1966

Multiple-input distortionless filters for estimating signals corrupted by noise

Roger David Benning
Iowa State University

This dissertation has been
microfilmed exactly as received 66-10,403

BENNING, Roger David, 1941-
MULTIPLE-INPUT DISTORTIONLESS FILTERS FOR ESTIMATING SIGNALS CORRUPTED BY NOISE.

Iowa State University of Science and Technology, Ph.D., 1966
Engineering, electrical

University Microfilms, Inc., Ann Arbor, Michigan

MULTIPLE-INPUT DISTORTIONLESS FILTERS

FOR ESTIMATING SIGNALS CORRUPTED BY NOISE

by

Roger David Benning

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Electrical Engineering

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1966

TABLE OF CONTENTS

		page
I.	INTRODUCTION	
	A. Statement of the Problem	1
	B. Review of Literature	9
II.	A USEFUL RELATIONSHIP FOR AN n+1 BY n MATRIX	11
	A. Notation	11
	B. Theorems 1 and 2	12
III.	THE OPTIMUM LINEAR DISTORTIONLESS FILTER	25
	A. Comments on Finite Operating Time Filters	25
	B. Derivation of the Integral Equations	28
IV.	THE INTUITIVE FILTER	38
	A. Generalized r-Dimensional Wiener Filter	38
	B. The Linear, Algebraic Operator	41
	C. The Integral Equations	49
V.	COMPARISON OF THE OPTIMUM AND INTUITIVE SYSTEMS	54
VI.	AN EXAMPLE USING THE WIENER FILTER	66
	A. The Optimum Filter	66
	B. First "Intuitive" System	72
	C. Second "Intuitive" System	75
VII.	AN EXAMPLE USING THE KALMAN FILTER	80
	A. The Kalman Filter Equations	82
	B. The Example	86
VIII.	SUMMARY	96
IX.	LITERATURE CITED	106
X.	ACKNOWLEDGMENTS	107

I. INTRODUCTION

A. Statement of the Problem

A problem sometimes encountered in systems work is that of making an optimum estimate of a signal s(t) when two or more independent sources, each of which is corrupted by noise, are available. If it is assumed that nothing is known a priori about the signal and that the system used to estimate it is nonadaptive, then no matter what optimization scheme is used, it may not depend in any way on the nature of the signal s(t).

As an example of this problem, suppose that we have available the two independent sources or inputs $s(t) + n_1(t)$ and $s(t) + n_2(t)$, where the noise functions $n_1(t)$ and $n_2(t)$ are assumed to be time-stationary random functions with known spectral density functions. Any linear, constant parameter system used to estimate s(t) from these inputs may be represented by the system shown in Figure 1. A simple way to avoid having s(t) appear in the optimization equations is to restrict ourselves to the class of systems which reproduce s(t) exactly in the absence of noise. For the simple system of Figure 1, this places the following constraint between Y_1 and Y_2:

$$Y_2(s) = 1 - Y_1(s) \tag{1.1}$$

The expression for the output is then

$$X = Y_1(S + N_1) + Y_2(S + N_2)$$

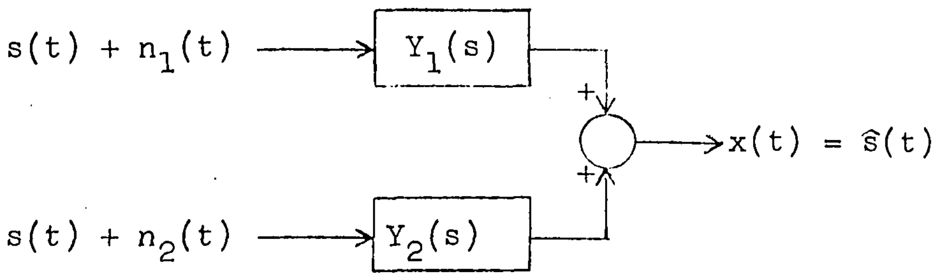


Figure 1. Linear system used to estimate s(t)

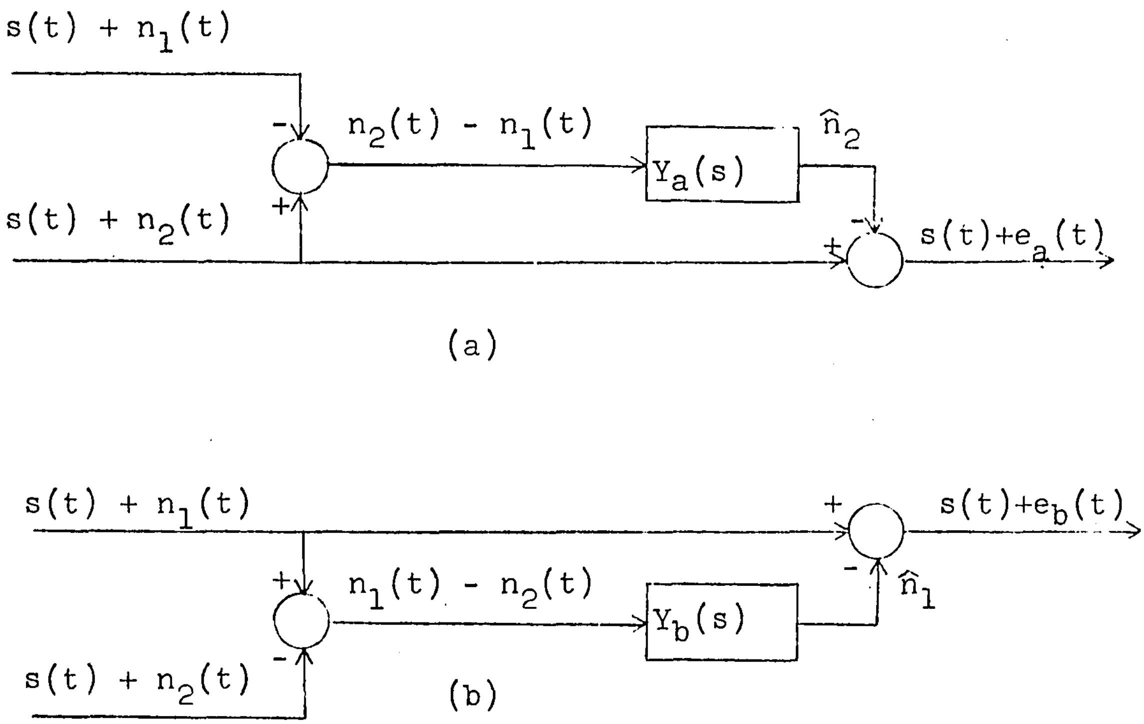


Figure 2. The two "intuitive" systems for estimating s(t)

$$= S + [N_1Y_1 + N_2(1 - Y_1)] \tag{1.2}$$

With the error defined as the difference between x(t) and s(t), it is observed that the quantity within the brackets of 1.2 is the transform of the error and that the choice of Y_1 will not affect the signal portion of the output. Furthermore, in the complete absence of noise, the output is exactly equal to the signal, as desired. Thus, we do not have to accept signal distortion as a consequence of smoothing the noise. For this reason we might refer to this as "distortionless" filtering[1].

An alternate, and perhaps more intuitive, way of estimating s(t) from the same inputs is shown in Figure 2(a). Letting $E_a(s)$ represent the Laplace transform of the error associated with estimating s(t) for this system, the expression for E_a is given by

$$\begin{aligned} E_a &= - [(N_2 - N_1)Y_a - N_2] \\ &= N_1Y_a + N_2(1 - Y_a) \end{aligned} \tag{1.3}$$

And, the over-all transfer functions between input line 1 and the output and between input line 2 and the output are given by $Y_a(s)$ and $1 - Y_a(s)$, respectively. Comparing Equations 1.3 and the error terms of 1.2, we see that merely specifying the same optimization criterion for the two systems involved will insure that $Y_a(s) = Y_1(s)$. This in turn implies that the systems of Figure 1 and Figure 2(a) are equivalent, even

---

[1]See for example Chapter 15 of Brown and Nilsson (5).

though they may differ in their physical configuration.

Still another method of estimating $s(t)$ is shown in Figure 2(b). The expression for $E_b(s)$ is given by

$$E_b = - [(N_1-N_2)Y_b - N_1]$$
$$E_b = N_1(1-Y_b) + N_2Y_b \tag{1.4}$$

The overall transfer functions between lines 1 and 2 and the output are given by $1 - Y_b(s)$ and $Y_b(s)$, respectively. Using the same optimization criterion as before, it can be shown that $Y_b(s) = Y_2(s)$. Consequently, this system is equivalent to both the system of Figure 1 and the system of Figure 2(a).

In particular, if minimization of the mean-square error is chosen as the optimization criterion, $Y_a(s)$ is the Wiener filter associated with estimating $n_2(t)$ from the input $n_2(t) - n_1(t)$. Similarly, $Y_b(s)$ is the Wiener filter associated with estimating $n_1(t)$ from the input $n_1(t) - n_2(t)$. The two systems of Figure 2 may then be thought of as reducing the original problem, which involved an unspecified signal in both the inputs and the output, to the more familiar Wiener filter problem.

The purpose of this thesis can be described, approximately, as the extension of the above concepts to higher dimension. To formalize the statement of the problem, consider the problem of estimating the signals $s_1(t),\ldots,s_m(t)$ from the n available input lines shown in Figure 3. We will make the

$$f_1(t) = a_{11}(t)s_1(t) + \ldots + a_{1m}(t)s_m(t) + n_1(t) \longrightarrow$$

$$- - - - - - - - - - - - - - - - - - - -$$

$$f_m(t) = a_{m1}(t)s_1(t) + \ldots + a_{mm}(t)s_m(t) + n_m(t) \longrightarrow$$

$$- - - - - - - - - - - - - - - - - - - -$$

$$f_n(t) = a_{n1}(t)s_1(t) + \ldots + a_{nm}(t)s_m(t) + n_n(t) \longrightarrow$$

Figure 3. The available input lines

assumption that nothing is known about signals $s_1(t),\ldots,s_m(t)$ and that $n_1(t),\ldots,n_n(t)$ are nonstationary, random noise inputs with known autocorrelation functions. The noises are assumed mutually independent. It is also assumed that $a_{ij}(t)$ for $i = 1, 2,\ldots,n$ and $j = 1, 2,\ldots,m$ are known functions of time and that $n > m$.

Any system that might be used to estimate the signals $s_1(t),\ldots,s_m(t)$ may be represented by the n input, m output "black box" of Figure 4. In this thesis the filter shown is constrained to be linear, physically realizable[1], and distortionless and is allowed to operate on only a finite amount of past data. By physically realizable we mean simply that it is not allowed to operate on any future data in making the estimates for time t. The distortionless constraint requires that the system reproduce $s_1(t),\ldots,$ $s_m(t)$ exactly in the event that all the noises are identically

[1] Causal is a more modern term for this.

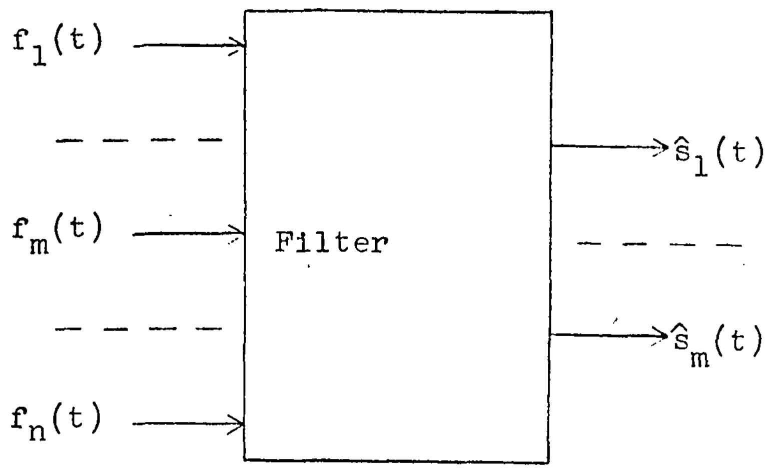


Figure 4. Block diagram of the general multiple-input, multiple-output filter for this problem

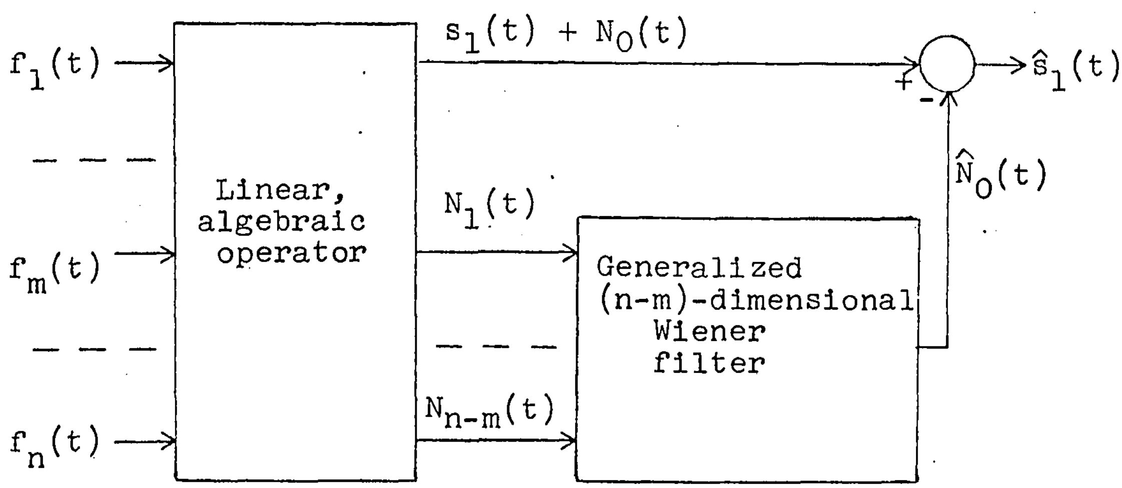


Figure 5. Block diagram of the "intuitive" system

zero. The desired filter is to be optimum in the sense that it minimizes each of the mean-square errors associated with estimating $s_1(t),\ldots,s_m(t)$. Since the errors are in general nonstationary, the mean or averaging here is to be taken in the ensemble sense. In particular the set of integral equations for that part of the filter of Figure 4 which satisfies the above requirements and estimates $s_1(t)$ is developed in Chapter III. The filter specified by these integral equations shall hereafter be referred to as the "optimum" filter. If desired the integral equations for the filter which estimates $s_k(t)$ may be found from the above mentioned integral equations by an appropriate change of subscripts, but for the purposes of the discussion in this thesis it is sufficient to talk about that part of the filter which estimates $s_1(t)$.

An alternate system for estimating $s_1(t)$ from the inputs shown in Figure 3 is now suggested and is shown in block diagram form in Figure 5. This system employs a linear, algebraic operator having the n-m+1 outputs shown in Figure 5, where $N_i(t) = \sum_{j=1}^{n} c_{ij}(t)n_j(t)$ for $i = 0,1,\ldots,n-m$, and each $c_{ij}(t)$ is a known function of time. The filter part of the system has as its n-m inputs the nonstationary, random noises $N_1(t),\ldots,N_{n-m}(t)$ and is assumed to make a minimum mean-square estimate of $N_0(t)$. Thus the filter is a generalized (n-m)-dimensional Wiener filter (see the next section for what is meant by this term). This system shall be referred to as the "intuitive" system.

The objective of the thesis is to show that under suitable assumptions on the A matrix, where $A = [a_{ij}(t)]$, $i = 1,\ldots,n$, $j = 1,\ldots,m$, the "intuitive" system is an optimum system for estimating $s_1(t)$. The assumptions on A amount to certain conditions of linear independence on the rows of A and are discussed later.

It is usually possible to choose the linear, algebraic operator shown in Figure 5 in quite a number of different ways, with the number of ways depending on the A matrix. To each choice of the linear, algebraic operator there corresponds a generalized (n-m)-dimensional Wiener filter. Since each of these choices constitutes a different system, we see that there are usually quite a number of possible "intuitive" systems. If we can show that each of these represents an optimum solution to the original problem, then it will follow that all these "intuitive" systems have the same mean-square error. Or, in other words, all the "intuitive" systems are equally good. This is an important result in its own right and, as a matter of fact, is what motivated this thesis.

It is interesting to point out at this time that for continuous operating systems the "intuitive" system does not offer any real advantage over the "optimum" system. That is, the set of integral equations that we get for the former are just as difficult to solve as the set we get for the latter. The real advantage of the "intuitive" system is that it lends itself to the discrete analog of the multidimensional, generalized Wiener filter, namely the Kalman filter, whereas the problem in its original form does not lend itself to this technique. This, of course, assumes that the noises are such that they can be generated by the use of shaping filters with white-noise inputs. The Kalman filter is devised specifically for a digital computer solution and has the advantage of handling a multiple-input problem almost as simply as a single-input problem, the only complexity added being the size of the matrices involved. This technique is discussed in Chapter 7 where a fairly general example with 2 signals and 3 input lines is treated.

B. Review of Literature

Quite a number of books and articles have been written on random processes in the years since World War II. The basic filter of the type that is of interest here was first developed by Wiener in 1942 and published in a classified report to Section D_2, National Defense Research Committee. It was later released for general use and published in a book entitled "Extrapolation, Interpolation and Smoothing of Stationary Time Series" by Norbert Wiener (9). In this book, Wiener treated only the case of time stationary inputs and considered constant parameter, linear, infinite operating time filters.

The basic type of filter considered by Wiener was later generalized in varying degrees by several authors as to the type of input(s) allowed and the constraints imposed on the filter. A convenient table of these generalizations is presented on page 150 of Bendat (3). Of primary interest here is the most general of these, namely the time varying parameter, linear, finite operating time filter with nonstationary random noise inputs. This case was first treated by Dolph and Woodbury (6), but was also considered by Zadeh (11) and Bendat (2). It is interesting to note that there is not any great difference in developing the integral equations for the various cases, but that each new generalization brought with it certain inherent difficulties in solving these equations. This was the primary reason for treating the various cases separately. For the purposes of this thesis, all of the above types of systems are referred to as "generalized Wiener filters".

Other than containing the basic theory in one form or another, most of the books and articles deal with topics that are of interest only to certain phases of the problem treated here and will be referred to throughout the thesis as the need arises.

II. A USEFUL RELATIONSHIP FOR AN n+1 BY n MATRIX

In this chapter a relationship involving the determinants of certain n by n and n-1 by n-1 submatrices of an n+1 by n matrix will be stated in the form of a theorem and proven. This relationship will be very useful later in reducing the form of the integral equations specifying the filter for the "intuitive" system. Before proceeding to this theorem, it is convenient to introduce some notation which will be used throughout the remainder of this thesis.

A. Notation

Following the usual matrix notation, an upper case letter will be used to represent a matrix (not necessarily square) and the corresponding small letter with two subscripts will denote one of its entries. Thus, a_{ij} is the entry in the i^{th} row and j^{th} column of the matrix A.

The determinant of the m by m matrix formed by selecting rows $i_1, \ldots, i_m$ and columns $j_1, \ldots, j_m$ from matrix A will be denoted by

$$\left| A \begin{matrix} j_1, \ldots, j_m \\ i_1, \ldots, i_m \end{matrix} \right|$$

where $i_1 < i_2 < \ldots < i_m$ and $j_1 < j_2 < \ldots < j_m$. When the integers $i_1, \ldots, i_m$ are consecutive with $1 \leq k \leq m$ it will be convenient to use the notation

$$\left|A^{j_1,\dots,j_m}_{i_1,\dots,i_m,i_{m+1}}\right|_{\text{exc. } i_k}$$

to denote the determinant of the matrix made up of rows $i_1,\dots,$ i_{k-1}, $i_{k+1},\dots,i_m$, i_{m+1} and columns $j_1,\dots,j_m$ of A. This notation will be extended to indicate that two or more rows are left out or to indicate that one or more columns are left out from a selection of consecutive rows or columns of A, respectively.

