral Laplace transform of the auto and cross-correlation function is defined as the auto or cross power density spectrum, $\Phi_{xx}(s)$ or $\Phi_{xy}(s)$, respectively. The notion of power density arises in the following fashion:

The mean square value of a random signal x is envisioned as a generalized form of average energy because of its quadratic nature, and is equal by definition to $\varphi_{xx}(0)$. If $\varphi_{xx}(0)$ is finite, it is equal to the

sum of the residues of either the left-half or right-half plane poles of the transform $\overline{\Phi}_{xx}(s)$, as seen directly from a partial fraction expansion of $\overline{\Phi}_{xx}(s)$ and term-by-term inversion. But by the residue theorem of complex variable theory, the evaluation of a closed contour up the imaginary axis of the s-plane and enclosing the left-half-plane at infinity will yield $2\pi j$ x summation of residues, providing the contour is of the order of no less than $\frac{1}{s^2}$ as $s \longrightarrow \infty$. That is, $\overline{\Phi}_{xx}(s)$ must contain at least two more poles than zeros for a finite mean square value, $\varphi_{xx}(0)$, to exist.

Thus,

$$\varphi_{xx}(0) = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} ds \quad \overline{\Phi}_{xx}(s) \tag{2.7}$$

Let $s = j\omega$

$$\varphi_{xx}(0) = \frac{1}{2\pi} \int_{\infty}^{\infty} d\omega \quad \overline{\Phi}_{xx}(\omega) \tag{2.8}$$

The mean square value (or power) of a signal is thus seen to be proportional to the integral of $\overline{\Phi}_{xx}(\omega)$ over all ω, and $\overline{\Phi}_{xx}(\omega)$ quite naturally is visualized as a power density per unit ω. Most authors have included the $\frac{1}{2\pi}$ in the definition of the power density spectrum so that the integral over all ω yields the total average power, but this appears to be less natural than retaining the pure transform relationship, especially since the name "power" is a misnomer in itself. The ω notation is the most common encountered in past literature on random processes, and brings to mind a weighting of harmonic content, considering the random process to be a superposition of an infinite number of infinitely small simusoidal waves.

It might be argued that the choice of nomenclature is a trivial matter, but in as much as it influences basic conceptualization of a random process, it is very important and deserves elaboration.

Ten years ago in automatic control literature, the transfer function

of a linear system was invariably written as $G(j\omega)$, and much was made of plots of frequency response on polar or logarithmic coordinates. Frequency response was almost regarded as an end in itself, and design specification in terms of these characteristics helped propagate this belief. However, the acceptance of Evan's root-locus method[17] and the strong emphasis by Truxal[9] and others towards use of the Laplace transform helped unify the differential equation, frequency response, and transient response approaches to dynamic behavior of linear systems. In proper perspective, frequency response is an often desirable experimental description of a system and provides, on logarithmic coordinates, a rapid means for design of simple control systems when specifications on transient behavior are loose. Frequency response is perhaps best visualized as an imaginary axis scan on the complex plane, as shown in Figure 2.3, where the function is evaluated by the complex product of vectors from all system zeroes divided by vectors from all system poles to the particular $s = j\omega$ point under consideration.

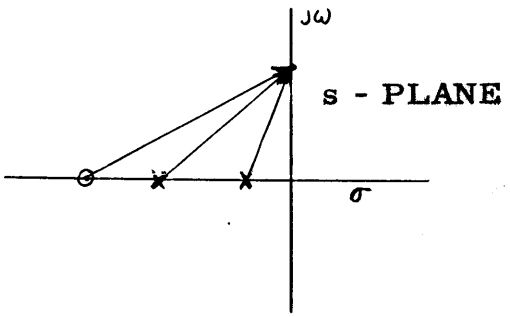


Figure 2.3 Frequency response viewed as imaginary axis scan

Since linear systems were previously regarded in terms of how they altered the magnitude and phase of an input sinusoidal signal, essentially a communications engineering viewpoint, it is natural that random processes should have been described in terms of relative frequency content. But now that the Laplace transform -- high-lighting the system poles and zeroes -- has emerged as perhaps the best index to the properties of a linear system, it is necessary to take the viewpoint in a random process that the characteristics of interest are the poles and zeroes of the power density spectrum $\Phi_{xx}(s)$, and not necessarily the spectrum

shape. A useful conception of the spectral representation, as a function of ω, is shown in Figure 2.4, where again the magnitude of the spectrum is determined as the resultant of vectors from all poles and zeroes of $\Phi_{xx}(s)$ to the $s = j\omega$ point. Note that an auto power density spectrum

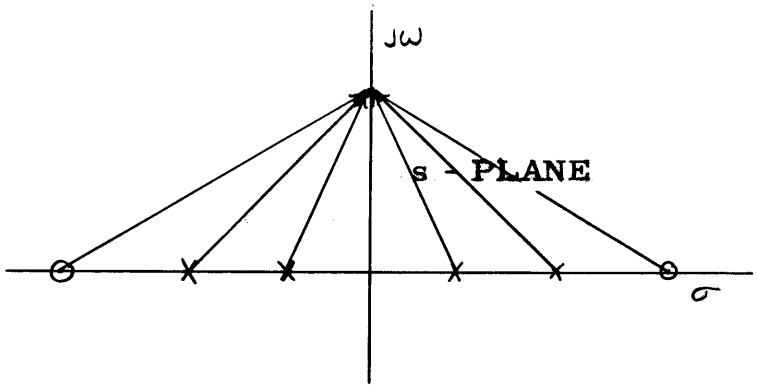


Figure 2.4 Power density spectrum viewed as imaginary axis scan

has a symmetrical distribution of poles and zeroes, since the relationship $\varphi_{xx}(\tau) = \varphi_{xx}(-\tau)$ becomes $\Phi_{xx}(s) = \Phi_{xx}(-s)$ in the frequency domain which means that, term for term, the LHP poles and zeroes must equal the RHP poles and zeroes.

The basic tools for the examination of random process have been presented -- the correlation functions and their transform mates, the power density spectra. It now remains to specify how these characteristics are altered by passage through a linear system.

2.4 A general formula for power density spectra transformations

A derivation is made in this section of a compact expression of the cross (or auto) power density spectra between any two signals in a linear system as a function of the cross power density spectra of the system inputs. This resulting formula will be used consistently in this and remaining chapters because of its generality and simplicity.

The general problem to be considered is pictured in Figure 2.5. x and y are two variables (x may equal y) which are the linear responses to two sets of random inputs, x_i and y_j, each individual input being operated on by a system weighting function, $g_i(t)$ or $h_j(t)$. The desired quan-

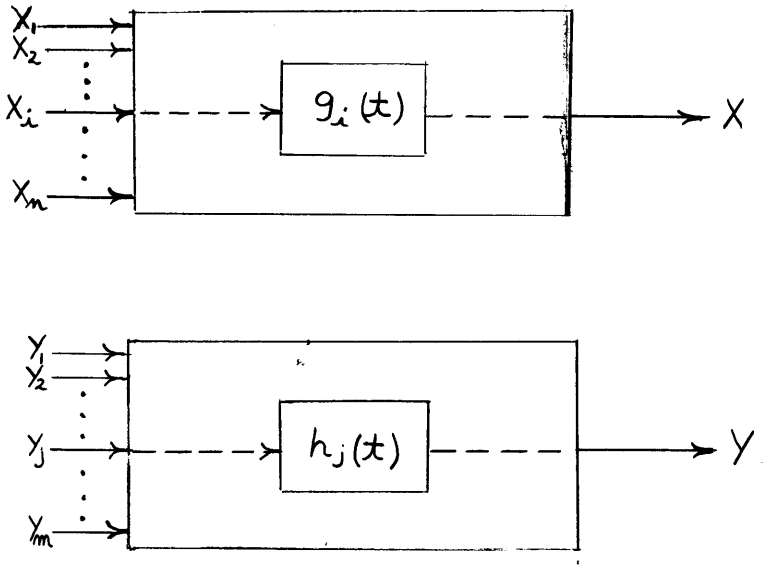


Figure 2.5 General model for linear system

tity is $\Phi_{xy}(s)$; the known quantities are the cross-power density spectra between any two of the inputs x_i and y_j.

$$x(t) = \sum_{i=1}^{n} g_i(t) * x_i(t) \quad ; \quad y(t) = \sum_{j=1}^{m} h_j(t) * y_j(t)$$

where "$*$" is a symbolic operator expressing convolution. Transforming,

$$X(s) = \sum_{i=1}^{n} G_i(s)\, X_i(s) \quad ; \quad Y(s) = \sum_{j=1}^{m} H_j(s) \cdot Y_j(s)$$

$$\varphi_{xy}(\tau) \triangleq E_t \left\{ x(t) \cdot y(t+\tau) \right\}$$

$$\Phi_{xy}(s) = \mathcal{L}_{\tau} \left\{ \varphi_{xy}(\tau) \right\} = E_t \left\{ x(t)\, \mathcal{L}_{\tau} \left[y(t+\tau) \right] \right\}$$

assuming that the integration involved with averaging in time can commute with the integration of the Laplace transform. The subscripts on the operator indicate the time variable which is used in the operation.

Consider a length of signal which exists for duration 2T, where T is arbitrarily large but finite, and which is zero elsewhere.

$$\Phi_{xy}(s) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} dt\, x(t) \int_{-T}^{T} y(t+\tau)\, e^{-s\tau}\, d\tau$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} dt\, x(t)\, e^{st}\, y(s)$$

$$= \lim_{T\to\infty} \frac{1}{2T} \; X(-s) \cdot Y(s)$$

which is a standard result found, for example, in Rice[34] and Solodovnikov[35].

But, substituting the values of X(-s) and Y(s),

$$\Phi_{xy}(s) = \lim_{T\to\infty} \frac{1}{2T} \sum_{i=1}^{m} \sum_{j=1}^{m} G_i(-s)\, H_j(s)\, Y_j(s)$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} G_i(-s)\, H_j(s) \lim_{T\to\infty} \frac{X_i(-s)\, Y_j(s)}{2T}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} G_i(-s)\, H_j(s)\, \Phi_{x_i y_j}(s) \qquad (2.9)$$

which is the desired result. Several examples will illustrate the convenience of this formula.

Consider first the system of Figure 2.6.

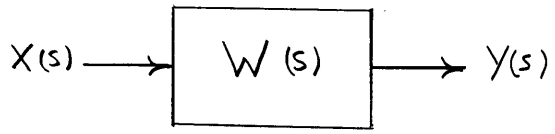


Figure 2.6 A simple linear system

$$X(s) = X(s); \qquad Y(s) = X(s) \cdot W(s)$$

$$\Phi_{xy}(s) = W(s)\, \Phi_{xx}(s) \qquad (2.10)$$

a basic result which has immediate practical consequences. If x and y are the available input and output signals of an otherwise inaccessible system, the system transfer function can be determined without introducing test disturbances by analyzing the cross-correlation between x and y.

Also,

$$\Phi_{yy}(s) = \Phi_{xx}(s)\, W(s)\, W(-s) \tag{2.11}$$

Next, Figure 2.7 shows a typical summing operation.

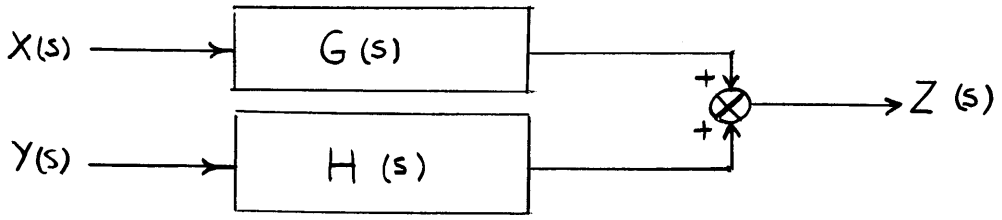


Figure 2.7 A typical summing operation

$$Z(s) = X(s) \cdot G(s) + Y(s)\, H(s)$$

$$\Phi_{zz}(s) = \Phi_{xx}(s)\, G(-s)\, G(s) + \Phi_{xy}(s)\, G(-s)\, H(s) + \Phi_{yx}(s)\, H(-s)\, G(s) + \Phi_{yy}(s)\, H(-s)\, H(s) \tag{2.12}$$

which is obtained by inspection by performing the necessary cross-multiplication and observing the proper sign of s.

2.5 Single-dimensional optimum systems

The classical Wiener theory of an optimum linear system to operate on a random process will now be derived using transform expressions wherever possible. This clear and direct approach is useful in its own right but is basically intended to provide an introduction to a similar development for multi-dimensional systems to follow.

Figure 2.8 shows the basic configuration to be studied. The stationary input random signal v in general will contain a signal to be operated on and an extraneous noise component. The ideal signal i is

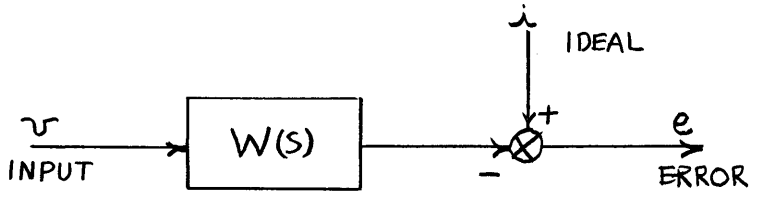


Figure 2.8 Configuration of an optimum system

the mathematical result of some desired operation on the signal component of the input, such as filtering, prediction, or some linear function of the signal. Figure 2.9 shows an elaboration of this structure, where

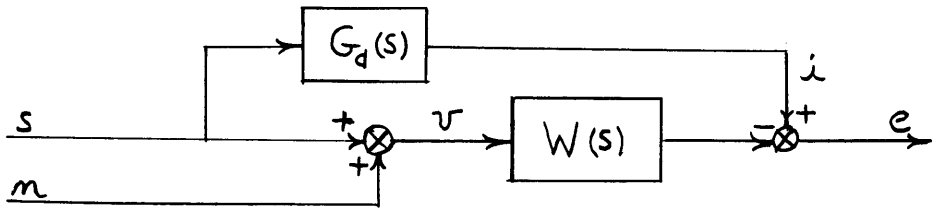


Figure 2.9 Formation of the ideal signal

the signal component s is operated on by some not-necessarily physically realizable transfer function, $G_d(s)$, such as 1, e^{st}, or s. If Φ_{ss} and Φ_{ns} are known, and since

$$\Phi_{vi}(s) = \Phi_{ss}\, G_d(s) + \Phi_{ns}\, G_d(s) \qquad (2.13)$$

$\Phi_{vi}(s)$ is as equally valid a statistical description of the desired operation as is specification of $G_d(s)$.

The error signal, e, is the difference between the actual response of the system to be determined, W(s), and the ideal signal. The optimum system will minimize the mean value of error squared, $\overline{e^2} = \varphi_{ee}(0)$, which is a satisfactory error criteria for many purposes. The use of the variance of the first probability distribution of error is a natural choice when longtime properties of signals are being examined, as a more complex error criterion besides being mathematically intractable would require more statistical knowledge of the processes involved.[13]

$$\overline{e^2} = \varphi_{ee}(0) = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} ds \; \Phi_{ee}(s)$$

$$E(s) = I(s) - V(s)\, W(s)$$

$$\Phi_{ee}(s) = \Phi_{ii}(s) - \Phi_{iv}(s)\, W(s) - \Phi_{vi}(s)\, W(-s) + \Phi_{vv}(s)\, W(s)\, W(-s)$$

The determination of the optimum W(s) in order to minimize the integral expression is the standard problem of the calculus of variations. If a perturbation in W(s) is made, called a variation, a resulting perturbation or variation in $\overline{e^2}$ results. More formally, W(s) is replaced by $W(s) + \epsilon\, \delta W(s)$, where ϵ is a "small" constant and the variation $\delta W(s)$ is any allowable change in W(s), or alternately any system which could be paralleled with W(s). This restricts $\delta W(s)$ to have the properties of physically-realizable and stable systems, that is, with no poles in the right-half plane. Also, for a finite $\overline{e^2}$, $\delta W(s)$ must not be of such order as to provide a component of white noise at e when excited by v. $\overline{e^2}$ is then expanded as a power series in ϵ around $\epsilon = 0$. The optimum system will have been found when the coefficient of the first power of ϵ is zero regardless of the form of $\delta W(s)$ -- in other words, no small allowable change in W(s) tends to decrease the value of the integral.

$$\delta \overline{e^2} \triangleq \frac{\partial}{\partial \epsilon} \left[\overline{e^2} \right]_{\epsilon = 0}$$

$$\delta \overline{e^2} = \frac{1}{2\pi j} \int_{-j}^{j} ds \; \delta \left[\Phi_{ee}(s) \right]$$

assuming that differentiation may be performed under the integral sign.

The variational notation will now be shown to follow the usual rules for differentiation, considering the individual terms of $\Phi_{ee}(s)$ consecutively.

$$\delta \left[\Phi_{ii}(s) \right] = 0$$

$$\delta\left[\Phi_{iv}(s)\,W(s)\right] = \Phi_{iv}(s)\,\frac{\partial}{\partial\epsilon}\left[W(s)+\epsilon\,\delta W(s)\right]_{\epsilon=0} = \Phi_{iv}(s)\,\delta W(s)$$

$$\delta\left[\Phi_{vi}(s)\,W(-s)\right] = \Phi_{vi}(s)\,\frac{\partial}{\partial\epsilon}\left[W(-s)+\epsilon\,\delta W(-s)\right]_{\epsilon=0} = \Phi_{vi}(s)\,\delta W(-s)$$

$$\delta\left[\Phi_{vv}(s)\,W(s)\,W(-s)\right] = \Phi_{vv}(s)\,\frac{\partial}{\partial\epsilon}\left\{\left[W(s)+\epsilon\,\delta W(s)\right]\left[W(-s)+\epsilon\,\delta W(-s)\right]\right\}_{\epsilon=0}$$

$$= \Phi_{vv}(s)\left[\delta W(s)\,W(-s) + W(s)\,\delta W(-s)\right]$$

The only restriction on this analogy with differentiation is that the variation of W(s) or W(-s) must carry the proper sign of s.

$$\delta\,\overline{e^2} = 0 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\left\{-\Phi_{iv}(s)\,\delta W(s) - \Phi_{vi}(s)\,\delta W(-s) + \Phi_{vv}(s)\left[W(-s)\,\delta W(s) + W(s)\,\delta W(-s)\right]\right\}$$

To simplify this expression, several auxiliary results are needed. (1) $\Phi_{iv}(-s) = \Phi_{vi}(s)$, from the fact that $\varphi_{iv}(-\tau) = \varphi_{vi}(\tau)$, Equation 2.6. (2) The sign of s may be changed in any single term of the above integral, without affecting its value, since the limit exchange and the sign change of the differential ds have cancelling effects.

Changing the sign of terms as necessary to be able to factor $\delta W(-s)$ and identifying $\Phi_{iv}(-s)$ as $\Phi_{vi}(s)$ and $\Phi_{vv}(s)$ as $\Phi_{vv}(-s)$.

$$\delta\,\overline{e^2} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\;\delta W(-s)\left[-\Phi_{vi}(s) + \Phi_{vv}(s)\,W(s)\right] = 0$$

If the integral exists and the contour is selected so as to enclose the left-half plane, the LHP residues must sum to zero for arbitrary $\delta W(-s)$, which has all its poles in the right-half plane. Obviously, the function $\left[\Phi_{vv}(s)\,W(s) - \Phi_{vi}(s)\right]$ must have no simple poles in the LHP, say at $s = -a_i$, for the sum of residues is $\sum \delta W(a_i)$, an arbitrary number for arbitrary $\delta W(-s)$. If this function has a multiple pole, say $\frac{1}{(s+a)^m}$, then $\delta W(-s)$ could be selected so as to include a $m-1^{th}$ order zero at $s = -a$ (only poles must be in the RHP), leaving the first

order case. Thus, it has been shown that $\left[\Phi_{vv}^{(s)} W(s) - \Phi_{vi}(s)\right]$ can have no poles in the left half plane, or

$$\mathcal{LF}^{-1}\left[\Phi_{vv}^{(s)} W(s)\right] = \mathcal{LF}^{-1}\left[\Phi_{vi}(s)\right] \tag{2.14}$$

where $\mathcal{LF}^{-1}$ is a picturesque operator used by Smith[11] to indicate the operation of inverting a transform into its positive and negative time parts ($\mathcal{F}^{-1}$ or the inverse Fourier transform) and using only the positive time part in taking the unilateral Laplace transform, $\mathcal{L}$. Despite a possible question as to the uni- or bi-lateral nature of $\mathcal{L}$, this compact notation will be used subsequently to denote the casting out of RHP poles.

A functional equality of LHP poles, such as in equation 2.14 above, is not affected by multiplication of both sides by the same arbitrary transform having poles in the RHP. For example, $\mathcal{L}W(-s)$ is such a function. Now, $\Phi_{vv}(s)$, because of its even nature, can always be factored into $\Phi^{+}_{vv}(-s)\,\Phi^{+}_{vv}(s)$, where $\Phi^{+}_{vv}(-s)$ contains only RHP poles and zeroes and $\Phi^{+}_{vv}(s)$ contains only LHP poles and zeroes.

$$\mathcal{LF}^{-1}\left\{\frac{1}{\Phi^{+}_{vv}(-s)}\,\Phi_{vv}(s)\,W(s)\right\} = \mathcal{LF}^{-1}\left\{\frac{1}{\Phi^{+}_{vv}(-s)}\,\Phi_{vi}(s)\right\}$$

$$\Phi^{+}_{vv}(s)\,W(s) = \mathcal{LF}^{-1}\left\{\frac{\Phi_{vi}(s)}{\Phi^{+}_{vv}(-s)}\right\}$$

$$W(s) = \frac{1}{\Phi^{+}_{vv}(s)}\,\mathcal{LF}^{-1}\left\{\frac{\Phi_{vi}(s)}{\Phi^{+}_{vv}(-s)}\right\} \tag{2.15}$$

This is the desired solution for an optimum system under a mean-square error criterion.

To review the derivation procedure,

(1) $\Phi_{ee}(s)$ was found using Equation 2.9. (2) $\mathcal{L}\Phi_{ee}(s)$ was expressed in the compact variational notation. (3) $\mathcal{L}\,\overline{e^2}$ was placed in the following form: $\mathcal{L}\,\overline{e^2} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\; 2\,\mathcal{L}W(-s)\left[\Phi_{vv}^{(s)} W(s) - \Phi_{vi}(s)\right] \overset{!}{=} 0$ (4) The left-half plane poles of $\Phi_{vv}^{(s)} W(s)$ were shown equal to those of $\Phi_{vi}(s)$,

and both sides of this equality were multiplied by the inverse factor of $\Phi^{+}_{vv}(-s)$.

An example of this procedure is given next to illustrate the ease in derivation of extensions to the basic theory. This modification is due to Newton[8, 10] and is an attempt to control saturation in a given power transducer. Figure 2.10 shows that a signal driving a fixed element $G_p(s)$ is

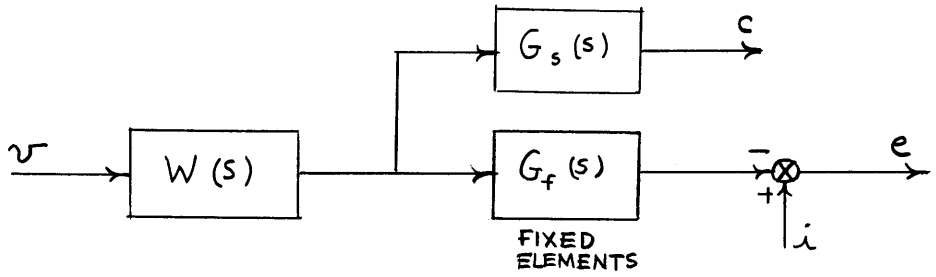


Figure 2.10 Control of saturation in fixed elements

to have some linear function ($G_s(s)$ usually equals 1, s, or s^2) of itself reproduced as the hypothetical signal c, which will have its mean-square value constrained by a Lagrange multiplier as the error is minimized in order to control the probability of saturation.

$$\delta\left\{\overline{e^2} + \lambda \overline{c^2}\right\} = 0 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds \left\{\delta \Phi_{ee}(s) + \lambda\, \delta \Phi_{cc}(s)\right\}$$

$$E(s) = I(s) - V(s) \cdot W(s) \cdot G_f(s)$$

$$C(s) = V(s)\, W(s)\, G_s(s)$$

$$\Phi_{ee}(s) = \Phi_{ii}(s) - \Phi_{vi}(s)\, W(-s)\, G_f(-s) + \Phi_{vv}(s)\, W(s)\, W(-s)\, G_f(s)\, G_f(-s) - \Phi_{iv}(s)\, W(s)\, G_f(s)$$

$$\Phi_{cc}(s) = \Phi_{vv}(s)\, W(s)\, W(-s)\, G_s(s)\, G_s(-s)$$

$$\delta\, \Phi_{ee}(s) = -\Phi_{iv}(s)\, G_f(s)\, \delta W(s) - \Phi_{vi}(s)\, G_f(-s)\, \delta W(-s) + \Phi_{vv}(s)\, G_f(s)\, G_f(-s)\left[W(s)\, \delta W(-s) + W(-s)\, \delta W(s)\right]$$

$$\delta\, \Phi_{cc}(s) = \Phi_{vv}(s)\, G_s(s)\, G_s(-s)\left[W(s)\, \delta W(-s) + W(-s)\, \delta W(s)\right]$$

$$\delta\left\{\overline{e^2} + \lambda \overline{c^2}\right\} = 0 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\, \delta W(-s)\left\{-2\, \Phi_{vi}(s)\, G_f(-s) + 2\, \Phi_{vv}(s)\left[G_f(s)\, G_f(-s) + \lambda\, G_s(s)\, G_s(-s)\right] W(s)\right\}$$

$$\mathcal{L}\mathcal{L}^{-1}\left\{\Phi_{vv}(s)\left[G_f(s)\,G_f(-s)+\lambda\,G_s(s)\,G_s(-s)\right]\,W(s)\right\}=\mathcal{L}\mathcal{L}^{-1}\left\{G_f(-s)\;\Phi_{vi}(s)\right\}$$

$$W(s)=\frac{1}{\Phi^{+}_{vv}(s)\left[G_f(s)\,G_f(-s)+\lambda\,G_s(s)\,G_s(-s)\right]^{+}}\mathcal{L}\mathcal{L}^{-1}\left\{\frac{G_f(-s)\;\Phi_{vi}(s)}{\Phi^{+}_{vv}(-s)\left[G_f(s)G_f(-s)+\lambda G_s(s)\,G_s(-s)\right]^{-}}\right\}$$

where the + and - symbols indicate LHP and RHP factors.

This same result, obtained through standard time-domain techniques, requires a formidable use of tedious multiple integrals plus the complex reasoning behind the time-domain motivation of spectral factoring.

It is interesting to note that factoring the input power density spectrum, $\Phi_{vv}(s) = \Phi^{+}_{vv}(s) \cdot \Phi^{+}_{vv}(-s)$, determines the system which could produce the observed statistics when excited by "white noise" with a unity power density spectra, as was pointed out by Bode and Shannon[4].

In Figure 2.11, a white noise signal, with $\Phi_{\omega\omega}(s) = 1$, passes through a linear system with a transfer function of $\Phi^{+}_{vv}(s)$. $\Phi_{vv}(s) = \Phi^{+}_{vv}(s)\,\Phi^{+}_{vv}(-s)$ from equation 2.11.

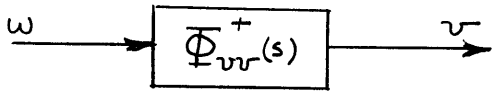


Figure 2.11 Reproduction of observed statistics from white noise.

White noise is a useful abstraction, since it is a totally random signal having uniform energy content at all frequencies, or alternately, an impulse auto-correlation function. It will be one of the major purposes of this thesis to stress the visualization of a random process, single or multi-dimensional, in terms of the linear mathematical model which could create the process. This has the effect of partitioning the process into two parts: (1) The white noise excitation, which is totally random and thus unknowable, and (2) The hypothetical physical system, which is completely known and which has instantaneous internal signal levels which completely define the entire past history of the white noise excitation for future use.

2.6 Multi-dimensional optimum systems

The class of system considered in this section is pictured in Figure 2.12.

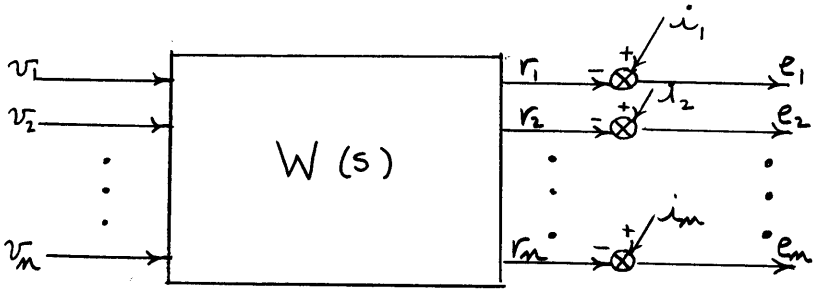


Figure 2.12 A multi-dimensional system

Here a set of n input signals, v, each of which may contain a signal and noise component which can be correlated with any other signal or noise is to be processed by a linear multi-dimensional system **W(s)**. The n outputs, r, are to be compared with ideal or desired signals, i, and the set of differences constitute the error signals. As will be shown at the end of this chapter, the ideal signals result from a linear operation on the signal components of the input signals, and specification of $\Phi_{i_j v_k}(s)$ for j and k = 1, 2, . . . n is enough to uniquely specify this relationship, as was shown to be true in the one-dimensional case.

The criterion for optimum performance is that the mean-square value of every error signal is to be minimized simultaneously.

W(s) is best described in matrix notation:

$$\underline{r(s)} = \underline{\mathbf{W}(s)}\ \underline{v(s)} \qquad (2.17)$$

where $W_{ij}(s)$ is the transmission linking the i^{th} output and the j^{th} input.

Consider the i^{th} error signal.

$$E_i(s) = I_i(s) - \sum_{j=1}^{n} W_{ij}(s)\, V_j(s)$$

$$\overline{e_i^2} = \varphi_{e_i e_i}(0) = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\ \Phi_{e_i e_i}(s)$$

$$\delta\,\overline{e_i^2} = 0 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\ \delta\,\Phi_{e_i e_i}(s)$$

From equation 2.9,

$$\Phi_{e_i\, e_i}(s) = \Phi_{i_i\, i_i}(s) - \sum_{j=1}^{m} \Phi_{i_i\, v_j}(s)\, W_{ij}(s) - \sum_{j=1}^{m} \Phi_{v_j\, i_i}(s)\, W_{ij}(-s) + \sum_{j=1}^{m} W_{ij}(-s) \sum_{k=1}^{m} W_{ik}(s)\, \Phi_{v_j\, v_k}(s)$$

In matrix notation, let $\underline{W_i(s)}$ be the i^{th} row of $\underline{W(s)}$. The scalar $\Phi_{e_i\, e_i}(s)$ is then seen to be expressed by

$$\Phi_{e_i\, e_i}(s) = \Phi_{i_i\, i_i}(s) - \left[\underline{W_i(s)}\, \Phi_{i_i\, v_j}(s)\right] - \left[\underline{W_i(-s)}\, \Phi_{v_j\, i_i}(s)\right] + \left[\underline{W_i(-s)}\, \left[\Phi_{vv}(s)\right]\, W_i(s)\right]$$

Let $\underline{W_i(s)}$ be replaced by $\underline{W_i(s)} + \epsilon\, \underline{\delta W_i(s)}$, where ϵ is a scalar and the variation $\underline{\delta W_i(s)}$ is an arbitrary row vector, each element of which satisfies the same physical realizability condition as in the one-dimensional case.

$\delta\, \Phi_{e_i\, e_i}(s)$ will be evaluated term by term to show that the matrix variation is found by an analog to matrix differentials.

$$\delta\, \Phi_{i_i\, i_i}(s) = 0$$

$$\delta\left\{\left[\underline{W_i(s)}\, \Phi_{i_i\, v_j}(s)\right]\right\} = \frac{\partial}{\partial \epsilon}\left\{\left\{\underline{W_i(s)} + \epsilon\, \underline{\delta W_i(s)}\right\} \cdot \Phi_{i_i\, v_j}(s)\right\}_{\epsilon = 0}$$

$$= \left[\underline{\delta W_i(s)}\, \Phi_{i_i\, v_j}(s)\right]$$

$$\delta\left\{\left[\underline{W_i(-s)}\, \Phi_{v_j\, i_i}(s)\right]\right\} = \frac{\partial}{\partial \epsilon}\left\{\left\{\underline{W_i(-s)} + \epsilon\, \underline{\delta W_i(-s)}\right\} \cdot \Phi_{v_j\, i_i}(s)\right\}_{\epsilon = 0}$$

$$= \left[\underline{\delta W_i(-s)}\, \Phi_{v_j\, i_i}(s)\right]$$

$$\delta\left\{\left[\underline{W_i(-s)}\, \left[\Phi_{vv}(s)\right]\, W_i(s)\right]\right\} =$$

$$\frac{\partial}{\partial \epsilon}\left\{\left\{\underline{W_i(-s)} + \epsilon\, \underline{\delta W_i(-s)}\right\} \cdot \left[\Phi_{vv}(s)\right] \cdot \left\{W_i(s)\right] + \epsilon\, \underline{\delta W_i(s)}\right\}_{\epsilon = 0}$$

$$= \underline{W_i(-s)}\left[\underline{\Phi}_{vv}(s)\right]\underline{\mathcal{L}W_i(s)} + \underline{\mathcal{L}W_i(-s)}\left[\underline{\Phi}_{vv}(s)\right]W_i(s)$$

$$\mathcal{L}\overline{e_i^2} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\left\{-\underline{\mathcal{L}W_i(s)}\,\underline{\Phi}\,i_i\,v_j(s)\right] - \underline{\mathcal{L}W_i(-s)}\,\underline{\Phi}\,v_j\,i_i(s)\right] + \underline{W_i(-s)}\left[\underline{\Phi}_{vv}(s)\right]\mathcal{L}W_i(s)\right] + \underline{\mathcal{L}W_i(-s)}\left[\underline{\Phi}_{vv}(s)\right]W_i(s)\right]\Bigg\}$$

Each term under the integral sign is a scalar and can be transposed at will, and the sign of s changed as was described in the single-dimensional case. Also, $\underline{\Phi}^{T}_{vv}(-s) = \underline{\Phi}_{vv}(s)$ since $\underline{\Phi}\,v_j\,v_i(-s) = \underline{\Phi}\,v_i\,v_j(s)$, Equation 2.6.

$$\mathcal{L}\overline{e_i^2} = 0 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\;\underline{\mathcal{L}W_i(-s)}\left\{-\underline{\Phi}\,i_i\,v_j(-s)\right] - \underline{\Phi}\,v_j\,i_i(s)\right] + \left[\Phi^{T}_{vv}(-s)\right]W_i(s)\right] + \left[\Phi_{vv}(s)\right]W_i(s)\right]\Bigg\}$$

$$= \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\;2\,\underline{\mathcal{L}W_i(-s)}\left\{-\underline{\Phi}\,v_j\,i_i(s)\right] + \left[\Phi_{vv}(s)\right]W_i(s)\right]\Bigg\}$$

This scalar integral expression is identical with the sum of n one-dimensional cases and the same reasoning, element by element, can be applied to the column vector as was applied to the single dimensional case. That is, there can be no net LHP poles in any element of $\left[\Phi_{vv}(s)\right]W_i(s)\right] - \underline{\Phi}\,v_j\,i_i(s)\right]$ since they are separately multiplied by arbitrary functions having RHP poles only.

Therefore,

$$\mathcal{L}\mathcal{L}^{-1}\left\{\left[\Phi_{vv}(s)\right]W_i(s)\right]\right\} = \mathcal{L}\mathcal{L}^{-1}\left\{\underline{\Phi}\,v_j\,i_i(s)\right]\right\} \quad (\mu = 1, 2 \ldots m)$$

where the $\mathcal{L}\mathcal{L}^{-1}$ operator is understood to apply to each element in the column vector.

An expression involving the matrix $\left[W(s)\right]$ is thus found:

$$\mathcal{L}\mathcal{L}^{-1}\left\{\left[\Phi_{vv}(s)\right]\left[W^{T}(s)\right]\right\} = \mathcal{L}\mathcal{L}^{-1}\left\{\left[\underline{\Phi}\,v_j\,i_i(s)\right]\right\}$$

The remainder of this work will need to express compactly the cross-power density spectra which exist between signals in two sets or

vectors in a random process. The following convention will be observed. $\underline{\Phi}_{xy}(s)$ will have an i_j^{th} element $\Phi_{x_i y_j}(s)$. Thus

$$\mathcal{L}\mathcal{F}^{-1}\left\{\underline{\Phi}_{vv}(s)\ \underline{W}^{T}(s)\right\} = \mathcal{L}\mathcal{F}^{-1}\left\{\underline{\Phi}_{vi}(s)\right\} \qquad (2.18)$$

This is the implicit solution to the optimum multi-dimensional system under mean-square error criteria. In the special case of a single-dimensional system, the result is identical to that derived previously in Equation 2.15.

Unfortunately, $\underline{W}(s)$ is not directly obtainable from this expression since $\underline{\Phi}_{vv}(s)$ contains poles in both the LHP and the RHP. This defining equation implicitly involving $\underline{W}(s)$ has been previously obtained independently by Amara[15] and Hsieh and Leondes[16], and the next section will outline and analyze their proposed methods of solution for this set of intercoupled equations.

2.7 Past attempts to determine optimum multi-dimensional system

Hsieh and Leondes[16] employed time-domain concepts in derivation of the optimum system. Their solution will be expressed in the matrix notation of the previous section. The basic problem is to determine W(s) from the equation

$$\mathcal{L}\mathcal{F}^{-1}\left\{\underline{\Phi}_{vv}(s)\ \underline{W}(s)\right\} = \mathcal{L}\mathcal{F}^{-1}\left\{\underline{\Phi}_{vi}(s)\right\} \qquad (2.18)$$

Hsieh and Leondes added an undetermined matrix $\underline{F}(s)$ to the above equation so as to provide an equality of both LHP and RHP poles.

$$\underline{\Phi}_{vv}(s) \cdot \underline{W}^{T}(s) = \underline{\Phi}_{vi}(s) + \underline{F}(s) \qquad (2.19)$$

$\underline{F}(s)$ contains the RHP poles of $\underline{\Phi}_{vv}(s)\ \underline{W}^{T}(s) - \underline{\Phi}_{vi}(s)$

Thus the $\mathcal{L}\mathcal{F}^{-1}$ operator is no longer applicable, it being understood that $\underline{W}^{T}(s)$ will have no poles in the RHP.

$$\underline{W}^{T}(s) = \underline{\Phi}_{vv}^{-1}(s)\left\{\underline{\Phi}_{vi}(s)_{+} - \underline{F}(s)\right\}$$

$$\underline{W}^{T}(s) = \frac{1}{\Delta^{+}\ \Delta^{-}} \left[\text{Adj } \underline{\Phi}_{vv}(s)\right] \left\{\underline{\Phi}_{vi}(s) + \underline{F}(s)\right\} \qquad (2.20)$$

where Δ^{+} and Δ^{-} are the LHP and RHP factors of $\left|\underline{\Phi}_{vv}(s)\right|$ respectively and $\left[\text{adj } \underline{\Phi}_{vv}(s)\right]$ is the adjoint matrix of $\underline{\Phi}_{vv}(s)$.

$$\Delta^{+}\ \underline{W}^{T}(s) = \frac{1}{\Delta^{-}} \left[\text{Adj } \underline{\Phi}_{vv}(s)\right] \underline{\Phi}_{vi}(s) + \frac{1}{\Delta^{-}} \left[\text{Adj } \underline{\Phi}_{vv}(s)\right] \underline{F}(s)$$

$$\text{Let } \frac{1}{\Delta^{-}} \left[\text{Adj } \underline{\Phi}_{vv}(s)\right] \underline{\Phi}_{vi}(s) = \underline{H}^{+}(s) + \underline{H}^{-}(s) \qquad (2.21)$$

where $\underline{H}^{+}(s)$ is known and contains only LHP poles, obtained by performing a partial fraction expansion of each element of $\frac{1}{\Delta^{-}} \left[\text{Adj } \underline{\Phi}_{vv}(s)\right] \underline{\Phi}_{vi}$ $\underline{H}^{-}(s)$ contains only RHP poles.

Each element of $\left[\text{Adj } \underline{\Phi}_{vv}(s)\right]$ can contain only as its LHP poles the LHP poles of $\underline{\Phi}_{vv}(s)$. $\frac{1}{\Delta^{-}}\ \underline{F}(s)$ will contain only RHP poles. Thus,

$$\frac{1}{\Delta^{-}} \left[\text{Adj } \underline{\Phi}_{vv}(s)\right] \underline{F}(s) = \sum_{i=1}^{m} \frac{1}{s + P_i}\ \underline{C}^{i} + \underline{J}^{-}(s) \qquad (2.22)$$

where $-P_i$ is the i^{th} LHP pole location of $\underline{\Phi}_{vv}(s)$, having an undetermined matrix coefficient $\underline{C}^{i}$, and $\underline{J}^{-}(s)$ is a matrix with only RHP poles which is not considered further. Accordingly,

$$\underline{W}^{T}(s) = \frac{1}{\Delta^{+}} \left\{\underline{H}^{+}(s) + \sum_{i=1}^{m} \frac{1}{s + P_i}\ \underline{C}^{i}\right\} \qquad (2.23)$$

At this point, it is claimed by Hsieh and Leondes that the undetermined matrix coefficients can be obtained by substituting $\underline{W}^{T}(s)$ into the basic equation, 2.18.

$$\mathcal{L}\mathcal{L}^{-1} \left\{\underline{\Phi}_{vv}(s) \cdot \frac{1}{\Delta^{+}} \left[\underline{H}^{+}(s) + \sum_{i=1}^{m} \frac{1}{s + P_i}\ \underline{C}^{i}\right]\right\} = \mathcal{L}\mathcal{L}^{-1} \left\{\underline{\Phi}_{vi}(s)\right\} \qquad (2.24)$$

No proof is offered as to the sufficiency of the resulting equations.

The non-generality of this method will now be demonstrated by considering a particular example, a multi-dimensional predictor, and

showing that the resulting equations are insufficient to determine the $\underline{C}^i$ coefficient matrices.

A multi-dimensional predictor example

The input signals, v_i, have no noise superimposed, and are represented by the known matrix $\underline{\Phi}_{vv}(s)$. The i^{th} ideal signal is a prediction of the i^{th} input signal τ seconds in the future.

$$V_i(s) = V_i(s) \; ; \qquad I_j(s) = e^{s\tau} V_j(s)$$

$$\Phi_{vi\ ij}(s) = e^{s\tau}\, \Phi_{vi\ vj}(s)$$

$$\underline{\Phi}_{vi}(s) = e^{s\tau}\, \underline{\Phi}_{vv}(s)$$

From equation 2.21,

$$\underline{H}^+(s) = \mathcal{LZ}^{-1}\left\{\left[\frac{1}{\Delta^-}\ \text{Adj}\ \underline{\Phi}_{vv}(s)\right]\ \underline{\Phi}_{vi}(s)\right\}$$

$$= \mathcal{LZ}^{-1}\left\{\Delta^+\ \underline{\Phi}_{vv}^{-1}(s)\ .\ e^{s\tau}\ \underline{\Phi}_{vv}(s)\right\}$$

$$= \mathcal{LZ}^{-1}\left\{\Delta^+\ e^{s\tau}\underline{I}\right\}$$

From equation 2.23,

$$\underline{W}^\tau(s) = \frac{1}{\Delta^+}\left\{\mathcal{LZ}^{-1}(\Delta^+\ e^{s\tau}\underline{I}) + \sum_{i=1}^{m} \frac{1}{s+P_i}\ \underline{C}^i\right\}$$

where $\underline{C}^i$ is determined from the equality of Equation 2.24.

$$\mathcal{LZ}^{-1}\left\{\underline{\Phi}_{vv}(s)\cdot\frac{1}{\Delta^+}\left\{\mathcal{LZ}^{-1}(\Delta^+ e^{s\tau}\underline{I}) + \sum_{i=1}^{m}\frac{1}{s+P_i}\underline{C}_i\right\}\right\} = \mathcal{LZ}^{-1}\left\{e^{s\tau}\underline{\Phi}_{vv}(s)\right\}$$

Next it will be shown that a partial fraction expansion of $\underline{\Phi}_{vv}(s)\ \dfrac{\mathcal{LZ}^{-1}(\Delta^+ e^{s\tau})}{\Delta^+}$ in the poles of $\underline{\Phi}_{vv}(s)$ is equal identically to the expansion of $\mathcal{LZ}^{-1}\left\{e^{s\tau}\ \underline{\Phi}_{vv}(s)\right\}$. This will be done by proving that

$$\left[\frac{\mathcal{LZ}^{-1}(\Delta^+(s)\, e^{s\tau})}{\Delta^+(s)}\right]_{s=-P_j} = e^{-P_j\tau}$$

which ensures that the external factors outside both the $\underline{\Phi}_{vv}(s)$ matrices are identical for each pole.

It is assumed for simplicity, and since this example is designed

to be essentially a counter-example, that only simple poles exist in $\Delta^+(s)$.

$$\Delta^+(s) \triangleq \frac{P(s)}{Q(s)} = \frac{P(s)}{\prod_{i=1}^{m} (s+P_i)}$$

$$\mathcal{L}^{-1}\left(\Delta^+(s)e^{s\tau}\right) = \mathcal{L}^{-1}\left\{\frac{P(s)\,e^{s\tau}}{\prod_{i=1}^{m}(s+P_i)}\right\} = \sum_{i=1}^{m} \frac{P(-P_i)\,e^{-P_i\tau}}{(s+P_i)\prod_{k\neq i}(P_k-P_i)}$$

$$\left.\frac{\mathcal{L}^{-1}\left(\Delta^+(s)\,e^{s\tau}\right)}{\Delta^+(s)}\right|_{s=-P_j} = \frac{\prod_{k=1}^{m}(s+P_k)}{P(s)}\left(\sum_{i=1}^{m}\frac{P(-P_i)\,e^{-P_i\tau}}{(s+P_i)\prod_{k\neq i}(P_k-P_i)}\right)\Bigg|_{s=-P_j}$$

$$= \sum_{i=1}^{m} \frac{P(-P_i)}{\prod_{k\neq i}(P_k-P_i)} \frac{e^{-P_i\tau}}{P(s)} \prod_{k\neq i}(s+P_k)\Bigg|_{s=-P_j}$$

Each term of this summation equals zero when $i \neq j$

$$\left.\frac{\mathcal{L}^{-1}\left(\Delta^+(s)\,e^{s\tau}\right)}{\Delta^+(s)}\right|_{s=-P_j} = \frac{P(-P_j)}{\prod_{k\neq i}(P_k-P_j)} \frac{e^{-P_j\tau}}{P(-P_j)} \prod_{k\neq i}(P_k-P_j) = e^{-P_j\tau}$$

Therefore, equations involving the LHP poles of $\underline{\Phi}_{vv}(s)$, which can determine $\underline{C}^i$, are

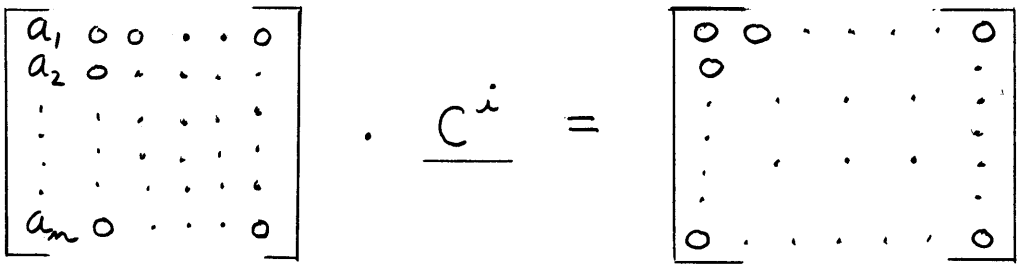
$$\mathcal{L}^{-1}\left\{\underline{\Phi}_{vv}(s)\cdot\frac{1}{\Delta^+(s)}\left[\sum_{i=1}^{m}\frac{1}{s+P_i}\,\underline{C}^i\right]\right\} = \sum_{i=1}^{m}\underline{0}$$

$$\frac{1}{s+P_i}\cdot\frac{1}{\Delta^+(-P_i)}\,\underline{\Phi}_{vv}(-P_i)\,\underline{C}^i = \underline{0}$$

$$\frac{1}{\Delta^+(-P_i)}\,\underline{\Phi}_{vv}(-P_i)\,\underline{C}^i = \underline{0} \qquad (i = 1, 2, \ldots m)$$

where $\underline{0}$ is the null matrix.

The matrix $\frac{1}{\Delta^+(P_i)}\,\underline{\Phi}_{vv}(-P_i)$ will have non-zero values only in elements where the scalar $\Phi_{v_i v_j}(s)$ has a pole at $s = -P_i$. Since simple poles were assumed in $\underline{\Phi}_{vv}(s)$, and since a determinant is formed with each separate term containing only one element from each column and row, it is clear that the i^{th} LHP pole will in general lie along one column or row of $\underline{\Phi}_{vv}(s)$ (actually a column, as will become clear in Chapter 3). Thus, an equation of the form



is patently not enough to determine $\underline{C}^i$.

Through the medium of an example involving a multi-dimensional predictor, it has been demonstrated by counter-example that the procedure of Hsieh and Leondes is not generally applicable. A method of undetermined coefficients is only valid when it can be proven that the coefficients can in fact be determined.

Amara[15] approached the same problem, but attempted to find a closed-form solution and was successful for a quite restricted class of multi-dimensional random processes. Unfortunately, in his derivation of the optimum system he chose to minimize instead of the mean square error of each output the mean square value of the total sum of all the errors, which could allow undesirable cancellation effects between the individual errors and in general is not the best quadratic error criterion. It is interesting to note that his implicit solution is identical with that obtained by considering each error separately, as in section 2.6.

Amara considered the class of random processes characterized by a matrix of power density spectra, $\underline{\Phi}_{vv}(s)$, which can be transformed to a diagonal form by pre- and post-multiplication by matrices with numerical elements, such that

$$\underline{U} \cdot \underline{\Phi}_{vv}(s) \cdot \underline{U}^{T} = \left[D_{ij}(s)\, \delta_{ij} \right]$$

where δ_{ij} is the Kronecker delta, ($\delta_{ij} = 0,\ i \neq j;\ \delta_{ij} = 1,\ i = j$)

$$\underline{\Phi}_{vv}(s) = \underline{U}^{-1} \left[D_{ij}(s)\, \delta_{ij} \right] \left[\underline{U}^{T} \right]^{-1}$$

$$D_{ij}(s) = D_{ij}^{+}(-s)\ D_{ij}^{+}(s)$$

Thus, the optimum system is given by

$$\mathcal{L}^{-1}\left\{\underline{U}^{-1}\left[D_{ij}^{+}(-s)\,\delta_{ij}\right]\left[D_{ij}^{+}(s)\,\delta_{ij}\right]\left[\underline{U^{T}}\right]^{-1}\ \underline{W^{T}(s)}\right\}$$

$$=\mathcal{L}^{-1}\left\{\underline{\Phi}_{vi}(s)\right\}$$

$$\mathcal{L}^{-1}\left\{\left[D_{ij}^{+}(-s)\,\delta_{ij}\right]\left[D_{ij}^{+}(s)\,\delta_{ij}\right]\left[\underline{U^{T}}\right]^{-1}\ \underline{W^{T}(s)}\right\}=\mathcal{L}^{-1}\left\{U\ \underline{\Phi}_{vi}(s)\right\}$$

If $\left[\underline{U^{T}}\right]^{-1}\ \underline{W}(s)$ is considered as another optimum system, the above equation in similar to n one-dimensional optimum systems, and

$$\mathcal{L}^{-1}\left\{\left[D_{ij}^{+}(s)\,\delta_{ij}\right]\left[\underline{U^{T}}\right]^{-1}\ \underline{W^{T}(s)}\right\}=\mathcal{L}^{-1}\left\{\left[\frac{1}{Dij(-s)}\,\delta_{ij}\right]\underline{U}\cdot\underline{\Phi}_{vi}(s)\right\}$$

$$\underline{W^{T}(s)}=\underline{U^{T}}\left[\frac{1}{D_{ij}(s)}\,\delta_{ij}\right]\mathcal{L}^{-1}\left\{\left[\frac{1}{D_{ij}^{+}(-s)}\,\delta_{ij}\right]\ \underline{U}\ \underline{\Phi}_{vi}(s)\right\}$$

The requirement that the power density spectra matrix be diagonalized by a numerical matrix is a severe limitation on random processes in general, as will become more clear in Chapter 3.

In summary, there is no hitherto published satisfactory solution for the optimum n-dimensional system. The next section will consider a more general approach to this problem, which will yield physical insight into random processes and bypass the restrictions of the previously described methods.

2.8 A new closed-form solution for an optimum multi-dimensional system

In the solution of the single-dimensional optimum system, where from Equation 2.14,

$$\mathcal{L}^{-1}\left\{\Phi_{vv}(s)\cdot W(s)\right\}=\mathcal{L}^{-1}\left\{\Phi_{vi}(s)\right\}$$

$\Phi_{vv}(s)$ was factored into RHP and LHP terms

$$\Phi_{vv}(s) = \Phi_{vv}^{+}(-s)\ \Phi_{vv}^{+}(s) \qquad (2.25)$$

and both sides of the equation were multiplied by $\frac{1}{\Phi_{vv}(-s)}$, maintaining

the $\mathcal{LT}^{-1}$ equality.

If the matrix $\underline{\Phi}_{vv}(s)$ could be factored into two matrices,

$$\underline{\Phi}_{vv}(s) = \underline{\Phi}^{-}_{vv}(s) \cdot \underline{\Phi}^{+}_{vv}(s)$$

where $\underline{\Phi}^{-1}_{vv}(s)$ and its inverse contains only RHP poles, it is logical to inquire whether multiplying both sides of the $\mathcal{LT}^{-1}$ matrix equality $\underline{\Phi}^{-}_{vv}(s)^{-1}$ would preserve this identity.

More generally, if the matrix equation is given

$$\mathcal{LT}^{-1}\left\{\underline{A(s)}\right\} = \mathcal{LT}^{-1}\left\{\underline{B(s)}\right\}$$

does

$$\mathcal{LT}^{-1}\left\{\underline{C(s)}\ \underline{A(s)}\right\} = \mathcal{LT}^{-1}\left\{\underline{C(s)}\ \underline{B(s)}\right\}$$

where $\mathbf{C}(s)$ has only RHP poles in every element? The ij^{th} elements of $\underline{C(s)}\ \underline{A(s)}$ and $\underline{C(s)}\ \underline{B(s)}$ are, respectively,

$$\sum_{k=1}^{m} C_{ik} A_{kj} \quad \text{and} \quad \sum_{k=1}^{m} C_{ik} B_{kj}$$

From the previous arguments of this chapter,

$$\mathcal{LT}^{-1}\left\{C_{ik} A_{kj}\right\} = \mathcal{LT}^{-1}\left\{C_{ik} B_{kj}\right\}$$

since

$$\mathcal{LT}^{-1}\left\{A_{kj}\right\} = \mathcal{LT}^{-1}\left\{B_{kj}\right\}$$

Obviously, the addition of n equalities of LHP poles is still a valid equality.

Thus it has been demonstrated that multiplying a matrix $\mathcal{LT}^{-1}$ equality by a matrix with all poles in the RHP preserves the $\mathcal{LT}^{-1}$ equality.

$$\mathcal{LT}^{-1}\left\{\left[\underline{\Phi}^{-}_{vv}(s)\right]^{-1}\ \underline{\Phi}^{-}_{vv}(s)\ \underline{\Phi}^{+}_{vv}(s)\ \underline{W}^{T}(s)\right\} = \mathcal{LT}^{-1}\left\{\left[\underline{\Phi}^{-}_{vv}(s)\right]^{-1}\underline{\Phi}_{vi}(s)\right\}$$

$$\mathcal{LT}^{-1}\left\{\underline{\Phi}^{+}_{vv}(s)\ \underline{W}^{T}(s)\right\} = \underline{\Phi}^{+}_{vv}(s)\ \underline{W}^{T}(s) = \mathcal{LT}^{-1}\left\{\left[\underline{\Phi}^{-}_{vv}(s)\right]^{-1}\ \underline{\Phi}_{vi}(s)\right\}$$

$$\underline{W}^{T}(s) = \left[\underline{\Phi}^{+}_{vv}(s)\right]^{-1} \mathcal{LT}^{-1}\left\{\left[\underline{\Phi}^{-}_{vv}(s)\right]^{-1}\ \underline{\Phi}_{vi}(s)\right\} \qquad (2.26)$$

In the above steps, $\underline{\Phi}^{-}_{vv}(s)^{-1}$ must contain only RHP poles, to justify the operation under the $\mathcal{LT}^{-1}$ operator, and $\underline{\Phi}^{+}_{vv}(s)$ must contain

only LHP poles, to justify the removal of $\mathcal{L}^{-1}$.

Further restrictions must obviously be placed on $\underline{\Phi}^{+}_{vv}(s)$ and $\underline{\Phi}^{-}_{vv}(s)$. It has been shown in Section 2.6 that $\underline{\Phi}^{T}_{vv}(s) = \underline{\Phi}^{+}_{vv}(-s)$

Therefore, let $\underline{\Phi}_{vv}(s) = \underline{G}(-s) \cdot \underline{G}^{T}(s)$ (2.27)

where $\underline{G}(s)$ and $\underline{G}^{-1}(s)$ are both physically realizeable. Thus

$$\underline{W}^{T}(s) = \left[\underline{G}^{T}(s)\right]^{-1} \mathcal{L}^{-1}\left\{\left[\underline{G}(-s)\right]^{-1} \underline{\Phi}_{vi}(s)\right\} \quad (2.28)$$

In section 2.5 it was pointed out that factoring the single dimensional $\Phi_{vv}(s)$ into $\Phi^{+}_{vv}(-s) \cdot \Phi^{+}_{vv}(s)$ determined $\Phi^{+}_{vv}(s)$, the transfer function of a linear system which could reproduce the observed signals when excited by white noise with unit power density. This is the Bode-Shannon approach[4]. It is natural to inquire if a similar interpretation can be placed on the factoring of $\underline{\Phi}_{vv}(s)$.

Suppose a set of n uncorrelated unity white noise excitations, w_j, are applied to a physical matrix filter, $\underline{G}(s)$, as shown in Figure 2.13.

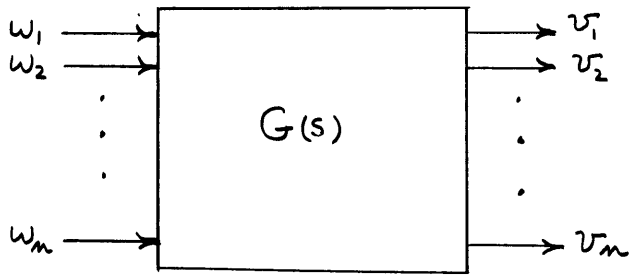


Figure 2.13. A random process created by n white noise sources

$$V_i(s) = \sum_{j=1}^{n} G_{ij}(s)\, W_j(s)$$

$$\Phi_{v_i v_j}(s) = \sum_{l=1}^{n} G_{il}(-s) \sum_{k=1}^{n} G_{jk}(s)\, \Phi_{w_l w_k}$$

$$\text{but } \Phi_{w_l w_k} = 1 \quad l = k$$
$$= 0 \quad l \neq k$$

$$\Phi_{v_i v_j}(s) = \sum_{l=1}^{n} G_{il}(-s)\, G_{jl}(s)$$

In matrix notation,

$$\underline{\Phi}_{vv}(s) = \underline{G}(-s)\, \underline{G}^{T}(s) \quad (2.27)$$

which is the desired result. That is, the process of matrix factoring, which leads to a closed-form solution to the optimum multi-dimensional system, is identical to the problem of finding a physical system which can produce the observed statistics with white noise excitation.

Thus, the multi-dimensional problem has been shown to parallel exactly the single-dimensional case in notation and meaning, if the matrix expression is substituted for the one-dimensional transfer function.

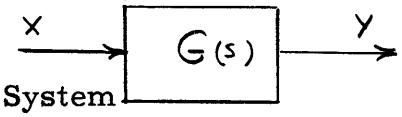
Chapter III will present various approaches and a complete solution to the formidable matrix factorization problem. It should be pointed out again that this matrix approach produces the first general closed-form solution to the optimum multi-dimensional system in the Wiener sense.

2.9 Statistical transformations on random vectors

A great similarity has been demonstrated between the scalar and the matrix representation of random processes. For example, $\Phi_{xx}(s)$ describes a single random process just as $\underline{\Phi}_{xx}(s)$ describes a set or "vector" of n random processes. Some of the simpler relations to be derived were earlier presented by Summers[18], but in view of the simplicity of derivation using equation 2.9, they will be repeated here.

Consider first the simple configuration of Figure 2.14

Figure 2.14



Multi-dimensional System

$$\left[Y(s) \right] = \underline{G(s)} \left[X(s) \right]$$

$$Y_i(s) = \sum_{j=1}^{m} G_{ij}(s)\, X_j(s)$$

$$\Phi_{y_i y_j}(s) = \sum_{l=1}^{m} G_{il}(-s) \sum_{k=1}^{m} G_{jk}(s)\, \Phi_{x_l x_k}(s)$$

$$\underline{\Phi_{yy}(s)} = \underline{G(-s)}\ \underline{\Phi_{xx}(s)}\ \underline{G^{T}(s)} \qquad (2.29)$$

In the special case where $\underline{x}$ is a set of uncorrelated white noise signals with unit power density,

$$\underline{\Phi}_{xx}(s) = \underline{I}$$

$$\underline{\Phi}_{yy}(s) = \underline{G}(-s)\ \underline{G}^{T}(s)$$

verifying Equation 2.27.

$$\Phi_{x_i y_j}(s) = \sum_{k=1}^{n} G_{jk}(s)\ \Phi_{x_i x_k}(s)$$

$$\underline{\Phi}_{xy}(s) = \underline{\Phi}_{xx}(s)\ \underline{G}^{T}(s)$$

Next, the summing operation of Figure 2.15 is examined.

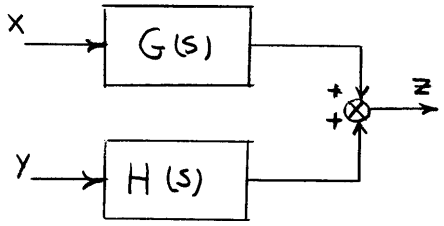


Figure 2.15 A Multi-Dimensional Summing Operation

$$Z_i(s) = \sum_{j=1}^{n} G_{ij}(s)\ X_j(s) + \sum_{k=1}^{n} H_{ik}(s)\ Y_k(s)$$

$$\Phi_{z_i z_i}(s) = \sum_{j=1}^{n} G_{ij}(-s) \sum_{k=1}^{n} G_{ik}(s)\ \Phi_{x_j x_k}(s)$$

$$+ \sum_{j=1}^{n} G_{ij}(-s) \sum_{k=1}^{n} H_{ik}(s)\ \Phi_{x_j y_k}(s) + \sum_{k=1}^{n} H_{ik}(-s) \sum_{j=1}^{n} G_{ij}(s)\ \Phi_{x_k y_j}(s)$$

$$+ \sum_{k=1}^{n} H_{ik}(-s) \sum_{l=1}^{n} H_{il}(s)\ \Phi_{y_k y_l}(s)$$

$$\underline{\Phi}_{zz}(s) = \underline{G}(-s)\ \underline{\Phi}_{xx}(s)\ \underline{G}^{T}(s) + \underline{G}(-s)\ \underline{\Phi}_{xy}(s)\ \underline{H}^{T}(s)$$

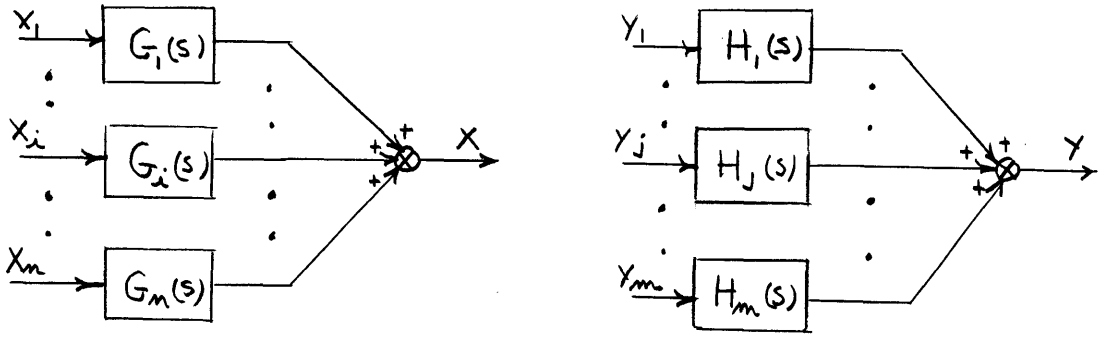
$$+ \underline{H}(-s)\ \underline{\Phi}_{yx}(s)\ \underline{G}^{T}(s) + \underline{H}(-s)\ \underline{\Phi}_{yy}(s)\ \underline{H}^{T}(s) \qquad (2.31)$$

The preceeding configurations were examined in deliberate similarity to the scalar results of Section 2.4. It inferentially appears that a general formula for vector random processes can be expressed just as Equation 2.9 applies to scalar processes.

$$\underline{\Phi}_{xy}(s) = \sum_{i=1}^{m} \sum_{j=1}^{m} \underline{G_i(-s)} \; \underline{\Phi_{x_i y_j}(s)} \; \underline{H_j^T(s)} \qquad (2.32)$$

where the vector $\underline{X(s)} = \sum_{i=1}^{m} \underline{G_i(s) X_i(s)}$ and $\underline{Y(s)} = \sum_{j=1}^{m} \underline{H_j(s) Y_j(s)}$. The ij subscript of $\underline{\Phi_{x_i y_j}(s)}$ refers to the i^{th} vector input, $\underline{x_i}$, making up $\underline{x}$, and similarly for the j^{th} vector excitation of $\underline{y}$.

To prove this formula, which is believed to be the most general expression of statistical transformations in linear systems, consider the system of Figure 2.16.



MATRIX	ELEMENT
$G_i(s)$	$G^i_{pq}(s)$
$H_j(s)$	$H^j_{tu}(s)$
$X_i(s)$	$X^i_q(s)$
$Y_j(s)$	$Y^j_u(s)$
$X(s)$	$X_p(s)$
$Y(s)$	$Y_t(s)$

Figure 2.16. A general multi-dimensional system

$$X_p(s) = \sum_{i=1}^{m} \left(\sum_{q=1}^{r} G^i_{pq}(s) \; . \; X^i_q(s) \right)$$

$$Y_t(s) = \sum_{j=1}^{m} \left(\sum_{u=1}^{r} H^j_{tu}(s) \; Y^j_u(s) \right)$$

From the basic equation, 2.9.

$$\Phi_{x_p y_t}(s) = \sum_{i=1}^{m} \left(\sum_{q=1}^{r} G^i_{pq}(-s) \right) \sum_{j=1}^{m} \left(\sum_{u=1}^{r} H^j_{tu}(s) \right) \Phi_{x^i_q \, y^j_u}(s)$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} \left[\sum_{q=1}^{r} G^{i}_{pq}(-s) \sum_{u=1}^{r} H^{j}_{tu}(s) \, \Phi_{x^i_q \, y^j_u}(s) \right]$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} \; p,t \text{ th element of} \left[G^{i}(-s) \, \Phi_{x^i \, y^j}(s) \left[H^{j}(s) \right]^{T} \right]$$

$$\underline{\Phi}_{xy}(s) = \sum_{i=1}^{m} \sum_{j=1}^{m} \underline{G^{i}(-s)} \; \underline{\Phi_{x^i \, y^j}(s)} \; \underline{\left[H^{j}(s) \right]^{T}} \qquad (2.32)$$

With the use of this formula, statistical relationships in multi-dimensional system variables are swiftly expressed. An example will prove the previous statement that $\underline{\Phi}_{vi}(s)$ is a sufficient description of the ideal signal in multi-dimensional optimization. Consider Figure 2.17,

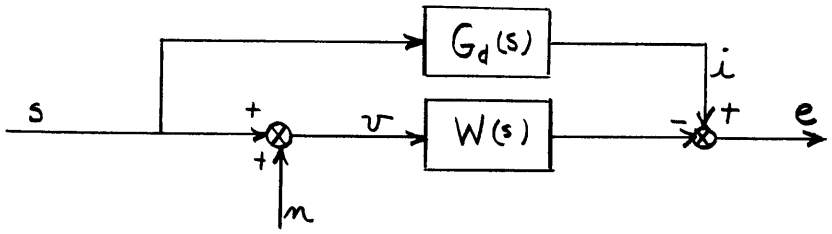


Figure 2.17. Calculation of $\underline{\Phi}_{vi}(s)$

where all variables are random vectors and all systems are matrix operators.

$$\left[v(s) \right] = \left[S(s) \right] + \left[N(s) \right]$$

$$\left[I(s) \right] = \left[G_d(s) \right] \left[S(s) \right]$$

$$\underline{\Phi}_{vi}(s) = \underline{\Phi}_{ss}(s) \; \underline{G_d^{T}(s)} + \underline{\Phi}_{ns}(s) \; \underline{G_d^{T}(s)} \qquad (2.33)$$

Thus, $\underline{\Phi}_{vi}(s)$ is equivalent to $\underline{G_d(s)}$ if the input statistics are known.