B. Theorems 1 and 2

Theorem 1. Let B be an n+1 by n matrix where $n \geq 2$ and p, i, and q be integers such that $1 \leq p < i < q \leq n+1$. Then,

$$\left|B^{2,\dots,n}_{1,\dots,n+1}\right|_{\text{exc. } i,q} \cdot \left|B^{1,\dots,n}_{1,\dots,n+1}\right|_{\text{exc. } p} + \left|B^{2,\dots,n}_{1,\dots,n+1}\right|_{\text{exc. } p,i} \cdot \left|B^{1,\dots,n}_{1,\dots,n+1}\right|_{\text{exc. } q}$$

$$- \left|B^{2,\dots,n}_{1,\dots,n+1}\right|_{\text{exc. } p,q} \cdot \left|B^{1,\dots,n}_{1,\dots,n+1}\right|_{\text{exc. } i} = 0 \qquad (2.1)$$

Proof: Define the matrix D in terms of the matrix B as follows:

row 1 of D = row p of B

rows $2,\dots,p$ of D = rows $1,\dots,p-1$ of B, respectively

rows $p+1,\dots,q-1$ of D = rows $p+1,\dots,q-1$ of B, respectively

rows $q,\dots,n$ of D = rows $q+1,\dots,n+1$ of B, respectively

row n+1 of D = row q of B

In terms of matrix D, Equation 2.1 is true if

$$\begin{aligned}&\underset{\text{exc. } i}{\left|D^{2,\ldots,n}_{1,\ldots,n}\right|}\cdot\left|D^{1,\ldots,n}_{2,\ldots,n+1}\right| + \underset{\text{exc. } i}{\left|D^{2,\ldots,n}_{2,\ldots,n+1}\right|}\cdot\left|D^{1,\ldots,n}_{1,\ldots,n}\right|\\ &- \left|D^{2,\ldots,n}_{2,\ldots,n}\right|\cdot\underset{\text{exc. } i}{\left|D^{1,\ldots,n}_{1,\ldots,n+1}\right|} = 0\end{aligned} \tag{2.2}$$

Consequently, to prove 2.1 it is sufficient to prove statement 2.2. Expanding the determinants of the three n by n matrices of 2.2 about column 1 yields

$$\begin{aligned}&\underset{\text{exc. } i}{\left|D^{2,\ldots,n}_{1,\ldots,n}\right|}\left\{\sum_{j=2}^{n+1}(-1)^{j}d_{j1}\underset{\text{exc. } j}{\left|D^{2,\ldots,n}_{2,\ldots,n+1}\right|}\right\}\\ &+\underset{\text{exc. } i}{\left|D^{2,\ldots,n}_{2,\ldots,n+1}\right|}\left\{\sum_{j=1}^{n}(-1)^{j+1}d_{j1}\underset{\text{exc. } j}{\left|D^{2,\ldots,n}_{1,\ldots,n}\right|}\right\}\\ &-\left|D^{2,\ldots,n}_{2,\ldots,n}\right|\left\{\sum_{j=1}^{i-1}(-1)^{j+1}d_{j1}\underset{\text{exc. } i,j}{\left|D^{2,\ldots,n}_{1,\ldots,n+1}\right|}\right.\\ &\left.+\sum_{j=i+1}^{n+1}(-1)^{j}d_{j1}\underset{\text{exc. } i,j}{\left|D^{2,\ldots,n}_{1,\ldots,n+1}\right|}\right\} = 0\end{aligned} \tag{2.3}$$

By simply collecting the coefficients of d_{j1} such that the left-hand side of 2.3 is written as

$$\sum_{j=1}^{n+1}(\text{coefficient of } d_{j1})d_{j1}$$

it is found that the coefficients of d_{11}, $d_{n+1,1}$ and d_{i1} are zero. This completes the proof of Theorem 1 for n = 2. For $n \geq 3$, j included in the set of integers $(2,\ldots,n)$ and

$j \neq i$, the coefficient of d_{j1} is given by the left-hand side of 2.4 below. Therefore, statement 2.3 is an equality if

$$\left|D^{2,\ldots,n}_{\substack{2,\ldots,n+1\\ \text{exc. } i}}\right| \cdot \left|D^{2,\ldots,n}_{\substack{1,\ldots,n\\ \text{exc. } j}}\right| - \left|D^{2,\ldots,n}_{\substack{2,\ldots,n+1\\ \text{exc. } j}}\right| \cdot \left|D^{2,\ldots,n}_{\substack{1,\ldots,n\\ \text{exc. } i}}\right|$$

$$\mp \left|D^{2,\ldots,n}_{2,\ldots,n}\right| \cdot \left|D^{2,\ldots,n}_{\substack{1,\ldots,n+1\\ \text{exc. } j,i}}\right| = 0 \qquad (2.4)$$

where the minus sign is used if $j < i$ and the plus sign is used if $j > i$. If statement 2.4 with the minus sign assumed can be shown to be an equality for $j < i$, it will follow that 2.4 with the plus sign assumed is an equality for $i < j$ by simply interchanging i and j. Since column 1 of D does not appear at all in 2.4, it is convenient to define a new $n+1$ by $n-1$ matrix A to be matrix D with column 1 deleted and to state and prove the following theorem for matrix A. Note that the proof of Theorem 2 will imply that 2.4 is an equality, which in turn will complete the proof of Theorem 1.

<u>Theorem 2</u>. Let A be an $n+1$ by $n-1$ matrix ($n \geq 3$) and p and q be integers such that $2 \leq q < p \leq n$. Then

$$\left|A^{1,\ldots,n-1}_{\substack{2,\ldots,n+1\\ \text{exc. } p}}\right| \cdot \left|A^{1,\ldots,n-1}_{\substack{1,\ldots,n\\ \text{exc. } q}}\right| - \left|A^{1,\ldots,n-1}_{\substack{2,\ldots,n+1\\ \text{exc. } q}}\right| \cdot \left|A^{1,\ldots,n-1}_{\substack{1,\ldots,n\\ \text{exc. } p}}\right|$$

$$- \left|A^{1,\ldots,n-1}_{2,\ldots,n}\right| \cdot \left|A^{1,\ldots,n-1}_{\substack{1,\ldots,n+1\\ \text{exc. } q,p}}\right| = 0 \qquad (2.5)$$

Proof: The proof of Theorem 2 is accomplished by mathematical induction by first proving that 2.5 is an equality for n = 3 and n = 4, then expanding 2.5 by the Laplace expansion for $n \geq 5$, and finally showing that 2.5 is true for n if it is true for n-2.

(1) Statement 2.5 can easily be shown to be true for n = 3 by direct expansion. Notice that p and q must be chosen as q = 2, p = 3. In the interest of brevity, the expansion is not shown here. But the conclusion is that for any 4 by 2 matrix A,

$$\left|A^{1,2}_{2,4}\right| \cdot \left|A^{1,2}_{1,3}\right| - \left|A^{1,2}_{3,4}\right| \cdot \left|A^{1,2}_{1,2}\right| - \left|A^{1,2}_{2,3}\right| \cdot \left|A^{1,2}_{1,4}\right| = 0 \qquad (2.6)$$

(2) For n = 4, we can again prove statement 2.5 by direct expansion. We choose to expand each determinant of the 3 by 3 matrices in 2.5 along the i^{th} row, where $i \neq 1$, q, p, 5. For example, if we choose q = 2 and p = 4, then i = 3, and the left-hand side of 2.5 becomes

$$\begin{aligned}
&\left(a_{31}\left|A^{2,3}_{2,5}\right| - a_{32}\left|A^{1,3}_{2,5}\right| + a_{33}\left|A^{1,2}_{2,5}\right|\right)\left(a_{31}\left|A^{2,3}_{1,4}\right| - a_{32}\left|A^{1,3}_{1,4}\right| + a_{33}\left|A^{1,2}_{1,4}\right|\right) \\
&-\left(a_{31}\left|A^{2,3}_{4,5}\right| - a_{32}\left|A^{1,3}_{4,5}\right| + a_{33}\left|A^{1,2}_{4,5}\right|\right)\left(a_{31}\left|A^{2,3}_{1,2}\right| - a_{32}\left|A^{1,3}_{1,2}\right| + a_{33}\left|A^{1,2}_{1,2}\right|\right) \\
&-\left(a_{31}\left|A^{2,3}_{2,4}\right| - a_{32}\left|A^{1,3}_{2,4}\right| + a_{33}\left|A^{1,2}_{2,4}\right|\right)\left(a_{31}\left|A^{2,3}_{1,5}\right| - a_{32}\left|A^{1,3}_{1,5}\right| + a_{33}\left|A^{1,2}_{1,5}\right|\right)
\end{aligned} \qquad (2.7)$$

After multiplying and collecting terms, the coefficients of a_{31}^2, a_{32}^2, and a_{33}^2 are all of the form of 2.6 and are therefore zero. The coefficient of $a_{31}a_{32}$ is

$$- \left|A^{2,3}_{2,5}\right| \cdot \left|A^{1,3}_{1,4}\right| - \left|A^{2,3}_{1,4}\right| \cdot \left|A^{1,3}_{2,5}\right| + \left|A^{2,3}_{4,5}\right| \cdot \left|A^{1,3}_{1,2}\right|$$

$$+ \left|A^{2,3}_{1,2}\right| \cdot \left|A^{1,3}_{4,5}\right| + \left|A^{2,3}_{2,4}\right| \cdot \left|A^{1,3}_{1,5}\right| + \left|A^{2,3}_{1,5}\right| \cdot \left|A^{1,3}_{2,4}\right| \qquad (2.8)$$

Notice that row 3 of A does not appear in 2.8; therefore, it is convenient to define a 4 by 3 matrix E which is made up of rows 1, 2, 4, and 5 of A. In terms of matrix E, 2.8 can be written

$$- \left|E^{a,b}_{2,4}\right| \cdot \left|E^{c,d}_{1,3}\right| - \left|E^{a,b}_{1,3}\right| \cdot \left|E^{c,d}_{2,4}\right| + \left|E^{a,b}_{3,4}\right| \cdot \left|E^{c,d}_{1,2}\right|$$

$$+ \left|E^{a,b}_{1,2}\right| \cdot \left|E^{c,d}_{3,4}\right| + \left|E^{a,b}_{2,3}\right| \cdot \left|E^{c,d}_{1,4}\right| + \left|E^{a,b}_{1,4}\right| \cdot \left|E^{c,d}_{2,3}\right| \qquad (2.9)$$

where a = 2, b,d = 3, and c = 1.

The coefficient of $a_{31}a_{33}$ is given by the negative of the quantity in 2.9 with a,d = 2, b = 3, and c = 1 and the coefficient of $a_{32}a_{33}$ is given by 2.9 with a,c = 1, b = 3, and d = 2.

In all three of the above cases it is easy to show by direct expansion that 2.9 is equal to zero. In fact, if it is assumed that a, b, c, and d must include all three integers 1, 2, and 3 among them, then the three cases above exhaust the possible selections for a, b, c, and d which make 2.9 different. Therefore, under the above assumptions, the quantity in 2.9 is identically zero. This relationship among the determinants of certain 2 by 2 submatrices of any 4 by 3 matrix will be useful again in the next section where we treat statement 2.5 for $n \geq 5$.

(3) For $n \geq 5$, the general procedure is to expand each of the determinants in 2.5 by the Laplace expansion. According to the Laplace expansion[1], the determinant of any n by n square matrix A can be expanded along the m rows $i_1, \ldots, i_m$ as

$$|A| = \sum_{w} (-1)^s \left| A^{j_1, \ldots, j_m}_{i_1, \ldots, i_m} \right| \cdot \left| A^{j_{m+1}, \ldots, j_n}_{i_{m+1}, \ldots, i_n} \right|$$

where (1) $s = i_1 + \ldots + i_m + j_1 + \ldots + j_m$

(2) w = the $\binom{n}{m}$ minors of the form $\left| A^{j_1, \ldots, j_m}_{i_1, \ldots, i_m} \right|$ that can be formed by choosing all possible combinations of m columns from n columns

(3) The indices in each of the four sets $(i_1, \ldots, i_m)$, $(i_{m+1}, \ldots, i_n)$, $(j_1, \ldots, j_m)$, and $(j_{m+1}, \ldots, j_n)$ are arranged in order of increasing magnitude.

In order to simplify the notation of 2.5 somewhat, define an $n + 1$ by $n - 1$ matrix A' which is formed from A as follows:

row 1 of A' = row 1 of A

row 2 of A' = row q of A

rows 3,..., q of A' = rows 2,...,q-1 of A, respectively

---

[1]See for example Ayres (1, p. 33).

rows q+1,...,p-1 of A'=rows q+1,...,p-1 of A, respectively

row p,...,n-1 of A' = rows p+1,...,n of A, respectively

row n of A' = row p of A

row n+1 of A' = row n+1 of A

Then replace A' by A. In terms of this new matrix A, Equation 2.5 is true if

$$\left|A\begin{matrix}1,\ldots,n-1\\ 2,\ldots,n-1,n+1\end{matrix}\right| \cdot \left|A\begin{matrix}1,\ldots,n-1\\ 1,3,\ldots,n\end{matrix}\right| - \left|A\begin{matrix}1,\ldots,n-1\\ 3,\ldots,n+1\end{matrix}\right| \cdot \left|A\begin{matrix}1,\ldots,n-1\\ 1,\ldots,n-1\end{matrix}\right|$$

$$- \left|A\begin{matrix}1,\ldots,n-1\\ 2,\ldots,n\end{matrix}\right| \cdot \left|A\begin{matrix}1,\ldots,n-1\\ 1,3,\ldots,n-1,n+1\end{matrix}\right| = 0 \qquad (2.10)$$

Thus, to prove statement 2.5, it is sufficient to prove statement 2.10.

Inspection of 2.10 shows that rows 3,...,n-1 are common to all of the matrices involved. This suggests that each of the determinants in 2.10 be expanded about rows 3,...,n-1 by the Laplace expansion. For example,

$$\left|A\begin{matrix}1,\ldots,n-1\\ 2,\ldots,n-1,n+1\end{matrix}\right| = \sum_{w}(-1)^{s}\left|A\begin{matrix}\text{exc. } j_1,j_2\\ 1,\ldots,n-1\\ 3,\ldots,n-1\end{matrix}\right| \cdot \left|A\begin{matrix}j_1,j_2\\ 2,n+1\end{matrix}\right|$$

where (1) $s = 3 + \ldots + (n-1) + 1 + \ldots + (n-1) - j_1 - j_2$

or $s = 2 + (n+1) + j_1 + j_2$

(2) w = the $\binom{n-1}{n-3}$ minors $\left|A\begin{matrix}\text{exc. } j_1,j_2\\ 1,\ldots,n-1\\ 3,\ldots,n-1\end{matrix}\right|$ that can be formed by choosing all possible combinations of n-3 columns from n-1 columns.

Making a similar expansion for each of the other determinants, and at the same time writing each product of sums as a double summation, the left-hand side of 2.10 becomes

$$\sum_{w_1}\sum_{w_2}(-1)^{s_1+s_2}\left|A_{3,\ldots,n-1}^{1,\ldots,n-1}\right|_{\text{exc. } j_1,j_2}\cdot\left|A_{3,\ldots,n-1}^{1,\ldots,n-1}\right|_{\text{exc. } k_1,k_2}\cdot\left|A_{2,n+1}^{j_1,j_2}\right|\cdot\left|A_{1,n}^{k_1,k_2}\right|$$

$$-\sum_{w_1}\sum_{w_2}(-1)^{s_3+s_4}\left|A_{3,\ldots,n-1}^{1,\ldots,n-1}\right|_{\text{exc. } j_1,j_2}\cdot\left|A_{3,\ldots,n-1}^{1,\ldots,n-1}\right|_{\text{exc. } k_1,k_2}\cdot\left|A_{n,n+1}^{j_1,j_2}\right|\cdot\left|A_{1,2}^{k_1,k_2}\right|$$

$$-\sum_{w_1}\sum_{w_2}(-1)^{s_5+s_6}\left|A_{3,\ldots,n-1}^{1,\ldots,n-1}\right|_{\text{exc. } j_1,j_2}\cdot\left|A_{3,\ldots,n-1}^{1,\ldots,n-1}\right|_{\text{exc. } k_1,k_2}\cdot\left|A_{2,n}^{j_1,j_2}\right|\cdot\left|A_{1,n+1}^{k_1,k_2}\right|$$

where (1) $s_1 + s_2 = 2 + (n+1) + j_1 + j_2 + 1 + n + k_1 + k_2$

$= 2n + 4 + j_1 + j_2 + k_1 + k_2$

$s_3 + s_4 = n + (n+1) + j_1 + j_2 + 1 + 2 + k_1 + k_2$

$= 2n + 4 + j_1 + j_2 + k_1 + k_2$

$s_5 + s_6 = 2 + n + j_1 + j_2 + 1 + (n+1) + k_1 + k_2$

$= 2n + 4 + j_1 + j_2 + k_1 + k_2$

(2) w_1 and w_2 are defined the same as w was defined above.

Factoring out the quantities that are common to all three double summations this can be written

$$\sum_{w_1}\sum_{w_2}(-1)^{j_1+j_2+k_1+k_2}\left|A_{3,\ldots,n-1}^{1,\ldots,n-1}\right|_{\text{exc. } j_1,j_2}\cdot\left|A_{3,\ldots,n-1}^{1,\ldots,n-1}\right|_{\text{exc. } k_1,k_2}\cdot \tag{2.11}$$

$$\left\{\left|A_{2,n+1}^{j_1,j_2}\right|\cdot\left|A_{1,n}^{k_1,k_2}\right|-\left|A_{n,n+1}^{j_1,j_2}\right|\cdot\left|A_{1,2}^{k_1,k_2}\right|-\left|A_{2,n}^{j_1,j_2}\right|\cdot\left|A_{1,n+1}^{k_1,k_2}\right|\right\}$$

Note that w (and consequently w_1 and w_2) can also be thought of as being the $\binom{n-1}{2}$ minors $\left|A^{1,\ldots,n-1 \text{ exc. } j_1,j_2}_{3,\ldots,n-1}\right|$ that can be formed by selecting all possible combinations of two columns, namely j_1 and j_2, to be excluded from n-1 columns, and in fact, it is more convenient to think of w in this way. With this in mind, let us take a closer look at the terms in 2.11.

Some of the terms in the double summation have the pair (j_1,j_2) equal to the pair (k_1,k_2). For each of these terms the quantity within the braces in 2.11 involves only four rows and two columns, namely rows 1, 2, n, and n+1 and columns j_1 and j_2. Comparison with Equation 2.6 shows that the quantity within the braces in 2.11 is zero for each term of this type.

For the terms remaining in 2.11, the pair (j_1,j_2) is not equal to the pair (k_1,k_2) which means that all of these terms fall into one of two categories. For one of these categories, call it v_1, the set (j_1,j_2,k_1,k_2) will contain exactly three different integers; the other category, call it v_2, will have exactly four different integers in the set (j_1,j_2,k_1,k_2). Furthermore the double summation of 2.11 can be replaced by a single summation by noting that the quantity outside the braces in 2.11 is the same for $(j_1,j_2) = (a,b)$ and $(k_1,k_2) = (c,d)$ as it is for $(j_1,j_2) = (c,d)$ and $(k_1,k_2) = (a,b)$ and choosing the new summation properly. In view of the above, the terms remaining in 2.11 can be represented by

$$\sum_{v_1+v_2} (-1)^{j_1+j_2+k_1+k_2} \left| A^{\overset{\text{exc. } j_1, j_2}{1,\ldots,n-1}}_{3,\ldots,n-1} \right| \cdot \left| A^{\overset{\text{exc. } k_1, k_2}{1,\ldots,n-1}}_{3,\ldots,n-1} \right| \cdot$$

$$\left\{ \left| A^{j_1,j_2}_{2,n+1} \right| \cdot \left| A^{k_1,k_2}_{1,n} \right| - \left| A^{j_1,j_2}_{n,n+1} \right| \cdot \left| A^{k_1,k_2}_{1,2} \right| - \left| A^{j_1,j_2}_{2,n} \right| \cdot \left| A^{k_1,k_2}_{1,n+1} \right| \right.$$

$$\left. + \left| A^{k_1,k_2}_{2,n+1} \right| \cdot \left| A^{j_1,j_2}_{1,n} \right| - \left| A^{k_1,k_2}_{n,n+1} \right| \cdot \left| A^{j_1,j_2}_{1,2} \right| - \left| A^{k_1,k_2}_{2,n} \right| \cdot \left| A^{j_1,j_2}_{1,n+1} \right| \right\} \qquad (2.12)$$

where (1) v_1 = the $\dfrac{\binom{n-1}{2}\left[\frac{(n-3)(2)+2(n-3)}{2}\right]}{2!}$ products of the minors $\left| A^{\overset{\text{exc. } j_1, j_2}{1,\ldots,n-1}}_{3,\ldots,n-1} \right| \cdot \left| A^{\overset{\text{exc. } k_1, k_2}{1,\ldots,n-1}}_{3,\ldots,n-1} \right|$ that can be formed by choosing two groups of 2 columns each from n-1 columns, the order of the groups being indistinguishable and either $k_1 = (j_1 \text{ or } j_2)$ or $k_2 = (j_1 \text{ or } j_2)$, but not both

(2) v_2 = the $\dfrac{\binom{n-1}{2}\binom{n-3}{2}}{2}$ products of the minors $\left| A^{\overset{\text{exc. } j_1, j_2}{1,\ldots,n-1}}_{3,\ldots,n-1} \right| \cdot \left| A^{\overset{\text{exc. } k_1, k_2}{1,\ldots,n-1}}_{3,\ldots,n-1} \right|$ that can be formed by choosing two groups of 2 columns each from n-1 columns, the order of the groups being indistinguishable and j_1, j_2, k_1, and k_2 being 4 distinct integers.

For each term that falls into category v_1, the quantity within the braces in 2.12 involves the determinants of certain 2 by 2 submatrices of a 4 by 3 matrix. By making a suitable association between the matrix E of 2.9 and the rows and columns of A that are present in the quantity within the braces of 2.12, it is possible to conclude from 2.9 and the discussion which follows it that the quantity within the braces of 2.12 is identically zero for this case. Thus the summation over v_1 contributes nothing, and 2.12 can be written as just the summation over v_2.

It is convenient at this point to symbolize the quantity within the braces in 2.12 by

$$\{(j_1,\ j_2;\ k_1,\ k_2)\}$$

to save some writing. Then 2.12 can be written as

$$\sum_z (-1)^{j_1+j_2+k_1+k_2} \left\{ \left| A^{\substack{1,\ldots,n-1 \\ \text{exc. } j_1, j_2}}_{3,\ldots,n-1} \right| \cdot \left| A^{\substack{1,\ldots,n-1 \\ \text{exc. } k_1, k_2}}_{3,\ldots,n-1} \right| \{(j_1,\ j_2;\ k_1,\ k_2)\} \right.$$
$$+ \left| A^{\substack{1,\ldots,n-1 \\ \text{exc. } j_1, k_1}}_{3,\ldots,n-1} \right| \cdot \left| A^{\substack{1,\ldots,n-1 \\ \text{exc. } j_2, k_2}}_{3,\ldots,n-1} \right| \{(j_1,\ k_1;\ j_2,\ k_2)\}$$
$$\left. + \left| A^{\substack{1,\ldots,n-1 \\ \text{exc. } j_1, k_2}}_{3,\ldots,n-1} \right| \cdot \left| A^{\substack{1,\ldots,n-1 \\ \text{exc. } j_2, k_1}}_{3,\ldots,n-1} \right| \{(j_1,\ k_2;\ j_2,\ k_1)\} \right\} \qquad (2.13)$$

where z = the $\binom{n-1}{4}$ possible combinations of 4 different columns, call them columns j_1, j_2, k_1, and k_2, from n-1 columns.