CHAPTER III.

MATRIX FACTORIZATION

3.1 Statement of the problem

This chapter is concerned with factoring a matrix of cross-power spectra between signals in a multi-dimensional random process. Chapter II has shown that solution of this problem will yield two significant results:

(1) A closed-form solution can be found for an optimum multi-dimensional configuration in the Wiener sense.

(2) A multi-dimensional linear model is determined which can reproduce the observed statistics when excited by a number of uncorrelated white noise sources.

The basic equation is

$$\underline{\Phi}_{vv}(s) = \underline{G(-s)}\ \underline{G^{T}(s)} \qquad (2.27)$$

or, in expanded form,

$$\begin{bmatrix} \Phi_{v_1v_1}(s) & \Phi_{v_1v_2}(s) & \cdots & \Phi_{v_1v_n}(s) \\ \Phi_{v_2v_1}(s) & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \Phi_{v_nv_1}(s) & \cdot & \cdot & \Phi_{v_nv_n}(s) \end{bmatrix} =$$

$$\begin{bmatrix} G_{11}(-s) & G_{12}(-s) & \cdots & G_{1n}(-s) \\ G_{21}(-s) & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ G_{n1}(-s) & \cdot & \cdot & G_{nn}(-s) \end{bmatrix} \begin{bmatrix} G_{11}(s) & G_{21}(s) & \cdots & G_{n1}(s) \\ G_{12}(s) & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ G_{1n}(s) & \cdot & \cdot & G_{nn}(s) \end{bmatrix}$$

The $\underline{G(s)}$ which is found as a result of the factorization process is the matrix filter described in (2) above. Each element of $\underline{G(s)}$ and $\underline{G(s)}^{-1}$ must be physically realizable in order to meet the requirements given in section 2.8 for the solution of an optimum multi-dimensional configuration.

3.2 Realizability considerations

Before plunging into a solution of this thorny problem, it is necessary and useful to examine the properties of $\underline{\overline{\Phi}_{vv}(s)}$ which characterize a set of random processes which could actually be found in the real world.

The ij^{th} element of $\underline{\overline{\Phi}_{vv}(s)}$ is $\overline{\Phi}_{v_i v_j}(s)$, where v_i and v_j are members of an n-dimensional random process. Since $\overline{\Phi}_{v_i v_j}(-s) = \overline{\Phi}_{v_j v_i}(s)$, $\underline{\overline{\Phi}_{vv}(-s)} = \underline{\overline{\Phi}^{T}_{vv}(s)}$.

In addition, Kraus and Potzl[19] have proven that a necessary and sufficient condition for $\underline{\overline{\Phi}_{vv}(s)}$ to represent a valid multi-dimensional random process is that $\underline{\overline{\Phi}_{vv}(j\omega)}$ be positive definite for all ω. This arises quite naturally if the n signals are allowed to pass through a system $\underline{G}$ which multiplies each signal by an arbitrary constant and sums the total. The power density spectrum in ω of the single output is, from Eq. 2.29,

$$\left[\underline{G}\right]\left[\underline{\overline{\Phi}_{vv}(j\omega)}\right]\left[\underline{G}\right]$$

This spectrum must have a non-negative value for all values of ω, since a negative mean square value of power density cannot exist. Thus $\underline{\overline{\Phi}_{vv}(j\omega)}$ must be non-negative definite for all values of ω. The special case where $\left|\underline{\overline{\Phi}_{vv}(j\omega)}\right|$ equals zero for all values of ω will be considered separatedly in section 3.9, and a positive-definite limitation on $\underline{\overline{\Phi}_{vv}(j\omega)}$ will henceforth be considered a valid demonstration of the realizability of the random process. As will become more clear in the remainder of this section, the only other case where a zero value of power density can occur at a finite value of ω is the occurrance of a multiple even-order zero on the $j\omega$ axis in $\left|\overline{\Phi}_{vv}(s)\right|$.

Positive-definiteness is a property of a matrix which is capable of a number of separate verifications. For the purpose of this theory, a particular method indicated by Bellman[20] is preferable. He states that a necessary and sufficient test for positive-definiteness of a Hermitian matrix is that each of the diagonal elements be positive and that the determinant be also positive. In the power density spectra application, this criterion means that the power density spectrum of each of the n random signals must be positive, as well as $\left|\overline{\Phi}_{vv}(j\omega)\right|$, for all ω.

It is interesting to relate these requirements to known properties of the auto and cross-correlation functions. For simplicity, the 2X2 case will be examined.

$$\underline{\overline{\Phi}_{vv}(s)} = \begin{bmatrix} \overline{\Phi}_{11}(s) & \overline{\Phi}_{12}(s) \\ \overline{\Phi}_{12}(-s) & \overline{\Phi}_{22}(s) \end{bmatrix}$$

The requirements for realizability are that $\overline{\Phi}_{11}(j\omega)$, $\overline{\Phi}_{22}(j\omega)$, and $\overline{\Phi}_{11}(j\omega)\,\overline{\Phi}_{22}(j\omega) - \overline{\Phi}_{12}(j\omega)\,\overline{\Phi}_{12}(-j\omega)$ each be greater than zero for all ω.

$$\mathcal{F}^{-1}\left\{\underline{\overline{\Phi}_{vv}(s)}\right\} = \begin{bmatrix} \varphi_{11}(\tau) & \varphi_{12}(\tau) \\ \varphi_{12}(-\tau) & \varphi_{22}(\tau) \end{bmatrix}$$

Newton, Gould, and Kaiser[10] have presented some physical realizability requirements on the correlation functions, derived from initially setting the square of a linear function of the signals equal to or greater than zero:

$$\varphi_{ii}(0) \geq \varphi_{ii}(\tau) \quad (i = 1, 2) \quad -\infty < \tau < \infty \tag{3.1}$$

$$\varphi_{11}(0)\,\varphi_{22}(0) \geq \left|\varphi_{12}(\tau)\right|^2 \quad -\infty < \tau < \infty \tag{3.2}$$

A relationship between the power density spectra and the correlation realizability requirements will now be derived. Eq. 3.1 can be

expressed for i = 1, as

$$\frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\ \Phi_{11}(s) \;-\; \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\ e^{s\tau}\ \Phi_{11}(s) \;\geq\; 0$$

Replacing s by jω,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\ \left(1 - e^{j\omega\tau}\right)\ \Phi_{11}(j\omega) \;\geq\; 0$$

Since the real part of $1 - e^{j\omega\tau}$ is always equal to or greater than zero, this integral will always be greater than zero if $\Phi_{11}(j\omega) \geq 0$ for all ω. This relates the positiveness of $\Phi_{11}(j\omega)$ (or $\Phi_{22}(j\omega)$) to the fact that a signal has the highest correlation with itself as opposed to any time-shifted version of itself.

At this point, it is well to ask if the positivity of $\Phi_{11}(s)$ can be determined by inspection. It is not enough that $\Phi_{11}(s) = \Phi_{11}(-s)$. For example, $\Phi_{11}(s) = \frac{s^2 + 3}{(-s+2)(s+2)}$ satisfies this relationship but $\frac{-\omega^2 + 3}{\omega^2 + 4}$ is negative for $\omega > \sqrt{3}$.

The example above contains a conjugate pair of zeroes on the jω axis and is not factorable into $\Phi_{11}^{+}(s) \cdot \Phi_{11}^{+}(-s)$. This pair of simple zeroes are the only factors for which $\Phi_{11}(s) = \Phi_{11}(-s)$ and which cannot be factored with mirror symmetry about the jω axis. Thus, factorization is the only realizability requirement for a single power density spectra.

The second correlation function inequality, Eq. 3.2, may be written as

$$\frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds_1\ \Phi_{11}(s_1) \cdot \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds_2\ \Phi_{22}(s_2) \;-$$

$$\frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds_1\ \Phi_{12}(s_1)e^{s_1\tau} \cdot \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds_2\ e^{-s_2\tau}\Phi_{12}(-s_2) \;\geq\; 0$$

or, replacing s by jω,

$$\left(\frac{1}{2\pi j}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\omega_1\, d\omega_2 \left[\Phi_{11}(j\omega_1)\,\Phi_{22}(j\omega_2) - e^{j(\omega_1-\omega_2)\tau}\,\Phi_{12}(j\omega_1)\,\Phi_{12}(-j\omega_2)\right] \geq 0$$

The maximum value of $e^{j(\omega_1-\omega_2)}\ \Phi_{12}(j\omega_1)\ \Phi_{12}(-j\omega_2)$ will occur when $\omega_1 = \omega_2 = \omega$ and $\left|\Phi_{12}(j\omega)\right|$ is a maximum. Therefore, the minimum value of the integrand is

$$\Phi_{11}(j\omega)\ \Phi_{22}(j\omega) - \Phi_{12}(-j\omega)\ \Phi_{12}(j\omega)$$

for some value of ω. If this integrand is positive for all ω, which is the previously given realizability criterion, the integral will always be positive. Positivity of this integrand is again equivalent to factorizability of the 2X2 $\left|\underline{\Phi}_{vv}(s)\right|$, as was true for the auto-power density spectra.

In summary, the realizability criteria found in the literature for the existance of the correlation and spectral functions of random processes are related and the factorability of the individual power density spectra and the matrix determinant is enough to satisfy all requirements. The reason for the emphasis on this matter of realizability is that any method of finding a real system which can create the observed statistics must fail when either the diagonal elements or the determinant of $\underline{\Phi}_{vv}(s)$ cannot be factored, for otherwise the paradox of unrealizable signals being created by a realizable system would exist.

3.3 Two special cases

In this section, two special matrix configurations will be examined which can be readily factored. These particular cases are of importance since they provide goals for a more general factorization procedure.

When $\underline{\Phi}_{vv}(s)$ is a diagonal matrix, each element must be able to be factored into LHP and RHP terms, as shown in section 3.2. Therefore,

$$\underline{\Phi}_{vv}(s) = \underline{D}(-s)\ \ \underline{D}(s)$$

where $\underline{D}(s)$ is a diagonal matrix containing the LHP factors of all the diagonal elements of $\underline{\Phi}_{vv}(s)$.

The second example of an easily factored matrix is the numerical Hermitian matrix. Lee[21] has investigated this problem and has proved that a solution always exists providing that the matrix is positive definite. The problem is to factor $\underline{H}$ into $\underline{X} \cdot \underline{X}^T$, where $\underline{X}$ is a numerical matrix.

Lee shows that a canonical triangular form exists for this problem, $\underline{H} = \underline{T} \cdot \underline{T}^T$, where $\underline{T}$ is triangular and an entire family of solutions is generated by $\underline{T} \cdot \underline{U} \cdot \underline{U}^T \cdot \underline{T}^T$ where $\underline{U} \cdot \underline{U}^T = \underline{I}$, or $\underline{U}$ is a unitary matrix with real elements. In illustration of this result, suppose $\underline{H} = \begin{bmatrix} 13 & 5 \\ 5 & 2 \end{bmatrix}$

and $\underline{T} = \begin{bmatrix} T_{11} & 0 \\ T_{12} & T_{22} \end{bmatrix}$

The elements of $\underline{T}$ can be solved for consecutively because of the triangular form, yielding

$$\underline{H} = \underline{T} \cdot \underline{T}^T = \begin{bmatrix} \sqrt{13} & 0 \\ \frac{5}{\sqrt{13}} & \frac{1}{\sqrt{13}} \end{bmatrix} \begin{bmatrix} \sqrt{13} & \frac{5}{\sqrt{13}} \\ 0 & \frac{1}{\sqrt{13}} \end{bmatrix}$$

A general form for a 2X2 unitary matrix is

$$\underline{U} = \begin{bmatrix} a & \sqrt{1-a^2} \\ -\sqrt{1-a^2} & a \end{bmatrix} \qquad -1 \leq a \leq 1 \qquad (3.3)$$

This single degree of freedom reflects the difference between the number of unknowns, 4, and the number of independent equations which can be written, 3 (as the symmetrical form of $\underline{T}$ leads to identical equations for transpose pairs off the main diagonal). In the general case, $\frac{n(n-1)}{2}$ bounded variables can be adjusted independently in the factorization problem.

The particular significance of the numerical case is that the general factorization procedure to be presented in section 3.5 will reduce in the last stage to a matrix with only numbers. Another perhaps

more conceptual use of this special result is to visualize a matrix $\underline{\Phi}_{vv}(j\omega)$ as a Hermitian matrix which can be factored for every value of ω, providing that the matrix remains positive definite (the realizability requirement), and thus a matrix which is some function of ω does exist.

It might seem at first approach that a triangular form could be postulated for $\underline{\Phi}_{vv}(s)$ factorization, in analogy to the numerical case. This is unfortunately not true, as will be demonstrated below.

Referring to the general two-dimensional case,

$$\begin{bmatrix} \Phi_{11}(s) & \Phi_{12}(s) \\ \Phi_{12}(-s) & \Phi_{22}(s) \end{bmatrix} = \begin{bmatrix} G_{11}(-s) & 0 \\ G_{21}(-s) & G_{22}(-s) \end{bmatrix} \begin{bmatrix} G_{11}(s) & G_{21}(s) \\ 0 & G_{22}(s) \end{bmatrix}$$

$$G_{11}(-s)\, G_{11}(s) = \Phi_{11}(s) = \Phi^{+}_{11}(s)\, \Phi^{+}_{11}(s)$$

Suppose that

$$G_{11}(s) = \Phi^{+}_{11}(s)$$

$$G_{11}(s)\, G_{21}(-s) = \Phi_{12}(-s)$$

$$G_{21}(s) = \frac{\Phi_{12}(s)}{G_{11}(-s)}$$

If $G_{11}(s)$ has its zeroes in the LHP, $G_{21}(s)$ will have these as poles in the RHP. If $G_{11}(s)$ had been selected to have RHP zeroes and LHP poles, the inverse matrix $\underline{G(s)}^{-1}$ would be physically unrealizable.

$$\underline{G(s)}^{-1} = \begin{bmatrix} \dfrac{1}{G_{11}(s)} & 0 \\ -\dfrac{G_{21}(s)}{G_{11}(s)\, G_{22}(s)} & \dfrac{1}{G_{22}(s)} \end{bmatrix}$$

Accordingly, the triangular form does not yield both a solution with a realizable and inverse realizable $\underline{G(s)}$. However, it offers a use-

ful method of reproducing a multi-dimensional random process in an analog computer where inverse realizability is of no concern. To assure a realizable $\underline{G(s)}$, the elements of which may be solved for successively, it is only necessary to select the diagonal elements of $\underline{G(s)}$ with RHP zeroes and LHP poles.

3.4 Properties of matrix transformations

The next section will present a general method for solving the matrix problem

$$\underline{\Phi}_{vv}(s) = \underline{G(-s)} \cdot \underline{G^T(s)} \qquad (2.27)$$

The philosophy of approach will be to multiply $\underline{\Phi}_{vv}(s)$ by a succession of simple matrices, transforming it at every step, until the numerical form is reached. In this section, the properties of simple matrix transformations will be presented, emphasizing the viewpoint that a matrix multiplication can be used as a tool to mold a given matrix into a desired form.

There are three basic matrix manipulations to be considered:

(1) Multiplying a row by a function of s and adding it to another row.

(2) Multiplying a row by a function of s.

(3) Exchanging rows.

In the above list and in the discussion to follow, operations on rows by premultiplication are investigated. The results are equally applicable to column operations through post-multiplication, however.

First, any row operation on a matrix can be accomplished by premultiplying the matrix by an identity matrix on which the desired row operations have been performed. The properties of interest in these transformations include the value of the determinant of the transforming identity matrix, and the realizability and inverse realizability of this matrix. In this particular application, as will be shown in the next sec-

tion, row operations are performed with matrices whose elements must have only RHP poles and whose inverse must also only have RHP pole elements.

(1) Multiplying a row by a function of s and adding it to another row.

$$\underline{T(-s)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ A(-s) & B(-s) & C(-s) & 1 \end{bmatrix}$$

The above matrix multiplies the first row by A(-s), the second row by B(-s), and the third row by C(-s), and adds the total to the last row. $\left|\underline{T(-s)}\right| = 1.$

$$\underline{T(-s)}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -A(-s) & -B(-s) & -C(-s) & 1 \end{bmatrix}$$

The simple form of the inverse will result for all matrices which add to or from only one row. If A(-s), B(-s), and C(-s) have RHP poles or no poles the matrices are proper for this application, regardless of the location of the element zeroes.

(2) Multiplying a row by a function of s.

$$\underline{T(-s)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & D(-s) \end{bmatrix}$$

The above matrix multiplies the last row by D(-s). $\left|\underline{T(s)}\right| = D(-s).$

$$\underline{T}^{-1}(-s) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{D(-s)} \end{bmatrix}$$

D(-s) must have both RHP poles and zeroes to be a proper transformation.

(3) Exchanging rows.

$$\underline{T} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The above matrix exchanges the third and fourth row. $|\underline{T}| = -1$. $\underline{T}^{-1} = \underline{T}$.

The matrices described above perform simple transformations, possess simple inverses, and in the second case can modify the determinant of the transformed matrix by other than a constant.

3.5 Matrix factorization: A general solution

A procedure is to be described in this section which will always yield a solution to the matrix factorization problem regardless of order, providing realizability criteria are satisfied. Because of the complexity of the problem, no easy solution appears to exist. However, the method of factorization to be presented here has been broken down into several separate phases with each phase consisting of simple matrix transformations and each having a well-defined goal.

Each transformation step can be presented in the following fashion:

$$\underline{T}_i(-s) \cdot \underline{\Phi}^{i-1}(s) \cdot \underline{T}_i^{T}(s) = \underline{\Phi}^{i}(s) \qquad (3.4)$$

The relationship between the pre and post-multiplication matrice is specified in order to ensure that $\underline{\Phi}^{i}(-s) = \left[\underline{\Phi}^{i}(s)\right]^{T}$ for all i.

The overall objective of this procedure is to produce a matrix with numerical elements, $\underline{\Phi}^{r}$, after a number of successive transformations. Thus, if $\underline{\Phi}_{vv}(s) \triangleq \underline{\Phi}^{0}(s)$

$$\underline{T_r(-s)} \ldots \underline{T_2(-s)}\ \underline{T_1(-s)}\ \underline{\Phi^0(s)}\ \underline{T_1^T(s)}\ \underline{T_2^T(s)} \ldots \underline{T_r^T(s)} = \underline{\Phi^r}$$

can be factored into two numerical matrices, $\underline{N} \cdot \underline{N}^{T}$. Inverting,

$$\underline{\Phi_{vv}(s)} = \underline{\Phi^0(s)} = \underline{T_1^{-1}(-s)} \cdot \underline{T_2^{-1}(-s)} \ldots \underline{T_{r(-s)}^{-1}} \cdot \underline{N}\left[\underline{T_1^{-1}(s)} \cdot \underline{T_2^{-1}(s)} \ldots \underline{T_r^{-1}(s)} \cdot \underline{N}\right]^{T}$$

$$= \underline{G(-s)} \cdot \underline{G^{T}(s)}$$

$$\underline{G(s)} = \underline{T_1^{-1}(-s)} \cdot \underline{T_2^{-1}(s)} \cdots \underline{T_r^{-1}(s)} \cdot \underline{N} \qquad (3.5)$$

$$\underline{G^{-1}(s)} = \underline{N^{-1}} \cdot \underline{T_r(s)} \cdots \underline{T_2(s)} \cdot \underline{T_1(s)} \qquad (3.6)$$

A proper solution of the problem will yield a physically-realizable $\underline{G(s)}$ and $\underline{G^{-1}(s)}$. This will obviously occur if $\underline{T_i(s)}$ and $\underline{T_i^{-1}(s)}$ have LHP poles only for all i. In other words, as $\underline{\Phi}^{i}(s)$ is manipulated into various configurations the realizability requirements on the solution will be met if each transforming matrix meets these requirements. Drawing on the results of section 3.4, the following constraints exist on the elementary matrix transformation $\underline{T_i(-s)}$:

(1) If $\underline{T_i(-s)}$ multiplies one row of $\underline{\Phi}^{i-1}(s)$ with a function of s and adds it to another, this function must have no poles in the LHP.

(2) If $\underline{T_i(-s)}$ multiplies a row of $\underline{\Phi}^{i-1}(s)$ with a function of s, this function must have no poles, or zeroes in the LHP.

Since the equation

$$\underline{\Phi^{i}(s)} = \underline{T_i(-s)}\ \underline{\Phi^{i-1}(s)}\ \underline{T_i^{T}(s)}$$

is in the form of equation 2.29, $\underline{T_i(s)}$ can be interpreted as a physical

system with an input random process having a matrix of cross power density spectra $\underline{\Phi}^{i-1}(s)$, and with an output spectra of $\underline{\Phi}^{i}(s)$. The succession of matrix transformations then is equivalent to cascading a series of physical systems until an output spectra involving only white noise -- the numerical matrix $\underline{N} \cdot \underline{N}^{T}$ - is achieved. This white noise random process is operated on by $\underline{N}^{-1}$ to produce a unit-valued uncorrelated set, whose spectra is given by the identity matrix. The total cascaded system thus operates on the given random process and produces uncorrelated white noise, and is naturally envisioned as the inverse of the hypothetical physical system creating the random process.

There are three general phases to this matrix factorization solution:

(1) Pole removal. The pole removal phase starts with the given matrix and removes the poles of every element.

(2) Determinant reduction. The determinant reduction phase converts a matrix with polynomial elements and with a determinant which is also a polynomial in s, into another matrix which still has polynomial elements but which has a unity determinant.

(3) Element order reduction. This phase operates on a matrix of polynomial elements having a unit determinant until a numerical matrix is reached.

To illustrate the central ideas of this method, a 2X2 example of a simple yet non-trivial case of matrix factorization will be solved. Then, the general case will be examined and each step justified.

EXAMPLE

Suppose a simple two-dimensional system is given by

$$\underline{G(s)} = \begin{bmatrix} \frac{1}{s+3} & \frac{1}{s+2} \\ \frac{1}{s+1} & \frac{1}{s+4} \end{bmatrix} .$$

and is excited by two uncorrelated unit-valued white noise sources. The matrix of output power density spectra is

$$\underline{\Phi}_{vv}(s) = \underline{G}(-s) \cdot \underline{G}^{T}(s) = \begin{bmatrix} \dfrac{-2s^2+13}{(-s+2)(-s+3)(s+2)(s+3)} & \dfrac{-2s^2+11}{(-s+2)(-s+3)(s+1)(s+4)} \\ \dfrac{-2s^2+11}{(-s+1)(-s+4)(s+2)(s+3)} & \dfrac{-2s^2+17}{(-s+1)(-s+4)(s+1)(s+4)} \end{bmatrix}$$

The inverse of the matrix $\underline{G}(s)$ is

$$\underline{G}(s)^{-1} = \frac{(s+1)(s+2)(s+3)(s+4)}{(-4s+10)} \begin{bmatrix} \dfrac{1}{s+4} & -\dfrac{1}{s+2} \\ -\dfrac{1}{s+1} & \dfrac{1}{s+3} \end{bmatrix}$$

which is unstable or unrealizable. Thus, the question is posed: Can a matrix $\underline{G}(s)$ be found which is realizable and inverse realizable, and is a solution to

$$\underline{G}(-s)\ \underline{G}^{T}(s) = \underline{\Phi}_{vv}(s) = \underline{\Phi}^{o}(s) \quad ?$$

(1) Pole removal phase

The objective of this phase is to remove all the poles in every element. This is quite easily done by row and column multiplication.

$$\underline{T}_1(-s) = \begin{bmatrix} (-s+2)(-s+3) & 0 \\ 0 & (-s+1)(-s+4) \end{bmatrix}$$

$$\underline{\Phi}^{1}(s) = \underline{T}_1(-s)\ \underline{\Phi}^{o}(s)\ \underline{T}_1^{T}(s) = \begin{bmatrix} -2s^2+13 & -2s^2+11 \\ -2s^2+11 & -2s^2+17 \end{bmatrix}$$

(2) Determinant reduction phase

In this phase we first desire to manipulate the $\underline{\Phi}$ matrix so it has a determinant which is constant and independent of s. This will be the

case if $\left|\underline{T_2(-s)}\right| \cdot \left|\underline{T_2}^{T}(s)\right| = \dfrac{1}{\left|\underline{\Phi}^{1}(s)\right|}$

Hence,

$$\left|\underline{\Phi}^{1}(s)\right| = -16s^2 + 100 = (-4s + 10)(4s + 10)$$

$$\underline{T_2(-s)} = \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{1}{-4s+10} \end{bmatrix}$$

$$\underline{\Phi}^{2}(s) = \underline{T_2(-s)}\ \underline{\Phi}^{1}(s)\ \underline{T_2}^{T}(s) = \begin{bmatrix} -2s^2 + 13 & \dfrac{-2s^2 + 11}{4s + 10} \\ \dfrac{-2s^2 + 11}{-4s + 10} & \dfrac{-2s^2 + 17}{(-4s+10)(4s+10)} \end{bmatrix}$$

It is now desired to remove the poles in $\underline{\Phi}^{2}(s)$ without affecting its determinant. An adding transformation is thus called for.

$$\Phi^{2}_{21}(s) = \frac{-2s^2 + 11}{-4s + 10} = .5s + 1.25 - \frac{.375}{-s + 2.5}$$

If the first row of $\underline{\Phi}^{2}(s)$ is multiplied by $\dfrac{k}{-s+2.5}$ and added to the second row, the pole will be cancelled if

$$k \cdot (-2s^2 + 13)\Big|_{s = 2.5} = +.375$$

$$\text{or} \quad k = .75$$

The total added quantity will be

$$\frac{.75}{-s+2.5}(-2s^2 + 13) = 1.50s + 3.75 + \frac{.375}{-s + 2.5}$$

$$\underline{T_3(-s)} = \begin{bmatrix} 1 & 0 \\ \dfrac{.75}{-s+2.5} & 1 \end{bmatrix}$$

$$\underline{\Phi}^{3}(s) = \underline{T_3(-s)}\ \underline{\Phi}^{2}(s)\ \underline{T_3}^{T}(s) = \begin{bmatrix} -2s^2 + 13 & -2s + 5 \\ 2s + 5 & 2 \end{bmatrix}$$

$$\left|\underline{\Phi}^{3}(s)\right| = \left|\underline{T_2}(-s)\right| \cdot \left|\underline{T_3}(-s)\right| \cdot \left|\underline{\Phi}^{1}(s)\right| \cdot \left|\underline{T_3}^{T}(s)\right| \cdot \left|\underline{T_2}^{T}(s)\right|$$

$$= \frac{1}{-4s+10} \cdot (-4s+10)(4s+10) \cdot \frac{1}{4s+10} = 1$$

(3) Element order reduction phase

Consider the array of the highest-order powers of s in each element of $\underline{\Phi}^{3}(s)$:

$$\begin{matrix} -2s^2 & -2s \\ 2s & 2 \end{matrix}$$

Note that the first row is equal to the second row multiplied by -s. This is no accident, and arises because the determinant is independent of s. Performing a reduction transformation,

$$\underline{T_4}(-s) = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$$

$$\underline{\Phi}^{4}(s) = \underline{T_4}(-s) \cdot \underline{\Phi}^{3}(s) \cdot \underline{T_4}^{T}(s) = \begin{bmatrix} 13 & 5 \\ 5 & 2 \end{bmatrix}$$

(4) Solution

This numerical matrix is, oddly enough, the example considered in section 3.3, for which the canonical triangular factorization is

$$\underline{N} \cdot \underline{N}^{T} = \begin{bmatrix} \sqrt{13} & 0 \\ \frac{5}{\sqrt{13}} & \frac{1}{13} \end{bmatrix} \begin{bmatrix} \sqrt{13} & \frac{5}{\sqrt{13}} \\ 0 & \frac{1}{\sqrt{13}} \end{bmatrix}$$

From equations 3.5 and 3.6,

$$\underline{G}(s) = \underline{T_1}^{-1}(s) \cdot \underline{T_2}^{-1}(s) \cdot \underline{T_3}^{-1}(s) \cdot \underline{T_4}^{-1}(s) \cdot \underline{N}$$

$$\underline{G}^{-1}(s) = \underline{N}^{-1} \cdot \underline{T_4}(s) \cdot \underline{T_3}(s) \cdot \underline{T_2}(s) \cdot \underline{T_1}(s)$$

$$\underline{G(s)} = \frac{1}{\sqrt{13}} \begin{bmatrix} \dfrac{5s+13}{(s+3)(s+2)} & \dfrac{s}{(s+3)(s+2)} \\ \dfrac{5s+11}{(s+1)(s+4)} & \dfrac{s+10}{(s+1)(s+4)} \end{bmatrix}$$

$$\underline{G(s)}^{-1} = \frac{1}{52\,(s+2.5)} \begin{bmatrix} (s+10)(s+2)(s+3) & -s(s+1)(s+4) \\ -(5s+11)(s+2)(s+3) & (5s+13)(s+1)(s+4) \end{bmatrix}$$

This example has illustrated the significant features of the general factorization procedure:

(1) Poles removed by row and column multiplications.

(2) Factors removed from a determinant of a polynomial matrix through successive introduction and removal of the inverse factor as a pole.

(3) Reduction of a unit-determinant polynomial matrix by operating on the highest powers of s in each element.

The general nxn case will now be examined.

(1) Pole removal phase

In the previous example, all RHP poles were identical in a single row, and all LHP poles were identical in a single column. This configuration facilitated the efficient removal of these poles by row or column multiplications, but did not occur coincidentally. In the general case the ij^{th} element of $\underline{\Phi_{vv}(s)}$ is $\left[\underline{G_i(-s)}\ \underline{G_j(s)}\right]$, where $\underline{G_k(s)}$ is the k^{th} row of $\underline{G(s)}$. All elements of the i^{th} row of $\underline{\Phi_{vv}(s)}$ will have the same RHP poles, which are the poles of $\underline{G_i(-s)}$, and the LHP poles in the j^{th} column of $\underline{\Phi_{vv}(s)}$ will also be similar, except for occasional cancellation effects in both cases.

(2) Determinant reduction phase

The resulting $\underline{\Phi}(s)$ matrix, which has elements which are polynomials is s, must have a factorable determinant with the RHP factors a mirror image of the LHP factors about the $j\omega$ axis. If not, the random process is not realizable according to the discussion of section 3.2. Considering the RHP factors, there is in general a constant and a number of not necessarily distinct zeroes in this determinant.

Suppose that one determinant factor, $-s + a$, is selected. The transformation matrix

$$\underline{T(-s)} = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ & & & \ddots \\ 0 & & & \dfrac{1}{-s+a} \end{bmatrix}$$

divides each element of the last row by $-s + a$. Let each of the resulting last row terms be expanded by partial fractions. The residue of the pole term in the nj^{th} element is $\Phi_{nj}(a)$. The important question now under consideration is: Can each of the first n - 1 rows of $\underline{\Phi}(s)$ be multiplied by a term $\dfrac{k_i}{-s+a}$ and added to the last row so as to eliminate simultaneously all the poles in the last row?

The added pole from the i^{th} row in the j^{th} column is $\dfrac{k_i \cdot \Phi_{ij}(a)}{-s+a}$. Accordingly, the equation to be solved for n - 1 values of k_i is

$$\sum_{i=1}^{n-1} k_i \quad \Phi_{ij}(a) = - \Phi_{nj}(a) \qquad (j = 1, 2 \ldots . n)$$

This in effect requires that the last row be a linear function of the first n - 1 rows of $\underline{\Phi}(a)$. Since $|\underline{\Phi}(a)| = 0$, because $-s + a$ is a factor, the last row of $\underline{\Phi}(a)$ is always a function of at most the first n - 1 rows and the above equations can always be solved.

The n - 1 element vector k is found from

$$\underline{\Phi}_{n-1}(a) \quad k \Big] = - \underline{\Phi}_{nj}(a) \Big]$$

where $\underline{\Phi}_{n-1}(a)$ is the square matrix of the first n - 1 rows and columns of $\underline{\Phi}(a)$, and $\underline{\Phi}_{nj}(a)\Big]$ contains the first n - 1 elements of the n^{th} row of $\underline{\Phi}(a)$.

The pair of premultiplication transformation matrices is thus

$$\begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots \\ \frac{k_1}{-s+a} & \frac{k_2}{-s+a} & \cdots \frac{k_{n-1}}{-s+a} \quad 1 \end{bmatrix} \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots \\ & & \frac{1}{-s+a} \end{bmatrix} = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots \\ \frac{k_1}{-s+a} & \frac{k_2}{-s+a} & \cdots \frac{k_{n-1}}{-s+a} \quad \frac{1}{-s+a} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \frac{1}{-s+a} \end{bmatrix} \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots \\ k_1 & k_2 & \cdots k_{n-1} \quad 1 \end{bmatrix}$$

From a computational point of view, $k\Big]$ should be determined and first used to transform the polynomial matrix with the right hand numerical matrix in the last expression above. Then, the - s + a factor should be removed from each element in the last row by synthetic division.

The same transformation, only with the transposed LHP matrices in post-multiplication, will remove the s + a term from $\left|\underline{\Phi}(s)\right|$. Thus, the order of $\left|\underline{\Phi}(s)\right|$ has been decreased by two. This procedure can obviously be iterated for all factors, single or repeated, until the determinant is only a positive constant K. Then, multiplying the last row and column by $\frac{1}{\sqrt{K}}$ will produce a matrix with polynomial elements in s and a unit determinant.

The only case in which the procedure will not be applicable is when the last row and column of $\underline{\Phi}(s)$ is zero except for the diagonal element. But in this configuration, the diagonal element can always be factored and the problem immediately degenerates to an n - 1 factorization problem.

In summary, it has been demonstrated that a polynomial matrix can always be reduced by simple transformations to a form which has a unit determinant, independent of s. This is an original contribution to the general theory of matrices with algebraic elements, and is the key to the solution of the matrix factorization problem.

(3) Element order reduction phase

The starting point of this phase is a matrix having a unit determinant and the goal is to produce by successive transformations a numerical matrix. An algebraic matrix having a constant determinant is called in the monumental work of Cullis[22] an "impotent" matrix.

Cullis proves that any unit-determinant impotent matrix can be obtained by successive multiplying-and-adding transformations on an identity matrix. Since the inverse of these transformations always exist, this means that there exists at least one set of transformation matrices which can operate on the given impotent matrix to achieve the identity matrix. Unfortunately, no method has been previously presented for determining this sequence but the procedure to be given next appears to be completely general and achieves the desired reduction.

Suppose an array is formed of the highest powered terms in s of each element. Obviously, the terms in the determinantal expansion which have the highest power of s will all be formed from these terms and must sum to zero because the determinant is independent of s. In this array identify the terms which make up the highest power of s in the determinant. Replace the other terms in the array by zero. For example, suppose the highest terms are

$$\begin{matrix} 1\,s^4 & -2\,s^3 & 3\,s^2 \\ 2\,s^3 & -4\,s^2 & 6\,s \\ 3\,s^2 & -6\,s & -2\,s^2 \end{matrix}$$

The highest power of s in the determinant is s^8. Replacing the terms not involved in the s^8 term by zero,

$$\begin{matrix} 1\,s^4 & -2\,s^3 & 0 \\ 2\,s^3 & -4\,s^2 & 0 \\ 0 & 0 & -2\,s^2 \end{matrix}$$

The determinant of this matrix must be zero, so one row can always be expressed as a function of the other rows. In other words, a transformation can be readily found which will reduce the highest power of s in the determinental expansion. In the above example, this transformation is obviously performed by multiplying the second row by $-\frac{1}{2}s$ and adding it to the first row.

Iteration of this reduction of the highest ordered terms can be continued until no element contains a power of s.

In the special case of the 2X2 impotent matrix, the determinant of the highest powered terms of all four elements is always equal to zero, and thus a series of simple operations of multiplying one row and adding it to another will speedily reduce the 2X2 matrix to numerical form.

(4) Solution of the numerical matrix

The only requirement that a solution exist to the factoring of the resulting numerical matrix is that it be positive definite. In the preceding steps, the factorizability of the determinant was the only realizability criterion needed. If one of the original diagonal elements had not been factorable, this would not in general have impeded any of the steps up to this point even though it would indicate an unrealizable system. However, referring to the matrix factorization procedure as a succession of linear systems operating on the random process, as was discussed in the beginning of this section, it is obvious that "unrealizability" and "realizability" are both properties of a set of signals which are not affected by passage through a linear system. Therefore, a positive definite numerical matrix will result if the original power density spectra matrix satisfied the realizability criteria. A non-positive definite matrix implies a set

of white noise having imaginary auto or cross correlation.

Appendix II gives a complete solution to a more complicated 3X3 factorization example.

Summary

This section has presented a general method for factoring a matrix of power density spectra, providing that the statistics arise from a multi-dimensional random process observable in the real world. Alternately, it has been proven that a linear multi-terminal system, excited by white noise, can always be found which (1) is stable, (2) has a stable inverse, and (3) reproduces the observed statistical interrelationships in a random process.

3.6 Matrix factorization: An iterative solution

The method presented in the previous section is always valid, and invariably leads to an answer which satisfies all requirements. This section discusses an iterative procedure which will often yield a valid and speedy solution without the need to determine and factor the determinant of $\underline{\Phi}_{vv}(s)$. This becomes especially valuable when the dimension of $\underline{\Phi}_{vv}(s)$ is high, and when digital computers are used.

The pole removal phase of the general procedure is readily accomplished, and the real factorization problem deals with the resulting matrix with polynomial elements. Let this matrix be designated as $\underline{\Phi}(s)$, which can be expressed as a power series in s with numerical matrix coefficients

$$\underline{\Phi}(s) = \sum_{k=0}^{2m} s^k \quad \underline{\Phi}_k \tag{3.7}$$

The problem considered is to find a matrix $\underline{H(s)}$ which satisfies the equation

$$\underline{H(-s)} \cdot \underline{H^T(s)} = \underline{\Phi}(s) \tag{3.8}$$

where

$$\underline{H(s)} = \sum_{k=0}^{m} s^k \, \underline{H_k} \tag{3.9}$$

$$\underline{H(-s)} \cdot \underline{H^T(s)} = \left(\sum_{k=0}^{m} (-1)^k \, s^k \, \underline{H_k} \right) \left(\sum_{j=0}^{m} s^j \, \underline{H_j}^T \right)$$

Equating coefficients,

$$\underline{\Phi}_r = \sum_{k} (-1)^k \, \underline{H_k} \cdot \underline{H_{r-k}} \tag{3.10}$$

where the range of k is bound by $0 \le k \le m$, $0 \le r-k \le m$.

The matrix factorization problem is, as an alternate interpretation, 2 m + 1 non-linear matrix equations. Suppose an approximate solution, $\underline{H(s)}$, is known. If a small perturbation in $\underline{H(s)}$ is made with $\underline{dH(s)}$ and the resulting product is to equal $\underline{\Phi}(s)$,

$$\underline{\Phi}_r = \sum_{k} (-1)^k \, (\underline{H_k} + \underline{dH_k})(\underline{H_{r-k}} + \underline{dH_{r-k}})^T$$

Neglecting the product $\underline{dH_k} \cdot \underline{dH_{r-k}}$ as of second order, the proper perturbation of $\underline{H(s)}$ is given by solution of the linear equations

$$\underline{\Phi}_r - \sum_{k} (-1)^k \, \underline{H_k} \, \underline{H_{r-k}^T} = \sum_{k} (-1)^k \left\{ \underline{dH_k} \, \underline{H_{r-k}^T} + \underline{H_k} \, \underline{dH_{r-k}^T} \right\} \tag{3.11}$$

The left-hand side of this equation is recognized as the matrix coefficient of the r^{th} power of s in the error: $\underline{\Phi(s)} - \underline{H(-s)} \cdot \underline{H^T(s)}$. After these equations are solved for $\underline{dH_i}$, the remaining r^{th} error will be

$$\sum_{k} \underline{dH_k} \cdot \underline{dH^T_{r-k}}$$

and the procedure may be iterated until the error becomes negligible, providing that the original guess was "close enough".

Besides needing an approximate solution to commence this procedure, another drawback is that the resulting solution $\underline{H(s)}$ is not guaranteed to have a realizable inverse -- that is, $\underline{H(s)}$ may contain RHP factors. To handle both of these requirements, a good initial solution for $\underline{H(s)}$ will often be the LHP factors of the diagonal elements of $\underline{\Phi(s)}$.

This first trial, while obviously in error if there are any non-zero off-diagonal elements in $\underline{\Phi}(s)$, will be close to the solution if the cross-correlation among the signals is weak. Also, it definitely has a determinant which has all LHP factors, which a small perturbation in the coefficients of $\underline{H(s)}$ will not appreciably modify.

Having some promise of solving the matrix factorization problem with successive linear equations, it is useful to consider these equations in more detail. The set of equations to be solved is, from Eq. 3.11,

$$\underline{\Phi}_r^e = \sum_k (-1)^k \left\{ \underline{\delta H_k} \cdot \underline{H}^T_{r-k} + \underline{H_k} \cdot \underline{\delta H}^T_{r-k} \right\} \quad \begin{matrix} 0 \le k \le m \\ 0 \le r-k \le m \end{matrix}$$

$$(r = 0, 1 \ldots 2m)$$

where $\underline{\Phi}_r^e$ is derived from

$$\underline{\Phi}(s) - \underline{H}(-s)\,\underline{H}^T(s) = \sum_{r=0}^{2m} s^r\, \underline{\Phi}_r^e$$

The new $\underline{H(s)}$ equals the original $\underline{H(s)}$ plus $\underline{\delta H(s)}$.

The total number of independent variables is the number of independent elements of $\underline{\Phi}_r^e$. For r even, where $\underline{\Phi}_r^e$ is symmetric, these are the diagonal and above-diagonal elements. For r odd, where $\underline{\Phi}_r^e$ is skew-symmetric with zero-valued diagonal elements, these are the above-diagonal elements. The total number of independent elements is thus

$$\frac{(m+1)(n)(n+1)}{2} + \frac{(m)(n)(n-1)}{2} = (m+1)n^2 - \frac{n(n-1)}{2}$$

The number of unknown variables is the number of coefficients of $\underline{\delta H(s)}$, which is $(m+1)n^2$. Therefore, $\frac{n(n-1)}{2}$ elements of $\underline{\delta H(s)}$ can be arbitrarily selected, which reflects the degrees of freedom of the imbedded numerical matrix in the complete rigorous solution. One way of removing this excess is to specify that $\underline{\delta H_o}$ be symmetric.

To illustrate these ideas and to indicate the expected degree of convergence, the sample problem solved in section 3.5 will be re-solved iteratively.

After the pole removal phase,

$$\underline{\Phi}(s) = \begin{bmatrix} -2s^2 + 13 & -2s^2 + 11 \\ -2s^2 + 11 & -2s^2 + 17 \end{bmatrix}$$

The assumed solution for $\underline{H(s)}$ is the LHP factors of the diagonal elements of $\underline{\Phi}(s)$

$$\underline{H(s)} = \begin{bmatrix} 1.414\,s + 3.61 & 0 \\ 0 & 1.414\,s + 4.12 \end{bmatrix}$$

$$\underline{H_o} = \begin{bmatrix} 3.61 & 0 \\ 0 & 4.12 \end{bmatrix} \qquad \underline{H_1} = \begin{bmatrix} 1.414 & 0 \\ 0 & 1.414 \end{bmatrix}$$

The equations to be solved are, from Eq. 3.11,

$$\underline{\Phi_o}^e = \underline{dH_o}\,\underline{H_o}^T + \underline{H_o}\,\underline{dH_o}^T$$

$$\underline{\Phi_1}^e = \underline{dH_o}\,\underline{H_1}^T + \underline{H_o}\,\underline{dH_1}^T - \underline{dH_1}\,\underline{H_o}^T - \underline{H_1}\,\underline{dH_o}^T$$

$$\underline{\Phi_2}^e = -\underline{dH_1}\,\underline{H_1}^T - \underline{H_1}\cdot\underline{dH_1}^T$$

$$\underline{\Phi}^e(s) = \begin{bmatrix} 0 & -2s^2 + 11 \\ -2s^2 + 11 & 0 \end{bmatrix}$$

As an example of the appearance of these equations,

$$\underline{\Phi_o}^e = \begin{bmatrix} \boxed{0} & \boxed{11} \\ 11 & \boxed{0} \end{bmatrix} = \begin{bmatrix} dh_{11}^\circ & dh_{12}^\circ \\ dh_{12}^\circ & dh_{22}^\circ \end{bmatrix} \begin{bmatrix} 3.61 & 0 \\ 0 & 4.12 \end{bmatrix} + \begin{bmatrix} 3.61 & 0 \\ 0 & 4.12 \end{bmatrix} \begin{bmatrix} dh_{11}^\circ & dh_{12}^\circ \\ dh_{12}^\circ & dh_{22}^\circ \end{bmatrix}$$

where $\underline{dH_o}$ was selected as symmetric. The boxed elements of $\underline{\Phi_o}^e$ indicate a set of independent equations. This set of equations can be solved directly, yielding $dh_{11}^\circ = 0$, $dh_{22}^\circ = 0$, $dh_{12}^\circ = 1.424$.

Solving the four remaining independent equations in $\underline{\Phi_1}^e$, and $\underline{\Phi_2}^e$ yields

$$\underline{dH(s)} = \begin{bmatrix} 0 & .660\,s + 1.424 \\ .753s + 1.424 & 0 \end{bmatrix}$$

The new $\underline{H(s)}$ is thus

$$\begin{bmatrix} 1.414\,s + 3.61 & .660\,s + 1.424 \\ .753\,s + 1.424 & 1.414\,s + 4.12 \end{bmatrix}$$

$$\underline{H(-s)}\,\underline{H^T(s)} = \begin{bmatrix} -2.435\,s^2 + 15.03 & -2\,s^2 + 11.02 \\ -2\,s^2 + 11.02 & -2.568\,s^2 + 18.93 \end{bmatrix}$$

$$\underline{\Phi}^e(s) = \begin{bmatrix} .435\,s^2 - 2.03 & -.02 \\ -.02 & +.568\,s^2 - 1.93 \end{bmatrix}$$

Repeating the solution of the seven linear equations, the new $\underline{H(s)}$ is given by

$$\begin{bmatrix} 1.22\,s + 3.285 & .7456\,s + 1.5348 \\ .913\,s + 1.5348 & 1.128\,s + 3.844 \end{bmatrix}$$

$$\underline{H(-s)}\,\underline{H^T(s)} = \begin{matrix} -2.047\,s^2 + 13.15 & -1.957\,s^2 - .012\,s + 10.95 \\ -1.957\,s^2 + .012\,s + 10.95 & -2.105\,s^2 + 17.16 \end{matrix}$$

which is compared with the actual $\underline{\Phi}(s)$

$$\underline{\Phi}(s) = \begin{bmatrix} -2\,s^2 + 13 & -2\,s^2 + 11 \\ -2\,s^2 + 11 & -2\,s^2 + 17 \end{bmatrix}$$

This solution is probably within the accuracy of measurement of $\underline{\Phi}(s)$, and no further iteration is made. $|\underline{H(s)}| = .697\,s^2 + 5.862\,s + 10.66$ which is stable.

In high order problems evaluating and factoring the determinant of $\underline{H(s)}$ can be a very difficult step. If an indication of inverse realizability is desired, an approach similar to that used by the Nyquist stability criterion is very useful. $\frac{1}{|H(j\omega)|}$ is evaluated, possibly by a digital computer, for a sequence of various values of ω and plotted on a complex plane. The presence of RHP factors will then be detected by any net number of encirclements of the origin.

To summarize, the described iterative method presents an attractive alternative to the complete factorization procedure, especially when digital computation is employed. In an example of this method, two iterations solved the problem to an acceptable accuracy level. The price which must be paid for this computational advantage is the possibility of a non-converging solution or one which converges on a solution having an unrealizable inverse.

3.7 Matrix factorization: A lightning solution

This section considers a very special case of matrix factorization, but one which is quite simple to solve. The central requirement is that each non-zero element of any single row of $\underline{G(s)}$, where $\underline{G(-s)} \cdot \underline{G^T(s)} = \underline{\Phi_{vv}(s)}$, must have separate and distinct poles and must have a denominator of higher order than the numerator.

The ij^{th} off-diagonal element of $\underline{\Phi_{vv}(s)}$ is $\sum_{k=1}^{n} G_{ik}(-s)\, G_{jk}(s) = \frac{P_{ij}(s)}{Q_{ij}(s)}$

The first question to be considered is whether each of the n terms $G_{ik}(-s)\, G_{jk}(s)$ can be recovered from a knowledge of $\frac{P_{ij}(s)}{Q_{ij}(s)}$. Alternately, if a partial fraction of $\frac{P_{ij}(s)}{Q_{ij}(s)}$ is made, can all poles belonging to a single element of $\underline{G(-s)}$ or $\underline{G^T(s)}$ be grouped together, and can these groups be further separated into LHP-RHP product pairs?

The key to this grouping is that any scalar function A(-s) B(s), where A(-s) and B(s) have RHP and LHP poles respectively, has an in-

verse time-domain transform $f(\tau)$ which is continuous across the origin as long as the order of the denominator is at least two degrees higher than that of the numerator of A(-s) B(s). This can be proven by showing that $f(0^+) = f(0^-)$ which, by using the contour integral to sum residues, becomes

$$\frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds \quad A(-s)\,B(s) \;=\; \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\; A(s)\,B(-s)$$

The latter equation is valid since the right hand side is merely the left-hand side with the sign of the integrating variable changed, as the negative sign of the differential is cancelled by the limit exchange.

As an example, let

$$A(-s)\,B(s) \;=\; \frac{3}{(-s+7)(s+2)(s+4)}$$

$$f(\tau) = \frac{1}{33}\; e^{7\tau} \qquad (\tau \le 0)$$

$$= \frac{1}{6}\; e^{-2\tau} - \frac{3}{22}\; e^{-4\tau} \qquad (\tau \ge 0)$$

$$f(o^-) = f(o^+) = \frac{1}{33}$$

Thus, the residue of the LHP poles must sum to those of the RHP poles in the partial fraction expansion of any such function as A(-s) B(s).

Therefore, the partial fraction expansion of $\frac{P_{ij}(s)}{Q_{ij}(s)}$ can be grouped to show this residue equality between, in general, n sets of LHP and RHP poles, providing that all elements have distinct poles. If i = j, each LHP pole has an equal RHP pole in the partial fraction expansion, and this grouping is impossible.

Suppose that n such sets of RHP poles have been determined in one element of the first row of $\underline{\Phi}_{vv}(s)$. Under the assumptions of the form of $\underline{G(s)}$, these sets should satisfy residue equality requirements in every off-diagonal element of the first row of $\underline{\Phi}_{vv}(s)$. The first diagonal element is similarly grouped, and the corresponding LHP and RHP terms

of each set are multiplied together. The resulting n terms are $\Phi_{11}(s) = \sum_{j=1}^{m} G_{1j}(-s) G_{ij}(s)$, and thus when individually factored, yield the first row of $\underline{G(s)}$ which may be placed in any desired order.

Having fixed the first row of $\underline{G(-s)}$, given by $\underline{G_1(-s)}$, the j^{th} element of the first row of $\Phi_{vv}(s)$ is $\underline{G_1(-s)}\ G_j(s)$, and thus $G_j(s)$ can be found directly for all j, since the residue equality requirement associates each element of $\underline{G_j(s)}$ with the known set of poles in an element of $\underline{G_1(-s)}$. $\underline{G(-s)}\ \underline{G^T(s)}$ is then evaluated, and under the restrictions of distinct poles of $\underline{G(s)}$, will equal $\underline{\Phi_{vv}(s)}$.

As an example, the much-battered veteran of this chapter will be resolved.

$$\underline{\Phi_{vv}(s)} = \begin{bmatrix} \dfrac{-2s^2+13}{(-s+2)(-s+3)(s+2)(s+3)} & \dfrac{-2s^2+11}{(-s+2)(-s+3)(s+1)(s+4)} \\ \dfrac{-2s^2+11}{(-s+1)(-s+4)(s+2)(s+3)} & \dfrac{-2s^2+17}{(-s+1)(-s+4)(s+1)(s+4)} \end{bmatrix}$$

$$\Phi_{v_1v_2}(s) = \frac{-2s^2+11}{(-s+2)(-s+3)(s+1)(s+4)} = \frac{\frac{1}{6}}{-s+2} + \frac{\frac{1}{4}}{-s+3} + \frac{\frac{1}{4}}{s+1} + \frac{\frac{1}{6}}{s+4}$$

$$= \frac{1}{(-s+3)(s+1)} + \frac{1}{(-s+2)(s+4)}$$

The poles at s = 3 and s = 2 satisfy separate residue equality requirements, lending support to the hope that $\underline{\Phi_{vv}(s)}$ can result from a $\underline{G(s)}$ with distinct elements.

$$\Phi_{v_1v_1}(s) = \frac{-2s^2+13}{(-s+2)(-s+3)(s+2)(s+3)} = \frac{1}{(-s+3)(s+3)} + \frac{1}{(-s+2)(s+2)}$$

Let $G_{11}(-s) = \dfrac{1}{-s+3}$ and $G_{12}(-s) = \dfrac{1}{-s+2}$

$$\Phi_{v_1v_2}(s) = G_{11}(-s)\ G_{21}(s) + G_{12}(-s)\ G_{22}(s)$$

$$G_{21}(s) = \frac{1}{s+1} \qquad G_{22}(s) = \frac{1}{s+4}$$

and the resulting $\underline{G(s)}$ is given by

$$\underline{G(s)} = \begin{bmatrix} \frac{1}{s+3} & \frac{1}{s+2} \\ \frac{1}{s+1} & \frac{1}{s+4} \end{bmatrix}$$

and $\underline{G(-s)}\ \underline{G^T(s)}$ yields the given $\underline{\Phi_{vv}(s)}$.

But the problem is not yet complete, and this example was purposely chosen to illustrate a significant defect of this simplified attack. As given in section 3.5, $\underline{G^{-1}(s)}$ contains a RHP pole and is unrealizable. Generally, the resulting solution in this method may or may not be inverse realizable, but its simplicity makes the attempt worthwhile as a preliminary to the increasing rigor, generality, and computational complexity of the methods given in sections 3.6 and 3.5.

3.8 Statistical degrees of freedom of a multi-dimensional random process

Up to this point it has been assumed that $\underline{\Phi_{vv}(s)}$ is a non-singular nxn matrix. If $|\underline{\Phi_{vv}(s)}| = 0$, this implies that one or more rows of a hypothesized nxn G(s) is a linear function (not necessarily numerical) of the remaining rows. Suppose the k^{th} row of G(s), $\underline{G_k(s)} = \sum_{i=1}^{m} C_i(s) \cdot \underline{G_i(s)}$ $(k > m)$. $v_k(s) = \underline{G_k(s)}\ W(s)]$, where $W(s)]$ is the hypothesized transform of the white noise excitation vector over a finite interval.

$$V_R(s) = \sum_{i=1}^{m} C_i(s)\ \underline{G_i(s)}\ W(s)] = \sum_{i=1}^{m} C_i(s)\ V_i(s)$$

Therefore, $v_k(s)$ is a redundant member of the set of signals and can contribute no additional statistical information on the multi-variable random process. At this point the representation of $\underline{G(s)}$ as an nxn matrix excited by n uncorrelated white noise sources is open to question, since there are less than n "useful" outputs.

Suppose, by striking out pairs of rows and columns, the highest order non-singular matrix contained in $\underline{\Phi}_{vv}(s)$ is found. Denote this matrix as $\underline{\Phi}_{m}(s)$, representing a set of m independent components of the set v. It has been shown in this chapter that if physical realizability criteria are satisfied $\underline{\Phi}_{m}(s)$ can be factored into $\underline{G}_{m}(-s) \cdot \underline{G}_{m}^{T}(s)$, with $G_m(s)$ excited by m white noise sources. It appears logical that the remaining n - m dependent signals can be derived from these m white noise sources, as shown in Figure 3.1.

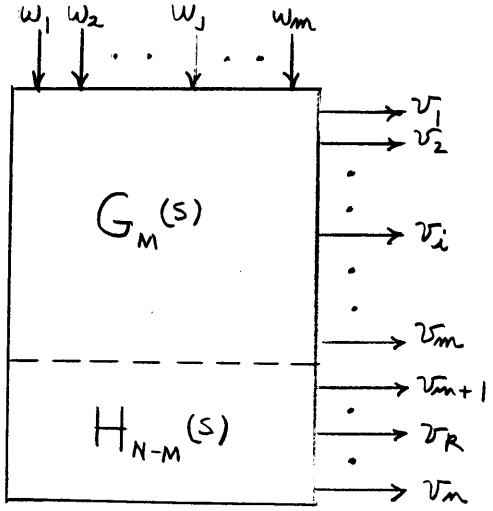


Figure 3.1. Formation of a multi-dimensional random process with redundant elements.

The adequacy of this model will be proved in the following steps: a $\underline{H}_{n-m}(s)$ will be found which satisfies the cross power density spectra relationship between every v_k and v_i. It will then be shown that this $\underline{H}_{n-m}(s)$ produces signals v_k which have the proper cross power density spectra among themselves. Thus every signal will be related as indicated in the original $\underline{\Phi}_{vv}(s)$ matrix and Fig. 3.1 will indeed be a valid representation of a multi-dimensional random process with a matrix $\underline{\Phi}_{vv}(s)$ of rank m.

To better picture the following steps, $\underline{\Phi}_{vv}(s)$ is shown in partitioned form.

$$\underline{\Phi}_{vv}(s) = \left[\begin{array}{c|c} \underline{\Phi}_{m}(s) & \underline{\Phi}_{v_i v_R}(s) \\ \hline \underline{\Phi}_{v_R v_i}(s) & \underline{\Phi}_{v_R v_R}(s) \end{array}\right] \stackrel{?}{=} \left[\begin{array}{c} G_m(-s) \\ \hline H_{n-m}(-s) \end{array}\right] \left[G_m^T(s) \,\middle|\, H_{n-m}^T(s)\right] \tag{3.13}$$

From Eq. 2.32,

$$\underline{\Phi}_{v_i v_k}(s) = \underline{G_m}(-s)\ \underline{H_{n-m}^T}(s)$$

$$\underline{H_{n-m}^T}(s) = \underline{G_m^{-1}}(-s)\ \underline{\Phi}_{v_i v_k}(s) \tag{3.14}$$

$\left[\underline{G_m}(-s)\right]^{-1}$ exists because of its non-singularity. But also, $\underline{H_{n-m}}(s)$ must satisfy the equality.

$$\underline{\Phi}_{v_k v_k}(s) = \underline{H_{n-m}}(-s) \cdot \underline{H_{n-m}^T}(s)$$

From Eq. 3.14, the following relations must hold for the partitioned sub-matrices of $\underline{\Phi}_{vv}(s)$

$$\underline{\Phi}_{v_k v_k}(s) = \underline{\Phi}^T_{v_i v_k}(-s)\ \left[\underline{G_m^{-1}}(s)\right]^T \cdot \left[\underline{G_m^{-1}}(-s)\right] \cdot \underline{\Phi}_{v_i v_k}(s)$$

$$= \underline{\Phi}_{v_k v_i}(s)\ \underline{\Phi}_{m}^{-1}(s)\ \underline{\Phi}_{v_i v_k}(s) \tag{3.15}$$

Since $\underline{\Phi}_{vv}(s)$ is of rank m, each of the last n - m rows can be considered as a linear function of the first m rows. Let $\underline{\Phi_k(s)} = \sum_{i=1}^{m} A_{ki}(s)\ \underline{\Phi_i(s)}$ where $\underline{\Phi_k(s)}$ and $\underline{\Phi_i(s)}$ are row vectors of $\underline{\Phi}_{vv}(s)$ and $A_{ki}(s)$ is a scalar to be determined. Writing this equation in complete matrix form, and recognizing the resulting partitioned matrices,

$$\left[\underline{\Phi}_{v_k v_i}(s) \,\middle|\, \underline{\Phi}_{v_k v_k}(s)\right] = \left[\underline{A(s)}\right]\left[\underline{\Phi}_{m}(s) \,\middle|\, \underline{\Phi}_{v_i v_k}(s)\right]$$

$$\underline{\Phi}_{v_k v_i}(s) = \underline{A(s)}\ \underline{\Phi}_{m}(s)$$

and

$$\underline{\Phi}_{v_k v_k}(s) = \underline{A(s)}\ \underline{\Phi}_{v_i v_k}(s)$$

$$\underline{A(s)} = \underline{\Phi}_{v_k v_i}(s)\ \underline{\Phi}_{m}^{-1}(s)$$

$$\underline{\Phi}_{v_k v_k}(s) = \underline{\Phi}_{v_k v_i}(s)\ \underline{\Phi}_{m}^{-1}(s)\ \underline{\Phi}_{v_i v_k}(s)$$

Thus equation 3.15 is verified, providing that some matrix $\underline{A(s)}$ exists, and the assumed form for $\underline{H_{n-m}(s)}$ produces the observed statistics. Note that $\underline{H_{n-m}(s)}$ is fixed for a choice of $\underline{G_m(s)}$. Transposing Eq. 3.14,

$$\begin{aligned}
\underline{H_{n-m}(s)} &= \underline{\Phi}_{v_k v_i}(-s)\ \left[\underline{G_m^T(-s)}\right]^{-1} \\
&= \underline{A(-s)}\ \underline{\Phi_m(-s)}\ \left[\underline{G_m^T(-s)}\right]^{-1} \\
&= \underline{A(-s)}\ \underline{G_m(s)}\ \underline{G_m^T(-s)}\ \left[\underline{G_m^T(-s)}\right]^{-1} \\
&= \underline{A(-s)}\ \underline{G_m(s)} \qquad (3.16)
\end{aligned}$$

For $\underline{H_{n-m}(s)}$ to be physically realizable, $\underline{A(s)}$ must contain only RHP poles. $\underline{A(s)}$ was used as a row transformation to express the redundant rows as a function of the independent rows. The elements of $\underline{A(s)}$ can be used in an elementary transformation at the beginning of the factorization problem to eliminate all redundant rows and columns, leaving $\underline{\Phi}_m(s)$. Physically, this means that the random process v is passed through a matrix filter $\underline{B(s)}$, such that the resulting output power density spectra matrix

$$= \underline{B(-s)}\ \underline{\Phi}_{vv}(s)\ \underline{B^T(s)} = \left[\begin{array}{c|c} \Phi_m(s) & 0 \\ \hline 0 & 0 \end{array}\right]$$

where $\underline{B(-s)} = \left[\begin{array}{c|c} I & 0 \\ \hline -A(s) & I \end{array}\right]$

That is, $\underline{B(s)}$ weights and adds together the m independent signals of v and nulls out the redundant signals. As discussed in the first paragraph of this section, this dependence among signals, if observed in a stable random process, must arise in a physically realizable system. Therefore, $\underline{B(s)}$, containing all the elements of $\underline{A(-s)}$, must be physically

realizable or $\underline{\Phi_{vv}(s)}$ does not represent a real process.

Thus an index of randomness of a multi-dimensional random process is the rank of the matrix of cross-spectra or, alternately, the number of white noise sources needed to reproduce the statistics of the process. Also, a set of dependent random processes is physically realizable only if the redundant rows of the matrix of power density spectra can be removed by a row transformation with RHP pole factors.

CHAPTER IV.

NEW RESULTS IN OPTIMUM SYSTEM THEORY

4.1 Introduction

The previous chapters have been in a sense an introduction, although a useful one, to the main theme of this report. It has been demonstrated that a linear system excited by white noise can always be found to duplicate the basic statistical properties of any stationary random process, single or multi-dimensional.

In standard texts on random processes it is customary to note that a power density spectrum has the same form as the spectrum of the output of a linear filter excited by white noise. In the early paper by Bode and Shannon[4], which served to convert the highly mathematical approach of Wiener into a form more understandable to engineers, this white noise filter and its inverse were used as a means to remove all memory from the random process and to justify the use of a straight $\mathcal{L}^{-1}$ operation to obtain the optimum configuration. In this work the idea is carried one step further and the hypothesis is offered that within the confines of a linear theory a random process should be viewed as actually being the result of white noise exciting a linear system. Although this system in some cases cannot be physically represented and the white noise sources cannot be traced to microscopic random phenomena, it is possible to make measurements on the random process itself with complete mathematical assurance that there is such a linear system "upstream and around the bend".

This hypothesis would be only of mild interest by itself, but this chapter will show how this simple assumption makes the study of stationary random processes purely a measurement problem and how it tends to unify the conventional analysis techniques of linear systems and those of stochastic processes.

4.2 Matrix differential equations and system state

The heart of the description of a linear physical system is its "state", which effectively describes the condition of every internal energy storage element at every instant. Since a random process is to be analyzed in terms of its equivalent system, it is useful at this point to summarize the major features of the matrix theory of differential equations, such as is found in Bellman[20], in order to emphasize the state approach to the analysis of linear systems. In this case, matrices allow compact expression of ideas without regard to order and dimensionality of the system under consideration. The standard theory outlined in this section will provide a foundation for clear presentation of the original results to be presented in the remainder of this report.

The basic matrix representation for a linear system is presented in the following equation

$$\frac{d}{dt}\,\mathbf{x} = \mathbf{A}\,\mathbf{x} + \mathbf{D}\,\mathbf{u} \qquad (4.1)$$

where x is the n-dimensional state vector of a linear system, A is a constant nxn matrix, D is a constant nxm matrix, and u is the m-dimensional excitation vector. For example, consider the simple second-order system of Figure 4.1, where a spring-mass-dashpot system is being excited by an external force F.

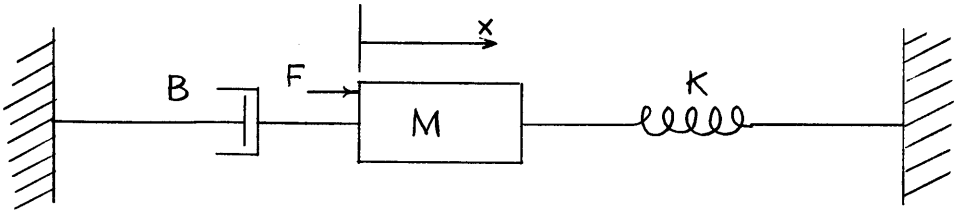


Fig. 4.1 A simple second-order system

Here the differential equation is

$$M\frac{d^2x}{dt^2} + B\frac{dx}{dt} + Kx = F$$

Defining one state variable, x_1, to be x, and x_2 to be $\frac{dx}{dt}$, is sufficient to fix the potential and kinetic energy of the system. The system equations are next cast into the general form of Eq. 4.1.

$$\frac{d}{dt}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{K}{M} & -\frac{B}{M} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{M} \end{bmatrix} F$$

The initial condition response or free behavior of linear systems will be of particular importance in the study of random processes in following sections. Given x (0), it is desired to find x (t) under conditions of no external excitation. It would obviously be desirable to have a solution in the form

$$x(t) = B(t)\ x(0)$$

Assuming this form and substituting into Eq. 4.1,

$$\frac{d}{dt}\ B(t)\ x(0) = A\ B(t)\ x(0)$$

or

$$\frac{d}{dt}\ B(t) = A\ B(t)$$

The series

$$B(t) = I + At + A^2\frac{t^2}{2} \ldots + A^n\frac{t^n}{n!} \ldots$$

where $\frac{d}{dt}\ B(t) = A + A^2 t \ldots + A^n \frac{t^{n-1}}{(n-1)!} \ldots = A\,B(t)$

satisfies this equality.

Thus,

$B(t) = \sum_{n=0}^{\infty} A^n \frac{t^n}{n!}$, where $A^0 \triangleq I$, is the desired solution and is known as the matrix exponential e^{At}, a quantity that is convergent for any value of A and t. It is analogous to the scalar exponential, and occupies a position of pivotal importance in linear systems analysis.