If it is assumed that $j_1 < j_2 < k_1 < k_2$, as is now perfectly permissible, than it can be shown by direct expansion that

$$\begin{aligned} \{(j_1, k_1; j_2, k_2)\} &= - \{(j_1, j_2; k_1, k_2)\} \\ \{(j_1, k_2; j_2, k_1)\} &= + \{(j_1, j_2; k_1, k_2)\} \end{aligned} \tag{2.14}$$

Substituting 2.14 into 2.13 the latter can be written as

$$\sum_z (-1)^{j_1+j_2+k_1+k_2} \{(j_1,j_2;k_1,k_2)\} \left\{ \left| A^{\substack{1,\ldots,n-1 \\ \text{exc. } j_1,j_2}}_{3,\ldots,n-1} \right| \cdot \left| A^{\substack{1,\ldots,n-1 \\ \text{exc. } k_1,k_2}}_{3,\ldots,n-1} \right| \right.$$

$$\left. - \left| A^{\substack{1,\ldots,n-1 \\ \text{exc. } j_1,k_1}}_{3,\ldots,n-1} \right| \cdot \left| A^{\substack{1,\ldots,n-1 \\ \text{exc. } j_2,k_2}}_{3,\ldots,n-1} \right| + \left| A^{\substack{1,\ldots,n-1 \\ \text{exc. } j_1,k_2}}_{3,\ldots,n-1} \right| \cdot \left| A^{\substack{1,\ldots,n-1 \\ \text{exc. } j_2,k_1}}_{3,\ldots,n-1} \right| \right\} \tag{2.15}$$

Let B be defined as the n-1 by n-3 matrix formed by taking the transpose of $\left[A^{1,\ldots,n-1}_{3,\ldots,n-1} \right]$, then interchanging rows j_1 and 1 and rows k_2 and n-1. Since the determinant of the transpose of a matrix is equal to the determinant of the matrix itself, the quantity within the braces in 2.15 is equal to zero if

$$\left| B^{1,\ldots,m-1}_{\substack{2,\ldots,m+1 \\ \text{exc. } j_2}} \right| \cdot \left| B^{1,\ldots,m-1}_{\substack{1,\ldots,m \\ \text{exc. } k_1}} \right| - \left| B^{1,\ldots,m-1}_{\substack{2,\ldots,m+1 \\ \text{exc. } k_1}} \right| \cdot \left| B^{1,\ldots,m-1}_{\substack{1,\ldots,m \\ \text{exc. } j_2}} \right|$$

$$+ \left| B^{1,\ldots,m-1}_{2,\ldots,m} \right| \cdot \left| B^{1,\ldots,m-1}_{\substack{1,\ldots,m+1 \\ \text{exc. } j_2,k_1}} \right| = 0 \tag{2.16}$$

where $n \geq 5$, $m = n-2$ and $2 \leq j_2 < k_1 \leq m$.

Equation 2.16 is just the negative of Equation 2.5 with n replaced by m ($m = n-2$), q replaced by j_2, and p replaced by k_1. It is also noted by the order of the above arguments that statement 2.5 is true if statement 2.16 is ture; that is, Equation 2.5 is true for n ($n \geq 5$) if it is true for n-2. Since 2.5 has already been proved for $n = 3$ and $n = 4$, the conclusion is reached by the principle of mathematical induction that 2.5 is true for all $n \geq 3$. This completes the proof of Theorem 2 which in turn completes the proof of Theorem 1. Theorem 1 is very useful in the work that follows.

III. THE OPTIMUM LINEAR DISTORTIONLESS FILTER

A. Comments on Finite Operating Time Filters

It is assumed that the reader has a basic understanding of the weighting function concept. However, since there are differences in notation among the various authors, a few comments about the notation used in this thesis and about finite operating time filters seem appropriate.

Consider the simple system shown in Figure 6 with input

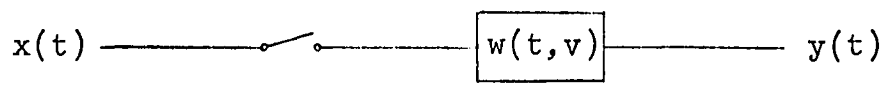


Figure 6. System illustrating weighting function

$x(t)$, output $y(t)$ and weighting function $w(t,v)$. In this thesis the weighting function $w(t,v)$ is defined as the output of the system shown in Figure 6 at time t due to a unit impulse applied at the input (with the switch closed, of course) at time $t-v$. The variable t is usually called the "running time variable" and v the "age variable".

If the system of Figure 6 possesses zero initial conditions and the switch is closed at $t = 0$, the output at time t can be written as

$$y(t) = \int_0^t w(t,v)x(t-v)dv \tag{3.1}$$

This system is a finite operating time filter since it weights only a finite amount of the past input. However, note that the interval over which the input is weighted increases as time goes on.

An alternate finite operating time filter is one in which the interval over which the input is weighted has a constant length T. In this case it is still convenient to think of the switch being closed at t = 0 and, providing $t \geq T$, the output at time t is given by

$$y(t) = \int_0^T w(t,v)x(t-v)dv \tag{3.2}$$

The corresponding equation for an infinite operating time filter can be obtained from 3.2 by simply letting $T \to \infty$.

If x(t) and y(t) are known functions of time, then 3.1, 3.2, and 3.3 are integral equations which specify their respective weighting functions. Notice that the problem of 3.1 is truly nonstationary in character; that is, the weighting function depends intrinsically upon t and v and a different solution w(t,v) is required for each t considered.

The situation is a little different in 3.2 where the weighting function depends not only on the variables t and v but also on the parameter T. With x(t) and y(t) given it is usually necessary to fix our attention on a specific value of time, say $t = t_1$, to solve the integral equation. By making the change of variable $t' = t - (t_1 - T)$, Equation 3.2 can be written as

$$y(t'+t_1-T) = \int_0^T w(t'+t_1-T,v)x(t'+t_1-T-v)dv$$

and specifying $t = t_1$ in 3.2 corresponds to specifying $t' = T$ in the above equation. Finally by defining

$$x_1(t') = x(t'+t_1-T)$$
$$y_1(t') = y(t'+t_1-T)$$
$$w_1(t',v) = w(t'+t_1-T,v)$$

Equation 3.2 evaluated at $t = t_1$ becomes

$$y_1(T) = \int_0^T w_1(T,v)x_1(T-v)dv \tag{3.3}$$

Consequently it is perfectly general to replace Equation 3.2 by Equation 3.3, but we should keep in mind that if we fix our attention on a different instant of time t we will in general get a different weighting function since x_1 and y_1 will in general be different. It is worth noting that since T is a constant in $w_1(T,v)$ above, T is more properly thought of as a parameter rather than a variable as the notation indicates; however, Equation 3.3 is left as it is because of its close relationship to 3.1.

In summary, from the preceding discussion and comparison of Equations 3.3 and 3.1, we can conclude that the integral equation for the type of filter represented by 3.2 can be obtained from the integral Equation 3.1 by simply fixing our attention on a fixed time instant t and identifying t with T. Therefore, the integral equations in the following sections will be developed for systems of the type characterized by Equation 3.1.

B. Derivation of the Integral Equations

In this section the integral equations are developed for the filter which optimizes the estimate of $s_1(t)$ from the available input lines shown in Figure 3. The criterion for optimization is the minimum mean-square error criterion. As stated below Figure 4, the filter is constrained to be linear, physically realizable, and distortionless. In addition, it is allowed, in general, to operate on only a finite amount of past data. There is one further assumption which is implicit in the development that follows, namely that the filter is not adaptive. In other words, the filter will not make use of the knowledge gained about $s_1(t)$ during the course of its operation to make a further improvement in itself. This is a subtle but important point and more will be said of it later.

Since the filter is constrained to be linear it may be represented by Figure 7 where $f_i(t)$ is the input signal on line i and $w_i(t,v)$ is the weighting function from line i to the output. Comparison of Figure 3 and Figure 7 shows that

$$f_i(t) = a_{i1}(t)s_1(t) + \ldots + a_{im}(t)s_m(t) + n_i(t) \qquad (3.4)$$

The constraint of physical realizability implies that $w_i(t,v) = 0$ for $v < 0$ and each $i = 1,\ldots,n$.

In view of the discussion in the preceeding section, the finite operating time filter is initially chosen to be the type discussed in connection with Equation 3.1 and can later

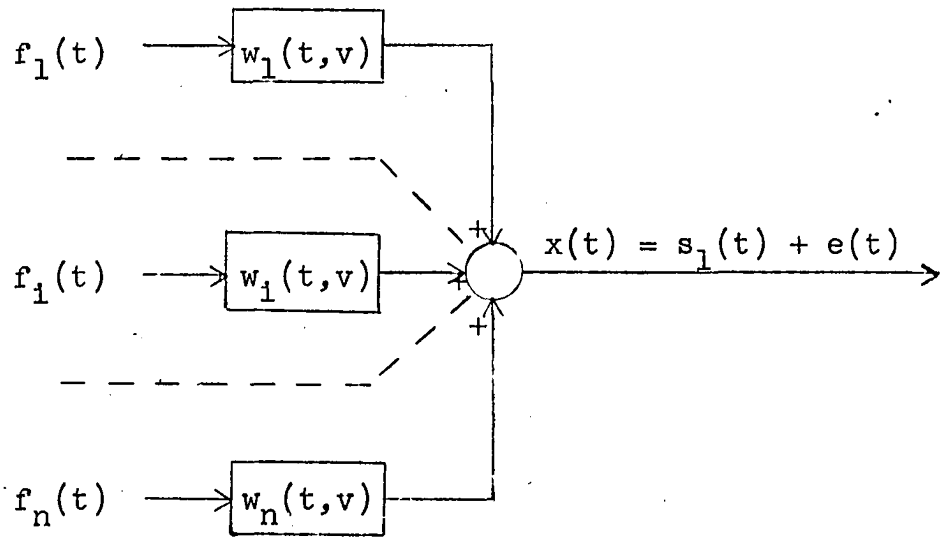


Figure 7. The general n-input, single output linear filter

be extended to another type if the need arises. Then the output $x(t)$ is given by

$$x(t) = \sum_{i=1}^{n} \int_0^t w_i(t,v)f_i(t-v)dv \qquad (3.5)$$

The distortionless constraint is defined to mean that the output at time t is identically $s_1(t)$ in the event that all of the noises are identically zero. Using 3.5 together with 3.1 this constraint implies that

$$G(t) \equiv \sum_{i=1}^{n} \int_0^t w_i(t,v)[\sum_{j=1}^{m} a_{ij}(t-v)s_j(t-v)]dv - s_1(t) = 0 \qquad (3.6)$$

where $G(t)$ is defined as shown for later use.

By inspection of Figure 7 the error, $e(t)$, associated with the estimate of $s_1(t)$ is given by

$$e(t) = x(t) - s_1(t)$$

Using Equation 3.5 for $x(t)$, along with 3.4 and 3.6, the expression for the error reduces to

$$e(t) = \sum_{i=1}^{n} \int_0^t w_i(t,v)n_i(t-v)dv$$

Squaring this expression for the error, writing the product of the two summations as a double sum and the product of two integrals as a double integral, and finally taking the ensemble average gives

$$E(t) \equiv \overline{e^2(t)} = \sum_{i=1}^{n} \sum_{j=1}^{n} \int_0^t \int_0^t w_i(t,u)w_j(t,v)\overline{n_i(t-u)n_j(t-v)}dudv$$

where $E(t)$ is defined as the mean-square error for notational convenience. Since the noises are assumed to be mutually independent,

$$\overline{n_i(t-u)n_j(t-v)} = \begin{cases} \varphi_i(t-u,t-v) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{3.7}$$

where $\varphi_i(t_1,t_2)$ is defined as the autocorrelation function of the nonstationary noise $n_i(t)$, i.e.

$$\varphi_i(t_1,t_2) = \overline{n_i(t_1)n_i(t_2)} \tag{3.8}$$

Utilizing 3.7 the expression for the mean-square error reduces to

$$E(t) = \sum_{i=1}^{n} \int_0^t \int_0^t w_i(t,u)w_i(t,v)\varphi_i(t-u,t-v)dudv \tag{3.9}$$

The problem of finding the weighting functions $w_1(t,v)$, ..., $w_n(t,v)$ which will minimize the mean-square error subject to the conditions imposed on the filter thus reduces to minimizing $E(t)$ subject to the constraint that $G(t) = 0$. This type of problem can be handled readily by using the Lagrange multiplier technique. To employ this technique note that $E(t)$ and $G(t)$ are really functions of $w_1(t,v),\ldots,w_n(t,v)$; i.e.,

$$\begin{aligned} E(t) &= E(w_1,\ldots,w_n) \\ G(t) &= G(w_1,\ldots,w_n) \end{aligned} \tag{3.10}$$

Then according to the Lagrange multiplier technique[1], in order that E attain an extreme value under the condition that $G = 0$ at a point $(w_1^*,\ldots,w_1^*)$, it is necessary that there be a number λ such that

---

[1]See for example Fulks (7), pp. 266.

$$\left.\frac{\partial E(w_1,\ldots,w_n)}{\partial w_i}\right|_{w_1^*,\ldots,w_n^*=0} + \lambda \left.\frac{\partial G(w_1,\ldots,w_n)}{\partial w_i}\right|_{w_1^*,\ldots,w_n^*=0} = 0 \text{ for } i = 1,\ldots,n$$

$$\text{and } G(w_1^*,\ldots,w_n^*) = 0 \tag{3.11}$$

Following the usual procedure of calculus of variations, $w_i(t,v)$ is replaced by $w_i^*(t,v) + \epsilon_i\eta_i(t,v)$ for each $i = 1,\ldots,n$ in Equation 3.11. Here, ϵ_i is an arbitrarily small parameter and $\eta_i(t,v)$ is an arbitrary perturbation for $0 \leq v \leq t$. Then 3.11 can be replaced by

$$\left.\frac{\partial E}{\partial \epsilon_i}\right|_{\epsilon_1,\ldots,\epsilon_n=0} + \lambda \left.\frac{\partial G}{\partial \epsilon_i}\right|_{\epsilon_1,\ldots,\epsilon_n=0} =0 \text{ for } i = 1,\ldots,n$$

$$\text{and } G(w_1^*,\ldots,w_n^*) = 0 \tag{3.12}$$

where the arguments of E and G are left out to save writing. The equivalence of Equations 3.11 and 3.12 can be shown by expanding E and G in a Taylor series about the point $(w_1^*,\ldots, w_n^*)$.

Substituting the expressions for E and G given by 3.9 and 3.6, respectively, into the first of Equations 3.12, performing the indicated differentiation, and using the fact that $\varphi_i(t_1,t_2) = \varphi_i(t_2,t_1)$ yields

$$\int_0^t \eta_i(t,u) \{2 \int_0^t w_i^*(t,v)\varphi_i(t-u,t-v)dv + \lambda \sum_{j=1}^m a_{ij}(t-u)s_j(t-u)\} du = 0$$

Since $\eta_i(t,u)$ is an arbitrary perturbation for $0 \leq u \leq t$, the above equation can be satisfied only if the quantity within the braces vanishes. Therefore,

$$2\int_0^t w_i^*(t,v)\varphi_i(t-u,t-v)dv + \lambda\sum_{j=1}^{m} a_{ij}(t-u)s_j(t-u) = 0 \quad (3.13)$$

for $0 \leq u \leq t$ and each $i = 1,\ldots,n$.

The second of Equations 3.12 is simply Equation 3.6 with $w_i(t,v)$ replaced by $w_i^*(t,v)$. If the assumption is made that the weighting functions $w_i^*(t,v)$, $i = 1,\ldots,n$, do not depend on the signals $s_j(t)$, $j = 1,\ldots,m$ (i.e., the filter is not adaptive), then this one equation implies the m conditions given by

$$\sum_{i=1}^{n} a_{ij}(t-v)w_i^*(t,v) = \begin{cases} \delta(v) & \text{for } j = 1 \\ 0 & \text{for } j = 2,\ldots,m \end{cases} \quad (3.14)$$

where $\delta(v)$ is the Dirac delta function.

The set of n equations represented by 3.13 together with the set of m equations represented by 3.14 are the necessary conditions that $w_1^*(t,v),\ldots,w_n^*(t,v)$ must satisfy for the mean-square error to be a minimum. These conditions are also sufficient if

$$\left.\frac{\partial^2 E}{\partial \epsilon_i^2}\right|_{\epsilon_1,\ldots,\epsilon_n=0} > 0 \quad \text{for each } i = 1,\ldots,n$$

The details of taking the second derivative indicated above are not shown here but the result is

$$\left.\frac{\partial^2 E}{\partial \epsilon_i^2}\right|_{\epsilon_1,\ldots,\epsilon_n=0} = 2\int_0^t\int_0^t \eta_i(u)\eta_i(v)\varphi_i(t-u,t-v)dudv$$

$$= 2\left[\int_0^t \eta_i(u)n_i(t-u)du\right]^2 > 0$$

for each $i = 1,\ldots,n$, which ensures that a minimum point has been achieved since the square is always positive.

Equations 3.13 and 3.14 together make up a set of m+n equations in the n+1 inknowns $w_1^*(t,v),\ldots,w_n^*(t,v)$, and λ which can be solved simultaneously for these unknowns. Indeed 3.14 represents a set of m equations in n unknowns and can be written in matrix notation as

$$[A(t-v)]^T \begin{bmatrix} w_1^*(t,v) \\ w_2^*(t,v) \\ \cdot \\ \cdot \\ w_n^*(t,v) \end{bmatrix} = \begin{bmatrix} \delta(v) \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \qquad (3.15)$$

where $A(t)$ is the n by m coefficient matrix with elements $a_{ij}(t)$, and $[A(t)]^T$ is the transpose of $A(t)$. Consequently, Equation 3.15 may be used to solve for m of the optimum weighting functions in terms of the other n-m weighting functions whenever $A(t-v)$ has m rows such that the determinant of the m by m submatrix consisting of these m rows is nonzero for all values of v in the interval $(0 \leq v \leq t)$ and all values of t of interest. Let us assume that A possesses m such rows and furthermore that these are rows $1,\ldots,m$; that is,

$$\left| A^{1,\ldots,m}_{1,\ldots,m}(t-v) \right| \neq 0 \text{ for all } v \text{ such that } 0 \leq v \leq t \qquad (3.16)$$

Under this assumption, it is possible to solve for $w_1^*(t,v)$, ..., $w_m^*(t,v)$ in terms of $w_{m+1}^*(t,v),\ldots,w_n^*(t,v)$. From either

3.14 or 3.15, one can write

$$\left[A^{1,\ldots,m}_{1,\ldots,m}(t-v)\right]^T \begin{bmatrix} w_1^*(t,v) \\ w_2^*(t,v) \\ \vdots \\ w_m^*(t,v) \end{bmatrix} = \begin{bmatrix} \delta(v) - \sum_{j=m+1}^{n} a_{j1}(t-v)w_j^*(t,v) \\ -\sum_{j=m+1}^{n} a_{j2}(t-v)w_j^*(t,v) \\ \vdots \\ -\sum_{j=m+1}^{n} a_{jm}(t-v)w_j^*(t,v) \end{bmatrix}$$

Using Cramers Rule, the fact that the determinant of a matrix transposed is equal to the determinant of the matrix itself, and finally that

$$\sum_{k=1}^{m} (-1)^k a_{jk} \left| A^{\substack{\text{exc. } k \\ 1,\ldots,m}}_{\substack{1,\ldots,m \\ \text{exc. } i}}(t-v) \right| = (-1)^m \left| A^{1,\ldots,m}_{\substack{1,\ldots,m,j \\ \text{exc. } i}}(t-v) \right| \quad (3.17)$$

for j included in the set of integers $(m+1,\ldots,n)$, the expression for $w_i^*(t,v)$ reduces to

$$w_i^*(t,v) = \frac{1}{\left| A^{1,\ldots,m}_{1,\ldots,m}(t-v) \right|} \{(-1)^{i+1}\delta(v) \left| A^{2,\ldots,m}_{\substack{1,\ldots,m \\ \text{exc. } i}}(t-v) \right|$$

$$- (-1)^{i+m} \sum_{j=m+1}^{n} \left| A^{1,\ldots,m}_{\substack{1,\ldots,m,j \\ \text{exc. } i}}(t-v) \right| w_j^*(t,v)\} \quad (3.18)$$

for $i = 1,\ldots,m$.

The set of n equations represented by 3.13 can be written in matrix notation as

$$\lambda[A(t-u)]\begin{bmatrix} s_1(t-u) \\ \cdot \\ \cdot \\ \cdot \\ s_m(t-u) \end{bmatrix} = \begin{bmatrix} -2\int_0^t w_1^*(t,v)\varphi_1(t-u,t-v)dv \\ \cdot \\ \cdot \\ \cdot \\ -2\int_0^t w_n^*(t,v)\varphi_n(t-u,t-v)dv \end{bmatrix}$$

which can be thought of as a set of n equations in the m unknown $\lambda s_1,\dots,\lambda s_m$. Because of the assumption 3.16, the first m of the above equations can be used to solve for λs_1, $\dots,\lambda s_m$ by using Cramer's Rule. The result is

$$\lambda s_i(t-u) = \frac{1}{\left|A^{1,\dots,m}_{1,\dots,m}(t-u)\right|}\left\{\sum_{k=1}^{m}(-2)(-1)^{i+k}\left|A^{\substack{\text{exc. } i\\ 1,\dots,m}}_{\substack{1,\dots,m\\ \text{exc. } k}}(t-u)\right| \cdot \int_0^t w_k^*(t,v)\varphi_k(t-u,t-v)dv\right\} \qquad (3.19)$$

for $i = 1,\dots,m$ and $0 \le u \le t$.

Equation 3.19 can now be substituted into the remaining n-m equations of 3.13. If Equation 3.18 is then used to eliminate $w_1^*(t,v),\dots,w_m^*(t,v)$, the result is

$$\int_0^t w_i^*(t,v)\varphi_i(t-u,t-v)dv$$

$$+ \frac{1}{\left|A^{1,\dots,m}_{1,\dots,m}(t-u)\right| \cdot \left|A^{1,\dots,m}_{1,\dots,m}(t-v)\right|}\left\{\sum_{j=1}^{m} a_{ij}(t-u)\left\{\sum_{k=1}^{m}(-1)^{j+k+1}\cdot\right.\right.$$

$$\left|A^{\substack{\text{exc. } j\\ 1,\dots,m}}_{\substack{1,\dots,m\\ \text{exc. } k}}(t-u)\right| \int_0^t [(-1)^{k+1}\left|A^{2,\dots,m}_{\substack{1,\dots,m\\ \text{exc. } k}}(t-v)\right| \delta(v)$$

$$\left.\left. - (-1)^{k+m}[\sum_{p=m+1}^{n}\left|A^{1,\dots,m}_{\substack{1,\dots,m,p\\ \text{exc. } k}}(t-v)\right| w_p^*(t,v)]\varphi_k(t-u,t-v)]dv\right\}\right\}$$

$$= 0$$

for $0 \leq u \leq t$ and each $i = m+1,\ldots,n$. The integration which includes $\delta(v)$ may then be carried out using the sifting property of the Dirac delta function. Then by using Equation 3.17 and freely interchanging the order of integration and summation the above equation can be put in the form

$$\int_0^t w_i^*(t,v)\varphi_i(t-u,t-v)dv$$

$$+ \sum_{j=m+1}^{n} \int_0^t w_j^*(t,v) \frac{1}{\left|A\begin{smallmatrix}1,\ldots,m\\1,\ldots,m\end{smallmatrix}(t-u)\right| \cdot \left|A\begin{smallmatrix}1,\ldots,m\\1,\ldots,m\end{smallmatrix}(t-v)\right|} \cdot$$

$$\left\{\sum_{k=1}^{m} \left|A\begin{smallmatrix}1,\ldots,m\\1,\ldots,m,i\\\text{exc. }k\end{smallmatrix}(t-u)\right| \cdot \left|A\begin{smallmatrix}1,\ldots,m\\1,\ldots,m,j\\\text{exc. }k\end{smallmatrix}(t-v)\right| \varphi_k(t-u,t-v)\right\}dv$$

$$+ (-1)^m \sum_{k=1}^{m} \frac{\left|A\begin{smallmatrix}1,\ldots,m\\1,\ldots,m,i\\\text{exc. }k\end{smallmatrix}(t-u)\right| \cdot \left|A\begin{smallmatrix}2,\ldots,m\\1,\ldots,m\\\text{exc. }k\end{smallmatrix}(t)\right|}{\left|A\begin{smallmatrix}1,\ldots,m\\1,\ldots,m\end{smallmatrix}(t-u)\right| \cdot \left|A\begin{smallmatrix}1,\ldots,m\\1,\ldots,m\end{smallmatrix}(t)\right|}\varphi_k(t-u,t) = 0 \tag{3.20}$$

for $0 \leq u \leq t$ and each $i = m+1,\ldots,n$. Notice that 3.20 is a set of n-m equations which together determine $w_{m+1}^*(t,v),\ldots,$ $w_n^*(t,v)$. Once this set of equations is solved, the remaining m weighting functions are given by 3.18. This completes the derivation of the integral equations for the "optimum" filter.

IV. THE INTUITIVE FILTER

The purpose of this chapter is to investigate the "intuitive" system. A form of the "linear, algebraic operator" referred to in the introduction will be constructed which will be fairly general yet specific enough to be handled without great notational difficulty. The integral equations for the generalized (n-m)-dimensional Wiener filter associated with this linear, algebraic operator will then be given. But before proceeding to this problem, it is convenient to develop the integral equations for the generalized r-dimensional Wiener filter for the type of input that will be of interest here.

A. Generalized r-Dimensional Wiener Filter

Consider the problem of finding the weighting functions $y_1(t,v),\ldots,y_r(t,v)$ which will minimize the mean-square error associated with estimating $N_0(t)$ in Figure 8. The weighting functions are assumed to be physically realizable and the system is "turned on" at time $t = 0$ with zero initial conditions. The nonstationary, random input noises $n_i(t)$, $i = 1, \ldots, n$, are mutually independent with known autocorrelation functions; that is,

$$\overline{n_i(t_1)n_j(t_2)} = \begin{cases} \varphi_i(t_1,t_2) & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \qquad (4.1)$$

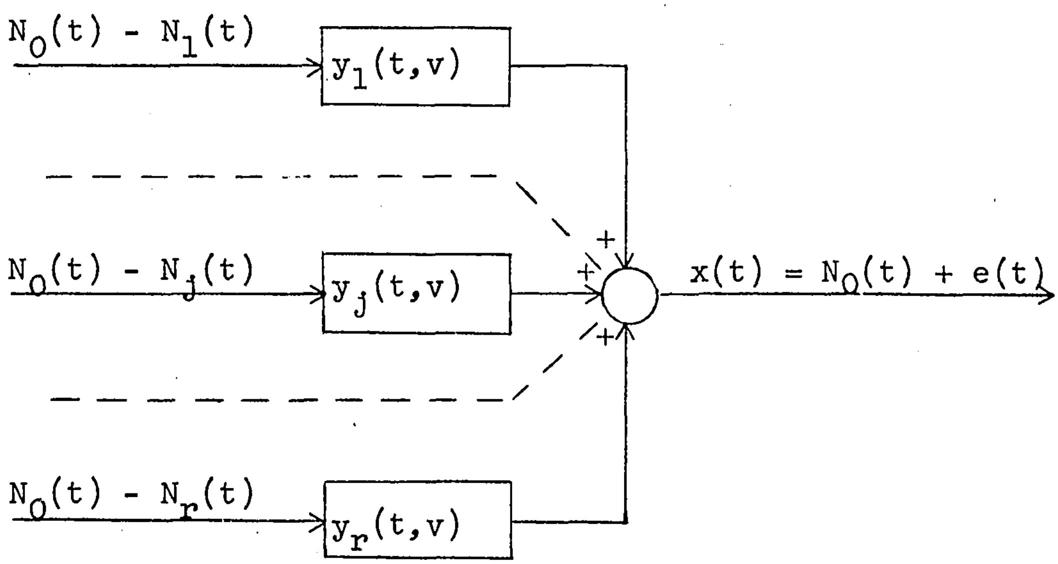


Figure 8. The "generalized r-dimensional Wiener filter"

It is also assumed that the r available input lines are of the form shown in Figure 8 where

$$N_i(t) = \sum_{j=1}^{n} c_{ij}(t)n_j(t) \tag{4.2}$$

and $c_{ij}(t)$ is a known function of time (possibly zero) for each $i = 0, 1, \ldots, r$ and $j = 1, \ldots, n$.

With the above given information and the error $e(t)$ defined as the difference between the output and $N_0(t)$, the error can be written as

$$e(t) = \sum_{i=1}^{r} \int_0^t y_i(t,v)[N_0(t-v) - N_i(t-v)]dv - N_0(t)$$

After squaring the expression for $e(t)$, substituting 4.2, and employing 4.1 when the mean is taken, the expression for $\overline{e^2(t)}$ can be reduced to

$$\overline{e^2(t)} = \sum_{i=1}^{r} \sum_{j=1}^{r} \int_0^t \int_0^t y_i(t,u)y_j(t,v)\cdot$$
$$\{\sum_{k=1}^{n} [c_{0k}(t-u) - c_{ik}(t-u)][c_{0k}(t-v) - c_{jk}(t-v)]\varphi_k(t-u,t-v)\}dudv$$
$$-2\sum_{i=1}^{r} \int_0^t y_i(t,u)\{\sum_{k=1}^{n} c_{0k}(t)[c_{0k}(t-u) - c_{ik}(t-u)]\varphi_k(t,t-u)\}du$$
$$+\sum_{k=1}^{n} c_{0k}^2(t)\varphi_k(t,t)$$

To find the set of weighting functions which minimizes $\overline{e^2(t)}$, the usual calculus of variations is used. That is, $y_i(t,v) + \epsilon_i\eta_i(t,v)$ is substituted for $y_i(t,v)$ where η_i is an arbitrary function for v in the interval $0 \leq v \leq t$, and $y_i(t,v)$ is understood to be the optimum weighting function from here on. A necessary condition for a minimum to occur at the point $(y_1, \ldots, y_r)$ is that

$$\left.\frac{\partial \overline{e^2(t)}}{\partial \epsilon_i}\right|_{\epsilon_1, \ldots, \epsilon_r = 0} = 0 \text{ for each } i = 1, \ldots, r$$

That a minimum indeed occurs can be shown by calculating the second derivative and showing that

$$\left.\frac{\partial^2 \overline{e^2(t)}}{\partial \epsilon_i^2}\right|_{\epsilon_1, \ldots, \epsilon_r = 0} > 0 \text{ for each } i = 1, \ldots, r$$

The details of the above calculations are not shown here but the result is that the optimum weighting functions are specified by the set of r integral equations

$$\sum_{j=1}^{r} \int_0^t y_j(t,v) \left\{ \sum_{k=1}^{n} [c_{0k}(t-u) - c_{qk}(t-u)] \cdot [c_{0k}(t-v) - c_{jk}(t-v)] \varphi_k(t-u, t-v) \right\} dv$$
$$- \sum_{k=1}^{n} c_{0k}(t)[c_{0k}(t-u) - c_{qk}(t-u)]\varphi_k(t, t-u) = 0 \qquad (4.3)$$

for $0 \leq u \leq t$ and each $q = 1, \ldots, r$. This set of integral equations is, of course, just a specialization of the generalized r-dimensional Wiener filter to the type of inputs and desired output shown in Figure 8.

B. The Linear, Algebraic Operator

As mentioned in the introduction there are usually quite a number of possible ways to construct the linear, algebraic operator shown in Figure 5. For example, with appropriate assumptions about the linear independence of the rows of the A matrix, there are 3 possible ways of choosing the linear, algebraic operator for m = 2 and n = 3, 6 possible ways for m = 3 and n = 4, and 16 possible ways for m = 2 and n = 4. The number of possible ways of choosing the linear, algebraic operator grows at a rather fantastic rate as m and n become larger.

Since it is not feasible to treat all the possible ways of constructing the general linear, algebraic operator shown in Figure 5, a different approach is chosen here. The approach here is to choose a form that is fairly general, yet specific enough so that the notation does not become too cumbersome and for which only a few assumptions need be made about the A matrix. Then the generalized (n-m)-dimensional Wiener filter which corresponds to this linear, algebraic operator is specified and if this system can be shown to be optimum, it will follow that whenever A permits another system of similar form to be constructed, it too will be optimum.

The form of the linear, algebraic operator chosen here is the portion of Figure 9 included inside the dotted box, where r = n-m. The notation

$$O_j\underset{\text{exc. } f_{p_j}}{(f_1,\ldots,f_m,f_{m+j})} \tag{4.4}$$

is used to denote a linear, algebraic operator (or sub-operator) having inputs $f_1(t),\ldots,f_{p_j-1}(t)$, $f_{p_j+1}(t),\ldots,$ $f_m(t)$, f_{m+j}, where p_j is included in the set of integers

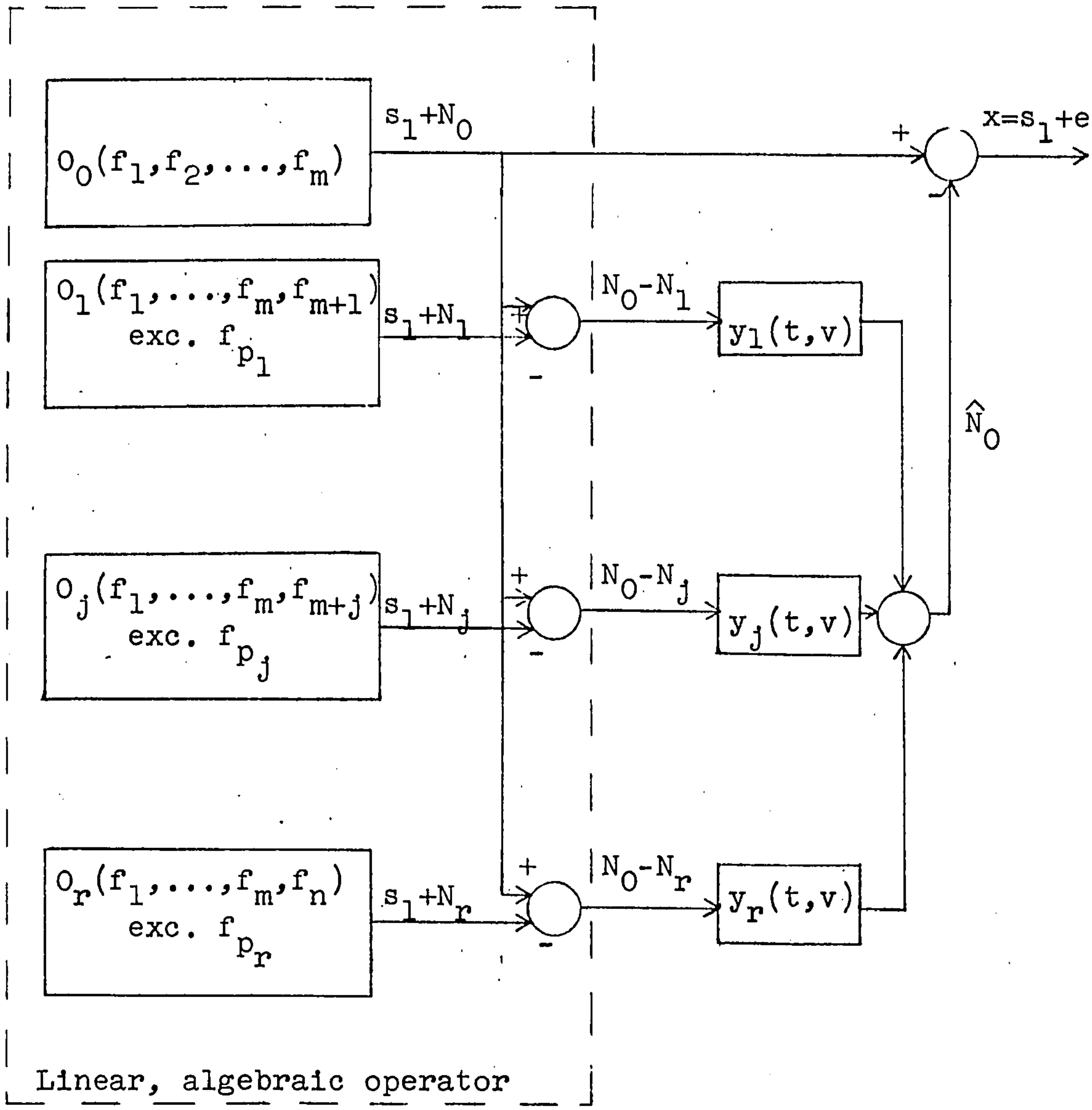


Figure 9. The "intuitive system"

(1,...,m) for $j = 1,\ldots,r$. To derive the specific form of O_j, first write the equations for input lines $1,\ldots,p_j-1$, $p_j+1,\ldots,m$, $m+j$ in matrix notation as (see Figure 3 and Equation 3.4)

$$\begin{bmatrix} A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j\\ \text{exc. } p_j}}(t) \end{bmatrix} \begin{bmatrix} s_1(t) \\ \vdots \\ s_m(t) \end{bmatrix} = \begin{bmatrix} f_1(t) \\ \vdots \\ f_{p_j-1}(t) \\ f_{p_j+1}(t) \\ \vdots \\ f_m(t) \\ f_{m+1}(t) \end{bmatrix} - \begin{bmatrix} n_1(t) \\ \vdots \\ n_{p_j-1}(t) \\ n_{p_j+1}(t) \\ \vdots \\ n_m(t) \\ n_{m+1}(t) \end{bmatrix} \quad (4.5)$$

Then for every value of time t for which the determinant of the matrix on the left is nonzero, Equation 4.5 may be solved for $s_1(t)$ using Cramer's Rule. The resulting linear combination of $f_i(t)$'s will be defined as $O_j(f_1,\ldots,f_m,f_{m+j})$. Indeed, because of the similarity of the two column vectors on the right-hand side of 4.5, it is noted that

$$s_1(t) = \underset{\text{exc. } f_{p_j}}{O_j(f_1,\ldots,f_m,f_{m+j})} - \underset{\text{exc. } n_{p_j}}{O_j(n_1,\ldots,n_m,n_{m+j})} \quad (4.6)$$

And, O_j is found to be given by

$$\underset{\text{exc. } f_{p_j}}{O_j(f_1,\ldots,f_m,f_{m+j})} =$$

$$\frac{1}{\left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j\\ \text{exc. } p_j}}(t)\right|}\left\{\sum_{i=1}^{p_j-1}(-1)^{i+1}\left|A^{2,\ldots,m}_{\substack{1,\ldots,m,m+j\\ \text{exc. } i,p_j}}(t)\right| f_i(t)\right.$$

$$+\sum_{i=p_j+1}^{m}(-1)^{i}\left|A^{2,\ldots,m}_{\substack{1,\ldots,m,m+j\\ \text{exc. } i,p_j}}(t)\right| f_i(t)$$

$$\left.+(-1)^{m+1}\left|A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } p_j}}(t)\right| f_{m+j}(t)\right\}. \tag{4.7}$$

for $j = 1,\ldots,r$. Also, O_0 is given by

$$O_0(f_1,\ldots,f_m) = \frac{1}{\left|A^{1,\ldots,m}_{1,\ldots,m}(t)\right|}\sum_{i=1}^{m}(-1)^{i+1}\left|A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } i}}(t)\right| f_i(t) \tag{4.8}$$

And, of course, $O_j(n_1,\ldots,n_m,n_{m+j})$ is given by 4.7 with $f_i(t)$ replaced by $n_i(t)$ for each $j = 1,\ldots,r$ and by 4.8 for $j = 0$. It is also observed by comparing 4.6 with Figure 9 that

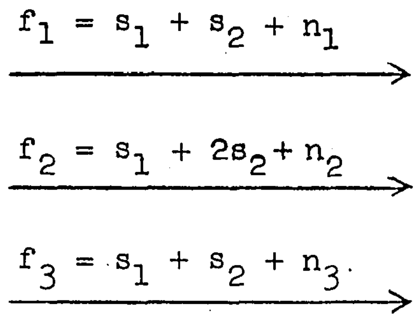
$$N_0(t) = O_0(n_1,\ldots,n_m)$$

and

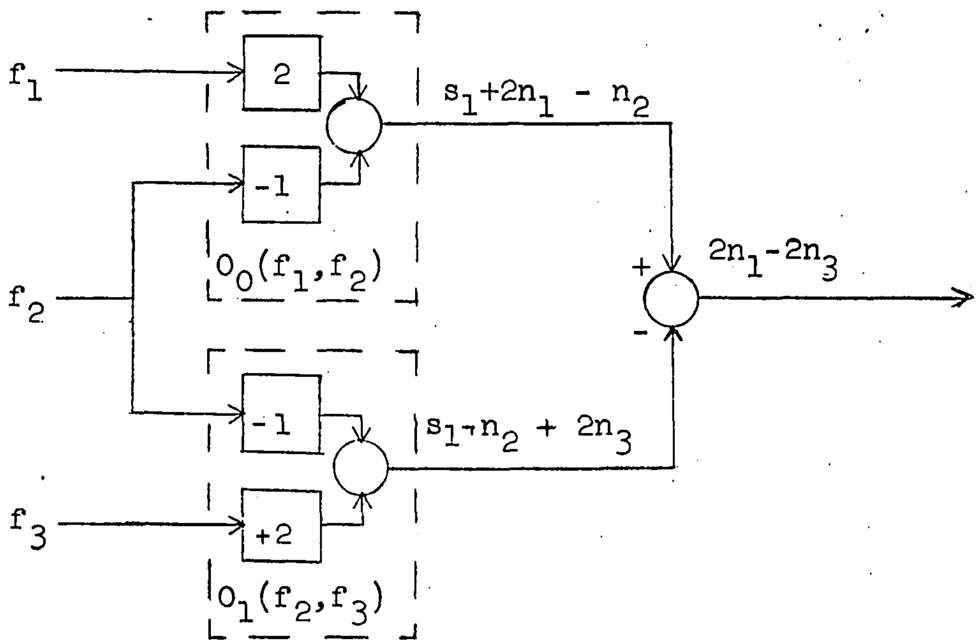
$$N_j(t) = O_j(\underset{\text{exc. } n_{p_j}}{n_1,\ldots,n_m,n_{m+j}}) \tag{4.9}$$

for $j = 1,\ldots,r$. Thus the expressions for $N_0(t)$ and $N_j(t)$ have been established and are given by 4.8 and 4.7, respectively, with $f_i(t)$ replaced by $n_i(t)$ in both equations.

As a simple example, consider constructing a linear, algebraic operator for the inputs shown in Figure 10(a). The quantities that can be measured are the signal levels



(a) The available inputs



(b) The linear, algebraic operator

Figure 10. Example illustrating the linear, algebraic operator

of the 3 input lines, namely f_1, f_2, and f_3, and the question is how to "operate" on these measurements to get the desired outputs of the linear, algebraic operator. The method suggested is to initially treat the noises as if they, too, were known and write the input as a set of 3 equations in the two unknowns s_1 and s_2; i.e.,

$$s_1 + s_2 = f_1 - n_1$$

$$s_1 + 2s_2 = f_2 - n_2$$

$$s_1 + s_2 = f_3 - n_3$$

In particular, the first two of the above equations are linearly independent and may be solved for s_1. This yields,

$$s_1 = 2(f_1 - n_1) - (f_2 - n_2)$$

Upon rearranging this can be written

$$2f_1 - f_2 = s_1 + 2n_1 - n_2$$

Comparing this to Figure 9, it is observed that the left-hand side of this equation defines the operator $O_0(f_1, f_2)$. Similarly, the last two equations are linearly independent and may also be solved for s_1 to yield the equation

$$- f_2 + 2f_3 = s_1 - n_2 + 2n_3$$

The left-hand side of this equation defines the operator $O_1(f_1, f_2, f_3)$, where $p_1 = 1$. The complete linear, algebraic
exc. p_1
operator is shown in Figure 10(b). Note that in this example, p_1 cannot be chosen as 2 since the first and third equations

of the above set are not linearly independent. Consequently, the number of possible linear, algebraic operators is reduced from 3 to 1, and the number of possible "intuitive" systems is reduced from 6 to 2. One of these "intuitive" systems consists of the linear, algebraic operator shown in Figure 10(b) together with the Wiener filter which makes an optimal estimate of $(2n_1 - n_2)$ from $(2n_1 - 2n_3)$; the other "intuitive" system consists of the same linear, algebraic operator together with the Wiener filter which makes an optimal estimate of $(2n_3 - n_2)$ from $(2n_1 - 2n_3)$.

Before proceeding to the integral equations which describe the generalized r-dimensional Wiener filter associated with the linear, algebraic operator proposed here, it seems appropriate to discuss the limitations of the chosen operator. From a close inspection of Figure 9, it is noted that there are two reasons for the proposed operator not being completely general. These are:

(1) There are n-m lines, namely lines m+1,...,n, which are included in only one operator. This means that $n_{m+1}(t)$ is included in only in $N_1(t)$,..., $n_{m+r}(t)$ is included only in $N_r(t)$.