If Eq. 4.1 is Laplace transformed,

$$s\ x(s) - x(0) = A\ x(s) + D\ u(s)$$

$$x(s) = \left[s\,I - A\right]^{-1} x(0) + \left[s\,I - A\right]^{-1} D\ u(s) \qquad (4.2)$$

The eigenvalues of the matrix A are thus the pole locations of the transform of the transient response. If, for example, A has only diagonal elements λ_i,

$$\left[s\,I - A\right]^{-1} = \left[\frac{1}{s-\lambda_i}\,\delta_{ij}\right] \quad \text{and} \quad x_i(s) = \frac{x_i(0)}{s-\lambda_i}$$

$$x_i(t) = x_i(0)\; e^{\lambda_i t} \tag{4.3}$$

In this case, the state variables refer to a system which, in Laplace transform terms, has been expanded by partial fractions into a series of simple poles. There is no unique state of a system, since any non-singular linear transformation can be made on a particular set of x. If $x \triangleq T\,y$, substituting in Eq. 4.1 yields

$$T\,\frac{d}{dt}\,y = A\,T\,y$$

$$\frac{d}{dt}\,y = T^{-1}\,A\,T\,y$$

A transformation on A, where $T^{-1}\,A\,T$ becomes a diagonal matrix, is always possible if A has distinct eigenvalues[20]. In this case, the general solution for a free system is, from Eq. 4.3

$$y(t) = \left[e^{\lambda_i t}\,\delta_{ij}\right]\; y(0)$$

$$T^{-1}\,x(t) = \left[e^{\lambda_i t}\,\delta_{ij}\right]\,T^{-1}\,x(0)$$

$$x(t) = T\left[e^{\lambda_i t}\,\delta_{ij}\right]\,T^{-1}\;x(0)$$

Therefore,

$$e^{At} = T\left[e^{\lambda_i t}\,\delta_{ij}\right]\,T^{-1}$$

for any A which has distinct eigenvalues λ_i and which is reduced to diagonal form by T. In the general case, from Eq. 4.2

$$e^{At} = \mathcal{L}^{-1}\;\left[s\,I - A\right]^{-1} \tag{4.4}$$

An alternate way to visualize the concept of state is to integrate and Laplace transform the basic equation, Eq. 4.1, yielding

$$x(s) = \frac{1}{s} x(0) + \frac{1}{s} A x(s) + \frac{1}{s} D u(s) \qquad (4.5)$$

This expression with integrals immediately yields a form suitable for direct mechanization on an analog computer. The number of integrators required would equal the dimension of x. The input to the ith integrator would be

$$\sum_{j=1}^{m} a_{ij} X_j(s) + \sum_{j=1}^{m} d_{ij} U_j(s)$$

and its output would be the ith state variable of the system $x_i(s)$. Relating a state to an output of an integrator lends a particularly clear meaning to this concept.

In summary, the state of a system is the set of numbers which at every instant is sufficient to define the signal level in every energy storage element. In a linear system which is not externally driven, the state trajectory is given by

$$x(t) = e^{At} x(0) \qquad (4.6)$$

4.3 Interpretation of the optimum linear predictor

The mathematical form for a linear predictor, optimum in the mean square sense, was one of the first significant results in random process theory, as presented by Wiener[1] and Kolmogorov[2]. This section will show that this predictor has a very simple interpretation in terms of the generating model for the process. For generality, the multi-dimensional case will be discussed, which of course includes the scalar or 1x1 problem.

Chapter 3 has shown that a random process can always be viewed as a generating matrix $\underline{G(s)}$, excited by a set of unit-valued uncorrelated white noise spectra. The optimum predictor for τ seconds in the future in a random process is given by

$$\underline{W^T(s)} = \left[\underline{G^T(s)}\right]^{-1} \mathcal{L}_+^{-1}\left\{\underline{G^{-1}(-s)}\ \underline{\Phi_{vi}(s)}\right\} \qquad (2.28)$$

where

$$\underline{\Phi}_{vi}(s) = \underline{\Phi}_{vv} \cdot e^{s\tau}\, \underline{I} = e^{s\tau}\, \underline{G(-s)}\, \underline{G^T(s)} \qquad (2.33)$$

After transposing, and substituting in Eq. 2.28

$$W(s) = \mathcal{L}\mathcal{F}^{-1} \left\{ e^{s\tau}\, \underline{G(s)} \right\} \underline{G^{-1}(s)} \qquad (4.7)$$

Figure 4.2 shows the resulting structure. The input white noise vector w is successively transformed into the given random process v,

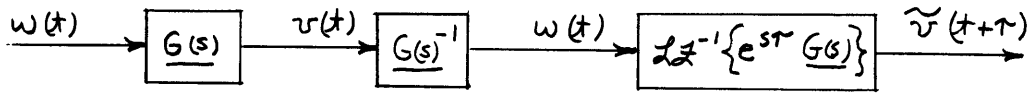


Fig. 4.2 Configuration of an optimum multi-dimensional predictor

back to the original white noise, which then passes through a system given by $\mathcal{L}\mathcal{F}^{-1}\left\{e^{s\tau}\underline{G(s)}\right\}$.

Suppose first, for simplicity, that the ijth element of $\underline{G(s)}$ contains only simple poles.

$$G_{ij}(s) = \sum_{i=1}^{m} \frac{k_i}{s + a_i}$$

This element can be portrayed in flow graph form, as shown in Figure 4.3.

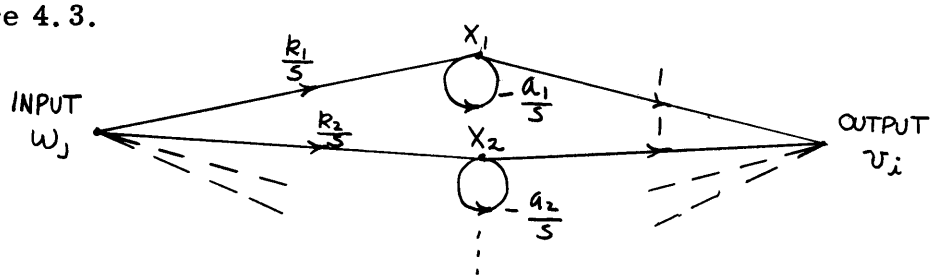


Fig. 4.3 Typical transmissions of the ijth element of G(s).

The set of values for x_i completely defines the state of the system and if white noise excitation should suddenly be cut off at t = 0, $x_i(t)$ would equal $x_i(0)\, e^{-a_i t}$.

The ijth element of $\mathcal{L}\mathcal{F}^{-1}\left\{e^{s\tau}\, \underline{G(s)}\right\}$ is then

$$\mathcal{L}\mathcal{F}^{-1}\left\{ \sum_{i=1}^{m} \frac{k_i\, e^{s\tau}}{s + a_i} \right\} = \sum_{i=1}^{m} \frac{k_i\, e^{-a_i \tau}}{s + a_i}$$

A flow graph of this system is given in Figure 4.4.

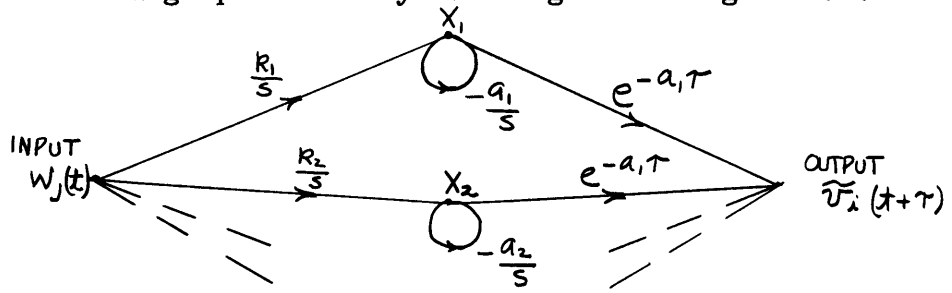


Fig. 4.4 The ijth element of $\mathcal{L}^{-1}\left\{ e^{s\tau} \underline{G(s)} \right\}$

A very significant interpretation can be made from comparison of Fig. 4.3 and Fig. 4.4. $G_{ij}(s)$ and $\mathcal{L}^{-1}\left\{e^{s\tau} G_{ij}(s)\right\}$ have the same excitation, and continually reproduce the same state variables, x_i. The difference is that the output from each first-order system which helps to form the prediction for $v_i(t+\tau)$ is weighted by the value of the unit initial condition response in τ of its own system. That is, just as the present value of v_i is a linear numerical function of the state variables, so is the optimum predictor the same linear function of these state variables after an initial condition decay of τ seconds.

More generally, the best prediction in a mean-square sense of the state of the random process τ seconds in the future is the initial condition response of the generating system from this state. Upon reflection, this seems to be a reasonable result when one views the state at time $t+\tau$ as the sum of the initial condition response from the state at time t and the results of white noise excitation from time t to $t+\tau$, the latter being essentially a zero-mean unknowable response.

The above demonstration included only the case of simple poles. As $G_{ij}(s)$ may contain multiple-poles, it is necessary to verify the decay of the state as contributing to the optimum predictor for this case. In all of linear transient analysis, the case of multiple poles is one handled with considerable difficulty. In the following proof, a canonic flow-graph

configuration will be postulated for a repeated pole transmission. The contribution to the optimum predictor will first be found using the straight-forward $\mathcal{L}^{-1}\left\{e^{s\tau}G_{ij}(s)\right\}$ expression from Figure 4.2. Then, the expression obtained by computing each state variable of $G_{ij}(s)$ and allowing each to decay as an initial condition will be found and manipulated into the same form.

Figure 4.5 shows a canonical configuration for a parallel transmission of $G_{ij}(s)$ involving m cascaded poles at s = - a. This form has

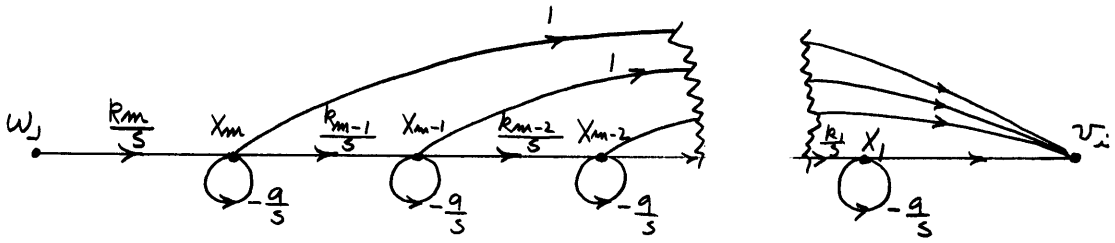


Fig. 4.5 Canonical form for a transmission involving multiple-order poles

internal node variables which are the system state variables.

The transmission from the state variable, x_j, to the output is given by the recurrence relation

$$T_j(s) = \frac{1}{\left(1 + \frac{a}{s}\right)}\left(1 + \frac{k_{j-1}}{s}\, T_{j-1}(s)\right) \quad (j \geq 1)$$

where $k_o \triangleq 0$.

Iterating this relation yields after simplification

$$T_j(s) = \frac{s}{s+a} + \frac{k_{j-1}\, s}{(s+a)^2} + \frac{k_{j-1}\, k_{j-2}\, s}{(s+a)^3} \cdots + \frac{\prod_{l=1}^{i-1} k_{j-l}\, s}{(s+a)^i} \cdots$$

which is also seen by inspection by tracing the paths from node j to the output in Figure 4.4.

The transmission from the input node to the output includes all the repeated pole terms in the partial fraction expansion of $G_{ij}(s)$ and is given by

$$T_{in}(s) = \frac{k_m}{s} T_m(s) = \sum_{i=0}^{m-1} \frac{\prod_{l=0}^{i} k_{m-l}}{(s+a)^{i+1}}$$

It is hypothesized that the contribution of this repeated pole transmission of $G_{ij}(s)$ to the optimum prediction of the ith variable is given by the sum of each of its state variables allowed to decay as initial conditions for τ seconds. Thus the cascaded system of Figure 4.2, which operates on the "recovered" white noise w_j, should supply a transmission from w_j to the r^{th} node, x_r, weight by the numerical factor of the unit initial condition response in τ from node r, and sum over all r. This system must then be equivalent to the result of applying the known solution $\mathcal{L}^{-1}\left\{e^{s\tau} G_{ij}(s)\right\}$ for this multi-pole leg.

The unit initial condition response $I_R(s)$ from node r to output v_i of Figure 4.4 is given by applying a unit step, $\frac{1}{s}$, to the rth node, yielding

$$I_r(s) = \frac{1}{s} T_r(s) = \frac{1}{s+a}\left[1 + \sum_{i=1}^{r-1} \frac{\prod_{l=1}^{i} k_{r-l}}{(s+a)^i}\right]$$

The inverse Laplace transform of $I_R(s)$ is the desired weighting for the rth state variable as a function of τ.

$$i_r(\tau) = e^{-a\tau} + \sum_{i=1}^{r-1} \frac{\prod_{l=1}^{i} k_{r-l}\, \tau^i e^{-a\tau}}{i!}$$

The transmission from input to node r is

$$\frac{\prod_{l=0}^{m-r} k_{m-l}}{(s+a)^{m-r+1}}$$

Therefore, the repeated pole part of the hypothesized optimum predictor is

$$\sum_{r=1}^{m} \frac{\prod_{l=0}^{m-r} k_{m-l}}{(s+a)^{m-r+1}} \left\{e^{-a\tau} + \sum_{i=1}^{r-1} \frac{\prod_{l=1}^{i} k_{r-l}}{i!} \tau^i e^{-a\tau}\right\} \qquad (4.8)$$

which should equal the known result

$$\mathcal{L}^{-1}\left\{e^{s\tau} T_{in}(s)\right\} = \mathcal{L}^{-1}\left\{\sum_{i=0}^{m-1} \frac{\prod_{l=0}^{i} k_{m-l}\; e^{s\tau}}{(s+a)^{i+1}}\right\}$$

The similarity in these two expressions is not staggering. The

quantity $\mathcal{L}^{-1}\left\{e^{s\tau}\, T_{in}(s)\right\}$ will now be manipulated into the form of Eq. 4.8.

$$\mathcal{L}\mathcal{L}^{-1}\left\{e^{s\tau}\, T_{in}(s)\right\} = \mathcal{L}\left\{\mathcal{L}^{-1}\left[e^{s\tau}\, T_{in}(s)\right]\right\}$$

$$= \mathcal{L}\left\{\sum_{i=0}^{m-1} \frac{\prod_{l=0}^{i} k_{m-l}}{i!} (t+\tau)^i\, e^{-a(t+\tau)}\right\}$$

$$= \mathcal{L}\left\{\sum_{i=0}^{m-1} \frac{\prod_{l=0}^{i} k_{m-l}}{i!}\, e^{-a\tau} \sum_{j=0}^{i} \binom{i}{j} t^{i-j}\, \tau^j\, e^{-at}\right\}$$

where $\binom{i}{j}$ is the binomial coefficient, $\dfrac{i!}{(i-j)!\, j!}$,

$$= \sum_{i=0}^{m-1} \prod_{l=0}^{i} k_{m-l}\; e^{-a\tau} \sum_{j=0}^{i} \frac{\tau^j}{j!} \frac{1}{(s+a)^{i-j+1}}$$

Expressing this series in terms of powers of $\dfrac{1}{s+a}$ where $j \triangleq i-p$

$$= \sum_{p=0}^{m-1} \frac{1}{(s+a)^{p+1}} \left\{\sum_{i=p}^{m-1} \frac{\prod_{l=0}^{i} k_{m-l}}{(i-p)!}\, \tau^{i-p}\, e^{-a\tau}\right\}$$

Replacing i by m-r+u and p by m-r,

$$= \sum_{r=m}^{r=1} \frac{1}{(s+a)^{m-r+1}} \left\{\sum_{u=0}^{r-1} \frac{\prod_{l=0}^{m-r+u} k_{m-l}}{u!}\, \tau^u\, e^{-a\tau}\right\}$$

$$= \sum_{r=1}^{m} \frac{\prod_{l=0}^{m-r} k_{m-l}}{(s+a)^{m-r+1}} \left\{\sum_{u=0}^{r-1} \frac{\prod_{l=m-r+1}^{m-r+u} k_{m-l}}{u!}\, \tau^u\, e^{-a\tau}\right\}$$

$$= \sum_{r=1}^{m} \frac{\prod_{l=0}^{m-r} k_{m-l}}{(s+a)^{m-r+1}} \left\{e^{-a\tau} + \sum_{u=1}^{r-1} \frac{\prod_{j=1}^{u} k_{r-j}}{u!}\, \tau^u\, e^{-a\tau}\right\}$$

which is equivalent to Eq. 4.8, completing the proof.

A more elegant proof can be made with the aid of relations developed in Section 4.2, where G(s) is considered a general system with a set of state variables, x, described by the matrix differential equation

$$\frac{d}{dt}\, x = A\, x + D\, w \qquad (4.1)$$

and where the output v is given by $v = R\, x$.

From Eq. 4.2, the input to output transfer function is implicitly given by

$$v = R \left[s I - A\right]^{-1} D\, w$$

It is desired to prove that

$$\mathcal{L}^{-1}\left\{e^{s\tau} R\left[s I - A\right]^{-1} D\right\} = R\ e^{A\tau}\left[s I - A\right]^{-1} D$$

which means that w is operated on by $\left[s I - A\right]^{-1} D$ to produce the current state variable vector, x(s), which is then weighted by its initial condition decay $e^{A\tau}$ and reproduced at the output by R.

$$\mathcal{L}^{-1}\left\{e^{s\tau} R\ \left[s I - A\right]^{-1} D\right\} = \mathcal{L}\left\{R\ e^{A(t+\tau)} D\right\}$$
$$= \mathcal{L}\left\{R\, e^{A\tau} e^{A t} D\right\}$$

according to a property of the matrix exponential proved by Bellman[20].

And, completing the proof, which applies for single and multiple roots alike in single and multi-dimensional systems,

$$\mathcal{L}\left\{R\ e^{A\tau}\ e^{A t} D\right\} = R\ \ e^{A\tau}\left[s I - A\right]^{-1} D$$

In sharp contrast to the arduous multiple-pole derivation made above with involved manipulations with series, the use of the general state equation provided the desired results with a minimum of effort.

Thus it has been proven that the optimum linear predictor for a stationary random process can be regarded in all cases as the result of computing the state of the random process and allowing these state variables to decay as initial conditions in the given model of the process.

The significant feature of a random process is then its state, which summarizes for use in the present and for future prediction all past behavior of the random signal or signals, using a compact number of variables. An expected trajectory of the state variables of the random process, and any system on which it may act, is then defined at every instant by these state variables just as a free determinate system settling to equilibrium is defined by its state variables at a single instant. This allows a wealth of known information concerning the behavior of unforced linear systems to become applicable to systems which are driven by random processes, especially in control applications. Chapter 5 will elaborate on this interesting by-product of the new approach to the repre-

sentation of random processes by the state concept.

The concept of state is only useful if the state variables are recoverable from operations on the random process alone. If the matrix model, G(s), has a realizable inverse, which accounted for most of the difficulty of Chapter 3, this is obviously necessary and sufficient in order to ensure that the state variables can be separately found by a stable system.

Having found that the future value of a random process is given by (1) the sum of present state values decaying as initial conditions, and (2) the response of an "empty" system to future values of white noise, it is now interesting to investigate the knowable properties of this white noise buildup.

The error of the optimum single-dimensional predictor will be solely due to future white noise excitation. Figure 4.6 shows this optimum configuration.

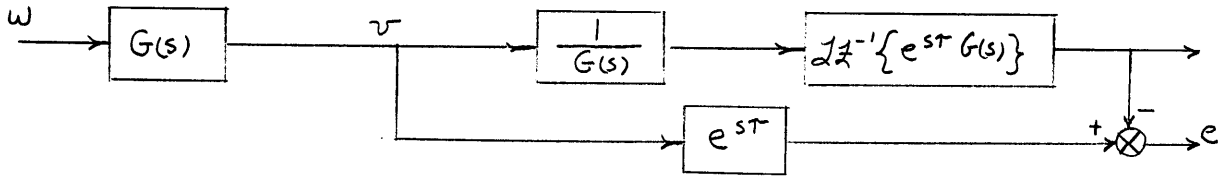


Fig. 4.6 Error configuration for an optimum single-dimensional predictor

The transmission from w to e is

$$H(s) = G(s)\, e^{s\tau} - \mathcal{L}^{-1}\left\{ e^{s\tau} G(s) \right\}$$

Suppose that the impulse response of G(s) is w(t).

$$H(s) = e^{s\tau} \int_0^\tau w(t)\, e^{-st}\, dt$$

$$\Phi_{ee}(s) = H(s)\, H(-s) = \int_0^\tau w(t_1)\, e^{-st_1}\, dt_1 \int_0^\tau w(t_2)\, e^{st_2}\, dt_2$$

$$\overline{e^2} = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} ds\ \Phi_{ee}(s) = \int_0^\tau dt_1 \int_0^\tau dt_2\, w(t_1)\, w(t_2)\, \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} ds\, e^{s(t_2 - t_1)}$$

assuming that the order of the integrations may be changed. But

$$\frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds \quad e^{s(t_2-t_1)} = \quad \mu_o\,(t_2 - t_1)$$

where $\mu_o(t)$ is the unit impulse at $t = 0$.

$$\overline{e^2(\tau)} = \int_0^{\tau} dt_1\; w(t_1) \int_0^{\tau} dt_2 \;\; w(t_2)\;\mu_o(t_2-t_1)$$

$$\overline{e^2(\tau)} = \int_0^{\tau} dt \;\; w^2(t) \qquad (4.9)$$

This general result indicates that the mean-square value of signal level at the output of a linear system, when the white noise is suddenly turned on at $t = 0$, is equal to the integral of the square of the impulse response from the excitation point to the output. As a check,

$$\overline{e^2(\infty)} = \int_0^{\infty} dt \;\; w^2(t) = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds \quad G(-s)\;\; G(s)$$

$$= \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} ds\; \Phi_{vv}(s) = \quad \overline{v^2}$$

Obviously, if more than one uncorrelated white noise source is driving a system, the resulting variance of an output signal is equal to the sum of the variances from each excitation point considered separately. The next section will use this result to motivate a quantitative replacement for the Nyquist sampling theorem.

4.4 A quantitative measure of sampling error for non-bandwidth limited signals

A classic problem in numerical analysis, pulse code modulation, and sampled-data control systems is the loss of information because of representing a continuous signal by a series of evenly-spaced samples. The conventional approach is to utilize the so-called Nyquist Sampling Theorem as given, for example, by Ragazinni and Franklin[23], which states in essence that a signal of absolute bandwidth Ω can be recovered if T, the sampling interval, is less than $\frac{2\pi}{\Omega}$.

In practice, since absolutely bandwidth-limited signals do not occur

in a random process, it is customary to apply a liberal factor of safety on the Sampling Theorem rate for the approximate signal bandwidth. This section will discuss a more basic and quantitative approach which considers the actual average mean-square error inherent in the sampling operation.

Suppose, for convenience, that the continuous random process is generated in the canonic models of Section 4.3, and sampled at the output. At every sampling instant each state variable is summed to form the output. The changes in the state variables at successive sample times arise from two separate effects: (1) The state variables decay as initial conditions for T seconds, and (2) White noise builds up for T seconds.

It is natural to postulate a discrete generating model for the process which has the same state variables as the continuous model at the sampling instants, and whose discrete transition is equivalent to T seconds of continuous initial condition decay. The discrete excitation of each state variable is then a random uncorrelated string of pulses which has the same mean square value as T seconds of white noise buildup to the particular node. In example, suppose a random process is generated as shown in Figure 4.7.

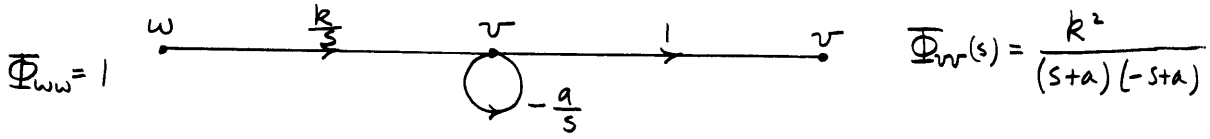


Fig. 4.7 A simple random process generating model

The unit decay of the state variable during a sampling interval is e^{-aT}. The white noise buildup is given by

$$\sigma^2 = \int_0^T dt\ w^2(t) = \int_0^T dt\ (k\, e^{-at})^2 = \frac{k^2}{2a}\ (1 - e^{-2aT})$$

Figure 4.8 shows the discrete model which creates a random process which is hypothesized to produce the same statistics as the sampled process of Fig. 4.7. Here $z(= e^{-sT})$ is a unit delay operator.

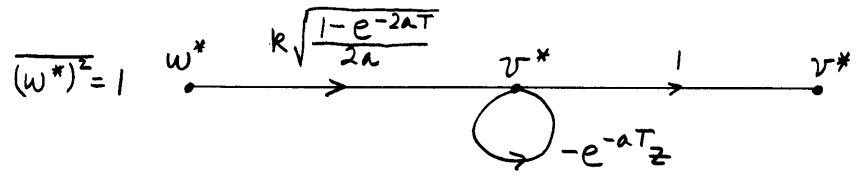


Fig. 4.8 Discrete model derived from Fig. 4.7

The power density spectrum of v*, realizing that $z = e^{-sT}$,

$$\Phi_{v^*v^*}(z) = \Phi_{ww}(z)\ \frac{k^2}{2a}\ (1 - e^{-2aT})\frac{1}{(1 + e^{-aT}z)} \cdot \frac{1}{(1 + e^{-aT}z^{-1})}$$

where $\Phi_{ww}(z) = 1$. (See Appendix II.)

Considering Fig. 4.7,

$$\varphi_{vv}(\tau) = \frac{k^2}{2a}\ e^{-a|\tau|}$$

$$\varphi_{v^*v^*}(nT) = \frac{k^2}{2a}\ e^{-a|n|T}$$

$$\Phi_{v^*v^*}(z) = \frac{k^2}{2a}\left[\frac{1}{1 + e^{-aT}z} + \frac{1}{1 + e^{-aT}z^{-1}} - 1\right]$$

$$= \frac{k^2}{2a}\ \frac{(1 - e^{-2aT})}{(1 + e^{-aT}z)(1 + e^{-aT}z^{-1})}$$

This example has illustrated the relation between discrete and continuous models for random processes, showing that the same discrete power density spectrum is obtained from considering either white-noise buildup over the sampling interval or through straightforward z-transform techniques.

The best estimate of the continuous variable v(nT+t) from its samples v(nT) is $v(nT)\ e^{-at}$ for $t < T$ since the future effect of the white noise cannot be predicted. In the general case, the best estimate has the current state variables decaying as initial conditions until the next set of state variables is computed. In analogy to the continuous case, a suitable inverse filter can always be found to recover these state variables if the continuous model is inverse realizable.

The reconstructed error of the random process is the difference

between the actual value between sampling instants and the initial condition decay -- in other words, the amount of white noise buildup at the output of the generating model over the sampling interval. This irreducible error is the fundamental penalty for representing a random process in terms of its samples.

The result of this discussion is that, from Eq. 4.9, the mean square error between sampling intervals is $\int_0^T dt\ w^2(t)$, where w(t) is the model impulse response. The average error is thus $\frac{1}{T}\int_0^T d\tau \int_0^\tau dt\ w^2(t)$. It is now proposed that a useful measure of the error due to sampling is the fractional error power, or the ratio of the mean square error to the mean square signal level

$$\text{F.E.P.} \triangleq \frac{\frac{1}{T}\int_0^T d\tau \int_0^\tau dt\ w^2(t)}{\int_0^\infty dt\ w^2(t) = v^2} \qquad (4.10)$$

This provides a quantitative measure of the inherent penalty for sampling any random process, regardless of the spectrum shape. An example will illustrate the utility of this approach.

Suppose the continuous model for an observed random process, v, is given by

$$G(s) = \frac{1}{(s+3)(s+4)}$$

$$w(t) = e^{-3t} - e^{-4t}$$

$$\overline{e^2} = \frac{1}{T}\int_0^T d\tau \int_0^\tau dt\ w^2(t) = \frac{1}{168} + \frac{2}{49\,T}\ (1 - e^{-7T}) - \frac{1}{36\,T}(1 - e^{-6T}) - \frac{1}{64\,T}\ (1 - e^{-8T})$$

In this form it is difficult to obtain the average square error for small T, and especially to solve for a T to meet a certain fraction of the mean square signal level. An alternate route is to expand G(s) in ascending powers of $\frac{1}{s}$

$$G(s) = \frac{\frac{1}{s^2}}{1 + \frac{7}{s} + \frac{12}{s^2}} = \frac{1}{s^2} - \frac{7}{s^3} \quad - - .$$

$$w(t) = t - \frac{7}{2} \quad t^2 \ . \ . \ .$$

$$w^2(t) = \quad t^2 - 7\,t^3 \ . \ . \ .$$

$$\overline{e^2} = \frac{1}{T} \int_0^T dt\ w^2(t) = \frac{T^3}{12} - \frac{7}{20} \quad T^4 \ . \ . \ .$$

$$FEP = \frac{\overline{e^2}}{\varphi_{vv}(0)} = \frac{\frac{T^3}{12} - \frac{7}{20} \quad T^4 \ . \ . \ .}{\frac{1}{168}} = 14\,T^3 - 58.8\,T^4 \ . \ . \ .$$

If the FEP is specified to be .01, an approximate value for T is given by

$$T \approx \left(\frac{.01}{14}\right)^{1/3} = \quad .089 \quad \text{Seconds}$$

This section has used the concept of white-noise buildup (1) to show the mechanism by which sampling of a random process always degrades knowledge of the signal, and (2) to present a quantitative measure of this error from which a rational decision can be made for a proper sampling interval.

4.5 New results and interpretations for the optimum filtering problem

A physical system which operates on a given random process can be viewed as a means of continuously extracting all possible information about future values of error from present values of input signals. An optimum system should result in an error signal e which is on the average unpredictable from and unrelated to past values of input signal v. In a linear statistical theory, this lack of relation can only be measured by a correlation function, which means that

$$E\left\{v_i(t-\tau)\ e_j(t)\right\} = \varphi_{v_i e_j}(\tau) = 0 \quad (\tau \geq 0)$$

$$(i, j = 1, 2 \ . \ . \ . \ n)$$

for a random process with n inputs, under this requirement. Accordingly,

$$\mathcal{L}\left[\underline{\varphi_{v_i e_j}(\tau)}\right]_{\tau \geq 0} = \mathcal{L}\mathcal{L}^{-1}\left\{\underline{\Phi_{ve}(s)}\right\} = \underline{0}$$

But $e(s)\rceil = i(s)\rceil - \underline{W(s)}\ v(s)\rceil$ where $\underline{W(s)}$ is the optimum system to be found, and $i(s)\rceil$ is the desired output vector. From Eq. 2.32,

$$\underline{\Phi}_{ve}(s) = \underline{\Phi}_{vi}(s) - \underline{\Phi}_{vv}(s)\ W^T(s)$$

Therefore,

$$\mathcal{L}\mathcal{L}^{-1}\left\{\underline{\Phi}_{vv}(s)\ \underline{W^T(s)}\right\} = \mathcal{L}\mathcal{L}^{-1}\left\{\underline{\Phi}_{vi}(s)\right\} \qquad (2.18)$$

which is an implicit statement of the optimum multi-dimensional system, which was obtained with considerable more difficulty (and perhaps more rigor) in Chapter 2 by an alternate route.

By either method, the basic statement of optimality of realizable linear systems is then

$$\mathcal{L}\mathcal{L}^{-1}\left\{\underline{\Phi}_{ve}(s)\right\} = 0 \qquad (4.11)$$

This result will be used to motivate a closer look at the properties of optimum single-dimensional systems. In particular, the filtering problem will be examined and an optimum unity feedback system will be derived which takes advantage of some not readily apparent properties of the standard mathematical solution given by

$$W(s) = \frac{1}{\Phi^{+}_{vv}(s)}\ \mathcal{L}\mathcal{L}^{-1}\left\{\frac{\Phi_{vi}(s)}{\Phi^{-}_{vv}(s)}\right\} \qquad (2.15)$$

Figure 4.9 shows the basic configuration to be examined. The following restrictions apply: (1) The signal s is derived from unit density

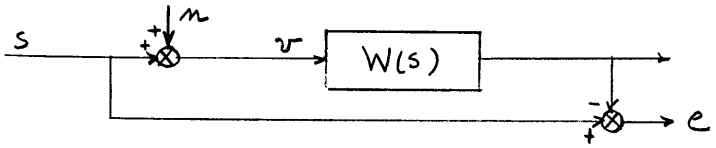


Figure 4.9 The basic filtering problem

white noise passing through a linear system $G_s(s)$, (2) The noise n is derived from unit density white noise, uncorrelated with the signal white noise, passing through a linear system $G_n(s)$, (3) The signal s is the desired quantity to be reproduced at the output of W(s), and (4) W(s) is to be a unity feedback system, with forward transference H(s), such that $W(s) = \frac{H(s)}{1+H(s)}$.

This model is of sufficient generality to include many filtering and control problems of practical interest, and its solution will later motivate a completely general solution.

From Eq. 2.9,

$$\Phi_{ve}(s) = \Phi_{ss}(s)\left[1 - W(s)\right] - \Phi_{nn}(s)\ W(s)$$

$$= \Phi_{ss}(s)\ \frac{1}{1 + H(s)} - \Phi_{nn}(s)\ \frac{H(s)}{1 + H(s)}$$

Hence, from the basic equation, Eq. 4.11,

$$\mathcal{L}^{-1}\left\{\Phi_{ss}(s)\ \frac{1}{1 + H(s)}\right\} = \mathcal{L}^{-1}\left\{\Phi_{nn}(s)\ \frac{H(s)}{1 + H(s)}\right\} \quad (4.12)$$

Two very important facts are revealed from this equality. Since the positive poles of $\Phi_{ss}(s)$ do not generally equal the positive poles of $\Phi_{nn}(s)$, this equation will only hold in general when (1) the poles of H(s), which are the zeroes of $\frac{1}{1 + H(s)}$, include all the positive poles of $\Phi_{ss}(s)$, and (2) the zeroes of H(s), which are the zeroes of $\frac{H(s)}{1 + H(s)}$, include all the positive poles of $\Phi_{nn}(s)$. If this were not so, then in the $\mathcal{L}^{-1}$ partial fraction expansion of both sides, there could not be pole-by-pole equality. Let

$$H(s) = \frac{N_p^+(s)}{S_p^+(s)}\ H^1(s)$$

where $N_p^+(s)$ and $S_p^+(s)$ are the LHP poles of $\Phi_{nn}(s)$ and $\Phi_{ss}(s)$, respectively, and $H^1(s)$ is an additional term which does not cancel any of the signal or noise pole terms.

The optimum system is, from Eq. 2.15,

$$\frac{S_p^+(s)\, N_p^+(s)}{V_z^+(s)} \left\{ \mathcal{L}\mathcal{I}^{-1} \left\{ \frac{\Phi_{v_i}(s) = \Phi_{ss}(s)}{\Phi_{vv}^-(-s)} \right\} \triangleq \frac{U(s)}{S_p^+(s)} \right\}$$

where $V_z^+(s)$ equals the LHP zeroes of $\Phi_{vv}(s)$.