(2) The fact that $N_0(t)$ appears as part of the input to each of the weighting functions $y_1,\ldots,y_r$.

In spite of being restricted in these two ways from the most general case, the linear algebraic operator proposed in this section is still fairly general and is of considerable practical importance.

C. The Integral Equations

With the inputs to the filter part of the "intuitive system" now known and of the basic form 4.2, the integral equations for the generalized (n-m)-dimensional Wiener filter corresponding to the linear, algebraic operator shown in Figure 9 can now be written by inspection. Each of the n-m inputs is of the form $N_0(t) - N_j(t)$, so it would be convenient to have an expression for this difference.

$$\sum_{k=1}^{n} [c_{0k}(t)-c_{jk}(t)]n_k(t) = N_0(t)-N_j(t)$$

$$= \sum_{\substack{k=1 \\ \text{exc. } p_j}}^{m} (-1)^{k+1} \left[\frac{\left| A^{2,\ldots,m}_{\substack{1,\ldots,m \\ \text{exc. } k}}(t) \right|}{\left| A^{1,\ldots,m}_{1,\ldots,m}(t) \right|} \pm \frac{\left| A^{2,\ldots,m}_{\substack{1,\ldots,m,m+j \\ \text{exc. } k,p_j}}(t) \right|}{\left| A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j \\ \text{exc. } p_j}}(t) \right|} \right] n_k(t)$$

$$+ (-1)^{p_j+1} \frac{\left| A^{2,\ldots,m}_{\substack{1,\ldots,m \\ \text{exc. } p_j}}(t) \right|}{\left| A^{1,\ldots,m}_{1,\ldots,m}(t) \right|} n_{p_j}(t)$$

$$+ (-1)^{m+1} \frac{\left| A^{2,\ldots,m}_{\substack{1,\ldots,m \\ \text{exc. } p_j}}(t) \right|}{\left| A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j \\ \text{exc. } p_j}}(t) \right|} n_{m+j}(t) \tag{4.10}$$

For the quantity inside the brackets in 4.10, the minus sign is used when $k < p_j$ and the plus sign is used when $k > p_j$. Note that 4.10 gives an explicit expression for the quantity $c_{0k}(t) - c_{jk}(t)$ for $k = 1,\ldots,n$ and $j = 1,\ldots,r$. Therefore 4.10 along with the expression for $c_{0k}(t)$

$$\sum_{k=1}^{n} c_{0k}(t)n_k(t) = N_0(t) = \sum_{k=1}^{m} (-1)^{k+1} \frac{\left|A^{2,\ldots,m}_{1,\ldots,m}(t)\right|_{\text{exc. } k}}{\left|A^{1,\ldots,m}_{1,\ldots,m}(t)\right|} n_k(t) \quad (4.11)$$

could be substituted directly into 4.3 to get the desired set of integral equations for the optimum weighting functions $y_1(t,v),\ldots,y_r(t,v)$ shown in Figure 9. The result would be a form for each of these integral equations which would not be at all convenient for later comparison to the "optimum" filter of Chapter III. Fortunately, however, the form of $c_{0k}(t) - c_{jk}(t)$ shown in 4.10 can be simplified considerably with the aid of Theorem 1 of Chapter II. Note that although Theorem 1 was stated and proved in terms of a matrix A having constant elements a_{ij}, the proof would not be altered by making $a_{ij} = a_{ij}(t)$. Consequently Theorem 1 can be extended to a matrix whose elements vary with time. By putting the two terms within the brackets in 4.10 over a common denominator and then using Equation 2.1 to reduce the numerator, the quantity within the brackets of 4.10 can be written

$$\frac{\left|A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } p_j}}(t)\right| \cdot \left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j\\ \text{exc. } k}}(t)\right|}{\left|A^{1,\ldots,m}_{1,\ldots,m}(t)\right| \cdot \left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j\\ \text{exc. } p_j}}(t)\right|} \tag{4.12}$$

for both $k < p_j$ and $k > p_j$. If this quantity is substituted back into 4.10, and both the numerator and denominator of the coefficient of $n_{p_j}(t)$ are multiplied by $\left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j\\ \text{exc. } p_j}}\right|$, then 4.10 may be reduced to

$$\sum_{k=1}^{n} [c_{0k}(t) - c_{jk}(t)]n_k(t) =$$

$$\sum_{k=1}^{m} (-1)^{k+1} \frac{\left|A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } p_j}}(t)\right| \cdot \left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j\\ \text{exc. } k}}(t)\right|}{\left|A^{1,\ldots,m}_{1,\ldots,m}(t)\right| \cdot \left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j\\ \text{exc. } p_j}}(t)\right|} n_k(t)$$

$$+ (-1)^{k+1} \frac{\left|A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } p_j}}(t)\right|}{\left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j\\ \text{exc. } p_j}}(t)\right|} n_{m+j}(t) \tag{4.13}$$

While the reduction of 4.10 to 4.13 may appear to be a minor step at this point, it is in reality a very important one.

The integral equations for $y_1(t,v),\ldots,y_r(t,v)$ can be obtained by substituting 4.13 and 4.11 into Equation 4.3. After some rearranging, the resulting set of integral equations can be written in the following form:

$$\int_0^t y_q(t,v)\,\frac{\left|A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } p_q}}(t-v)\right|}{\left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+q\\ \text{exc. } p_q}}(t-v)\right|}\cdot\frac{\left|A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } p_q}}(t-u)\right|}{\left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+q\\ \text{exc. } p_q}}(t-u)\right|}\cdot$$

$$\varphi_{m+q}(t-u,t-v)\,dv+\sum_{j=1}^{r}\int_0^t y_j(t,v)\cdot$$

$$\frac{\left|A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } p_j}}(t-v)\right|}{\left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j\\ \text{exc. } p_j}}(t-v)\right|}\cdot\frac{\left|A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } p_q}}(t-u)\right|}{\left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+q\\ \text{exc. } p_q}}(t-u)\right|}$$

$$\left\{\sum_{k=1}^{m}\frac{\left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+q\\ \text{exc. } k}}(t-u)\right|}{\left|A^{1,\ldots,m}_{1,\ldots,m}(t-u)\right|}\cdot\frac{\left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j\\ \text{exc. } k}}(t-v)\right|}{\left|A^{1,\ldots,m}_{1,\ldots,m}(t-v)\right|}\varphi(t-u,t-v)\right\}dv$$

$$-\sum_{k=1}^{m}\frac{\left|A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } k}}(t)\right|\cdot\left|A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } p_q}}(t-u)\right|\cdot\left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+q\\ \text{exc. } k}}(t-u)\right|}{\left|A^{1,\ldots,m}_{1,\ldots,m}(t)\right|\cdot\left|A^{1,\ldots,m}_{1,\ldots,m}(t-u)\right|\cdot\left|A^{1,\ldots,m}_{\substack{1,\ldots,m,m+q\\ \text{exc. } p_q}}(t-u)\right|}\cdot$$

$$\varphi_k(t,t-u)=0 \tag{4.14}$$

for $0 \le u \le t$ and each $q = 1,\ldots,r$ $(r = n - m)$.

This completes the development of the "intuitive" system. It is noted in passing that 4.14 turns out to be a very convenient form for the integral equations specifying the weighting functions $y_1, \ldots, y_r$. This fact will be appreciated more in the next chapter where the "intuitive" and "optimum" systems are compared. The reader is reminded of the importance of Theorem 1 in reducing Equation 4.10 to Equation 4.13, which in turn makes the relatively simple form of 4.14 possible.

V. COMPARISON OF THE OPTIMUM AND INTUITIVE SYSTEMS

A comparison of the "optimum" system developed in Chapter III with the "intuitive" system of Chapter IV shows that both possess the same general characteristics; that is, both are linear (possibly with time varying parameters), distortionless, and finite operating time filters. The "optimum" system was, of course, constrained to be distortionless, but it is interesting to note that the distortionless property of the "intuitive" system arose quite naturally. To make a further comparison of the two systems it is assumed that both systems have the same inputs (namely those of Figure 1), that these inputs are arranged in the same order, and that the assumption stated in 3.16 still holds, i.e.

$$\left| A^{1,\ldots,m}_{1,\ldots,m}(t-v) \right| \neq 0 \text{ for all } v \text{ such that } 0 \leq v \leq t. \quad (5.1)$$

Also, it is tacitly assumed throughout the remainder of this discussion that a solution of the set of integral equations 3.20 describing the "optimum" system exists. This is not to say that finding the expressions for $w_1^*(t,v),\ldots,w_n^*(t,v)$ which constitute the solution is an easy task, but merely that such expressions do exist if some means can be found to calculate them.

All that is necessary to demonstrate that the "intuitive" system is indeed an optimum solution is to show that the mean-square error associated with the "intuitive" system is the same as the mean-square error associated with the "optimum" system.

Certainly a sufficient condition for this is that the over-all weighting function from input line i to the output is the same for the "intuitive" as it is for the "optimum" system for each $i = 1,\ldots,n$. And, in fact, this latter method turns out to be easier than the direct calculation and comparison of the mean-square errors since it avoids having to solve the set of integral equations associated with each system.

Let $w_i(t,v)$ be defined as the over-all weighting function from input line i to the output of the "intuitive" system, and let $w_i^*(t,v)$ be retained as the symbol for the corresponding weighting function for the "optimum" system. Then in view of the above paragraph it is desired to show that $w_i(t,v) = w_i^*(t,v)$ for each $i = 1,\ldots,n$. But the "intuitive" system is distortionless so $w_1(t,v),\ldots,w_m(t,v)$ are expressed in terms of $w_{m+1}(t,v),\ldots,w_n(t,v)$ by Equation 3.18 with $w_i^*(t,v)$ replaced by $w_i(t,v)$. Therefore, $w_i(t,v) = w_i^*(t,v)$ for $i = 1,\ldots,m$ if $w_i(t,v) = w_i^*(t,v)$ for $i = m+1,\ldots,n$, and the problem of showing the "intuitive" solution is an optimum one reduces to showing that $w_i(t,v) = w_i^*(t,v)$ for $i = m+1,\ldots,n$.

Since line m+ j enters only into the linear algebraic operator O_j, the output due to line m+j alone is given by

$$\int_0^t y_j(t,v)(-1)^{m+1} \frac{\left| A^{2,\dots,m}_{\substack{1,\dots,m\\ \text{exc. } p_j}}(t-v) \right|}{\left| A^{1,\dots,m}_{\substack{1,\dots,m,m+j\\ \text{exc. } p_j}}(t-v) \right|} f_{m+j}(t-v)\,dv \qquad (5.2)$$

for $j = 1,\dots,m-n$. The above expression can be written by inspecting Figure 9 together with Equation 4.7. It follows from 5.2 and the definition of the over-all weighting function $w_i(t,v)$ that

$$w_{m+j}(t,v) = (-1)^{m+1} \frac{\left| A^{2,\dots,m}_{\substack{1,\dots,m\\ \text{exc. } p_j}}(t-v) \right|}{\left| A^{1,\dots,m}_{\substack{1,\dots,m,m+j\\ \text{exc. } p_j}}(t-v) \right|} y_j(t,v) \qquad (5.3)$$

for $j = 1,\dots,n-m$. Note that 5.3 is true regardless of the behavior of the determinants in the numerator and denominator of the right-hand side of the equation.

An examination of 4.14 reveals that the quantity

$$\frac{\left| A^{2,\dots,m}_{\substack{1,\dots,m\\ \text{exc. } p_q}}(t-u) \right|}{\left| A^{1,\dots,m}_{\substack{1,\dots,m,m+q\\ \text{exc. } p_q}}(t-u) \right|} \qquad (5.4)$$

appears in every term on the left-hand side of the equation and does not involve the variable of integration or any of the indices of summation. Therefore, this quantity may be factored out of each term. At the same time Equation 4.14 may be multiplied through by $(-1)^{m+1}$ and Equation 5.3 used to replace $y_j(t,v)$ by $w_{m+j}(t,v)$ for each $j = 1,\ldots,n-m$. Thus, in terms of its over-all weighting functions, the set of integral equations describing the "intuitive" system can be written as

$$\frac{\left|A^{2,\ldots,m}_{1,\ldots,m}(t-u)\right|_{\text{exc. } p_{i-m}}}{\left|A^{1,\ldots,m}_{1,\ldots,m,i}(t-u)\right|_{\text{exc. } p_{i-m}}} \left\{\int_0^t w_i(t,v)\varphi_i(t-u,t-v)dv\right.$$

$$+ \sum_{j=m+1}^{n} \int_0^t w_j(t,v) \frac{1}{\left|A^{1,\ldots,m}_{1,\ldots,m}(t-u)\right| \cdot \left|A^{1,\ldots,m}_{1,\ldots,m}(t-v)\right|} \cdot$$

$$\left\{\sum_{k=1}^{m} \left|A^{1,\ldots,m}_{1,\ldots,m,i}(t-u)\right|_{\text{exc. } k} \cdot \left|A^{1,\ldots,m}_{1,\ldots,m,j}(t-v)\right|_{\text{exc. } k} \varphi_k(t-u,t-v)\right\}dv$$

$$\left. + (-1)^m \sum_{k=1}^{m} \frac{\left|A^{1,\ldots,m}_{1,\ldots,m,i}(t-u)\right|_{\text{exc. } k} \cdot \left|A^{2,\ldots,m}_{1,\ldots,m}(t)\right|_{\text{exc. } k}}{\left|A^{1,\ldots,m}_{1,\ldots,m}(t-u)\right| \cdot \left|A^{1,\ldots,m}_{1,\ldots,m}(t)\right|} \varphi_k(t-u,t)\right\} = 0 \qquad (5.5)$$

for $0 \leq u \leq t$ and each $i = m+1,\ldots,n$. Note that in going from 4.14 to 5.5 the dummy index q has been replaced by $i-m$ and the dummy index j has been replaced by $j-m$. These changes of indices are made so that 5.5 may be more conveniently compared to 3.20.

For the remainder of the discussion, it is assumed that there exists at least one set of integers $(p_1,\ldots,p_r)$, where as before $r = n-m$, such that for each of these integers the determinants

$$\left| A^{1,\ldots,m}_{\substack{1,\ldots,m,m+j \\ \text{exc. } p_j}}(t-v) \right| \tag{5.6}$$

and
$$\left| A^{2,\ldots,m}_{\substack{1,\ldots,m \\ \text{exc. } p_j}}(t-v) \right| \tag{5.7}$$

goes to zero at most at only a finite number of isolated points in the interval $0 \leq v \leq t$. This assumption on 5.6 along with 5.1 insures that there are at least $(n-m+1)$ m by m submatrices of the matrix $A(t-v)$ which have nonzero determinants almost everywhere in the interval $0 \leq v \leq t$, which in turn implies that Equation 4.5 can be solved for $s_1(t)$ almost everywhere in the n-m+1 different ways suggested in the "intuitive" approach. The above assumption on 5.7 insures line m+j is not given zero weight over any subinterval of the interval $0 \leq v \leq t$ by the rather arbitrary choice of the linear, algebraic operator. If the determinant 5.7 were zero on some subinterval, it could be argued intuitively that any system incorporating this linear, algebraic operator could not, in general, be expected to be an optimum system since the decision to give line m+j zero weight over that subinterval would be based on the arbitrary choice of the linear, algebraic operator and not on any property of the noise $n_{m+j}(t)$. Notice that the assumptions at the beginning of this paragraph are just the properties that one would normally expect these determinants to have when picking a linear, algebraic operator, so these assumptions amount to assuming that at least one "reasonable intuitive" system exists. From this point on, unless otherwise stated, when an "intuitive" system is mentioned it is understood that, at a minimum, the determinants 5.6 and 5.7 are nonzero except for a finite number of isolated points in the interval $0 \leq v \leq t$ for each integer included in the set of integers $p_1,\ldots,p_r$ associated with the particular linear, algebraic operator.

Proceeding to the direct comparison of the integral equations which describe the "optimum" and "intuitive" systems, it is observed that the two systems are most easily compared when neither the determinant in the numerator nor the determinant in the denominator of 5.4 is zero anywhere in the interval $0 \leq v \leq t$. In this case the quantity in front of the braces in 5.5 is nonzero which forces the quantity within the braces to be zero. If this is true for each $i = m+1,\ldots,n$ (recall that $q = i-m$), then comparison of 5.5 and 3.20 reveals that $w_i(t,v) = w_i^*(t,v)$ for each value of i. Therefore, the "intuitive" system is indeed an optimum one for this case. Furthermore, Equation 5.3 can be solved for $y_j(t,v)$ in terms of $w_{m+j}(t,v)$, and for this case, it is observed that for any fixed values of t and v, $y_j(t,v)$ is just a nonzero constant times $w_{m+j}(t,v)$. This means that not only is the existence of a solution to the set of integral Equations 4.14 assured, but also $y_j(t,v)$ is as "well behaved" as $w_{m+j}(t,v)$.

The situation considered in the previous paragraph is of considerable practical interest, but it seems as though, at least under certain assumptions, the "intuitive" system might be an optimum one under less restrictive conditions than assuming both the numerator and denominator of 5.4 nonzero everywhere for all values of $q = 1,\ldots,r$. To see how these restrictions might be relaxed, consider all the noises $n_1(t),\ldots,n_n(t)$ to have smooth, bounded autocorrelation functions. Under this assumption, each noise has a timewise correlation with itself and something can be said about the value of $n_i(t)$ from a measurement of this noise at time $t + \epsilon$, where ϵ is small. Consequently, if a measurement of $n_i(t)$ is unreliable or not available, all is not lost if a measurement of $n_i(t+\epsilon)$ is available.

In addition to the above assumption about the noises, consider t to be fixed in the following discussion. Also make the assumption that the optimum weighting function $w^*_{m+1}(t,v)$ is a smooth, bounded function in the open interval $0 < v < t$. Notice that it might not be necessary to assume this property, that $w^*_{m+1}(t,v)$ might possess it quite naturally from the solution of the set of integral Equations 3.20.

But the property is assumed here in the absence of a solution to 3.20.

For the first case in relaxing the restrictions consider the determinant in the numerator of 5.4 to be nonzero everywhere for all values of q and the determinant in the denominator to be nonzero everywhere for $q = 2,\ldots,r$. For $q = 1$, the determinant

$$\left| A_{\substack{1,\ldots,m,m+1 \\ \text{exc. } p_1}}^{1,\ldots,m}(t-u) \right| \tag{5.8}$$

is assumed nonzero for all u in the interval 0 to t except at the point $u_1 (0 < u_1 < t)$ where it is zero. If the set of integral Equations 5.5 specifying the over-all weighting functions of the "intuitive" system is compared to 3.20 for this case, it is observed that the quantity within the braces in 5.5 is forced to be zero which means that over-all weighting functions for the intuitive system are forced to obey the same set of integral equations as the "optimum" weighting functions. Consequently, for this case too, it is found that $w_i(t,v) = w_i^*(t,v)$ for each $i = m+1,\ldots,n$, and the "intuitive" system is again optimum. This case is interesting for two reasons. One is that at first glance it appears as though there might be some doubt as to whether a solution exists to the set of integral Equations 4.14 since the integrands involving the quantity 5.8 are unbounded. However, 5.3 and the assumption on $w_{m+1}^*(t,v)$ assure us that

$y_j(t,v)$ does exist for each $j = 1,\ldots,r$. The second is that the operator O_1 does not exist at time $t-u_1$ and it might appear at first that the solution could not be an optimum. However, it must be remembered that a finite amount of past data, not just the data at one instant, goes into making the estimate of s_1 at time t, and furthermore the determinant 5.8 being zero at $u = u_1$ effects the optimum solution, too, although not in such an obvious way. This concept may be extended to cases where 5.8 goes to zero at several points in the interval $0 < u < t$, and then also to similar situations for other values of q.

Next, let's try to relax the restriction that the determinant

$$\left| A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } p_q}}(t-v) \right| \tag{5.9}$$

is nonzero everywhere for each value of $q = 1,\ldots,r$. To examine this case, consider the denominator of 5.4 nonzero everywhere for each $q = 1,\ldots,r$, and the numerator of 5.4 to be nonzero everywhere for all values of q except $q = 1$. For $q = 1$, the determinant

$$\left| A^{2,\ldots,m}_{\substack{1,\ldots,m\\ \text{exc. } p_1}}(t-u) \right| \tag{5.10}$$

is assumed nonzero for all u in the interval 0 to t except at the point $u_1 (0 < u_1 < t)$ where it is zero. Comparison of the set of integral Equations 5.5 specifying the over-all weighting functions for the "intuitive" system to the set 3.20 for the "optimum" system shows that for $i = m+2,\ldots,n$ the quantity within the braces in 5.5 must be equal to zero, and therefore these $n-m-1$ integral equations are the same for both systems. For $i = 1$ there is one value of u, namely u_1, for which the quantity outside the braces in 5.5 is zero and consequently the quantity within the braces is not forced to be zero. But there is nothing wrong with setting it equal to zero anyway at this point and if this is done, the set of integral equations describing the over-all weighting functions for the "intuitive" system is again the same as the set for the "optimum" system. There is still one difficulty though, which is that even though a solution to the "optimum" system is assumed to exist, and consequently $w_{m+1}(t,v)$ exists, there is no assurance that $y_1(t,v)$ exists at $v = u_1$. However, since it is also assumed that $w_{m+1}(t,v)$ is smooth and bounded for $0 < v < t$, which seems a reasonable assumption, then 5.3 shows that $y_1(t,v)$ becomes unbounded at $v = u_1$ in such a way that the limit of the product

$$\frac{\left| A^{2,\ldots,m}_{\substack{1,\ldots,m \\ \text{exc. } p_1}}(t-v) \right|}{\left| A^{1,\ldots,m}_{\substack{1,\ldots,m,m+1 \\ \text{exc. } p_1}}(t-v) \right|} \; y_1(t,v) \tag{5.11}$$

remains bounded as $v \to u_1$. All of this suggests a way to solve the set of integral Equations 4.14, for this case, in such a way as to "force" the "intuitive" system to be optimum. That is, instead of solving directly for $y_1(t,v)$, $y_2(t,v)$, $\ldots,y_r(t,v)$, solve for $w_{m+1}(t,v)$, $y_2(t,v),\ldots,y_r(t,v)$ and then get $y_1(t,v)$ from 5.3.