Equating this to $\frac{H(s)}{1 + H(s)}$ and solving,

$$H(s) = \frac{N_p^+(s)\, U(s)}{V_z^+(s) - N_p^+ U(s)}$$

Although this is not obvious by inspection, the polynomial $S_p^+(s)$ must be a factor of $V_z^+(s) - N_p^+(s)\, U(s)$ in order that Eq. 4.12 be satisfied, according to previous arguments.

$$H(s) = \frac{1}{\frac{V_z^+(s)}{N_p^+(s)\, S_p^+(s)} - \frac{U(s)}{S_p^+(s)}} \quad \frac{U(s)}{S_p^+(s)} = \frac{\mathcal{L}\mathcal{I}^{-1}\left\{ \frac{\Phi_{ss}(s)}{\Phi_{vv}^-(-s)} \right\}}{\Phi_{vv}^+(s) - \mathcal{L}\mathcal{I}^{-1}\left\{ \frac{\Phi_{ss}(s)}{\Phi_{vv}^-(-s)} \right\}}$$

This leads to the interesting conclusion that $\mathcal{L}\mathcal{I}^{-1}\left\{ \frac{\Phi_{ss}(s)}{\Phi_{vv}^+(s)} \right\}$, which contains only the signal poles, is equal in this case to the sum of signal poles in a partial fraction expansion of $\Phi_{vv}^+(s)$, since no cancellation of $S_p^+(s)$ is allowed. A more general proof of this important identity will be made later in this section.

Therefore,

$$H(s) = \frac{\sum \text{ Signal Poles of } \Phi_{vv}^+(s)}{\sum \text{ Noise Poles of } \Phi_{vv}^-(s)} \triangleq \frac{S(s)}{N(s)} \qquad (4.13)$$

$$W(s) = \frac{H(s)}{1 + H(s)} = \frac{S(s)}{S(s) + N(s)} \qquad (4.14)$$

This result is of considerable practical and theoretical interest and applies to all single-dimensional filtering problems, when noise and signal are uncorrelated. The optimum system determined, W(s), has the following significance:

The best estimate of an input signal under a mean square error criterion is that the signal originated from signal poles of a single system, with transfer function $\Phi_{vv}^+(s)$ and excited by unit-density white noise.

The optimum system then merely determines and sums the canonic state variables of the signal portion of the random process generating model. The optimum predictor in this noisy case is intuitively the result of allowing these instantaneous state variables to decay as initial conditions for the desired τ seconds. This is verified by noting that, where the signal poles of the generating model are $\mathcal{LL}^{-1}\left\{\frac{\Phi_{ss}(s)}{\Phi^{+}_{vv}(s)}\right\} = \sum_i \frac{k_i}{s+p_i}$, the optimum predictor is given by

$$\frac{1}{\Phi^{+}_{vv}(s)}\,\mathcal{LL}^{-1}\left\{\frac{\Phi_{ss}(s)\,e^{s\tau}}{\Phi^{+}_{vv}(s)}\right\} = \frac{1}{\Phi^{+}_{vv}(s)}\,\frac{k_i\,e^{-p_i\tau}}{s+p_i}$$

which computes and weights state variables for τ seconds of initial condition decay.

The above simple interpretation of an optimum system was obtained through rather a roundabout method, and holds only for uncorrelated signal and noise and a one-dimensional random process. But having this result, it becomes simple to extend it to the general multi-dimensional filtering and prediction problem with all possible correlations existing between signals and noise.

The basic equation defining the optimum multi-dimensional system is the transpose of Eq. 2.28

$$\underline{W(s)} = \mathcal{LL}^{-1}\left\{\underline{\Phi}^{T}_{vi}(s)\,\left[G^{T}(-s)\right]^{-1}\right\}\underline{G^{-1}(s)} \qquad (2.28)$$

Also

$$\underline{\Phi}^{T}_{vi}(s) = \underline{G_d(s)}\;\underline{\Phi}^{T}_{ss}(s) + \underline{G_d(s)}\cdot\underline{\Phi}^{T}_{ns}(s) \qquad (2.33)$$

The perfect operation on the signal, $\underline{G_d(s)}$, is $\underline{I}$ for the filtering problem. Thus,

$$\underline{W(s)} = \mathcal{LL}^{-1}\left\{\underline{\Phi}^{T}_{ss}(s)\,\left[G^{T}(-s)\right]^{-1} + \underline{\Phi}^{T}_{ns}(s)\,\left[G^{T}(-s)\right]^{-1}\right\}\underline{G^{-1}(s)}$$

In analogy with the simple case discussed earlier, it is desired now to prove the $\mathcal{LL}^{-1}$ term is merely the result of expanding each element of $\underline{G(s)}$ in partial fractions and retaining only those with signal poles.

First it is necessary to identify the signal poles. Figure 4.10 shows a multi-dimensional model for the formation of correlated signals and noise.

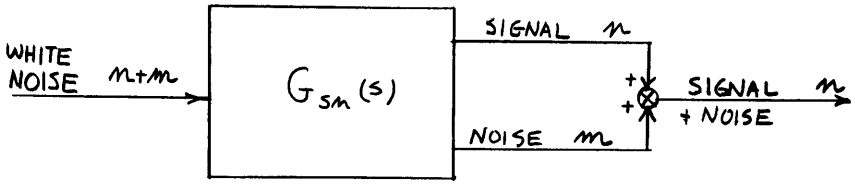


Fig. 4.10 A hypothetical model for the creation of correlated signals and noise

Given the auto and cross power density spectra of the signal and noise vectors, where the noise vector can be of less dimension than the signal, a (n+m) x (n+m) realizable and inverse realizable matrix filter $\underline{G_{sn}(s)}$ can always be found which can reproduce the observed statistics of the separate signal and noise components. The poles of the ith signal are the poles of the ith row of $\underline{G_{sn}(s)}$.

The matrix of power density spectra of the signal and noise signals is given by $\underline{G_{sn}(-s)}\ \underline{G_{sn}^T(s)}$ which in partitioned form is

$$\underline{G_{sn}(-s)}\ \underline{G_{sn}^T(s)} = \left[\begin{array}{c|c} \underline{\Phi}_{ss}(s) & \underline{\Phi}_{sn}(s) \\ \hline \underline{\Phi}_{ns}(s) & \underline{\Phi}_{nn}(s) \end{array}\right]$$

The positive poles of the ith row of $\underline{G_{sn}(s)}$ appear only in the ith column of the above partitioned matrix. That is, the positive signal poles appear only in the sub-matrices $\underline{\Phi_{ss}(s)}$ and $\underline{\Phi_{ns}(s)}$, and the positive noise poles only appear in $\underline{\Phi_{sn}(s)}$ and $\underline{\Phi_{nn}(s)}$.

Considering the observed random process v, where v = s + n and zero elements are permissible in n

$$\underline{\Phi}_{vv}(s) = \underline{\Phi}_{ss}(s) + \underline{\Phi}_{ns}(s) + \underline{\Phi}_{sn}(s) + \underline{\Phi}_{nn}(s) \qquad (2.31)$$

$$\underline{\Phi}_{vv}(s) = \underline{G(-s)}\ \underline{G^T(s)} \qquad (2.27)$$

$$\underline{G^T(s)} = \underline{G^{-1}(-s)}\left\{\underline{\Phi}_{ss}(s) + \underline{\Phi}_{ns}(s)\right\} + \underline{G^{-1}(-s)}\left\{\underline{\Phi}_{sn}(s) + \underline{\Phi}_{nn}(s)\right\}$$

$$\underline{G(s)} = \left\{\underline{\Phi}_{ss}^{(s)\,T} + \underline{\Phi}_{ns}^{(s)\,T}\right\}\left[\underline{G^{-1}(-s)}\right]^{T} + \left\{\underline{\Phi}_{sn}^{T}(s) + \underline{\Phi}_{nn}^{T}(s)\right\}\left[\underline{G^{-1}(-s)}\right]^{T}$$

But $\underline{G(s)}$ has no RHP poles.

$$\mathcal{LT}^{-1}\left\{\underline{G(s)}\right\} = \underline{G(s)} = \mathcal{LT}^{-1}\left\{\left[\underline{\Phi}_{ss}^{T}(s) + \underline{\Phi}_{ns}^{T}(s)\right]\left[\underline{G^{-1}(-s)}\right]^{T}\right\}$$
$$+\mathcal{LT}^{-1}\left\{\left[\underline{\Phi}_{sn}^{T}(s) + \underline{\Phi}_{nn}^{T}(s)\right]\left[\underline{G^{-1}(-s)}\right]^{T}\right\}$$

Since the first and second bracketed terms above have only positive signal and noise poles, respectively, they are immediately identified as the separate signal and noise terms in a partial fraction expansion of $\underline{G(s)}$, which is the desired proof.

Let $\underline{G(s)} = \underline{S(s)} + \underline{N(s)}$, where all the signal and noise poles are grouped together in $\underline{S(s)}$ and $\underline{N(s)}$, respectively. Of course, if one or more signal poles are identical to a noise pole, the contribution of these signal poles to $\underline{S(s)}$ would be obtained through their separate partial fraction expansion in $\mathcal{LT}^{-1}\left\{\underline{G^{-1}(-s)}\ \underline{\Phi}_{ss}(s) + \underline{G^{-1}(-s)}\ \underline{\Phi}_{ns}(s)\right\}^{T}$, since they could not be separated in a partial fraction expansion of $\underline{G(s)}$. The optimum filter is then, from Eq. 2.28,

$$\underline{W(s)} = \underline{S(s)}\ \left[\underline{S(s) + N(s)}\right]^{-1} \qquad (4.16)$$

A unity feedback system is readily seen to have a forward loop transmission

$$\underline{H(s)} = \underline{S(s)}\ \underline{N(s)}^{-1} \qquad (4.17)$$

and has the appearance of Fig. 4.11.

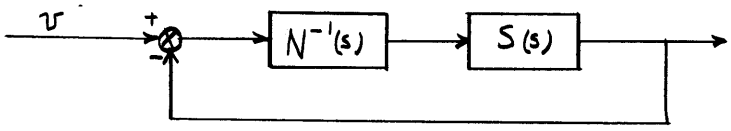


Fig. 4.11 A canonic optimal multi-dimensional filter

Fig. 4.11 is invalid if $\underline{N^{-1}(s)}$ is singular, which would be the case if one or more of the input signals is uncorrupted by noise. In this case, the canonic configuration of Fig. 4.12 is still applicable, providing

the trivial restriction of signal having to be present in all input components of v is satisfied.

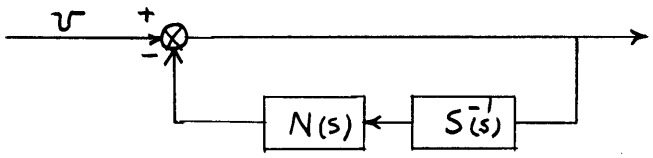


Fig. 4.12 An alternate optimal multi-dimensional filter

These optimal configurations have an interesting interpretation as systems which compute inner signal levels of an effective random process generating model, G(s) = S(s) + N(s). As shown in Figure 4.13, the optimal configurations merely act to reproduce quantities which exist at the input and outputs of the signal and noise portions of the model.

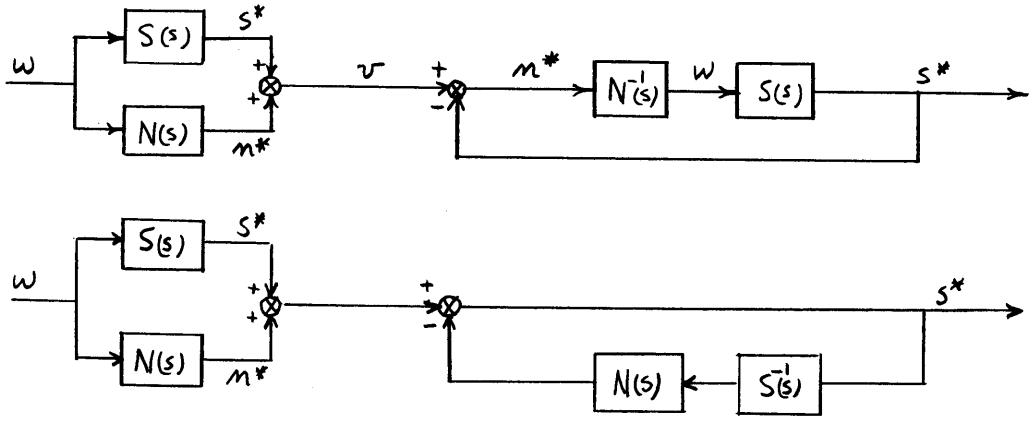


Fig. 4.13 Signal reproduction in optimum configuration

Kalman and Bucy[24] recently presented an approach to the optimum filtering problem which considered the special case of pure white noise corrupting all input signals, with no cross-correlation between signal or noise. They postulated a model of the original signal generating model which appeared in the forward path of a unity feedback system. In the light of the above analysis, it is easy to see why they were unable to extend their results, since, from Fig. 4.11, the model which should have been specified is the signal generating portion S(s) of the hypothetical model G(s) -- which creates the actual signal observed and not the pure

signal component.

The results of this section are particularly important, both in understanding and in operating on random processes with linear systems. In essence, it has been shown that the physical system found from factoring a matrix of input power density spectra contains in its signal levels all the knowable information about the random process which can be obtained by linear measurement of the random process. The optimum system has the simple form $\underline{S(s)}\left[\underline{S(s)} + \underline{N(s)}\right]^{-1}$, where $\underline{S(s)}$ contains all the signal poles (positive poles of $\underline{\Phi_{ss}(s)}$ and $\underline{\Phi_{ns}(s)}$) in a partial fraction expansion--element by element -- of $\underline{G(s)}$, the effective generating system. Figures 4.11 and 4.12 show canonic forms for optimum feedback systems to filter the multi-dimensional random process.

4.6 Correlation functions and initial condition responses

Auto and cross-correlation functions have an appearance similar to the dynamic behavior of linear systems, usually decaying to zero exponentially as τ becomes very large. This section will relate the correlation functions to the initial condition response of the white-noise driven linear model for the random process with an equation of considerable simplicity and generality.

First, suppose that the cross-correlation function $\varphi_{x_i x_j}(\tau)$ is known between two state variables, x_i and x_j, that are defined in a linear system by the general equation

$$\frac{d}{dt} \quad x = A\,x + D\,w \qquad (4.1)$$

where x is the n-dimensional state vector, and w is a r-dimensional white noise vector.

Since from Eq. 2.9,

$$\Phi_{x_i \dot{x}_j}(s) = s\,\Phi_{x_i x_j}(s)$$

$$\varphi_{x_i \dot{x}_j}(\tau) = \frac{d}{d\tau}\,\varphi_{x_i x_j}(\tau) = E\left\{x_i(t) \cdot \dot{x}_j(t+\tau)\right\}$$

But

$$\dot{x}_j(t) = \sum_{k=1}^{n} a_{jk}\, x_k(t) + \sum_{k=1}^{r} d_{jk}\, w_k(t)$$

$$\varphi_{x_i \dot{x}_j}(\tau) = \sum_{k=1}^{n} a_{jk}\, E\left\{x_i(t) \cdot x_k(t+\tau)\right\} + \sum_{k=1}^{m} d_{jk}\, E\left\{x_i(t)\, w_k(t+\tau)\right\}$$

$$E\left\{x_i(t) \cdot w_k(t+\tau)\right\} = 0 \qquad (\tau > 0)$$

since future values of white noise are not causally related (ie: correlated) to present values of system signal level (or, more formally, since $\Phi_{x_i w_k}(s)$ contains only RHP poles). Thus,

$$\frac{d}{d\tau}\, \varphi_{x_i x_j}(\tau) = \sum_{k=1}^{n} a_{jk}\, \varphi_{x_i x_k}(\tau) \qquad (\tau > 0)$$

Writing this equation in matrix notation,

$$\frac{d}{d\tau}\, \underline{\varphi_{xx}(\tau)} = \underline{\varphi_{xx}(\tau)}\ \underline{A^T} \qquad (\tau > 0)$$

Transforming,

$$s\ \mathcal{L}^{-1}\left\{\underline{\Phi_{xx}(s)}\right\} - \underline{\varphi_{xx}(0)} = \mathcal{L}^{-1}\left\{\underline{\Phi_{xx}(s)}\right\}\ \underline{A^T}$$

$$\mathcal{L}^{-1}\ \underline{\Phi_{xx}(s)} = \underline{\varphi_{xx}(0)}\ \left[\underline{I}\, s - \underline{A}^T\right]^{-1}$$

But, from Eq. 4.4

$$\left[I\, s - A\right]^{-1} = \mathcal{L}\left[e^{At}\right]$$

$$\underline{\varphi_{xx}(\tau)} = \underline{\varphi_{xx}(0)}\ \left[e^{A\tau}\right]^T \qquad (\tau > 0)$$

Transposing,

$$\underline{\varphi_{xx}{}^T(\tau)} = e^{A\tau}\ \underline{\varphi_{xx}{}^T(0)} \qquad (\tau > 0) \qquad (4.18)$$

This is the desired general relationship, which shows that the n^2 correlations between state variables in a linear system are mapped through time by the same transformation that governs the decay of the state variables in the linear model: $x(t) = e^{At}\, x(0)$.

Now, it remains to use this result in order to show the meaning of the correlation functions which would be measured at the outputs or

output of the random process. Suppose that the r-fold output vector v is obtained through multiplication of the state vector x by a rxn matrix $\underline{R}$

$$v_i(t) = \sum_{k=1}^{n} r_{ik}\, x_k(t)$$

The cross-correlation function of two output signals is thus

$$\varphi_{v_i v_j}(\tau) = \sum_{k=1}^{n} \sum_{l=1}^{n} r_{ik}\, r_{jl}\, \varphi_{x_k x_l}(\tau)$$

Or in matrix notation

$$\underline{\varphi_{vv}(\tau)} = \underline{R}\ \ \underline{\varphi_{xx}(\tau)}\ \ \underline{R}^T$$

For $\tau > 0$, using Eq. 4.18

$$\underline{\varphi_{vv}(\tau)} = \underline{R}\ \ \underline{\varphi_{xx}(0)}\ \ \underline{\left[e^{A\tau}\right]^T}\ \ \underline{R}^T$$

Transposing,

$$\underline{\varphi^T_{vv}(\tau)} = \underline{R}\ \ \underline{e^{A\tau}}\ \ \underline{\left[R\ \ \varphi_{xx}(0)\right]^T} \qquad (\tau > 0)$$

Since $\varphi_{v_i x_j}(\tau) = \sum_{k=1}^{n} r_{ik}\, \varphi_{x_k x_j}(\tau)$ or

$$\underline{\varphi_{vx}(\tau)} = \underline{R}\ \ \underline{\varphi_{xx}(\tau)}$$

then

$$\underline{\varphi^T_{vv}(\tau)} = \underline{R}\ \ \underline{e^{A\tau}}\ \ \underline{\varphi^T_{vx}(0)} = \underline{R}\ \ \underline{e^{A\tau}}\ \ \underline{\varphi^T_{vx}(0)} \quad (\tau > 0) \quad (4.19)$$

This equation is in proper form to permit interpretation of the output correlation functions. The initial condition response of the system, viewed at the output, is

$$v(t) = \underline{R}\ \ \underline{e^{At}}\ \ x(0)$$

Therefore, if the vector x(0) is set equal in the model to $\varphi_{x v_i}(0) = \varphi_{v_i x}(0)$ then the transient observed at the jth output terminal will be $\varphi_{v_i v_j}(\tau)$. In words, this means that the cross (or auto) correlation function, $\varphi_{v_i v_j}(\tau)$ between two signals in a random process is the transient which would be observed at the jth signal location when each of the system state variables, x_k, is initially set to $\varphi_{x_k v_i}(0)$ and the system released.

This result tends (1) to re-emphasize the basic nature of the hypothetical model which is capable of generating a given random process, and (2) to interpret the correlation function as a transient of this model.

4.7 Advantages of the state and model approach to random processes

This chapter has been written in the hope of altering current ways of approaching the visualization and study of random processes by the presentation of a simple explanation for the mathematically-complex results of contemporary theory. In a sense, the basic question is whether one should look at what a system does or whether one should look at what a system is.

It was necessary to first ensure that such a system can always be found from auto and cross-correlation functions of a multi-dimensional random process. This was the contribution of Chapter 3. With this assurance, the conventional Wiener theory could be reworked with complete generality.

Section 4.3 considered the optimum predictor configuration. It was shown that this problem is only a matter of continuously measuring the state variables and weighting them by their initial condition decay for τ seconds.

Section 4.5 dealt with the problem of filtering extraneous noise from a desired signal. In this case it was shown that the equivalent generating model was actually two systems in parallel, one associated with the signal and the other with noise. The optimum filter merely computed the output of the signal portion. With the recognition of this simple interpretation, two general canonic feedback arrangements were found which should be of considerable interest in control systems design.

In section 4.4 a quantitative measure of error due to sampling of a random process was presented. This was determined from the buildup

of white noise between sampling instants in the model.

Section 4.6 showed that correlation functions can be regarded as transient behavior of the effective model under certain initial conditions.

In all these results, the ideas of white-noise excited system and system state play the dominant role. "State" and "system" are far more general terms, however, then their use here would indicate. It is interesting to conjecture at this point how these concepts might aid the study of non-stationary and non-linear random processes.

First, in the case of non-stationary random processes it seems highly probable that the conceptual results derived in this chapter remain valid, providing that the effective linear time-varying model for the generation of the process is known or can be found. The optimum predictor could still neglect future values of white noise and use only present values of system state, but of course in this non-stationary case the initial condition decay would no longer be described with the matrix exponential. Also, the case of finding a time-varying inverse of the effective generating model in order to recover the state variables appears possible if extremely difficult. Further promise in this respect is lent by recent work by Kalman and Bucy[24] who have derived an optimum time-varying system which remains similar in form to the stationary case.

In the case of so-called non-linear random processes, which are distinguished by decidedly non-Gaussian probability distributions, it is appealing to hypothesize that they occur as the result of independent white noise driving a suitable non-linear system. Further, from current work in this field, for example by Bose[25], it appears possible that such a non-linear system might be a finite-state linear system driving a memoryless non-linear function generator. This is an interesting alternate approach to the study of non-linear random processes which is more appealing to the engineer than the more general and highly-mathematical

treatment of, for example, Wiener[26].

In short, it is hoped that the simple physical interpretation of the optimum linear systems presented in this chapter for a stationary random process will motivate a similar approach to more complex stochastic problems.

CHAPTER V.

RANDOM PROCESSES AND AUTOMATIC CONTROL

5.1 Introduction

Stationary random processes have been examined in the previous chapters with an eye toward delineating the recoverable information which exists as a result of optimum linear operations on the signals. The concept of a generating model, excited by white noise and possessing state variables, has been shown to be a particularly effective way to visualize the action of optimum systems -- that they perform essentially a measurement or signal recovery of certain quantities in the generating model.

The time has now come, however, to consider how this increased intuitive understanding of random processes can be of help when control decisions must be formulated as a result of the information received. The general control problem is of great interest to mathematicians and engineers alike, and most significant control problems involve signals, wanted and unwanted, which are random in nature. In this chapter we restrict attention to the following situation:

A fixed linear system exists whose output is to be forced to follow a stationary random input signal, which in the limiting case of a regulator is constant. Corruption of the command signal with noise is allowable. Also, load disturbances may be present which are stationary random processes uncorrelated with the input signals. Finally, the controller configuration is completely arbitrary as to the possible use of linear and non-linear elements, with the single important limitation that the controller output signal which drives the fixed system be limited in amplitude to correspond to the saturation level existing in the controlled system.

Section 5.2 considers the scalar problem and develops a design philosophy which appears to have considerable promise in the optimum

control of saturating systems. The particular problem of load disturbance in linear and saturating systems is treated in Section 5.3. With this foundation, contemporary approaches to full-throw control which can be found in the literature are critically analyzed in Section 5.4. Finally, Section 5.5 presents an extension to the determinate Second Method of Lyapunov to include random processes. This leads to a design procedure suitable for a multi-dimensional saturating control system, optimizing a quadratic error criterion.

In the past chapters general equations, simple proofs, and sweeping statements could be presented with mathematical aplomb because of the simplicity and power of linear methods of analysis. But in this chapter the spectre of saturation has arisen to confound our linear theory and the whole tenor of this thesis must change. No longer can general quantitative statements be made concerning system behavior; it is difficult enough to make useful qualitative observations. We must be content with small nibbles at this frontier of control theory and recognize that the verification of original ideas can only come with computer analysis and can only be valid for the specific cases investigated.

5.2 Saturation and control in a stochastic environment

It is profitable to consider again the optimum unity feedback configuration derived in section 4.5 for the recovery of one-dimensional signal from noise. This is shown in Fig. 5.1 where the input signal v is composed of two hypothetical components, signal s* and noise n*, which

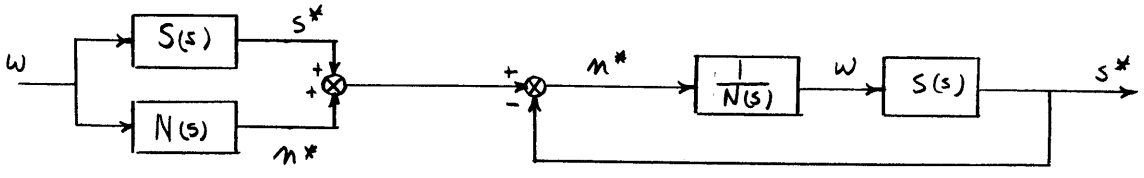


Fig. 5.1 Optimum filter configuration

are the best estimates of the actual signal and noise in a mean-square sense. This minimization of mean-square error means that s* is the expected value of the actual signal component, conditioned on a physically-realizable linear recovery. In the system depicted in Fig. 5.1, the expected value of error at every instant is equal to zero, since the output is the expected value of signal. Now, from section 4.3 it is known that the expected future value of signal in a linear system excited by white noise is derived from the decay of the state variables. Applying this fact to the optimum filter, it is seen that at every instant the expected value of error is zero for all future time because the output element S(s) has the same state variables as S(s) in the generating model and they both are not further excited (as w remains zero in both configurations). Therefore, an alternate statement of optimality in the linear filtering problem is that the expected value of all future error be zero at every instant. With this interpretation, the use of a mean-square error criterion is seen not to lend much emphasis to the squared-error per se, but rather it acts as a mechanism for reproducing expected values.

The reason for the emphasis on the particular use of a mean-square error criterion in the linear theory is that when saturation occurs in practical output equipment it does not necessarily mean that the optimum non-linear control system must be designed on a mean-square error basis to be consistent with linear random process theory. In other words, the random process generating models emphasized in this work contain internal signals which should be the recovery goals of a non-linear saturation-limited practical control system, but the measure of error in recovery is entirely at the discretion of the designer.

Since the optimum linear system is constructed so as to make the expected value of error zero for all future time, a logical choice for a saturating design criterion should obviously involve this expected future error, which is, of course, the best information available at any instant for future use. A convenient way of decomposing this future value of error

is to consider the initial condition decay of random process and fixed system state variables as one component, designated $e(\tau)$, and the response of an otherwise "empty" fixed system to the future output of the controller, $c(\tau)$, as the other. With this division, the job of the controller at any instant is to formulate and execute the initial action of a plan that will make $c(\tau)$ equal to $e(\tau)$ as rapidly and efficiently as possible -- a pursuit problem.

It is important that this viewpoint be understood in order to follow the presentation in this section. The effect of all past input and control signals is summarized in the state variables, which are in turn used to represent the expected value of future error without further control, $e(\tau)$. In most cases of practical interest the control plan $c(\tau)$ will start at zero and must lie somewhere on or within the boundaries formed by the application of either maximum positive or negative step inputs to the fixed system. Fig. 5.2 shows two possible control trajectories for a given

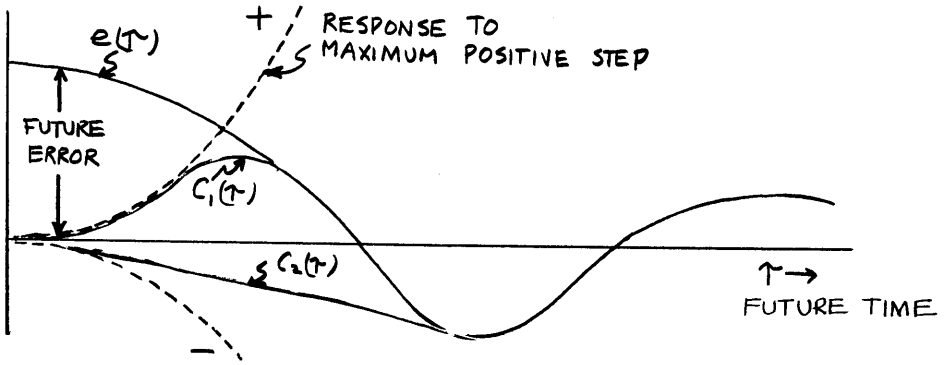


Fig. 5.2. Possible control trajectories

$e(\tau)$, where $c_1(\tau)$ is obviously better than $c_2(\tau)$ since it reduces the expected future error more quickly. In formulating this plan, the controller must select for each future instant some value of command signal within the saturation constraints, preferably to satisfy some design condition of optimality. Then it must execute the initial command of this sequence, and in the next instant the following changes will occur:

(1) The state variables that were previously in the fixed system and the random process generating model will decay as initial conditions, as indicated by $e(\tau)$.

(2) The controller command signal will have perturbed the fixed system state variables, as indicated by $c(\tau)$.

(3) White noise will enter the random process model and further change these state variables.

Because of the change in (3) above, the previously computed approach plan of the controller is no longer valid, and a new one must be computed. This frustrating need to solve for an optimum $c(\tau)$, use only the initial action, and then discard it an instant later is caused by the fact that we have imperfect knowledge of future events and must "muddle along" with the currently available information.

The use of expected future error is a very significant formulation of the problem of control in a random environment, for it transmutes a stochastic problem into a determinate one that is solely a function of the state variables of fixed system and random process generating models. Some possible criteria and general means of solution are presented next, followed by a more detailed look at a particular design which has the virtues of near optimal performance and easy mechanization.

The most general approach to this problem would employ the techniques of dynamic programming, which in this case would attempt to minimize some integral of a function of the state variables over all future time as the error approached the zero or equilibrium condition. To accomplish a valid solution by this means, thereby developing a control decision as a function of all the state variables, would require considerable ingenuity, very large amounts of digital computation, and is properly outside the scope of this report. The mechanization of the solution would in general involve a table lookup capability for the control system.

Another valid criterion for the design would be one of time-optimality. In analogy to the determinate or bang-bang regulator problem, which specifies that the time required to make all the state variables of the controlled system equal to zero should be minimized, one could demand that the future expected error and its defined derivatives be brought to zero in the quickest possible time. It has been proven in most determinate cases that full-throw or maximum effort control yields a minimum time solution.

Thus a set of transcendental equations could be easily written to equate the expected value of error and its n-1 derivatives to zero at some future time after n switching intervals, where n is the number of state variables in the controlled system. If these equations could be continuously solved to determine the duration of the first switching interval, then the switching time of the control system would occur when this switching interval became zero.

Unfortunately, the actual real-time solution of these transcendental equations appears quite difficult, assuming that a solution even exists. One source of difficulty is that the dependent variables, the switching times, must be constrained to be positive and in a certain order corresponding to successive sign changes of the control variable.

Another more abstract objection can be made to the criterion itself. First of all, the fact that the expected value of error and its defined derivatives are zero at a certain future time does not ensure that they will remain zero over the remainder of the interval, unlike the determinate case, since the saturation of the controlled system may prevent it from following exactly the further decay of the random process state variables. Next, the existence of a future value of zero of this expected error and associated derivatives does not necessarily mean that the intermediate values of error in transit were small. That is, the requirement that the error derivatives be brought to zero simultaneously may cause the con-

troller to select a trajectory which is obviously less desirable than one which approximately "matches up" at a considerably earlier time.

In the two approaches considered, the dynamic programming and the time-optimal, it is clear that there are very difficult analytical problems as yet unanswered, and that the sophistication (and consequently cost and size) of the control equipment must be relatively high. Is there then no way of practically utilizing the state variable approach to random processes in control? In the remainder of this section we shall discuss a proposed scheme of single-dimensional design which has many appealing features, not the least of which is the ease of instrumentation. Then, in Section 5.5 a comparatively simple multi-dimensional saturating controller will be described which is based on an extension of the Second Method of Lyapunov. The case of load disturbance will be dealt with in Section 5.3.

First of all, it is useful to reconsider the objectives of using the expected value of error in a design criterion. By constructing a non-linear system which would reduce the expected value of future error rapidly if white noise were suddenly cut off, it is hoped that the truly optimum linear system will be closely approximated. This hope is based on the observation that the optimum linear system produces, if white noise were cut off, a zero value of error for all future time. An alternate interpretation is that the best estimate of future error is its expected value. A decision scheme for control that always tends to reduce this expected future error in an efficient manner will, on the average, yield desired performance under the constraint of saturation and will best utilize the information about the random aspects of the problem available from linear theory.

Full throw or maximum effort control is selected in order to capitalize on, rather than linearize, the saturation in the output equipment. This will guarantee that the mean-square corrective effort is at an absolute maximum. Also, it has been proven an optimum mode in time-

optimal determinate control systems.

The simplest criterion to use would be that the future error become zero in the smallest time. This would be nothing more, if full-throw control were used and there were no noise at the input, than an error-controlled relay. This is patently not a very satisfactory solution, for the large error derivative which would usually result at the instant of zero error would ensure a large error before the next zero crossing -- possibly an unstable buildup would occur.

However, if it were specified that the future error and the error rate should be brought to zero simultaneously in the quickest possible time, then the result of non-coincident second and higher order derivatives between desired and actual output would have a definitely much smaller effect on the amount of later error. This specification would mean that the error would be brought to controllable proportions in the shortest time.