The concepts discussed in the above case may be extended to situations where 5.10 goes to zero at several values of u within the open interval $0 < u < t$, and then also to similar situations for other values of q.

The results of this chapter may be summarized as follows:

1. If the determinants in both the numerator and the denominator of 5.4 are nonzero for all values of u in the interval $0 \leq u \leq t$ and for each $q = 1,\ldots,n-m$, then the corresponding "intuitive" system is an optimum system.

2. Under the assumptions that the autocorrelation functions of the noises are smooth, bounded functions and the weighting functions $w^*_{m+1}(t,v),\ldots,w^*_n(t,v)$ are smooth, bounded functions in the open interval $0 < v < t$ (t is considered fixed here), it is possible to allow the determinants in 5.4 to be to zero at a finite number of isolated points. The "intuitive" solution will still be an optimum solution providing a certain care is used in solving the integral equations for the "intuitive" system.

Because of the assumption on the "optimum" weighting functions and the fact that the "intuitive" system is "forced" to be optimum, the latter result appears to be of limited usefulness, practically speaking. This difficulty could possibly be alleviated by deriving necessary and/or sufficient conditions for the existence of a solution to a set of integral equations of the form of 3.20 or 4.14. However, this would probably be quite a difficult task.

VI. AN EXAMPLE USING THE WIENER FILTER

An example with four input lines, two signals, time stationary Gaussian noises and a constant A matrix is considered in this chapter. The system or filter is allowed to operate on an infinite amount of the past data. Consequently the optimum filter turns out to be a constant parameter linear one. This type of example is chosen because the integral equations associated with it are much easier to solve than those for the more general type of problem treated in the previous chapters. The four available input lines are of the form shown in Figure 3 with

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & -1 \\ 1 & -2 \end{bmatrix} \tag{6.1}$$

and the autocorrelation functions of the noises given by

$$\begin{aligned} \varphi_1(t) &= e^{-|t|} \\ \varphi_2(t) &= \delta(t) \\ \varphi_3(t) &= e^{-2|t|} \\ \varphi_4(t) &= 4\delta(t) \end{aligned} \tag{6.2}$$

This completes the specification of the problem. In the following sections, the "optimum" filter and 2 of the 16 possible "intuitive" systems are considered.

A. The Optimum Filter

The set of integral equations describing the optimum filter for this example is given by 3.20 specialized to an infinite operating time, constant parameter filter, time stationary noises, and a constant A matrix. These changes allow letting $t \to \infty$, replacing $w_j^*(t,v)$ by $w_j^*(v)$, replacing $\varphi_k(t-u,\ t-v)$ by $\varphi_k(u-v)$ and $\varphi_k(t,t-u)$ by $\varphi_k(u)$, and dropping the arguments in the determinants appearing in 3.20, respectively. Also, $w_j^*(v)$ can be defined as being zero for $v < 0$ to satisfy the requirement of physical realizability, and the lower limit on the integrals changed to $-\infty$. Once 3.20 is put into the form described above, one can take the Fourier transform of both sides utilizing the convolution theorem from Fourier transform theory. Letting $m = 2$ and $n = 4$, the transformed set of integral equations describing the optimum filter can be written

$$W_i^*(s)\,\Phi_i(s) + \sum_{j=3}^{4} W_j^*(s)\left[\sum_{k=1}^{2} \frac{\left|A_{\substack{1,2,i\\ \text{exc. } k}}^{1,2}\right| \cdot \left|A_{\substack{1,2,j\\ \text{exc. } k}}^{1,2}\right|}{\left|A_{1,2}^{1,2}\right|^2}\,\Phi_k(s)\right]$$

$$+ \sum_{k=1}^{2} \frac{\left|A_{\substack{1,2,i\\ \text{exc. } k}}^{1,2}\right| \cdot \left|A_{\substack{1,2\\ \text{exc. } k}}^{2}\right|}{\left|A_{1,2}^{1,2}\right|^2}\,\Phi_k(s) = A_i(s) \text{ for } i = 3,4 \qquad (6.3)$$

where $W_i^*(s)$ = Fourier transform of $w_i^*(t)$ with $j\omega$ replaced by s

$\Phi_k(s)$ = Fourier transform of $\varphi_k(t)$ with $j\omega$ replaced by s

$A_i(s)$ = an unknown function which has all its poles in the right-half s-plane

Notice that $\Phi_k(s)$ is just the familiar power spectral density with $j\omega$ replaced by s, and for the two white- and two Markov-noises of this example are given by

$$\begin{aligned}\Phi_1(s) &= \frac{2}{-s^2+1}\\ \Phi_2(s) &= 1\\ \Phi_3(s) &= \frac{4}{-s^2+4}\\ \Phi_4(s) &= 4\end{aligned} \tag{6.4}$$

Also, the various determinants involved in the two integral equations are given by

$$\begin{aligned}\left|A^{1,2}_{1,2}\right| &= 1 & \left|A^{1,2}_{2,4}\right| &= -4\\ \left|A^{1,2}_{1,3}\right| &= -2 & \left|A^{2}_{2}\right| &= 2\\ \left|A^{1,2}_{2,3}\right| &= -3 & \left|A^{2}_{1}\right| &= 1\\ \left|A^{1,2}_{1,4}\right| &= -3\end{aligned} \tag{6.5}$$

Substituting these quantities into 6.3, the two transformed equations describing the optimum system become

$$W_3^*(s)\left[\frac{4}{-s^2+4} + \frac{9(2)}{-s^2+1} + 4\right] + W_4^*(s)\left[\frac{12(2)}{-s^2+1} + 6\right] - \frac{6(2)}{-s^2+1} - 2 = A_3(s) \tag{6.6}$$

$$\text{and } W_3^*(s)[\frac{12(2)}{-s^2+1} + 6] + W_4^*(s)[4 + \frac{16(2)}{-s^2+1} - 3]$$

$$- \frac{8(2)}{-s^2+1} - 3 = A_4(s)$$

The method of undetermined coefficients is used here to solve 6.6 for $W_3^*(s)$ and $W_4^*(s)$. This method is similar to that suggested by N. Wiener in Chapter 4 of (9) and is also discussed by way of example in Chapter 15 of Brown and Nilsson (5). The notation used here is similar to that used in the latter. The basic method involves solving for $W_3^*(s)$ and $W_4^*(s)$ from the two nonhomogeneous equations of 6.6 just as if $A_3(s)$ and $A_4(s)$ were known, then expanding the resultant expression for each by partial fraction expansion utilizing the fact that $A_3(s)$ and $A_4(s)$ have all their poles in the right-half s-plane whereas $W_3^*(s)$ and $W_4^*(s)$ have all their poles in the left-half s-plane. For example, $W_3^*(s)$ is given by

$$W_3^*(s) = \frac{(-2s^2+14)(-13s^2+45)(-s^2+4)-(-6s^2+30)(-3s^2+9)(-s^2+4)}{(-s^2+1)(16s^4-206s^2+540)}$$

$$+ \frac{(-13s^2+45)(-s^2+4)A_3(s) - (-6s^2+30)A_4(s)}{16s^4 - 206s^2 + 540}$$

Since $W_3^*(s)$ is allowed to have only poles in the left-hand s-plane, it is found that it must be of the form

$$W_3^*(s) = \frac{k_4}{s+1} + \frac{k_5}{s+3.19} + \frac{k_6}{s+1.915} + k_7 \qquad (6.7)$$

Similarly, $W_4^*(s)$ is found to be of the form

$$W_4^*(s) = \frac{k_1}{s+1} + \frac{k_2}{s+3.19} + \frac{k_3}{s+1.915} \tag{6.8}$$

A constant term needs to be included in each of these expressions because of improper fractions. However, since the noise on line 4 is white, the constant in $W_4^*(s)$ must be zero or the output from that line would be infinite which would obviously not result in a minimum mean-square error. The next steps are to substitute 6.7 and 6.8 into the first of Equations 6.6, multiply, perform a partial fraction expansion of the product terms, and collect the coefficients of similar terms. The resulting equation must be an identity in s, and $A_3(s)$ has no poles in the left-half s-plane. Therefore, the coefficients of all the terms whose poles are in the left-half s-plane, in this case the coefficients of $\frac{1}{(s+1)^2}$, $\frac{1}{s+1}$, $\frac{1}{s+2}$, $\frac{1}{s+3.19}$, and $\frac{1}{s+1.915}$, must vanish. This results in 5 equations in the 7 unknown constants $k_1, \ldots, k_7$. Substituting 6.7 and 6.8 into the second of Equations 6.6 results in 4 more equations, only 2 of which are independent of the 5 derived above. Omitting all the algebra, the resulting 7 independent equations are

$$4k_1 + 3k_4 = 0$$

$$21k_1 + 7.31k_2 + 17.48k_3 + 12k_4 + 5.49k_5 + 13.14k_6 + 12k_7 = 8$$

$$6k_1 + 2.74k_2 + 6.56k_3 + 9.14k_4 + 2.06k_5 + 4.91k_6 + 4.5k_7 = 3$$

$$9.11k_2 + 3.075k_5 = 0$$

$$0.942k_3 - 3.04k_6 = 0$$

$$-0.5k_4 + 0.42k_5 - 5.88k_6 + 0.5k_7 = 0, \text{ and } 2k_7 = 1$$

From these it is found that

$$k_1 = k_4 = 0$$
$$k_2 = 0.040$$
$$k_3 = 0.11$$
$$k_5 = -0.120$$
$$k_6 = 0.034$$
$$k_7 = 0.500$$

This completes the determination of the optimum transfer functions $W_3^*(s)$ and $W_4^*(s)$. The transfer functions $W_1^*(s)$ and $W_2^*(s)$ are given in terms of these by the Laplace transform of Equation 3.18 specialized to this problem, which becomes

$$W_i^*(s) = \frac{(-1)^{i+1}}{\left|A^{1,2}_{1,2}\right|} \left\{ \left| A^{2}_{\substack{1,2\\ \text{exc. } i}} \right| + \sum_{j=3}^{4} \left| A^{1,2}_{\substack{1,2,j\\ \text{exc. } i}} \right| W_j^*(s) \right\}$$

for $i = 1,2$. With this it is found that

$$W_1^*(s) = 0.500 + \frac{0.198}{s+3.19} - \frac{0.540}{s+1.915}$$

$$W_2^*(s) = -\frac{0.119}{s+3.19} + \frac{0.397}{s+1.915}$$

$$W_3^*(s) = 0.500 - \frac{0.120}{s+3.19} + \frac{0.034}{s+1.915} \qquad (6.9)$$

$$W_4^*(s) = \frac{0.040}{s+3.19} + \frac{0.110}{s+1.915}$$

If the weighting functions for the "optimum" system are desired, they can be found by taking the inverse Laplace transform of Equation 6.9.

Although not needed for this example, the mean-square error is of interest. It can be found by specialization of the equation for $\overline{e^2(t)}$ given in Chapter III (just before Equation 3.7) to this example, or by

$$\overline{e_0^2\ (t)} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{i=1}^{4} |W_i^*(j\omega)|^2\ \phi_i(j\omega)d\omega \tag{6.10}$$

since it is a steady state error[1]. The calculation is not shown here, but the result is

$$\overline{e_0^2(t)} = 0.424.$$

B. First "Intuitive" System

The first "intuitive" system is chosen to fit the form of Figure 9 with $p_1 = 1$ and $p_2 = 2$ and is shown in Figure 11. The integral equations for $y_1(t)$ and $y_2(t)$ are given by 4.14 specialized to an infinite operating time, constant parameter filter with time stationary noises and constant coefficients on $s_1(t)$ and $s_2(t)$. As in the previous section, the integral equation of 4.14 can be transformed. The quantities of interest are given in 6.4 and 6.5. Upon substitution of these quantities, the transfer functions $Y_1(s)$ and $Y_2(s)$ can be written (after multiplying the first through by $(-\frac{3}{2})$ and the second through by (-3))

---

[1]See for example Chapter 13 of Brown and Nilsson (5) for a discussion of this method and a short table for evaluating integrals of this form.

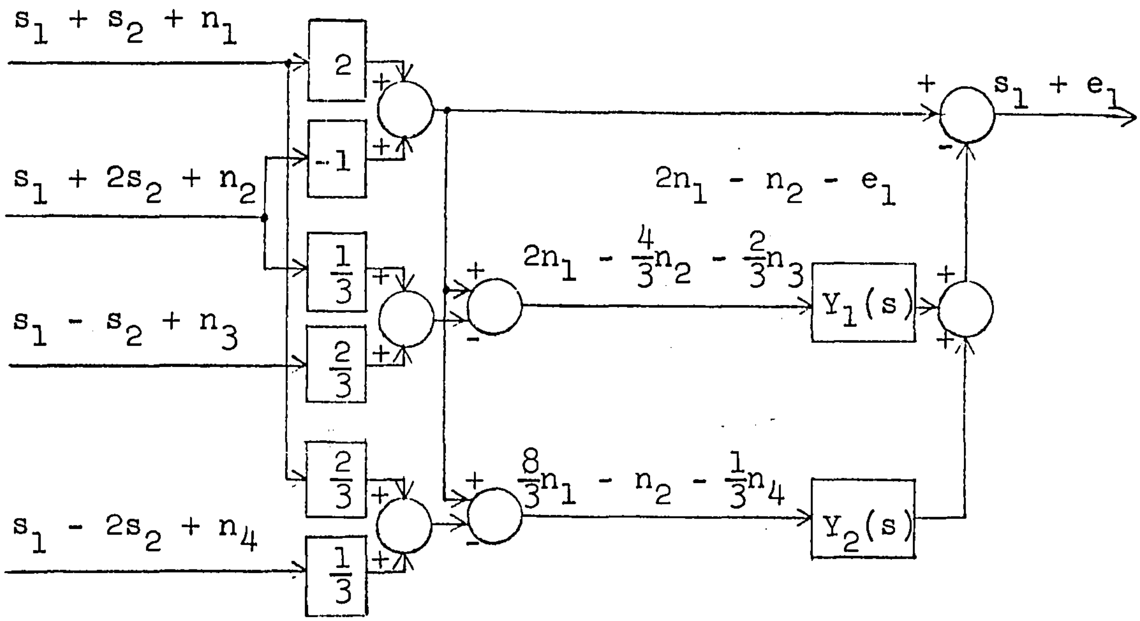


Figure 11. The first "intuitive" system for the Wiener example

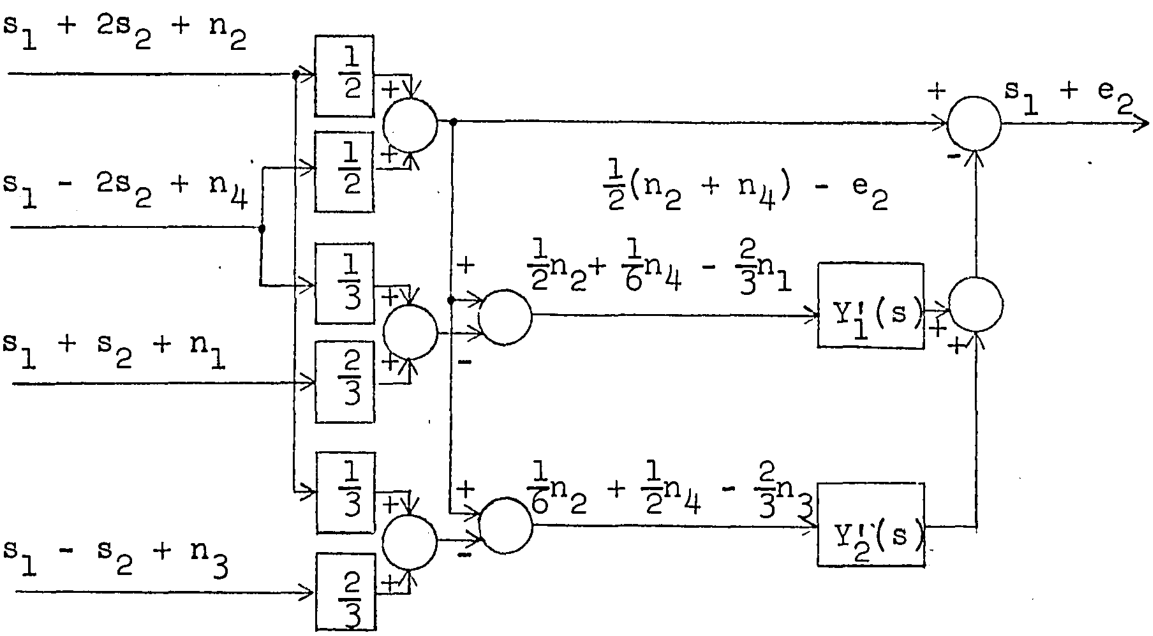


Figure 12. The second "intuitive" system for the Wiener example

$$- \frac{2}{3}Y_1(s)\left(\frac{4}{-s^2+4} + \frac{9(2)}{-s^2+1} + 4\right) + \left(- \frac{1}{3}Y_2(s)\right)\left(\frac{12(2)}{-s^2+1} + 6\right)$$

$$+ \frac{6(2)}{-s^2+1} + 2 = A_1(s)$$

and
$$- \frac{2}{3}Y_1(s)\left(\frac{12(2)}{-s^2+1} + 6\right) + \left(- \frac{1}{3}Y_2(s)\right)\left(4 + \frac{16(2)}{-s^2+1} - 3\right) \qquad (6.11)$$

$$+ \frac{8(2)}{-s^2+1} + 3 = A_2(s)$$

where $A_1(s)$ and $A_2(s)$ are unknown functions having all their poles in the right-half s-plane.

These equations can be solved for $Y_1(s)$ and $Y_2(s)$ very easily by comparison to Equations 6.6. Notice that since $A_1(s)$, $A_2(s)$, $A_3(s)$, and $A_4(s)$ are unknown functions, it cannot be said that $A_1(s) = A_3(s)$ and $A_2 = A_4(s)$. However, the effect of these unknown functions on their respective equations is such that it can be said from comparison of 6.6 and 6.11 that

$$\frac{2}{3}\, Y_1(s) = W_3^*(s)$$

$$\qquad (6.12)$$

$$\frac{1}{3}\, Y_1(s) = W_4^*(s)$$

The truth of this latter statement can easily be inferred from an article by Wong and Thomas (10) describing a general method of solving systems of equations of this type.

The easiest way to show that the mean-square error $\overline{e_1^2}$ associated with this "intuitive" system is the same as that of the "optimum" system is to compare the over-all transfer functions $W_i(s)$ of the "intuitive" system with the optimum transfer functions $W_i^*(s)$. It is observed from Figure 11 that

$$W_3(s) = \frac{2}{3} Y_1(s)$$

$$W_4(s) = \frac{1}{3} Y_2(s)$$

It follows from these equations and Equations 6.12 that $W_3(s) = W_3^*(s)$ and $W_4(s) = W_4^*(s)$. Also, from Figure 8, it is observed that

$$\begin{aligned} W_1(s) &= 2 - 2Y_1(s) - 2Y_2(s) + \frac{2}{3} Y_2(s) \\ &= 2 - 3W_3(s) - 4W_4(s) \\ &= W_1^*(s) \end{aligned}$$

and

$$\begin{aligned} W_2(s) &= -1 + Y_1(s) + Y_2(s) + \frac{1}{3} Y_1(s) \\ &= -1 + 2W_3(s) + 3W_4(s) \\ &= W_2^*(s) \end{aligned}$$

Since all the transfer functions are the same, the systems are equivalent, and obviously the mean-square errors are the same.

C. Second "Intuitive" System

For the second "intuitive" system the inputs are rearranged as shown in Figure 12. Notice that the "intuitive" system in Figure 12 fits the form shown in Figure 9 with $p_1 = 1$ and $p_2 = 2$. The integral equations for $y_1'(t)$ and $y_2'(t)$ are given by 4.14 specialized to the assumptions of this example. These integral equations can be transformed as before by defining $y_1'(t)$ and $y_2'(t)$ as being zero for $t < 0$. The quantities of interest are now given by

$$\begin{aligned}
\Phi_1'(s) &= \Phi_2(s) = 1 \\
\Phi_2'(s) &= \Phi_4(s) = 4 \\
\Phi_3'(s) &= \Phi_1(s) = \frac{2}{-s^2+1} \\
\Phi_4'(s) &= \Phi_3(s) = \frac{4}{-s^2+4}
\end{aligned} \tag{6.13}$$

and (leaving off the prime notation on the new A matrix)

$$\begin{aligned}
&\left|A_{1,2}^{1,2}\right| = -4 \qquad &&\left|A_{2,4}^{1,2}\right| = +1 \\
&\left|A_{1,3}^{1,2}\right| = -1 \qquad &&\left|A_{2}^{2}\right| = -2 \\
&\left|A_{2,3}^{1,2}\right| = 3 \qquad &&\left|A_{1}^{2}\right| = +2 \\
&\left|A_{1,4}^{1,2}\right| = -3
\end{aligned} \tag{6.14}$$

Upon substitution of these quantities, multiplication of the first equation through by (-24), and multiplication of the second equation through by (-8), the transformed integral equations specifying the transfer functions $Y_1'(s)$ and $Y_2'(s)$ can be written

$$\begin{aligned}
&-\frac{2}{3}Y_1'(s)\left(9 + 4 + \frac{16(2)}{-s^2+1}\right) - \frac{2}{3}Y_2'(s)(3 + 12) \\
&\qquad + 6 + 8 = A_1'(s) \\
\text{and } &-\frac{2}{3}Y_1'(s)(3 + 12) - \frac{2}{3}Y_2'(s)\left(1 + 36 + \frac{16(4)}{-s^2+4}\right) \\
&\qquad + 2 + 24 = A_2'(s)
\end{aligned} \tag{6.15}$$

where $A_1'(s)$ and $A_2'(s)$ are unknown functions having all their poles in the right-half s-plane.

Since it is not possible to solve for $Y_1'(s)$ and $Y_2'(s)$ by simply comparing Equations 6.15 to Equations 6.6, the method of undetermined coefficients will be used to solve for these quantities. The calculation proceeds much as it did in solving for $W_3^*(s)$ and $W_4^*(s)$ in the first section of this chapter, so most of the details are left out here. By solving the two nonhomogenous equations in 6.15 for $Y_1'(s)$ and $Y_2'(s)$ and looking at the significant terms, i.e., the ones with poles in the left-half s-plane, it is found that $Y_1'(s)$ and $Y_2'(s)$ are of the form

$$\frac{2}{3}Y_1'(s) = k_1 + \frac{k_2}{s+3.19} + \frac{k_3}{s+1.915}$$

and

$$\frac{2}{3}Y_2'(s) = k_4 + \frac{k_5}{s+3.19} + \frac{k_6}{s+1.915} \qquad (6.16)$$

When these equations are substituted back into the first of the equations in 6.15, the result must be an identity in s. Since $A_1'(s)$ has no poles in the left-half s-plane, the coefficients of all the left-half pole terms on the left-hand side of the resulting equation must be zero. The same argument can be used when Equations 6.16 are substituted into the second equation in 6.15. Together these yield six independent equations in six unknowns $k_1, \ldots, k_6$. The six equations are:

$$16k_1 + 7.31k_2 + 17.5k_3 = 0$$
$$9.125k_2 + 15k_5 = 0$$
$$0.942k_3 + 15k_6 = 0$$
$$16k_4 + 13.44k_5 - 188k_6 = 0$$
$$13k_1 + 15k_4 = 14$$
$$15k_1 + 37k_4 = 26$$

From these it is found that

$$\begin{aligned} k_1 &= 0.500 \\ k_2 &= 0.197 \\ k_3 &= -0.540 \\ k_4 &= 0.500 \\ k_5 &= -0.121 \\ k_6 &= 0.0339 \end{aligned} \tag{6.17}$$

This completes the determination of the second "intuitive" system, but it is desired, of course, to compare this system to the "optimum" system. One way to do this is to compare the over-all transfer functions $W_i'(s)$ of this system with those of the "optimum" system. The systems are equivalent if $W_1'(s) = W_2^*(s)$, $W_2'(s) = W_4^*(s)$, $W_3'(s) = W_1^*(s)$, and $W_4'(s) = W_3^*(s)$. From Figure 12, it is observed that

$$W_3'(s) = \frac{2}{3}Y_1'(s) = 0.500 + \frac{0.197}{s+3.19} - \frac{0.540}{s+1.915} = W_1^*(s)$$

$$W_4'(s) = \frac{2}{3}Y_2'(s) = 0.500 - \frac{0.121}{s+3.19} + \frac{0.0339}{s+1.915} = W_3^*(s)$$

$$W_1'(s) = \frac{1}{2} - \frac{1}{2}Y_1'(s) - (\frac{1}{2} - \frac{1}{3})Y_2'(s)$$

$$= \frac{-0.119}{s+3.19} + \frac{0.397}{s+1.915} = W_2^*(s)$$

$$
\begin{aligned}
W_2'(s) &= \frac{1}{2} - \left(\frac{1}{2} - \frac{1}{3}\right)Y_1'(s) - \frac{1}{2}Y_2'(s) \\
&= \frac{0.041}{s+3.19} + \frac{0.1096}{s+1.915} = W_4^*(s)
\end{aligned}
$$

Thus, the second "intuitive" solution is an optimum solution, too.

VII. AN EXAMPLE USING THE KALMAN FILTER

From the discussion in Chapter V, the reader may have the impression that the conclusion reached in Chapter V, namely that under appropriate assumptions the "intuitive" system is an optimum one, is of limited usefulness. The only advantage that the "intuitive" system has is that it reduces the original problem of finding the optimum, linear, distortionless filter for estimating $s_1(t)$ from the available inputs to a form whereby the estimate can be made by coupling a "ready made" filter into the system. For continuous systems, the "ready made" filter is the generalized (n-m)-dimensional Wiener filter. The only problem is that the integral equations for the generalized Wiener filter are at least as hard to solve as the integral equations describing the "optimum" system, and in addition, certain problems about the existence of the solution arise when some of the determinants in the linear, algebraic operator go to zero at one or more isolated points in the interval 0 to t. Consequently, now that the integral equations for the optimum, linear, distortionless filter have been derived and are given by 3.20, it seems advisable to solve them directly and forget about the "intuitive" system for the continuous case.

For discrete systems, the Kalman filter (8) can be chosen as the "ready made" filter (this assumes, of course, that noises $n_1(t),\ldots,n_n(t)$ can be generated by the use of shaping filters with white-noise inputs). Since the Kalman filter is the discrete analog of the generalized multidimensional Wiener filter, i.e., both minimize the mean-square error and have the same estimation properties for their respective input-output relationships, what was proved for continuous systems utilizing the latter should be true for discrete systems using the former. In other words, if only discrete measurements of the inputs shown in Figure 1 are available, then an "intuitive" system using the linear, algebraic operator of Chapter IV and the Kalman filter ought to be an optimum filter (under the linear, distortionless constraint) for estimating $s_1(t)$. This is a very useful result for discrete systems for the following reasons:

1. It eliminates the need for deriving the equations for an "optimum" discrete filter.

2. In contrast to the "ready made" filter for the continuous "intuitive" system, the "solution" to the Kalman filter is easily obtained. In fact, finding the "solution" consists of nothing more than straight forward calculation since the Kalman filter was designed specifically for a numerical, computer solution.

3. The fact that one or more of the determinants involved in the linear, algebraic operator vanish at certain isolated points in the interval 0 to t presents no difficulty since these determinants are known functions of time, and

each sampling instant can be chosen so that none of the determinants involved are zero at that time.

4. If two "intuitive" systems are constructed and have the same sampling times, it follows that they will have the same mean-square error since both will be optimum.

The example which follows in Section B of this chapter amounts to a proof of this last statement for m = 2 and n = 3. In addition it will serve as the example of the theory developed earlier for the general case where the noises are nonstationary, the A matrix is a function of time, and only a finite amount of past data is used in making the estimate.

A. The Kalman Filter Equations

The equations and presentation of the Kalman filter given here are taken largely from unpublished notes by R. G. Brown (4), but only a very brief outline of the method is offered here. The reader is referred to these notes or other publications for a more complete description.

Most of the notation in this section is the same as that used by Brown, the only exception being that a letter signifying a column vector is underlined here. The notation is summarized as follows:

1. An underlined, lower case letter denotes a column vector.

2. An upper case letter is used to denote a matrix, with the notable exceptions of b and φ which are also matrices.

3. A subscript n on any of the above quantities is used to show that the quantity is evaluated at time t_n: e.g., $b_n = b(t_n)$ and $\underline{z}_n = \underline{z}(t_n)$.

The mathematical model of the system is assumed to be of the form

$$\underline{z}_{n+1} = \varphi_n \underline{z}_n + \underline{g}_n \tag{7.1}$$

$$\underline{y}_n = M_n \underline{z}_n + \delta \underline{y}_n \tag{7.2}$$

where

$\underline{z}_n$ = state of the system at time t_n

φ_n = transition matrix

$\underline{g}_n$ = column vector of state responses due to all of the independent white-noise driving functions that occur in the interim between t_n and t_{n+1} (Note that only white-noise driving functions are allowed in the mathematical model.)

$\underline{y}_n$ = output vector (i.e., the "observable")

$\delta \underline{y}_n$ = observation noise

M_n = output matrix

Furthermore, the measurement errors are assumed to be uncorrelated (both component-wise and timewise) and unbiased, i.e.,

$$E[\delta \underline{y}_n \delta \underline{y}_i^T] = \begin{cases} V_n & \text{for } i = n \\ 0 & \text{for } i \neq n \end{cases} \tag{7.3}$$

where V_n is a diagonal matrix whose terms are the variances of the respective measurement errors, $\delta \underline{y}_i^T$ is the transpose of the column vector $\delta \underline{y}_i$, and the notation $E[x]$ indicates taking the expected or mean value of x.

For the filter, the linear relationship

$$\underline{\hat{z}}_n = \underline{\hat{z}}'_n + b_n(\underline{y}_n - \underline{\hat{y}}'_n) \tag{7.4}$$

is assumed where $\underline{y}_n$ is the observed quantity at t_n and

$\underline{\hat{z}}'_n$ = best estimate of $\underline{z}_n$ based on all past measurements up through $\underline{y}_{n-1}$ (the a priori estimate of $\underline{z}_n$)

$\underline{\hat{z}}_n$ = best estimate of $\underline{z}_n$ based on all measured data up through $\underline{y}_n$ (the a posteriori estimate of $\underline{z}_n$)

b_n = "weighting" matrix

Because the driving functions are white the a priori estimate $\underline{\hat{z}}'_n$ of $\underline{z}_n$ is given by

$$\underline{\hat{z}}'_n = \varphi_{n-1}\underline{\hat{z}}_{n-1} \tag{7.5}$$

Also, the output vector $\underline{y}$ corresponding to $\underline{\hat{z}}'_n$ is given by

$$\underline{\hat{y}}'_n = M_n\underline{\hat{z}}'_n \tag{7.6}$$

The weighting matrix b_n is then chosen to minimize the loss function L which is given by

$$\begin{aligned} L &= E[(\underline{\hat{z}}_n-\underline{z}_n)^T(\underline{\hat{z}}_n-\underline{z}_n)] \\ &= E[\underline{e}_n^T\ \underline{e}_n] \end{aligned} \tag{7.7}$$

where $\underline{e}_n$ is the estimation error. Note that L is a scalar and just the sum of the variances of the estimation errors in the elements of the state vector. It can be shown that minimizing this sum is equivalent to minimizing each individually, so the Kalman filter minimizes the mean-square error associated with the estimation of the elements of the state vector $\underline{z}_n$.

It is convenient to define two error-covariance matrices P_n and P_n^* as

$$P_n = E[\underline{e}_n \, \underline{e}_n^T] \tag{7.8}$$

and
$$P_n^* = E[\underline{e}_n' \, \underline{e}_n'^T] \tag{7.9}$$

where $\underline{e}_n' = \hat{\underline{z}}_n' - \underline{z}_n$ is the a priori estimation error.

The derivation of the expression for the optimum weighting matrix b_n is not shown here but the result is

$$b_n = P_n^* M_n^T (M_n P_n^* M_n^T + V_n)^{-1} \tag{7.10}$$

Once b_n is determined, the a posteriori estimate is given by (from 7.4 and 7.6)

$$\hat{\underline{z}}_n = \hat{\underline{z}}_n' + b_n(\underline{y}_n - M_n \hat{\underline{z}}_n') \tag{7.11}$$

The a posteriori error covariance matrix can be computed from

$$P_n = P_n^* - b_n(M_n P_n^* M_n^T + V_n) b_n^T \tag{7.12}$$

One can then extrapolate ahead $\hat{\underline{z}}_n$ and P_n to get $\hat{\underline{z}}_{n+1}'$ and P_{n+1}^* by the equations

$$\hat{\underline{z}}_{n+1}' = \varphi_n \hat{\underline{z}}_n \tag{7.13}$$

$$P_{n+1}^* = \varphi_n P_n \varphi_n^T + H_n \tag{7.14}$$

where H_n is the covariance matrix of the state responses due to the white-noise inputs, i.e.,

$$H_n = E[\underline{g}_n \underline{g}_n^T] \tag{7.15}$$

Equations 7.10 - 7.14 comprise the iterative solution for the Kalman filter. As is the case for any iterative process, one must know or assume some initial values to get started.

B. The Example

In this example, the problem of estimating $s_1(t_n)$ from three available input lines is considered by the two "intuitive" methods shown in Figures 13 and 14. The coefficients a_{ij}, the signals s_j, and the noises n_i are all explicit functions of time, but the time dependence is not shown in the figures for notational convenience. Also, the notation $\left|A^{1,2}_{i,j}\right| = \left|A_{i,j}\right|$ is used to save writing. Discrete measurements are made without error of y_a, y_b, and the "secondary observables" of both systems at the sampling times $t_1, t_2, \ldots, t_n$. (Actually, the physical situation might be that the input lines are measured, with the measurement error being included in the noises, and y_a, y_b, and the "secondary observables" are calculated without further error.) Each of these sampling times t_j is chosen so that all of the quantities a_{12}, a_{22}, a_{32}, $\left|A_{1,2}\right|$, $\left|A_{1,3}\right|$, and $\left|A_{2,3}\right|$ are nonzero when evaluated at t_j. The only assumptions made about the noises is that $n_i(t)$ is related to a white-noise function $f_i(t)$ by a first order, linear differential equation, and that the white noise functions $f_1(t)$, $f_2(t)$, and $f_3(t)$ are mutually independent.

For the system in Figure 13, the Kalman filter is to be used to estimate the quantity

$$\frac{1}{\left|A_{1,2}(t_n)\right|}\left[a_{22}(t_n)n_1(t_n) - a_{12}(t_n)n_2(t_n)\right] \qquad (7.16)$$

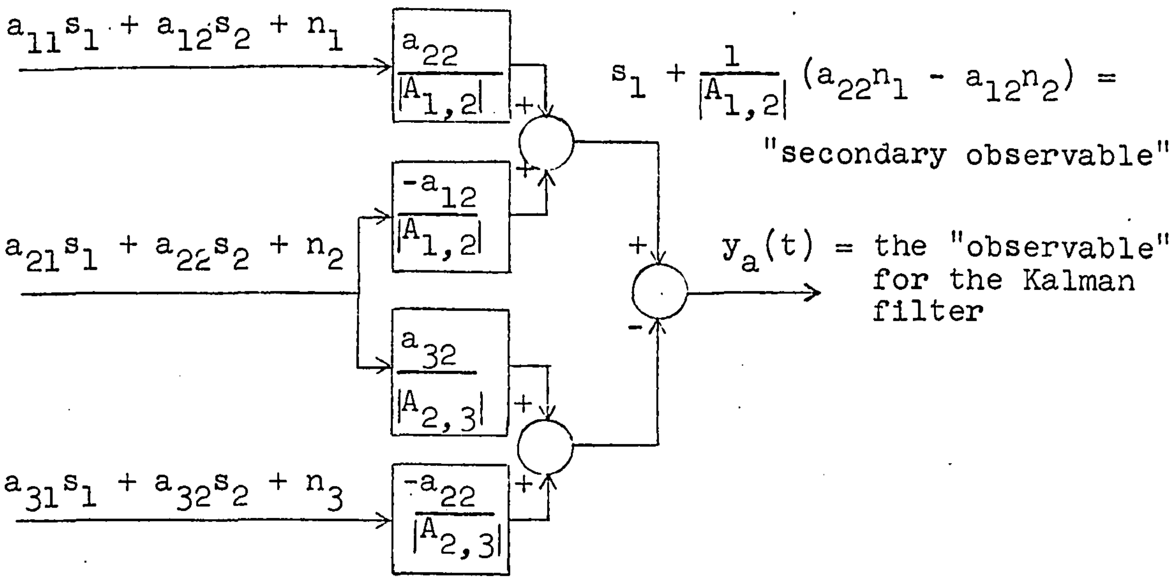


Figure 13. Linear, algebraic operator for "System a"

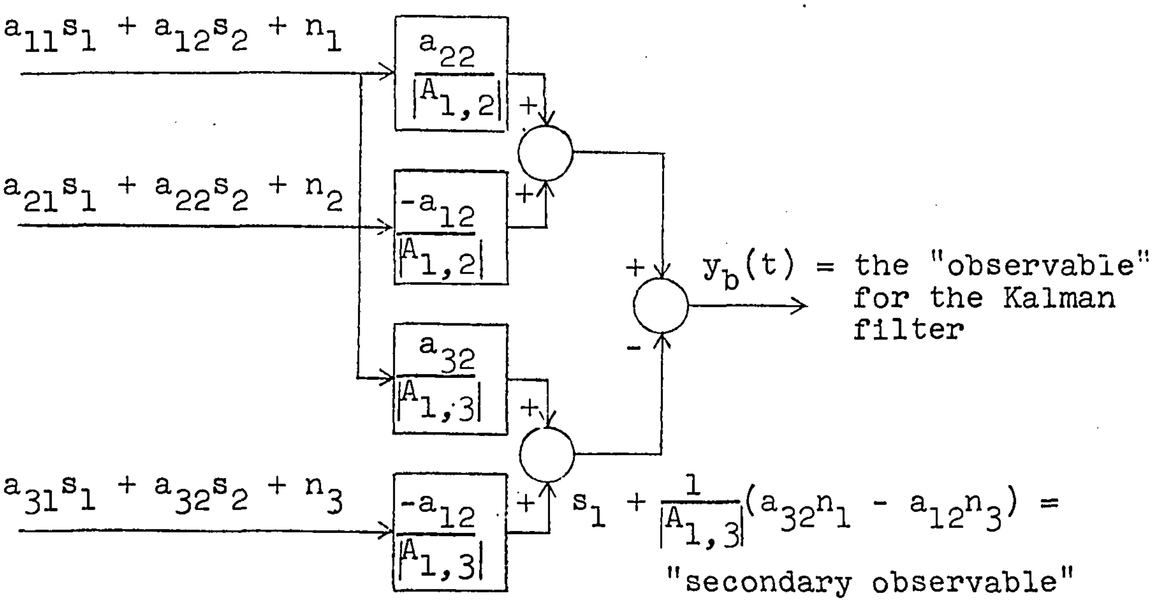


Figure 14. Linear, algebraic operator for "System b"

from the measurements of y_a at times $t_1, t_2, \ldots, t_n$. This estimate can then be subtracted from the measurement of the "secondary observable" at time t_n to get an estimate of $s_1(t_n)$. The error associated with this estimate of $s_1(t_n)$ is $e_a(t_n)$. Similarly, for the system of Figure 13, the Kalman filter is to be used to estimate the quantity

$$\frac{1}{|A_{1,3}(t_n)|}\left[a_{32}(t_n)n_1(t_n) - a_{12}(t_n)n_3(t_n)\right] \qquad (7.17)$$

from the measurements of y_b at times $t_1, t_2, \ldots, t_n$, and this estimate subtracted from the measurement of the "secondary observable" at time t_n to get an estimate of $s_1(t_n)$. The error associated with this estimate of $s_1(t_n)$ is $e_b(t_n)$.

Before proceeding farther, it is worthwhile to point out that the notation used in this section is chosen to agree with that used in the previous section. In addition the following notation is used:

1. $z_i(t_n)$ is the i^{th} component of the column vector $\underline{z}_n$. Later, it will be convenient to use the notation $z_i(t_n) = (z_i)_n$.

2. $p_{ij}(t_n)$ is the element in the i^{th} row and j^{th} column of the matrix P_n.

3. Similar notation is used for the elements of the other column vectors and matrices.

4. Many times it will be convenient to show the time dependence (or the specific fixed time of evaluation) on one side of an equation and not on the other. When this is done,

it is to be assumed that both sides are evaluated at the same point.

The state variables, z_1, z_2, and z_3, of the Kalman filter can be assigned as

$$z_i(t) = n_i(t) \quad \text{for } i = 1,\ 2,\ 3 \tag{7.18}$$

The assumptions on the noises insure that these three state variables are enough to completely describe the system and that the mathematical model has only independent white-noise driving functions. Notice that the specific expressions for the transition matrix φ_n and the covariance matrix of the state response due to the white-noise inputs, H_n, cannot be derived unless the differential equation relating $n_i(t)$ and $f_i(t)$ is known for each $i = 1$, 2, and 3. However, H_n is always a symmetrical matrix and φ_n is a diagonal matrix for this example since the noises $n_1(t)$, $n_2(t)$, and $n_3(t)$ are mutually independent.

Since the measurements of the "observables", y_a and y_b, are made without error, the matrix V_n is identically zero for both the Kalman filter of "System a" and the Kalman filter of "System b".

The only matrices (of the ones that are known) which are not the same for the two systems are the M_n matrix associated with "System a" and the M_n matrix associated with "System b", which are denoted M_n^a and M_n^b, respectively. From Figures 13 and 14 the expressions for y_a and y_b can be written as

$$y_a = \frac{a_{22}}{|A_{1,2}|} z_1 - \left(\frac{a_{32}}{|A_{2,3}|} + \frac{a_{12}}{|A_{1,2}|}\right) z_2 + \frac{a_{22}}{|A_{2,3}|} z_3 \quad (7.19)$$

and $$y_b = \left(\frac{a_{22}}{|A_{1,2}|} - \frac{a_{32}}{|A_{1,3}|}\right) z_1 - \frac{a_{12}}{|A_{1,2}|} z_2 + \frac{a_{12}}{|A_{1,3}|} z_3 \quad (7.20)$$

Since none of the quantities involved in the coefficients in 7.19 and 7.20 are zero at the sampling times, the identities

$$\frac{a_{32}}{|A_{2,3}|} + \frac{a_{12}}{|A_{1,2}|} = \frac{a_{22}\,|A_{1,3}|}{|A_{1,2}| \cdot |A_{2,3}|} \quad (7.21)$$

$$\frac{a_{22}}{|A_{1,2}|} - \frac{a_{32}}{|A_{1,3}|} = \frac{a_{12}\,|A_{2,3}|}{|A_{1,2}| \cdot |A_{1,3}|} \quad (7.22)$$

may be used to reduce two of these coefficients. Then the 1 by 3 matrix M_n^a can be written as

$$M_n^a = \frac{a_{22}}{|A_{1,2}| \cdot |A_{2,3}|} [\, |A_{2,3}| \quad -|A_{1,3}| \quad |A_{1,2}| \,]$$
$$= [(m_{11})_n \quad (m_{12})_n \quad (m_{13})_n] \quad (7.23)$$

And, it is found that at time t_n, y_b is just a constant k_n times y_a, where

$$k_n = \frac{a_{12}(t_n)\, |A_{2,3}(t_n)|}{a_{22}(t_n)\, |A_{1,3}(t_n)|} \quad (7.24)$$

Thus, the 1 by 3 matrix M_n^b can be written as

$$M_n^b = k_n[(m_{11})_n \quad (m_{12})_n \quad (m_{13})_n] \quad (7.25)$$

Let the error-covariance matrix P_n for "System a" be denoted by P_n^a and the error-covariance matrix P_n for "System b" be denoted by P_n^b. Then it is easy to show that $P_n^a = P_n^b$.