It is much easier to understand these ideas with the aid of typical control trajectories. Figure 5.3 shows a curve of expected error with no

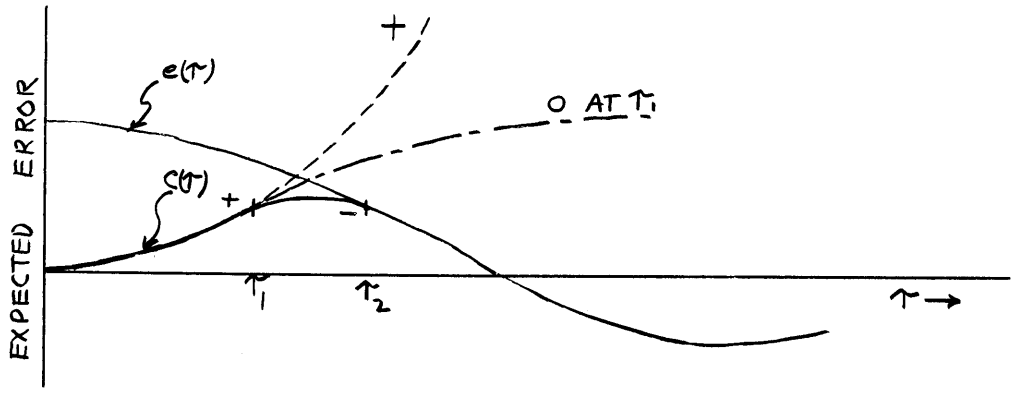


Fig. 5.3 An almost time-optimal control trajectory

further control, $e(\tau)$, and a superimposed planned control trajectory, $c(\tau)$. The controlled system of this example is assumed to include an integrator, and the initial path of $c(\tau)$ corresponds to the step response of this system to a positive saturation-constrained input command. At

time τ_1 the sign of the control variables is changed from + to - , and the expected future error, which is the difference between $e(\tau)$ and $c(\tau)$, is brought to zero with zero rate at time τ_2 .

Figure 5.4 shows a similar error plot, only the problem has advanced to time τ_1. The new $e(\tau)$ is the expected value of error with the new $c(\tau)$ set equal to zero for all time greater than τ_1, which corresponds to the difference between the $e(\tau)$ curve of Fig. 5.3 and the dashed path indicated by "0 at τ_1".

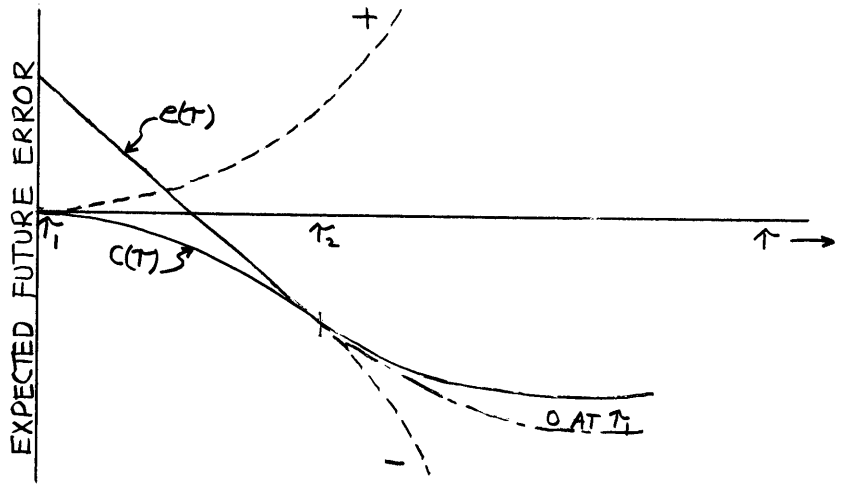


Fig. 5.4 Switching time determined by tangency

The very significant fact demonstrated in Fig. 5.4 is that the time to switch from + to - is τ_1 because at that time $e(\tau)$ first becomes tangent to a $c(\tau)$ representing the negative applied step. On the basis of this, we can postulate a control law for the proper sign of the current full-throw forcing variable, which is the desired output of the controller. If $c+(\tau)$ and $c-(\tau)$ are defined as the step responses of the controlled system under maximum positive and negative steps, respectively, then the current forcing function should be either + or - depending on whether the most future intersection of $e(\tau)$ is with $c+(\tau)$ or $c-(\tau)$. This switching law always yields an output which continually seeks to reduce large errors with maximum effort, and switches at the last moment (when the tangency first

occurs) in order to reduce the expected error and error rate to zero simultaneously. An intersection is always guaranteed, since (1) the random process models in this theory are stable, and (2) the step response of a system will always exceed the initial condition response as $\Upsilon \to \infty$.

Before proceeding on to a practical mechanization of this idea, it should be reemphasized that at every instant the control computer is dealing with expected future trajectories and its current decision is made as a function of present random process state variables. The planned approach to reduce future error to zero will in general never be completed exactly, for future white noise will enter the system and perturb the state variables. The design philosophy is, however, that the unpredicability of white noise means that on the average the decisions made will be the best for the conditions existing at that time.

Suppose a high-speed repetitive analog computer is used to generate (1) the expected error $e(\Upsilon)$ as a function of the current values of the state variables and (2) $c+(\Upsilon)$. Υ is now computer time. It is desired to determine whether $e(\Upsilon)$ intersects finally with $c+(\Upsilon)$ or $c-(\Upsilon)$ as τ becomes large. When $|e(\Upsilon)| - |c+(\Upsilon)| = 0$, or alternately, $e^2(\Upsilon) - c+^2(\Upsilon) = 0$, an intersection has taken place, and the sign of $e(\Upsilon)$ at that instant determines whether $c+(\Upsilon)$ or $c-(\Upsilon)$ has been crossed.

Fig. 5.5 shows the proposed analog instrumentation. The operation is as follows:

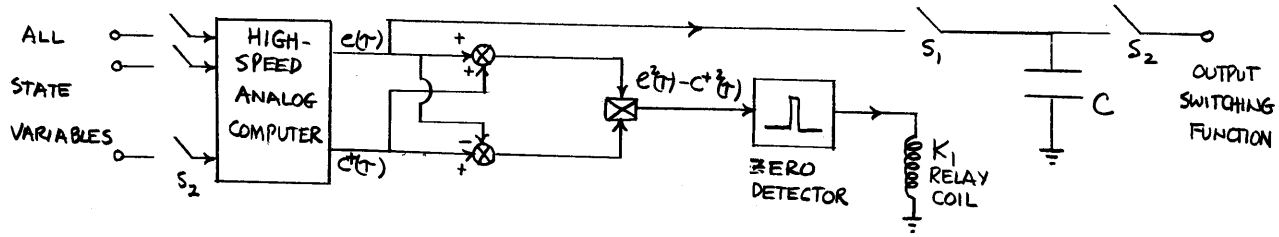


Fig. 5.5 A proposed full-throw controller

At the beginning of the computer cycle, current system and random process state variables are introduced as initial conditions in an analog

system which will reproduce $e(\tau)$ and $c(\tau)$ at its output when released. With the trivial identity, $e^2(\tau) - c^{+2}(\tau) = [e(\tau) + c^+(\tau)][e(\tau) - c^+(\tau)]$, the intersections of $e(\tau)$ and $c(\tau)$ result in an output from a zero-detecting device (perhaps a suitably configured relay with a small dead zone) which energizes coil K_1, momentarily closing switch S_1. Capacitor C then "remembers" the voltage $e(\tau)$ according to the previous zero crossing. After a suitable run, the computer is recycled, and the programmed closing of switch S_2 delivers the last $e(\tau)$ voltage at the output. This sampled signal has the sign of the desired polarity of the maximum command to the fixed system; further, it becomes zero when the present and future error becomes zero. This makes it a desirable switching function to drive a command relay with an arbitrarily small dead zone which will prevent, for example, a continuous cycling under zero error conditions. Alternately, a limiter with very high but finite gain near zero input can be used as the output command element.

The computer repetition rate is chosen so that an error of one cycle in switching will have small effect on the accuracy of control.

This configuration has the virtues of (1) being applicable to any scalar linear system which saturates and any random process, regardless of order, (2) being based on a design criteria which is intuitively satisfying, and (3) being the first practical design offered for a saturating control system which uses all the available statistical information and tends to exploit rather than linearize away the incontestable saturation phenomenon.

5.3 Optimum feedback configurations with load disturbance

The previous chapters have been mainly concerned with extracting useful information from an input signal. In a control system, one of the reasons for using feedback is that a disturbing signal often exists at the output equipment. Figure 5.6 shows the conventional means of manipulating a disturbance d inside a loop into a form which can be dealt with

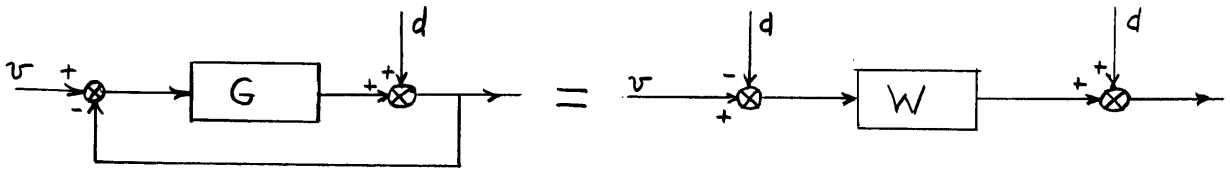


Fig. 5.6 Manipulation of load disturbance to obtain standard cascade configuration

in the standard theory. This is the approach taken by Newton, Gould and Kaiser[10] and by Smith[11]. There are two difficulties with this step however. The first is that the form of the feedback path must be assumed. Secondly, and much more important, the preliminary dilution of disturbance and input signal creates an unnecessary task for the optimum system in separating them again.

It will be demonstrated in this section that in a linear theory load disturbance does not affect the basic statistical design. As a start, one optimum system which theoretically reduces the effect of load disturbance to zero and yet operates optimally on the input signal is given in Figure 5.7. Here $\frac{S(s)}{S(s) + N(s)}$ is the optimum system proven in section 4.5, where $\Phi_{vv}(s) = G(s)\,G(-s)$ and $G(s) = S(s) + N(s)$, the signal and noise components, respectively.

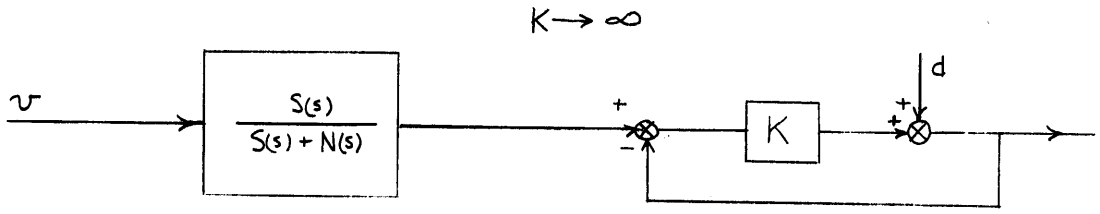


Fig. 5.7 Elimination of load disturbance with infinite gain amplifier

A more practical elaboration of this scheme is given in Figure 5.8, which shows an arbitrary transfer function H(s) enclosed with a minor loop with infinite gain. This configuration is of considerable practical significance since it is optimum, compensates any fixed minimum-phase transfer function, unless the excess of poles over zeroes of H(s) is such

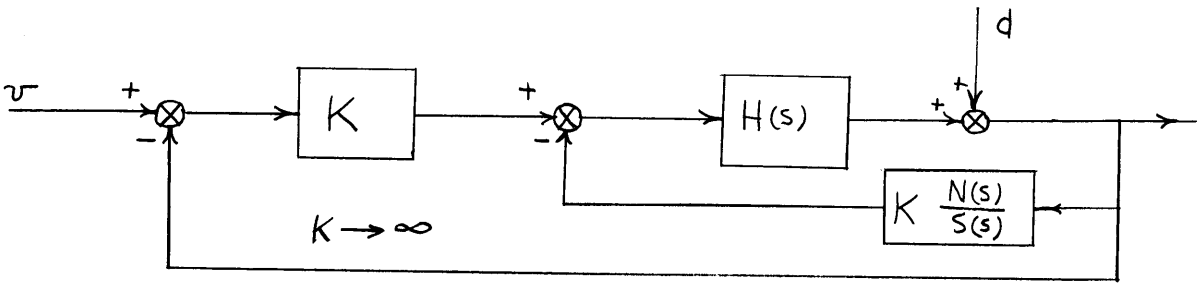


Fig. 5.8 General form of an optimum feedback system

to lead to instability as $K \longrightarrow \infty$, and eliminates any effect of load disturbance.

Unfortunately, these pleasant linear conjectures are often based on the principle that a mouse can pull an ox-cart if beaten hard enough. If H(s) in Fig. 5.8 has a saturating characteristic, the random process entering the system at d becomes significant, and must be separately operated on to compute its state variables, which contribute additively to $e(\tau)$, the expected value of future error used in the previous section.

5.4 Contemporary designs for full throw control of a system subject to a random process

Smith[11] has presented with his "predictor" controller the first fruitful attack on the problem of saturating control of a random process. His idea is quite simple. A fixed future time τ^* is selected for the prediction of a number of derivatives of the input random process equal to the number of state variables of the controlled system. Then, the controller is designed as a standard bang-bang servo in order to reduce the error between present position and this future command signal in the shortest possible time.

There is, of course, a glaring flaw in this reasoning. If τ^* is fixed, the only valid control decisions are made under the particular conditions when this "error" between present position and future command can be actually brought to zero exactly in τ^* seconds. Otherwise, and in the general case, the controller plans to drive toward the correct position, but at the wrong time. Fig. 5.9 shows how this disregard of the actual

time required to obtain a change in state can result in poor control decisions, using the display presented in section 5.2.

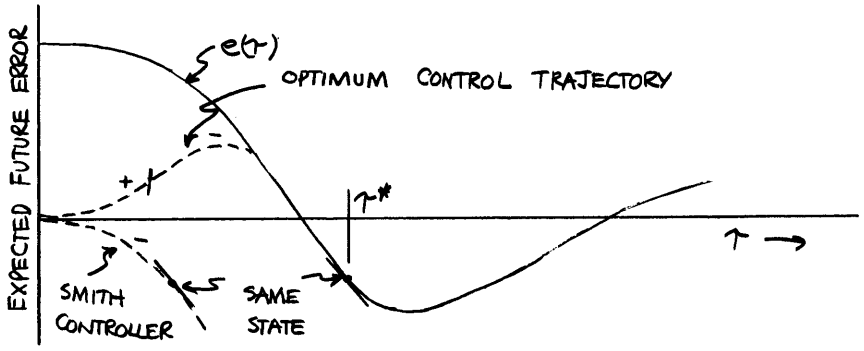


Fig. 5.9 Consequences of a fixed τ^* in the Smith predictor servo

Benedict[30] based his dissertation on this lack of optimality in an attempt to justify or discredit this approach with analog computer simulation. His results indicate that this Smith predictor servo is better than a bang-bang controller which ignores any future change in the control signal (ie: $\tau^* = 0$), which is to be expected. He also notes that increasing the value of τ^* when the input signal level is high improves performance, which again is logical since the actual time required to reach the specified state would tend to be larger.

Hopkin and Wang[31] have taken perhaps a more logical look at this problem. They make a Taylor's series expansion of the input random process signal, and attempt to find a set of control switching intervals which will reduce all the derivatives of the extrapolated future system error to zero simultaneously.

The two defects in this approach are:

(1) The intrinsic quantities of the random process, the state variables, are neglected in the Taylor series approximation, this providing a poor error prediction.

(2) The resulting transcendental equations are difficult to solve, if a solution exists at all.

In summary, it is felt that the two attempts discussed above have merit as beginning steps, but that the problem outline and approximate solution of Section 5.2 more clearly define the optimum system and best utilize the information contained in the input random process.

5.5 Multi-dimensional bang-bang control of systems subject to random process inputs

There are three general classes of power actuator in a control system. First, the output transducer may be conservatively rated and perform in essentially a linear manner, which allows use of the large body of design information on linear control systems. Secondly, it may operate in a partially saturated condition, the improvement of which case having been considered earlier in this chapter. Finally, the power actuator may be fairly inadequate and under-rated for the job presented by the input random process.

It is this latter case which will be considered in this section. The corrective action of the controller will not have a pronounced effect on the error, but it is desired to optimize the effect small as it may be. Essentially, what will be done is to define a figure of "badness" for the state of the controlled system and of the random process which is a measure of the expected future error. Then, the control or controls will be continuously thrown in such a direction so as to maximize the rate of decrease of this figure of "badness" at every instant.

The control system chosen for illustration of these ideas is a regulator, but the ideas are equally applicable to a servo application.

To structure this design procedure in an orderly fashion, it is first necessary to present some results of the venerable "Second Method of Lyapunov"[32]. Then, an original modification will be made in order to extend this determinate theory to include random processes. Finally, it will be shown how an optimal control law can be found as a linear function of

the state variables.

The Second Method of Lyapunov is not so much a method as it is a way of characterizing the free dynamic behavior of linear and non-linear systems. It uses a type of generalized energy expression, and examines the rate of change of this function for various states of the system. If this energy expression, called a Lyapunov function, tends to decrease everywhere except at the equilibrium point in a region of possible system states, then the system is considered stable in this region.

In the particular case of a free linear system with no external excitation, the standard differential equation form is

$$\frac{d}{dt} \quad x = A\,x \qquad (4.1)$$

A Lyapunov function, V(x), is chosen as a quadratic form $x^T Px$, where P must be positive definite and symmetric. From the results of section 3.3, it is known that, if P is positive definite, it can be factored into two matrices

$$P = N^T \cdot N$$

with N having real elements only. Thus,

$$x^T Px = (Nx)^T \; Nx$$

which is the square of some linear transformation N on x.

One choice for P might be such as to make V(x) equal the energy of the system. According to the Second Method[32], the system is stable if and only if $\frac{d}{dt}$ V(x) $<$ 0 for all x, where x $\neq$ 0

$$\frac{d}{dt} \; V(x) = \frac{d}{dt} \quad x^T Px \quad = \left\{\frac{d}{dt} \quad x^T\right\} \; P\,x + x^T P\left\{\frac{d}{dt} \quad x\right\}$$

but

$$\frac{d}{dt} \quad x = A\,x$$

$$\frac{d}{dt} \; V(x) = x^T A^T Px + x^T PAx = x^T \left[\, A^T P + PA \right] x$$

Thus,

$$A^T P + PA = -\,Q \qquad (5.2)$$

where Q is some positive semi-definite symmetric matrix, if $\frac{d}{dt}\ V(x)$ is always to be negative for any value of x.

The above relations are very important to the linear theory. Since

$$V(x) = \int_0^{V(x)} dV(x) = \int_{t=\infty}^{t} dt \quad \frac{dV(x)}{dt} = \int_{t}^{\infty} dt \left(- \frac{dv(x)}{dt} \right)$$

and

$$x^T P x = \int_t^{\infty} dt \quad (x^T Q x) \tag{5.3}$$

the two symmetric matrices, P and Q, provide quadratic forms which are related in an integral fashion. Accordingly, if $x^T Q x$ represents some measure of instantaneous error of a free system, $x^T P x$ is the integral of this error over all future time, which is a very useful error criterion.

Eq. 5.2, in this case, must be solved for P with $\frac{n(n+1)}{2}$ independent linear equations for a given Q.

Bass[33] suggested, in the case of a linear system settling to equilibrium, that one form of good full-throw control would attempt to maximize the negative rate of change of V(x) at every instant, V(x) being a suitably-defined error criterion for the system without further control. In this case

$$\frac{d}{dt}\ x = A x + D c$$

c being a control vector having an amplitude constraint on each component.

$$\frac{d}{dt} \left[x^T P x \right] = \left\{ \frac{d}{dt}\ x^T \right\} P x + x^T P \left\{ \frac{d}{dt}\ x \right\}$$

Substituting from the matrix differential equation

$$\begin{aligned} \frac{d}{dt} \left[x^T P x \right] &= \left[Ax + Dc \right]^T P x + x^T P \left[Ax + Dc \right] \\ &= x^T \left[A^T P + P A \right] x + c^T D^T P x + x^T P D c \\ &= x^T \left[A^T P + P A \right] x + 2\, c^T D^T P x \end{aligned}$$

since a scalar can be transposed at will and P is symmetric. Therefore to select c_i so as to maximize the negative rate of change of $x^T P x$

$$\text{sign } c_i = - \text{ sign } \left\{ D^T P x \right\}_i \qquad (5.4)$$

and the magnitude of c_i should be the maximum possible. P, being a measure of future error of the system, will be found by postulating a positive definite matrix Q, which represents the instantaneous error, and solving Eq. 5.2.

With this admittedly brief account of some of the available techniques from the determinate Second Method of Lyapunov, it is now desired to consider how this theory might be adapted to include stationary random processes, since the state concept has been extended in this work to stochastic inputs.

To fix ideas, the regulator problem will be considered. Without any control action, a physical system H(s) is shown in Fig. 5.10 being acted upon by a random process which is hypothesized to originate in a white-noise driven system G(s). The output e is an undesired error. From the results of this and the previous chapter, it is known that the expected value of $e(t+\tau)$ for $\tau > 0$ is completely specified by knowledge of the state variables of G(s) and H(s). Therefore it is logical to define a Lyapunov function, $x^T P x$, which represents an integral error criterion over all future time of the expected value of error from the total state vector. That is, the concept of system in the Lyapunov theory is enlarged to include the effective system which generates the random process.

The error e and its m - 1 derivatives are linear combinations of the m state quantities of H(s). Thus, e^2, $\dot{e}^2$, $\ddot{e}^2$, . . . can all be weighted with a non-negative measure of instantaneous undesirability.

If $e = B x_h$, where $e_i = \frac{d^{i-1}}{dt^{i-1}} e$, B is a mxm matrix, and x_h is the state vector of H(s), and the measure of undesirability of e is given by $e^T E e = x_h^T B^T E B x_h$, then the matrix Q to weight the instantaneous undesirability of the entire state vector x is given by $Q = \begin{bmatrix} B^T E B & 0 \\ 0 & 0 \end{bmatrix}$

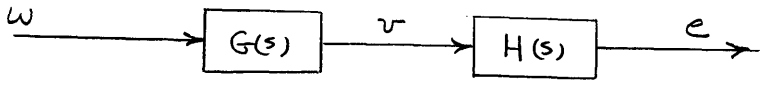


Fig. 5.10 System subject to random disturbance

The first m elements of x are identical to x_h.

The integral error criterion $\int_0^\infty x^T Q x \, dt = x^T P x$ is found from solution of

$$A^T P + P A = - Q \tag{5.2}$$

if A is the matrix of the differential equation which governs the entire system

$$\frac{d}{dt} x = A x$$

Now, suppose that it is desired to regulate the error with controls that saturate and have small effect on the physical system in comparison with the random process. Bass's approach, described previosuly in the determinate case, appears to have considerable promise in this problem.

The objective of the control system in this case is to maximize continuously the negative rate of change of the measure of future error. Eq. 5.4 is still valid

$$\text{sign } c_i = - \text{sign } \left\{ D^T P x \right\}_i$$

with D defined by

$$\frac{d}{dt} x = A x + D c$$

A simple example will illustrate the ease of application of these ideas:

EXAMPLE

A spring-mass-dashpot configuration is shown in Figure 5.11. F_R is the disturbing random process, with power density spectrum $\frac{R^2}{(s+a)(-s+a)}$

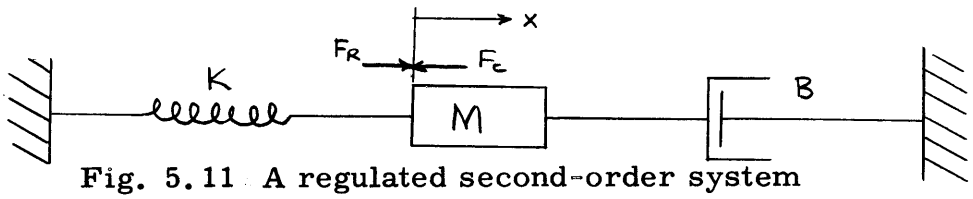


Fig. 5.11 A regulated second-order system

F_c is the regulating force, with maximum amplitude constrained to be ± A.

The random process is generated in an effective system with transfer function $\frac{R}{s+a} = \frac{\frac{R}{s}}{1+\frac{a}{s}}$ as shown in flow graph form in Figure 5.12.

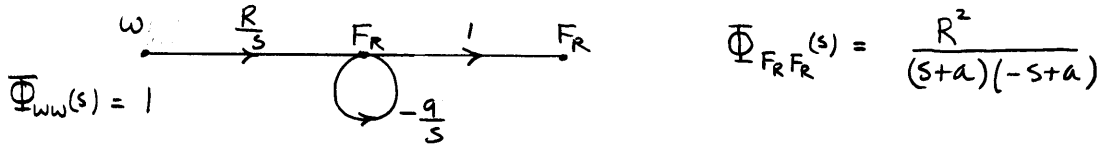


Fig. 5.12 Flow-graph representation of random process generation.

The differential equation of motion of the mass is

$$M \frac{d^2x}{dt^2} + B \frac{dx}{dt} + K x = F_R - F_c$$

Defining x_1 as x, x_2 as $\dot{x}$, and x_3 as F_R, the following matrix equation results:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -\frac{K}{M} & -\frac{B}{M} & \frac{1}{M} \\ 0 & 0 & -a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -\frac{1}{M} & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} F_c \\ \omega \end{bmatrix}$$

or

$$\frac{d}{dt} x = A x + D F \qquad (4.1)$$

It is desired to minimize the motion of the body, with the squared velocity given a weight of μ with respect to the displacement squared. One possible choice for μ, $\frac{M}{K}$, would weight equally the kinetic and potential energy of the system.

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Next Eq. 5.2 must be solved.

$$A^T P + P A = - Q \qquad (5.2)$$

In this example, this equation is readily solved for the 6 independent elements of the symmetrical P. Solving,

$$P_{11} = \frac{K}{2B}\left(\frac{M}{K} + \mu\right) + \frac{B}{2K} \; ; \; P_{22} = \frac{M}{B}\left(\frac{M}{2K} + \frac{\mu}{2}\right)$$

$$P_{33} = \frac{1}{2aM} \; \frac{BM + aM^2 + \mu AKM}{K^2 B + a^2 KBM + KB^2}$$

$$P_{12} = \frac{M}{2K} \; ; \quad P_{13} = (Ba + a^2 M) P_{33} - \frac{1}{2B}\left(\frac{M}{K} + \mu\right)$$

$$P_{23} = aM P_{33}$$

From Eq. 5.4

$$F_c = - A \operatorname{sgn} \left\{ D^T P x \right\}_1$$

$$= - A \operatorname{sgn} \left\{ - \frac{1}{M} P_{12} x_1 - \frac{1}{M} P_{22} x_2 - \frac{1}{M} P_{23} x_3 \right\}$$

$$F_c = A \operatorname{sgn} \left\{ \frac{1}{2K} x_1 + \frac{1}{2B}\left(\frac{M}{K} + \mu\right) x_2 + \frac{1}{2K} \frac{B + a(\mu K + M)}{K + a(aM + B)} x_3 \right\}$$

Here sgn is an operator which equals + 1 if the enclosed quantity is positive, and - 1 if it is negative.

This is the linear switching law which continuously tends to maximize the rate of decrease of the error criterion

$$\int_t^\infty \left[\left\{ E\left[x(\tau)\right] \right\}^2 + \left\{ E\left[\dot{x}(\tau)\right] \right\}^2 \right] d\tau$$

which is a function of the state variables.

If it is desired to discount future values of error with an exponential, e^{-bt}, the matrix A is replaced by A - bI, since the Laplace trans-

form of the system is given by $\left[sI - A\right]^{-1}$, which, when s is replaced by $s + \sigma$, becomes $\left[s\,I + \sigma I - A\right]^{-1}$.

This discounting becomes particularly significant in the case of a servo, where integrators often appear in the controlled system, because the integral of the squared initial condition response of an integrator is infinite, and the design impossible. Replacement of $\frac{1}{s}$ by $\frac{1}{s+\sigma}$ means a finite square response exists and this procedure is applicable.

To summarize the advantages of this proposed full-throw control of a random process:

(1) The system becomes more and more optimum as the inadequacy or non-linearity of the control transducer is emphasized.

(2) The design procedure is simple to execute and results in a completely linear system except for the output relay.

(3) The resulting system is guaranteed to be stable from the non-linear part of the Second Method, since $\dot{V}(x)$ is always negative.

(4) Multi-dimensional designs can be made with no more theoretical, computational, or hardware difficulty than the single-dimensional case.

These four advantages make this proposed design philosophy very promising for practical applications where power transducers are inadequate for their job in a stochastic environment.

CHAPTER VI.

SUMMARY AND CONCLUSIONS

6.1 Outline and summary

The results obtained in this thesis investigation intertwine throughout the entire theory of stationary random processes. However, there are three fundamental and significant contributions in this work:

(1) Development of a complete multi-variable theory

The random process with many separate but statistically related signals -- the so-called multi-dimensional random process -- has been studied to an extent that there is now no conceptual difference between single- and multi-dimensional theory. An important analytical tool in this respect is the equation

$$\underline{\Phi}_{xy}(s) = \sum_i \sum_j \underline{G_i}(-s) \quad \underline{\Phi}_{x_i y_j}(s) \quad \underline{H_j}^T(s) \qquad (2.32)$$

which compactly determines the statistical relations between signal vectors in a linear system as a function of the properties of the inputs.

The key to the solution of the optimum multi-dimensional system in the Wiener sense is the concept of matrix factorization, which separates a matrix of cross-power density spectra among the members of an input random process v, according to the equation

$$\underline{\Phi}_{vv}(s) = \underline{G}(-s) \quad \underline{G}^T(s) \qquad (2.27)$$

such that both $\underline{G}(s)$ and $\underline{G}^{-1}(s)$ represent transfer function of stable systems.

With this factorization, it is possible to represent the optimum multi-dimensional system with

$$\underline{W}^T(s) = \left[\underline{G}^T(s)\right]^{-1} \mathcal{L}^{-1} \left\{\underline{G}^{-1}(-s) \quad \underline{\Phi}_{vi}(s)\right\} \qquad (2.28)$$

The matrix factorization problem is one of great complexity. The general solution presented in Chapter 3 was reached only after many

other approaches aborted. It basically consists of a series of simple steps which manipulate the given matrix into desired forms, of which the final one is a numerical matrix which can be easily factored. An iterative method is also presented which shows promise of efficient and rapid solution when a digital computer is used to solve resulting sets of linear equations.

(2) Introduction of a theory of random processes based on physical models

The results of this thesis indicate that the simplest understanding of random processes and of the optimum systems which operate on them is obtained by hypothesizing that some linear system is being excited by white noise to produce the random process, v. To support this claim:

(1) The result of matrix factorization, $\underline{G(s)}$, is such a system, where $\underline{G(-s)}\ \underline{G^T(s)} = \underline{\Phi_{vv}(s)}$.

(2) The optimum predictor merely reproduces the individual state variables of $\underline{G(s)}$, and weights each by its reduction after τ seconds of initial condition decay in the model.

(3) If $\underline{G(s)}$ is separated into two parallel systems, $\underline{S(s)} + \underline{N(s)}$, associated with the signal and noise components, respectively, of a random process, then the optimum filter merely recovers the output of $\underline{S(s)}$. This optimum filter is, in canonic form, a unity feedback system with a forward loop transmission of $\underline{S(s)}\ \underline{N^{-1}(s)}$.

(4) Auto- and cross-correlation functions can be interpreted as the initial condition response of $\underline{G(s)}$.

Incidental to this approach, it was found that the fundamental statement of an optimum physically-realizable system is that any resulting error should be uncorrelated with the past values of any input signal.

(3) The state of a random process viewed as fundamental information for control use

The state approach has proven a powerful tool in the analysis of

determinate system behavior. A major contribution of this thesis is the extension of these techniques to include the study of stationary random processes.

The output of an optimum predictor minimizes the mean-square error of the estimate, or alternately, is the expected value of the future signal. It has been shown that this expected value is merely the initial condition decay of the current state variables of the effective generating model.

As was discussed in Chapter 5, the optimum filter at any time t has a configuration which causes the expected value of future error, $e(t+\tau)$, to be zero for all positive values of τ. Thus, for the purposes of control, the actual system will more closely approximate the optimum system as the expected value of future error is minimized.

A typical control problem has an input random process which is to be followed and perhaps filtered, a fixed linear system which is externally controlled, and a disturbance which acts on the fixed system. The expected value of future error is given by (1) initial condition decay of state variables of the input and disturbance generating models and of the fixed system, and (2) the effect of future control action on an otherwise "empty" fixed system. This problem of control becomes quite difficult when an amplitude constraint is placed on the control variable.

Thus, the controller of a saturating system must continuously select a control variable which is a function of all the state variables so as to tend to minimize, with some criterion, the expected value of future error.

Two general and feasible solutions to this problem have been given in Chapter 5. Both assume that the best operation of the system will result with full-throw control. The first solution selects for a criterion that the control variable should switch at an instant when the expected value of error and its derivative can be brought to zero simultaneously

along the next control trajectory.

The second solution considers the effect of future control as small, and always tends to minimize a quadratic measure of future error. This is accomplished through extension of the classical Second Method of Lyapunov to include stationary random processes. A particularly significant result of this approach is that it permits a rational and easily instrumented design procedure for a multi-dimensional saturating system.

6.2 Paths for future research

In the course of this thesis investigation, many problems were encountered which could not be satisfactorily dealt with in this report. The following discussion presents some of the more prominent of these in the hope that further interest and research can be stimulated.

(1) In many random processes of practical interest, for example, the national economy, there are available a great number of possible components for a multi-dimensional analysis. Assuming that this stationary theory might approximate the true behavior (which would probably not be valid) it is interesting to conjecture what might happen as the number of scalar processes used becomes very large. Since the error in prediction of a variable, for example, is always made less as the dimension of the random process increases, one intuitively feels that the prediction error could be made arbitrarily small by analyzing enough processes which cross-correlate with the variables of interest.

It would be interesting to obtain some measure or bound on the increase in precision obtainable by considering an additional correlated random process, without completing a refactorization. Also, a means of selecting the most useful (in the sense of reducing prediction error) members of a set from consideration of their correlation functions is needed.

(2) A general solution was presented in Chapter 3 for the matrix

factorization problem. The resulting answer for $\underline{G(s)}$ can be multiplied by any real unitary matrix $\underline{U}$, where $\underline{U} \cdot \underline{U}^T = \underline{I}$, without affecting the validity of the answer. Although existence has been proven, uniqueness has not. A useful further addition to this theory would be a proof of this uniqueness of $\underline{G(s)}$ -- or, by counter-example, that a multiplicity of answers exist besides the unitary transformation.

(3) Two significant alternate statements of optimality for linear systems have been found through analysis of the Wiener theory.

(a) Zero correlation exists between present error and all past values of input signal.

(b) The expected value of all future error is zero at any instant.

Considering (a), this could be generalized to the non-stationary and "non-linear" case by specifying that all measures to indicate statistical relation between signals be zero between past input and present error.

In the case of (b), this statement appears to be just as basic as requiring that the mean-square error be minimized. This appears to have an immediate application to random processes which are non-stationary and/or non-Gaussian.

The viewpoint of the author is at variance with much of the present work going on in non-linear random process theory. It is suggested that a possibly fruitful (although modest) line of attack would be to specify simple non-linear models for the creation of the process from independent white noise, and determine suitable statistical measurements which could fix the parameters of these models. This approach is in contrast with a theory which attempts to be totally general (ie: Wiener[26]) but which results in models which have an infinite number of state variables. If a finite-state model -- perhaps a linear system followed by a memeory-less non-linear function generator -- could be found to represent adequately a class of random processes -- then the optimum configurations for systems to operate on these processes could be found through extension of the state and model

concepts outlined in this work.

(4) Practical control of systems operating in a stochastic environment will generally involve considerations of saturation, unless the designer is willing to pay for the validity of the linear theory with increased power actuator size, weight, and cost. The optimum controller has been shown, in this case, to be one which continually plans to reduce the expected value of future error to zero in some optimal fashion.

There appears to be considerable promise in attempting a dynamic programming solution to this most important problem. If a maximum effort control system is specified, the desired solution is the delineation of the switching surface as a numerical function of the state variables of the random process and of the controlled system. A useful approximation to this surface would be its Taylor series expansion as linear and quadratic functions of the state variables.

The two proposed designs of Chapter 5 have the advantages of (1) comparatively uncomplicated instrumentation and (2) rational design theories which make use of the state concept and the known saturation limitations of the control signal. A complete analog computer investigation of their merit would be warranted for comparison with more conventional configurations.

APPENDICES

APPENDIX I.

OPTIMALITY IN DISCRETE LINEAR SYSTEMS

1. Introduction

The random signals which have been the concern of the main part of this report have been continuous in time, as have been the systems which act on them. However, many problems of physical interest deal with random sequences of numbers -- perhaps equally-spaced samples of a continuous random process - and associated linear discrete systems.

There are several good textbooks -- for example, Ragazzini and Franklin[23] -- which adequately present sampled-data theory, both determinate and stochastic. In all, the primary emphasis has been on automatic control applications, heightened by the increasing use of digital computers in control.

The purpose of this appendix is not to encapsulate this general discrete theory, but merely to show how the major results of this particular work in continuous random processes are easily extended to the case of stationary, ergodic, and discrete stochastic processes. Pertinent equations are preceded by the number of their analogous continuous equation in brackets.

2. Fundamental properties of discrete signals and systems

A discrete signal is a sequence of numbers, such as

$$\ldots\ f(n-1),\quad f(n),\quad f(n+1)\ \ldots$$

with n indicating discrete time. A "z-transform" is defined by

$$\ldots + f(n-1)\, z^{n-1} + f(n)\, z^{n} + f(n+1)\, z^{n+1} \ldots$$

where the transform variable z^k serves to index the associated f(k) to the proper time. For example, the sequence

$$1,\ a,\ a^2,\ a^3,\ \ldots\ldots,\ a^k,\ \ldots$$

has a z-transform

$$1 + a\,z + a^2 z^2 + \ldots + a^k z^k \ldots = \frac{1}{1 - a\,z}$$

The variable z acts as a unit delay operator. Discrete systems are constructed with this building block to delay, sum or multiply by constants the discrete variables which pass through it. A convenient way to visualize this process is to consider the discrete signal as a series of impulses with areas equal to the value of the variables. The discrete system is then denoted by a transfer function in z, where $z = e^{-sT}$. T is the time interval between impulses. This representation allows immediate extension of much of the continuous Laplace transform theory. In example, if x(n) is the input sequence, y(n) the output sequence, and the system transform is given by G(z), then

$$Y(z) = X(z) \cdot G(z)$$

3. Statistical relationships

The cross correlation function between two discrete signals, x(n) and y(n), is defined by [2.3]

$$\varphi_{xy}(k) = E\left\{x(n)\ y(n+k)\right\} \qquad \text{(A 1.1)}$$

and the discrete "cross power density spectrum" -- a misnomer, but used for continuity -- is given by

$$\Phi_{xy}(z) = \sum_{k=-\infty}^{\infty} \varphi_{xy}(k)\ z^k \qquad \text{(A 1.2)}$$

For later use, the general transformation of the statistical properties of the random sequence by linear systems will now be derived, in analogy with Section 2.4.

Suppose an arbitrarily large but finite length of a signal x(n) is available. Its z-transform is given by $x(z) = \sum_{n=-N}^{N} x(n)\, z^n$. Also over the same interval, $y(z) = \sum_{n=-N}^{N} y(n)\, z^n$.

Consider the product term $x(z^{-1})\, y(z)$

$$x(z^{-1})\ y(z) = \sum_{n=-N}^{N} x(n)\ z^{-n} \sum_{m=-N}^{N} y(m)\ z^{m}$$
$$= \sum_{n=-N}^{N} \sum_{m=-N}^{N} x(n)\ y(m)\ z^{m-n}$$

The coefficient of the kth term, which is multiplied by z^k, is

$$\sum_{m=-N}^{N} x(n)\ y(n+k)$$

But $$\varphi_{xy}(k) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{m=-N}^{N} x(n)\ y(n+k)$$

Therefore

$$\Phi_{xy}(z) = \lim_{N\to\infty} \frac{1}{2N+1}\ x(z^{-1})\ y(z)$$

If $X(z) = \sum_{i=1}^{p} X_i(z)\ G_i(z)$ and $Y(z) = \sum_{j=1}^{r} Y_j(z)$

$$\Phi_{xy}(z) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{i=1}^{p} X_i(z^{-1})\ G_i(z^{-1}) \sum_{j=1}^{r} Y_j(z)\ H_j(z)$$
$$= \sum_{i=1}^{p} \sum_{j=1}^{r} G_i(z^{-1})\ H_j(z) \lim_{N\to\infty} \frac{1}{2N+1} X_i(z^{-1})\ Y_j(z)$$

which yields [2.9]

$$\Phi_{xy}(z) = \sum_{i=1}^{p} \sum_{j=1}^{r} G_i(z^{-1})\ H_j(z)\ \Phi_{x_i y_j}(z) \qquad \text{(A 1.3)}$$

By steps identical to those of section 2.9, the matrix relationship [2.32]

$$\underline{\Phi_{xy}(z)} = \sum_{i=1}^{n} \sum_{j=1}^{m} \underline{G_i(z^{-1})}\ \underline{\Phi_{x_i y_j}(z)}\ \underline{H_j^{T}(z)} \qquad \text{(A 1.4)}$$

is easily obtained, where

$$\left[x(z)\right] = \sum_{i=1}^{n} \underline{G_i(z)}\ \left[x_i(z)\right]$$

and

$$\left[y(z)\right] = \sum_{j=1}^{m} \underline{H_j(z)}\ \left[y_j(z)\right]$$

4. Optimum configurations

From the arguments of Section 4.5, it is clear that the basic statement of optimality for a linear system to operate on this discrete signal is [4.11]

$$\varphi_{v_i e_j}(k) = 0 \qquad k \geq 0 \qquad (i, j = 1, 2 \ldots n)$$

or

$$\mathcal{Z}^{-1}\left\{\underline{\Phi}_{ve}(z)\right\} = \underline{0} \qquad \text{(A 1.5)}$$

where the $\mathcal{Z}^{-1}$ operator retains its conventional meaning, discarding all parts of the term in brackets with negative powers of z.

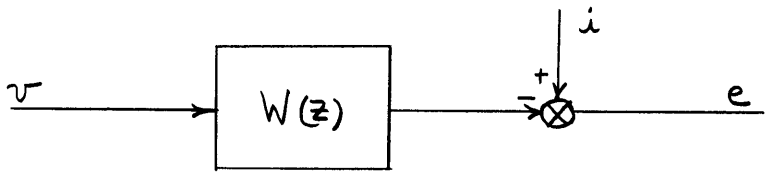


Fig. A-1.1 Configuration of optimum multi-dimensional system

From Eq. A-1.4 and Fig. A-1.1,

$$\underline{\Phi}_{ve}(z) = \underline{\Phi}_{vi}(z) - \underline{\Phi}_{vv}(z)\ \underline{W^T(z)}$$

Hence [2.18]

$$\mathcal{Z}^{-1}\left\{\underline{\Phi}_{vv}(z)\ \underline{W^T(z)}\right\} = \mathcal{Z}^{-1}\left\{\underline{\Phi}_{vi}(z)\right\} \qquad \text{(A 1.6)}$$

$\underline{\Phi}_{vv}(z)$ can be factored into $\underline{G(z^{-1})}\ \underline{G^T(z)}$ with the methods of Chapter 3, if functions of z^{-1} are regarded as functions of -s ($z^{-1} = e^{sT}$, $z = e^{-sT}$). $\underline{G(z)}$ and $\underline{G^{-1}(z)}$ will both be realizable. It is assumed that the elements of $\underline{\Phi}_{vv}(z)$ have polynomials in z in the numerator and denominator.

From the results of section 2.8, [2.28]

$$\underline{W^T(z)} = \left[\underline{G^T(z)}\right]^{-1} \mathcal{Z}^{-1}\left\{\underline{G^{-1}(z^{-1})}\ \underline{\Phi_{vi}(z)}\right\} \qquad \text{(A 1.7)}$$

5. Special interpretation of optimum systems

For simplicity, the following discussion refers to single-dimensional systems, but the results are readily extended to multi-dimensional problems.

Since $\Phi_{vv}(z) = G(z^{-1})\ G(z)$, G(z) is a linear discrete system which can reproduce the observed statistics when excited by a sequence

of uncorrelated numbers with unit variance. It will be shown first that the optimum predictor for k seconds in the future is a process of measuring the model state variables and weighting them for k units of initial condition decay, according to the results of Section 4.3.

For the predictor,

$$\Phi_{vi}(z) = z^{-k}\, \Phi_{vv}(z)$$

Then

$$W(z) = \frac{1}{G(z)}\, \mathcal{Z}^{-1} \left\{ z^{-k}\, G(z) \right\}$$

Since $\frac{1}{G(z)}$ recovers the excitation of the model G(z), $\mathcal{Z}^{-1}\left\{ z^{-k}\, G(z) \right\}$ must provide a transmission to each state variable, and weight by the transient decay for k units. Suppose, since this is more in the nature of a demonstration than a proof, that

$$G(z) = \sum_{i=1}^{r} \frac{k_i}{1 - a_i z}$$

Here, the partial fraction expansion into r poles gives an allowable set of state variables which act in each of the r sub-systems.

$$\mathcal{Z}^{-1} \left\{ z^{-k} \sum_{i=1}^{r} \frac{k_i}{1 - a_i z} \right\} = \sum_{i=1}^{r} \frac{k_i\, a_i^{k}}{1 - a_i z}$$

This shows the desired weighting by a_i^{k}.

In the case of an optimum filter,

$$\Phi_{vi}(z) = \Phi_{ss}(z) + \Phi_{ns}(z)$$

and

$$W = \frac{1}{G(z)}\, \mathcal{Z}^{-1} \left\{ \frac{\Phi_{ss}(z) + \Phi_{ns}(z)}{G(z^{-1})} \right\}$$

The arguments of section 4.5 indicate that $\mathcal{Z}^{-1}\left\{ \frac{\Phi_{ss}(z) + \Phi_{ns}(z)}{G(z^{-1})} \right\}$ is the partial fraction expansion of G(z) in the signal poles. Hence [4.5]

$$G(z) = S(z) + N(z)$$

the signal and noise parts, respectively, and

$$W(z) = \frac{S(z)}{S(z) + N(z)} \qquad \text{(A- 1.8)}$$

Fig. A-1.2 shows the canonic feedback filter.

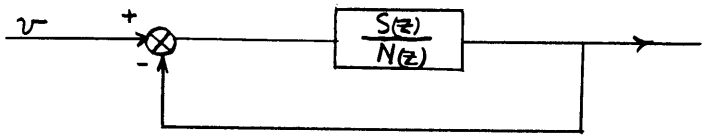


Fig. A-1.2 Optimum discrete filter.

6. Considerations for optimum linear sampled-data control systems

For practical end use, a theory of pure numbers such as is built up with z-transform is seldom applicable in control. Rather, it is necessary to convert the discrete information signal into a quantity with physical significance which will in turn be the input to a continuous physical system. A particular problem will be considered in this section, that of an error-sampled control system which must attempt to follow a noisy input signal.

The general approach in this thesis has been to emphasize the models which effectively create random processes, and have signal levels which are recoverable and useful. There are three models which could account for the sampled input signal, as shown in Fig. A-1.3.

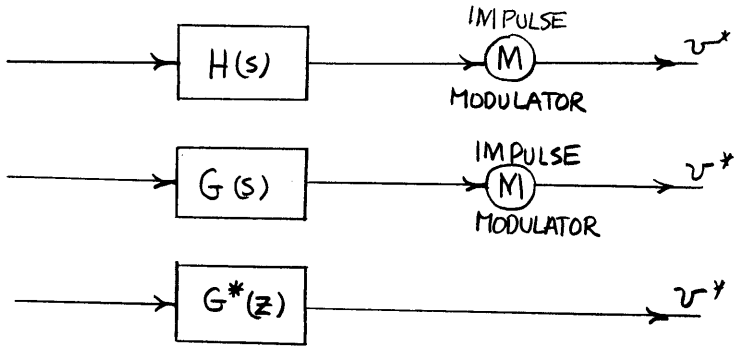


Fig. A-1.3 Various schemes for obtaining a random sequence of impulses, v^*

In (a), the actual random process is sampled. As was discussed

in section 4.4, the continuous values of the state variables between sampling instants are forever lost after sampling. The effect of white noise builds up over the sampling interval, and the best estimate of the continuous state variables is their initial condition decay. The configuration of (b) in Fig. A-1.3 reconstructs these state variables at the sampling instants, and they do decay as initial conditions until the next impulse is received. Accordingly, (b) represents a system which reproduces the desired statistics, and contains signal levels which are totally recoverable. The configuration of (c) neglects the knowable continuous portion of the random process. The various systems are related as follows:

Let $\Phi_{vv}(s)$ be the power density spectrum of the continuous input v, and $\Phi_{vv}^{*}(z)$ be the spectrum of the sampled v. Therefore,

$$H(s)\,H(-s) = \Phi_{vv}(s)$$

$$G^{*}(z)^{-1}\,G^{*}(z) = \Phi_{vv}^{*}(z)$$

$G^{*}(z)$ is a system which can be considered to have an impulse response which is the sampled version of one of a continuous system, G(s).

Figure A-1.4 shows the optimum unity feedback system to recover the knowable signal component of v, with G(s) = S(s) + N(s). $N^{*}(z)$ is the "starred" discrete system which is the sampled version of N(s). The impulse modulator is a mathematical fiction used to represent the process of sampling, converting the values of input signal at the sampling instant

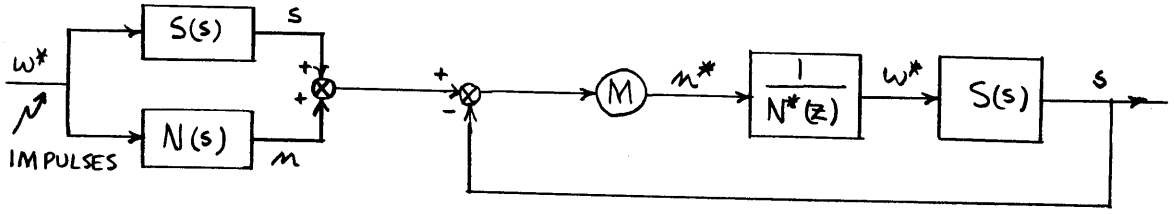


Fig. A-1.4 Optimum error-sampled noise filter

into areas of output impulses. One typical actual output from a "sampler" or a digital-to-analog converter is shown in Fig. A-1.5.

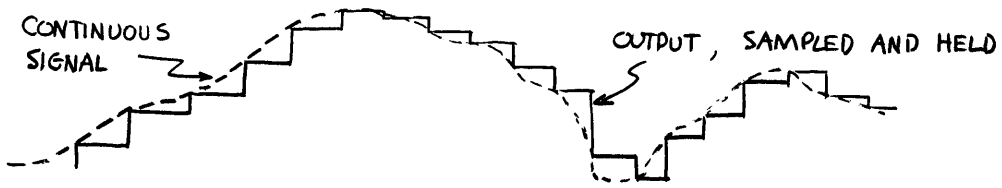


Fig. A-1.5 Typical output of practical sampling device

Cascading the impulse modulator with the transfer function, $\frac{1 - z}{s}$, can account theoretically for this stair-step signal. If discrete compensation is allowable to process the signal in number form, with an output of the sort pictured in Fig. A-1.5, the discrete compensation should be $\frac{1}{N(z)\ (1 - z)}$ and the continuous driven system should be $s\ S(s)$ in order to have the overall optimum forward-loop transmission of Fig. A-1.4.

7. Conclusions

The theory of discrete random processes and systems which act on them has been sketchily shown to be substantially equal to the continuous case. The major deviation occurs when it is necessary to reproduce an optimum continuous signal from samples of a random process. Section 6 of this appendix has presented the optimum feedback filter to perform this operation.

APPENDIX II.

A 3X3 EXAMPLE OF MATRIX FACTORIZATION

To generate the problem, a simple 3x3 system, $\underline{G^*(s)}$, is selected which has an unrealizable inverse, and whose output power density, when excited by white noise, is given by

$$\underline{\Phi_o(s)} = \underline{G^*(-s)} \cdot \underline{G^{*T}(s)} \qquad (2.27)$$

where

$$G^*(s) = \begin{bmatrix} \frac{3}{s+4} & 0 & 0 \\ 0 & \frac{1}{s+1} & \frac{2}{s+3} \\ \frac{2}{s+6} & \frac{1}{s+5} & \frac{1}{s+2} \end{bmatrix} \qquad |G^*(s)| = \frac{3\,(s+2.464)\,(-s+4.464)}{(s+4)\,(s+1)\,(s+2)\,(s+3)\,(s+5)}$$

$$\underline{\Phi_o} = \underline{G^*(-s)}\ \underline{G^{*T}(s)} = \begin{bmatrix} \frac{9}{(-s+4)(s+4)} & 0 & \frac{6}{(-s+4)(s+6)} \\ 0 & \frac{-5s^2+13}{(-s+1)(-s+3)(s+1)(s+3)} & \frac{-3s^2-7s+16}{(-s+1)(-s+3)(s+2)(s+5)} \\ \frac{6}{(-s+6)(s+4)} & \frac{-3s^2+7s+16}{(-s+2)(-s+5)(s+1)(s+3)} & \frac{6s^4-217s^2+1444}{(-s+2)(-s+5)(-s+6)(s+2)(s+5)(s+6)} \end{bmatrix}$$

The problem is to find some $\underline{G(s)}$, where both $\underline{G(s)}$ and $\underline{G^{-1}(s)}$ are physically realizable, such that

$$\underline{G(-s)}\ \underline{G^T(s)} = \underline{\Phi_o(s)} \qquad (2.27)$$

A general solution for this matrix factorization problem has been presented in Section 3.5. The following steps follow the notation and procedure given there.

1. Pole removal phase

$$\underline{T_1(-s)} = \begin{bmatrix} -s+4 & 0 & 0 \\ 0 & (-s+1)(-s+3) & 0 \\ 0 & 0 & (-s+2)(-s+5)(-s+6) \end{bmatrix}$$

$$\underline{\Phi}_1(s) = \underline{T_1}(-s)\,\underline{\Phi}_o(s)\,\underline{T_1}^T(s)$$

$$= \begin{bmatrix} 9 & 0 & 6(s+5)(s+2) \\ 0 & -5s^2 + 13 & (-3s^2 - 7s + 16)(s+6) \\ 6(-s+5)(-s+2) & (-3s^2 + 7s + 16)(-s+6) & 6s^4 - 217s^2 + 1444 \end{bmatrix}$$

2. Determinant reduction phase

$$\left|\underline{\Phi}_1(s)\right| = 9(-s+2.464)(-s+4.464)(-s+6)(s+2.464)(s+4.464)(s+6)$$

$$\underline{T_2}(-s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \dfrac{1}{-s+6} \end{bmatrix}$$

$$\underline{\Phi}_2(s) = \underline{T_2}(-s)\,\underline{\Phi}_1(s)\,\underline{T_2}^T(s)$$

$$= \begin{bmatrix} 9 & 0 & \dfrac{6s^2+42s+60}{s+6} \\ 0 & -5s^2 + 13 & -3s^2 - 7s + 16 \\ \dfrac{6s^2-42s+60}{-s+6} & -3s^2 + 7s + 16 & \dfrac{6s^4 - 217s^2 + 1444}{(s+6)(-s+6)} \end{bmatrix}$$

$$\frac{6s^2 - 42s + 60}{-s+6} = -6s + 6 + \frac{24}{-s+6} \quad ; \quad 9k_1 = -24; \quad k_1 = -\frac{8}{3}$$

$$\underline{T_3}(-s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \dfrac{-\frac{8}{3}}{-s+6} & 0 & 1 \end{bmatrix}$$

$$\underline{\Phi_3(s)} = \underline{T_3(-s)}\,\underline{\Phi_2(s)}\,\underline{T_3{}^T(s)}$$

$$= \begin{bmatrix} 9 & 0 & 6s+6 \\ 0 & -5s^2+13 & -3s^2-7s+16 \\ -6s+6 & -3s^2+7s+16 & -6s^2+33 \end{bmatrix}$$

$$\underline{T_4(-s)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \dfrac{1}{-s+4.464} \end{bmatrix}$$

$$\underline{\Phi_4(s)} = \underline{T_4(-s)}\,\underline{\Phi_3(s)}\,\underline{T_4{}^T(s)}$$

$$= \begin{bmatrix} 9 & 0 & \dfrac{6s+6}{s+4.464} \\ 0 & -5s^2+13 & \dfrac{-3s^2-7s+16}{s+4.464} \\ \dfrac{-6s+6}{-s+4.464} & \dfrac{-3s^2+7s+16}{-s+4.464} & \dfrac{-6s^2+33}{(-s+4.464)(s+4.464)} \end{bmatrix}$$

$$\frac{-6s+6}{-s+4.464} = 6 - \frac{20.784}{-s+4.464} \quad ; \quad 9k_1 = 20.784;\ k_1 = 2.304$$

$$-5s^2+13 \Big|_{s=4.464} = -86.55$$

$$\frac{-3s^2+7s+16}{-s+4.464} = 3s+6.392 - \frac{12.55}{-s+4.464} \quad ; \quad -86.55\,k_2 = 12.55$$

$$k_2 = -.1450$$

$$T_5(-s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \dfrac{2.304}{-s+4.464} & \dfrac{-.1450}{-s+4.464} & 1 \end{bmatrix}$$

$$\underline{\Phi_5^{(s)}} = \underline{T_2(-s)}\ \underline{\Phi_4^{(s)}}\ \underline{T_s{}^T(s)}$$

$$= \begin{bmatrix} 9 & 0 & 6 \\ 0 & -5s^2 + 13 & -2.275\,s + 3.156 \\ 6 & 2.275\,s + 3.156 & 5.227 \end{bmatrix}$$

$$\underline{T_6(-s)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{-s+2.464} \end{bmatrix}$$

$$\underline{\Phi_6^{(s)}} = \underline{T_6(-s)}\ \underline{\Phi_5^{(s)}}\ \underline{T_6{}^T(s)}$$

$$= \begin{bmatrix} 9 & 0 & \frac{6}{s+2.464} \\ 0 & -5\,s^2 + 13 & \frac{-2.275s + 3.156}{s+2.464} \\ \frac{6}{-s+2.464} & \frac{2.275\,s + 3.156}{-s+2.464} & \frac{5.227}{(-s+2.464)(s+2.464)} \end{bmatrix}$$

$$9\,k_1 = -6\ ; \quad k_1 = -\frac{2}{3}\ ; \quad -5\,s^2 + 13 \Big|_{s = 2.464} = -17.35$$

$$\frac{2.275\,s + 3.156}{-s + 2.464} = -2.275 + \frac{8.765}{-s + 2.464}\ ; \quad -17.35\,k_2 = -8.765$$

$$k_2 = .505$$

$$\underline{T_7(-s)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{-\frac{2}{3}}{-s+2.464} & \frac{.505}{-s+2.464} & 1 \end{bmatrix}$$

$$\underline{\Phi}_7(s) = \underline{T}_7(-s)\ \underline{\Phi}_6(s)\underline{T}_7^T(s)$$

$$= \begin{bmatrix} 9 & 0 & 0 \\ 0 & -5s^2 + 13 & -2.525s + 4.00 \\ 0 & 2.525s + 4.00 & 1.295 \end{bmatrix}$$

$$\left|\underline{\Phi}_7(s)\right| = 9$$

3. Element order reduction phase

$$\underline{T}_8(-s) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1.95s \\ 0 & 0 & 1 \end{bmatrix}$$

$$\underline{\Phi}_8(s) = \underline{T}_8(-s)\ \underline{\Phi}_7(s)\ \underline{T}_8^T(s)$$

$$= \begin{bmatrix} 9 & 0 & 0 \\ 0 & 13 & 4.0 \\ 0 & 4.0 & 1.295 \end{bmatrix} = \underline{N} \cdot \underline{N}^T$$

Using the canonic triangular form of Section 3.3,

$$\underline{N} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3.61 & 0 \\ 0 & 1.108 & 2.77 \end{bmatrix}$$

The solution for $\underline{G(s)}$ is given by

$$\underline{G(s)} = \underline{T}_1^{-1}(s)\ \underline{T}_2^{-1}(s) \ldots . \underline{T}_8^{-1}(s)\ \underline{N} \qquad (3.5)$$

All the inverse matrices can be determined by inspection. Alter multiplication,

$$\underline{G(s)} = \begin{bmatrix} \frac{3}{s+4} & 0 & 0 \\ 0 & \frac{2.16\,(s+1.67)}{(s+1)\ (s+3)} & \frac{.540\,s}{(s+1)\ (s+3)} \\ \frac{2}{s+6} & \frac{1.28\,(s+3.55)}{(s+2)\ (s+5)} & \frac{.776\,(s+3.93)}{(s+2)\ (s+5)} \end{bmatrix}$$

To check the accuracy of the solution,

$$\underline{G(-s)\ G^T(s)} = \begin{bmatrix} \frac{9}{(-s+4)(s+4)} & 0 & \frac{6}{(-s+4)(s+6)} \\ 0 & \frac{-4.962s^2+13.07}{(-s+1)(-s+3)(s+1)(s+3)} & \frac{-3.189s^2-6.866s+16.43}{(-s+1)(-s+3)(s+2)(s+5)} \\ \frac{6}{(-s+6)(s+4)} & \frac{-3.189s^2+6.866s+16.43}{(-s+2)(-s+5)(s+1)(s+3)} & \frac{6.242s^4-226.9s^2+1480}{(-s+2)(-s+5)(-s+6)(s+2)(s+5)(s+6)} \end{bmatrix}$$

which is compared with the original $\underline{\Phi_o}$ matrix

$$\begin{bmatrix} \frac{9}{(-s+4)(s+4)} & 0 & \frac{6}{(-s+4)(s+6)} \\ 0 & \frac{-5s^2+13}{(-s+1)(-s+3)(s+1)(s+3)} & \frac{-3s^2-7s+16}{(-s+1)(-s+3)(s+2)(s+5)} \\ \frac{6}{(-s+6)(s+4)} & \frac{-3s^2+7s+16}{(-s+2)(-s+5)(s+1)(s+3)} & \frac{6s^4-217s^2+1444}{(-s+2)(-s+5)(-s+6)(s+2)(s+5)(s+6)} \end{bmatrix}$$

The differences between the desired and actual results reflect the mortality of the author, and do not indicate the accuracy of the method. The resulting $\underline{G(s)}$ has a realizable inverse, since

$$\underline{G(s)}^{-1} = \underline{N}^{-1}\ \underline{T_8(s)} \ldots\ldots \underline{T_1(s)} \qquad (3.6)$$

and each of the $\underline{T_i(s)}$ is obviously realizable.

BIBLIOGRAPHY

1. N. Wiener, "The Extrapolation, Interpolation, and Smoothing of Stationary Time Series", Technology Press, Cambridge, Mass., 1949.

2. A.N. Kolmogorov, "Interpolyatsiya i ekstrapolyatsiya statsionarnykh sluchaynykh posledovatelnostey" (Interpolation and extrapolation of stationary random sequences), Izvestiya AN SSSR, ser. matem., No. 5, 1941.

3. L.A. Zadeh and J.R. Ragazzini, "An Extension of Wiener's Theory of Prediction", J. Appl. Phys., Vol. 21, pp. 645-655, 1950.

4. H.W. Bode and C.E. Shannon, "A Simplified Derivation of Linear Least-Squares Smoothing and Prediction Theory", Proc. IRE, Vol. 38, pp. 417-425, 1950.

5. M. Blum, "Generalization of the Class of Non-random Inputs of the Zadeh-Ragazzini Prediction Model", IRE PGIT Trans., Vol. IT-2, No. 2, pp. 76-81, 1956.

6. Y.W. Lee, "Application of Statistical Methods to Communications Problems", Research Lab. of Electronics Rept. No. 181, Mass. Inst. of Tech., Cambridge, Mass., 1950.

7. E.W. Pike, "A New Approach to Optimum Filtering", Proc. Nat. Electronics Conf., 1952, Vol. 8, pp. 407-418, 1953.

8. G.C. Newton, Jr., "Compensation of Feedback Control Systems Subject to Saturation", J. Franklin Inst., Vol. 254, pp. 281-286, 391 - 413, 1952.

9. J.G. Truxal, "Automatic Feedback Control System Synthesis", McGraw-Hill Book Co., New York, 1955.

10. G.C. Newton, Jr., L.A. Gould, and J.F. Kaiser, "Analytical Design of Linear Feedback Controls", John Wiley and Sons, Inc., New York, 1957.

11. O.J.M. Smith, "Feedback Control Systems", McGraw-Hill Book Co., New York, 1958.

12. W.W. Seifert and C.W. Steeg, Jr., Editors, "Control Systems Engineering", McGraw-Hill Book Co., New York, 1960.

13. J.H. Laning and R.H. Battin, "Random Processes in Automatic Control", McGraw-Hill Book Co., New York, 1956.

14. J.H. Westcott, "Design of Multi-Variable Optimum Filters", Trans. ASME, Vol. 80, pp. 463-467, 1958.

15. R.C. Amara, "Application of Matrix Methods to the Linear Least Squares Synthesis of Multi-Variable Systems", J. Franklin Inst., Vol. 268, pp. 1 - 16, 1959.

16. H.C. Hsieh and C.T. Leondes, "On the Optimum Synthesis of Multipole Control Systems in the Wiener Sense", IRE PGAC Trans., Vol. AC-4, No. 2, pp. 16-29, November 1959.

17. W.R. Evans, "Control System Dynamics", McGraw-Hill Book Co., New York, 1954.

18. R.A. Summers, "A Statistical Description of Large-Scale Atmospheric Turbulence", Rept. T-55, Instrumentation Lab., MIT, Cambridge, Mass., 1954.

19. G. Kraus and H. Potzl, "Limiting Conditions on the Correlation Properties of Random Signals", IRE PGCT Trans., Vol. CT-3, No. 4, Dec. 1956.

20. R. Bellman, "Introduction to Matrix Analysis", McGraw-Hill Book Co., New York, 1960.

21. H.C. Lee, "Canonical Factorization of Non-negative Hermitian Matrices", J. London Math. Soc., Vol. 23, pp. 100-110, 1948.

22. C.E. Cullis, "Matrices and Determinoids, Vol. III, Part 1", University Press, Cambridge, 1925.

23. J.R. Ragazzini and G.F. Franklin, "Sampled-Data Control Systems", McGraw-Hill Book Co., New York, 1958.

24. R.E. Kalman and R.S. Bucy, "New Results in Linear Filtering and Prediction Theory"(Preprint), ASME Paper 60 - JAC-12, presented at the Joint Automatic Controls Conf., Cambridge, Mass., Sept., 1960.

25. A.G. Bose, "A Theory of Nonlinear Systems", MIT Research Lab. of Electronics Report No. 309, Cambridge, Mass., 1956.

26. N. Wiener, "Nonlinear Problems in Random Theory", John Wiley and Sons, New York, 1958.

27. E.B. Lee, "Mathematical Aspects of the Syn thesis of Linear Minimum Response-Time Controllers", IRE PGAC Trans., Vol. AC-5, No. 4, Sept., 1960.

28. R. Bellman, I. Glicksberg, and O. Gross, "On the Bang-Bang Control Problem", Quart. J. Appl. Math., Vol. 14, pp. 11-18, 1956.

29. R. Bellman, "Dynamic Programming", Princeton University Press, Princeton, N.J., 1957.

30. T.R. Benedict, "Predictor-Relay Servos with Random Inputs", Proc. of Nat. Auto. Control Conf., 1959, IRE PGAC Trans., Vol. AC-4, No. 3, pp. 232-245, Dec., 1959.

31. A.M. Hopkin and P.K.C. Wang, "A Relay-Type Feedback Control System Design for Random Inputs", AIEE Trans., Appl. and Ind., No. 44, pp. 228-233, Sept., 1959.

32. R.E. Kalman and J.E. Bertram, "Control System Analysis of Design via the Second Method of Lyapunov", Pts. I and II, ASME Trans., Vol. 82, pp. 371-400, June, 1960.

33. R.W. Bass, discussion of a paper by A.M. Letov, Proc. Heidelberg Conf. on Automatic Control ("Regelungstechnik; Moderne Theorien und ihre Verwendbarkeit", by R. Oldendourf, Munich, 1957), pp. 209-210.

34. S.O. Rice, "Mathematical Analysis of Random Noise", Bell System Tech. J., Vol. 23, pp. 282-332, 1944, and Vol. 24, pp. 46-156, 1945.

35. V.V. Solodovnikov, "Introduction to the Statistical Dynamics of Automatic Control Systems", translated from Russian by J.B. Thomas and L.A. Zadeh, Dover Publications, Inc., New York, 1960.

BIOGRAPHICAL NOTE

Michael C. Davis was born in Fullerton, California on October 12, 1931. He is married to the former Beverly Citrano, and has two sons, Michael Jr., 5, and Mark, 3.

After completion of high-school and preparatory school in Long Beach, California, Lt. Davis attended the U.S. Naval Academy at Annapolis, Md., graduating in June 1953 with the degree of Bachelor of Science and with a commission in the U.S. Navy.

He served as Gunnery Officer aboard the destroyer USS SHELTON (DD 790) and as Missile Guidance Officer aboard the guided-missile submarine USS TUNNY (SSG 282). He was designated "qualified in Submarines" in June 1957. Subsequently, he reported to the Massachusetts Institute of Technology for instruction in Naval Construction and Engineering.

Upon graduation, Lt. Davis will be designated as an Engineering Duty officer. He is a member of the Institute of Radio Engineers, the American Institute of Electrical Engineers, and the Society of Naval Architects and Marine Engineers.