This may be done by observing from 7.10 and 7.12, together with the relationship between M_n^a and M_n^b, that $P_n^{*a}=P_n^{*b}$ implies $P_n^a = P_n^b$. Then, since $\varphi_n^a = \varphi_n^b$ and $H_n^a = H_n^b$, Equations 7.14 implies that $P_{n+1}^{*a} = P_{n+1}^{*b}$. Consequently, by assuming that $P_1^{*a} = P_1^{*b}$, which is the most logical choice anyway, it follows that

$$P_n^a = P_n^b = P_n \tag{7.26}$$

by mathematical induction. This result is what one would expect since the two measured quantities, y_a and y_b, each contain a linear combination of the noises with known, nonzero coefficients; therefore, the Kalman filter associated with "System a" and the Kalman filter associated with "System b" ought to make equally good estimates of the noises.

Since the "secondary observable" of "System a" is measured without error, $e_a(t_n)$ is given by

$$-e_a(t_n) = [\text{best estimate of } (\frac{a_{22}n_1 - a_{12}n_2}{|A_{1,2}|})] - (\frac{a_{22}n_1 - a_{12}n_2}{|A_{1,2}|}) \tag{7.27}$$

where the "best estimate" referred to is assumed to be the best estimate of the indicated sum that the Kalman filter is capable of giving, assuming y_a as the "observable". But it can be shown that for the independent state variables z_1 and z_2 of the Kalman filter equations, the best estimate of the sum, $(z_1 + z_2)$, is just equal to the sum of the best estimates; i.e.,

$$(z_1 \hat{+} z_2) = \hat{z}_1 + \hat{z}_2 \tag{7.28}$$

This result is most easily verified by recognizing that for the Kalman filter, $\underline{\hat{z}}_n = [(\hat{z}_1)_n \quad (\hat{z}_2)_n]^T$ is equal to the mean of the conditional density function $p(\underline{z}_n|\underline{y}_n)$.[1] Since z_1 and z_2 are independent and Gaussian, this conditional density function may be written as

$$p(\underline{z}_n|\underline{y}_n) = p((z_1)_n|\underline{y}_n)p((z_2)_n|\underline{y}_n) \tag{7.29}$$

where the two conditional density functions on the right-hand side are Gaussian with means of $(\hat{z}_1)_n$ and $(\hat{z}_2)_n$, respectively. One can then define a random variable $w_n = (z_1)_n + (z_2)_n$, and the mean of the conditional density function $p(w_n|\underline{y}_n)$ will be $(\hat{z}_1)_n + (\hat{z}_2)_n$. Then, if a Kalman filter were used to estimate w_n, it would pick as its best estimate $\hat{w}_n$, the mean of $p(w_n|\underline{y}_n)$. Consequently,

$$\hat{w}_n = [z_1(t_n) \hat{+} z_2(t_n)] = \hat{z}_1(t_n) + \hat{z}_2(t_n) \tag{7.30}$$

which is the desired result.

With the use of the above result, Equation 7.27 can be written

$$-e_a(t_n) = \frac{1}{|A_{1,2}|} (a_{22}(\hat{z}_1 - z_1) - a_{12} (\hat{z}_2 - z_2))$$

Squaring this and taking the mean yields

$$\overline{e_a^2(t_n)} = \frac{1}{|A_{1,2}|^2} (a_{22}^2 p_{11}(t_n) - 2a_{12}a_{22}p_{12}(t_n) + a_{12}^2 p_{22}(t_n)) \tag{7.31}$$

---

[1]See for example Brown (4), pp. 25.

Similarly, the mean-square error associated with "System b" can be written as

$$\overline{e_b^2(t_n)} = \frac{1}{|A_{1,3}|^2}(a_{32}^2 p_{11}(t_n) - 2a_{12}a_{32}p_{13}(t_n) + a_{12}^2 p_{33}(t_n)) \tag{7.32}$$

Notice that the result of Equation 7.26 has been employed to write the p_{ij}'s of Equation 7.31 and the p_{ij}'s of Equation 7.32 as elements of the same matrix, P_n, which can be calculated from either of the two systems.

The expression for P_n can be calculated in terms of the elements of P_n^* and M_n (which will be denoted by p_{ij}^{*n} and m_{li}, respectively, in the remaining equations) with the aid of Equations 7.10 and 7.12. Since P_n^* is a symmetrical matrix, it can be shown that the elements of P_n are given by

$$c_n p_{kq}(t_n) = \sum_{\substack{i=1 \\ i\neq k}}^{3} \sum_{\substack{j=1 \\ j\neq q}}^{3} m_{li} m_{lj} (p_{kq}^{*n} p_{ij}^{*n} - p_{ki}^{*n} p_{qj}^{*n}) \tag{7.33}$$

for k, q = 1, 2, 3, where

$$c_n = \sum_{i=1}^{3} \sum_{j=1}^{3} m_{li} m_{lj} p_{ij}^{*n} \tag{7.34}$$

Substituting the values of p_{11}, p_{12}, and p_{22} into Equation 7.31 and collecting terms yields

$$c_n \overline{e_a^2(t_n)} = m_{11}^2 \left(m_{12}^2 + \frac{a_{12}^2}{|A_{1,2}|^2} + \frac{2a_{12}m_{12}}{|A_{1,2}|}\right)(p_{11}^{*n} p_{22}^{*n} - (p_{12}^{*n})^2)$$
$$+ m_{11}^2 m_{13}^2 (p_{11}^{*n} p_{33}^{*n} - (p_{13}^{*n})^2 + \frac{a_{12}^2 m_{13}^2}{|A_{1,2}|^2}(p_{22}^{*n} p_{33}^{*n} - (p_{23}^{*n})^2)$$

$$+ 2m_{11}^2(m_{12}m_{13} + \frac{a_{12}m_{13}}{|A_{1,2}|})(p_{11}^{*n}p_{23}^{*m} - p_{12}^{*n}p_{13}^{*n})$$

$$+ 2m_{11}m_{13}\ (\frac{a_{12}^2}{|A_{1,2}|^2} + \frac{a_{12}m_{12}}{|A_{1,2}|})(p_{13}^{*n}p_{22}^{*n} - p_{12}^{*n}p_{23}^{*n})$$

$$- \frac{2m_{11}m_{13}^2a_{12}}{|A_{1,2}|}\ (p_{12}^{*n}p_{33}^{*n} - p_{13}^{*n}p_{23}^{*n}) \qquad (7.35)$$

Similarly, substituting the values of p_{11}, p_{13}, and p_{33} into Equation 7.32 and collecting similar terms yields

$$c_n\overline{e_b^2(t_n)} = \frac{a_{32}^2m_{12}^2}{|A_{1,3}|^2}(p_{11}^{*n}p_{22}^{*n} - (p_{12}^{*n})^2)$$

$$+ (\frac{a_{32}^2m_{13}^2}{|A_{1,3}|^2} + \frac{a_{12}^2m_{11}^2}{|A_{1,3}|^2} + \frac{2a_{12}a_{32}}{|A_{1,3}|^2}\ m_{11}m_{13})(p_{11}^{*n}p_{33}^{*n} - (p_{13}^{*n})^2)$$

$$+ \frac{a_{12}^2m_{12}^2}{|A_{1,3}|^2}(p_{22}^{*n}p_{33}^{*n} - (p_{23}^{*n})^2)$$

$$+ \frac{2a_{32}m_{12}}{|A_{1,3}|^2}\ (a_{12}m_{11} + a_{32}m_{13})(p_{11}^{*n}p_{23}^{*n} - p_{12}^{*n}p_{13}^{*n})$$

$$- \frac{2a_{12}a_{32}m_{12}^2}{|A_{1,3}|^2}\ (p_{13}^{*n}p_{22}^{*n} - p_{12}^{*n}p_{23}^{*n})$$

$$+ \frac{2a_{12}m_{12}}{|A_{1,3}|^2}\ (a_{12}m_{11} + a_{32}m_{13})(p_{12}^{*n}p_{33}^{*n} - p_{13}^{*n}p_{23}^{*n}) \qquad (7.36)$$

where it is implied that the coefficients in both 7.35 and 7.36 are evaluated at t_n. By simply utilizing Equations 7.21, 7.22, and 7.23, where appropriate, it can be shown that

$$\overline{e_a^2(t_n)} = \overline{e_b^2(t_n)} \tag{7.37}$$

Thus the two "intuitive" systems under consideration are equally good for estimating $s_1(t_n)$.

There are four more possible "intuitive" systems for this example. Under similar assumptions about the sampling times, it could be shown by an appropriate change of subscripts that the mean-square error for each of these systems is the same as the mean-square error for the two systems considered above. Consequently, this example could be considered as a general proof, for the given inputs, of the contention that all the possible "intuitive" systems are equally good.

It is interesting to note that the proof for this discrete system was accomplished by direct comparison of the mean-square errors, rather than the indirect approach that was used for continuous systems. Due to the matrix operations involved, it appears as though the extension of the above proof to the general n input line, m signal case would be very difficult at best. Consequently, proving the desired result first for continuous systems, then extending it to discrete systems appears to have been a work saving route.

VIII. SUMMARY

The goal of this thesis was to show that, with suitable assumptions about the A matrix of the input, $A(t)\underline{s}(t) + \underline{n}(t)$, and an appropriate linear, algebraic operator, an "intuitive" system having the general configuration shown in Figure 5 would give an optimum estimate of $s_1(t)$. The criterion chosen for the optimization was the minimum mean-square error criterion, with the system allowed to operate on only a finite amount of past data and constrained to be linear, physically realizable, and distortionless. With the linear, algebraic operator chosen as shown in Figure 9, the "intuitive" system was shown to be an optimum system if the determinants of Equation 5.4 did not go to zero for any values of their arguments which were of interest.

Although the linear, algebraic operator of Figure 9 is not completely general, it is sufficiently general to demonstrate that the particular choice of the linear, algebraic operator is not important. Consequently, it seems reasonable to extend the above result to the linear, algebraic operator of the general form shown in Figure 5. Sufficient conditions to insure that the "intuitive" system is then optimum would then appear to be that none of the input lines are given zero weight at any time due to the choice of the linear, algebraic operator and that the linear, algebraic operator is well defined at all values of time which are of interest. With proper regard for these conditions, we get the useful result that all the "intuitive" systems are equally good.

An attempt was made to show that the "intuitive" system was an optimum system for cases where one or the other of the determinants in 5.4 went to zero at a finite number of isolated points, but problems were encountered about the existence of a solution to the set of integral equations describing the generalized (n-m)-dimensional Wiener filter. Under certain assumptions on the optimum weighting functions and the noises, the extension appeared to be valid, but the demonstration of this result took the form of "forcing" the "intuitive" solution to be optimum rather than showing the solution existed on its own merits. This is certainly one area in which more work could be done, providing the "intuitive" solution is of sufficient value for continuous systems to merit the extra work.

As mentioned before, the results of this thesis are interesting, but of limited practical value for continuous systems. This is because the set of integral equations which describe the generalized Wiener filter associated with the "intuitive" system are just as difficult to solve as the set which describe the "optimum" system.

The results should be extendable to discrete systems which are analogous to the continuous systems above. One such discrete analog to the generalized, multidimensional Wiener filter is the Kalman filter, and therefore, if only discrete measurements of the inputs are to be used, it seems reasonable to replace the generalized Wiener filter by the Kalman filter in the "intuitive" system of Figure 5. Furthermore, for discrete systems, the sampling times can be chosen so that the determinants involved in the linear, algebraic operator are nonzero. Consequently, no restrictions need be made on the noises except that they can be generated by the use of shaping filters with white-noise inputs. The practical advantages of extending the above theory to discrete problems are:

1. It provides a convenient, optimum distortionless filter for the discrete problem.

2. It insures that one need not concern himself with trying to pick a "best" linear, algebraic operator. Simply choose one (with proper regard for not weighting any of the lines by zero); the theory insures that it will be as good as any other.

The first statement above is not meant to preclude the possibility of a direct derivation of an "optimum" distortionless filter for the discrete problem similar to what was done for continuous systems in Chapter III; it simply means that such a derivation is unnecessary. Of course, it is possible that the direct "optimum" system would offer computational advantages, and for this reason, such a derivation is suggested as a topic for further study.

It should be pointed out that it is really the distortionless requirement and the nonadaptive assumption on the "optimum" system which permit the "intuitive" system to make as good an estimate of $s_1(t)$ as the "optimum" system. The second of these conditions limits the "optimum" system to making its estimate from the knowledge of the noises; the first forces it, in effect, to operate on (n-m) independent linear combinations of the noises, even though these noises are originally unmixed.

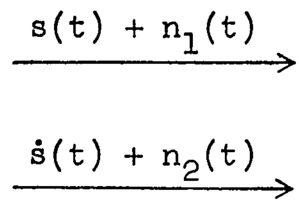
Another interesting topic for further study is suggested by considering the construction of an optimum, distortionless filter for estimating the signal s(t) from the available inputs shown in Figure 15(a), where $\dot{s}(t) = \frac{ds(t)}{dt}$. The signal s(t) is assumed to be differentiable everywhere, and the two noises are assumed to be mutually independent, nonstationary random functions. For the general linear system shown in Figure 15(b), the output can be written

$$x(t) = \int_0^t y_1(t,u)[s(t-u) + n_1(t-u)]du$$

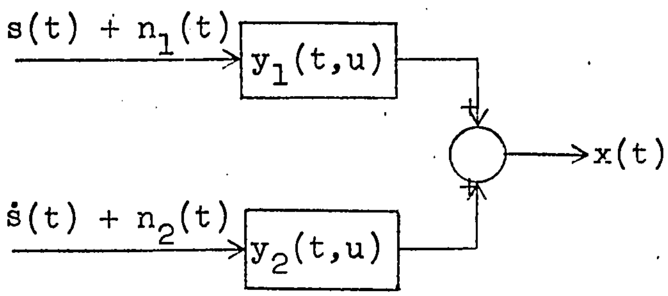
$$+\int_0^t y_2(t,u)[\frac{ds(t-u)}{d(t-u)} + n_2(t-u)]du \qquad (8.1)$$

The distortionless constraint requires that

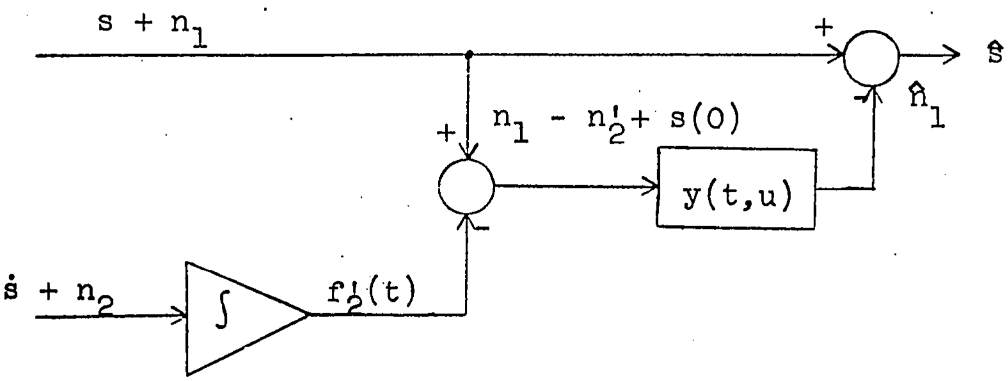
$$\int_0^t y_1(t,u)s(t-u)du + \int_0^t y_2(t,u)\,\frac{ds(t-u)}{d(t-u)}\,du - s(t) = 0 \qquad (8.2)$$



(a) The available inputs



(b) The "optimum" distortionless filter



(c) An "intuitive" system

Figure 15. A simple example with related signals

Applying the usual calculus of variations to minimize the mean-square error under the constraint 8.2 yields the following two equations which, along with 8.2, define the optimum weighting functions $y_1(t,v)$ and $y_2(t,v)$:

$$2\int_0^t y_1(t,v)\varphi_1(t-u,t-v)dv + \lambda s(t-u) = 0 \qquad \text{for } 0 \leq u \leq t$$

$$2\int_0^t y_2(t,v)\varphi_2(t-u,t-v)dv + \lambda \frac{ds(t-u)}{d(t-u)} = 0 \tag{8.3}$$

These two equations can be reduced to one by differentiating the first with respect to u and adding the resulting equation to the second of Equations 8.3. This yields

$$\int_0^t y_1(t,v)\frac{\partial}{\partial u} \varphi_1(t-u,t-v)dv + \int_0^t y_2(t,v)\varphi_2(t-u,t-v)dv = 0 \tag{8.4}$$

An "intuitive" system for estimating s(t) from the inputs of Figure 15(a) is shown in Figure 15(c). If the system is "turned on" at t = 0 with zero initial conditions, the output of the integrator at time t is given by

$$\begin{aligned} f_2'(t) &= \int_0^t [\dot{s}(u) + n_2(u)]du \\ &= s(t) - s(0) + \int_0^t n_2(u)du \\ &= s(t) - s(0) + n_2'(t) \end{aligned} \tag{8.5}$$

Assuming that s(0) is uncorrelated with both $n_1(t)$ and $n_2'(t)$, the integral equation which specifies the optimum value of the weighting function y(t,v) can be written as

$$\int_0^t y(t,v)[\varphi_1(t-u,t-v) + \varphi_2'(t-u,t-v) + s^2(0)]dv$$

$$-\varphi_1(t-u,t) = 0 \quad \text{for } 0 \le u \le t \tag{8.6}$$

Substituting the appropriate expression for the autocorrelation function of n_2' into this expression yields

$$\int_0^t y(t,v)[\varphi_1(t-u,t-v) + \int_0^{t-u}\int_0^{t-v} \varphi_2(x,z)dzdx + s^2(0)]dv$$

$$-\varphi_1(t-u,t) = 0 \quad \text{for } 0 \le u \le t \tag{8.7}$$

The over-all weighting function from input line 1 to the output of the "intuitive" system is given by

$$y_1(t,v) = \delta(v) - y(t,v) \tag{8.8}$$

Letting $f_2(t)$ represent the second input, the output due to the second input alone can be written as

$$\int_0^t y(t,u) \int_0^{t-u} f_2(x)dx\, du$$

$$= \int_0^t f_2(x) \int_0^{t-x} y(t,u)dudx$$

$$= \int_0^t f_2(t-v) \int_0^v y(t,u)dudv$$

From this it is observed that the equivalent weighting function from input line 2 to the output is given by

$$y_2(t,v) = \int_0^v y(t,u)du \tag{8.9}$$

To prove that the "intuitive" system is indeed an optimum system, it is sufficient to show that the over-all weighting functions of the "intuitive" system are a legitimate solution to Equations 8.2 and 8.4. Substitution of Equations 8.8 and 8.9 into 8.4 and integration of the Dirac delta function term yields

$$- \int_0^t y(t,v) \frac{\partial}{\partial u} \varphi_1(t-u,t-v)dv + \int_0^t [\int_0^v y(t,x)dx]\varphi_2(t-u,t-v)dv$$

$$+ \frac{\partial}{\partial u} \varphi_1(t-u,t) = 0 \tag{8.10}$$

as one condition that $y(t,v)$ must satisfy if the "intuitive" system is to be an optimum one. That $y(t,v)$ does indeed satisfy Equation 8.10 can be shown by taking the partial derivative with respect to u of Equation 8.7. This yields

$$\int_0^t y(t,v)[\frac{\partial}{\partial u} \varphi_1(t-u,t-v) - \int_0^{t-v} \varphi_2(t-u,z)dz]dv$$

$$- \frac{\partial}{\partial u} \varphi_1(t-u,t) = 0 \tag{8.11}$$

Comparison of Equations 8.10 and 8.11 shows that they are equivalent if

$$\int_0^t y(t,v) \int_0^{t-v} \varphi_2(t-u,z)dzdv = \int_0^t [\int_0^v y(t,x)dx]\varphi_2(t-u,t-v)dv \tag{8.12}$$

By replacing v by x and interchanging the order of integration, the left-hand side of 8.12 can be written as

$$\int_0^t \int_0^{t-z} y(t,x)\varphi_2(t-u,z)dxdz$$

Then, letting $v = t - z$, this becomes

$$\int_0^t [\int_0^v y(t,x)dx]\varphi_2(t-u,t-v)dv$$

which shows that 8.12 is an identity. Consequently, it can be concluded that the solution $y(t,v)$ to Equation 8.7 will satisfy Equation 8.10.

By inspection of Figure 15(c), it is noted that if the "intuitive" system is to be distortionless, the weighting function $y(t,v)$ must give zero weight to the constant $s(0)$ which appears in its input (i.e., in the frequency domain, $Y(t,\omega)$ must be zero at $\omega = 0$). This also turns out to be the only requirement of $y(t,v)$ that is necessary to satisfy the distortionless constraint, as can be demonstrated by substituting 8.8 and 8.9 into Equation 8.2. Furthermore, it is observed that the solution $y(t,v)$ to Equation 8.7 will satisfy this requirement since Equation 8.7 must be true for arbitrary values of $s(0)$. Thus, the over-all weighting functions of the "intuitive" system represent a valid solution to the integral equations specifying the "optimum" distortionless filter.

There are many possible generalizations of the above example which appear to be worth investigating. For example, for the 2 inputs in Figure 15(a), one might try replacing $\dot{s}(t)$ by $L[s]$, where L is some general linear operator. Then the integrator in Figure 15(c) would be replaced by the appropriate inverse operator of L, denoted by L^{-1}. It is noted that the derivative operator is an example of such a linear operator, and happens to be one for which the inverse operator is not single-valued. Defining the inverse operator to be the definite integral with limits 0 to t eliminated this problem but introduced the troublesome initial value s(0). The fact that, in the above example, the "intuitive" system was an optimum distortionless system in spite of this difficulty is encouraging. Consequently, one would expect a similar result for the situation where $\dot{s}(t)$ is replaced by L[s], at least for the "well-behaved" operators for which the operational products LL^{-1} and $L^{-1}L$ are the same. Note that the derivative operator and its inverse do not satisfy this requirement, so by this criterion are not "well-behaved".

The next obvious step is to try to extend the above generalization to the multiple-input problem; that is, investigate the "intuitive" system for the case where the inputs are of the form shown in Figure 3 with each algebraic coefficient $a_{ij}(t)$ replaced by a linear operator L_{ij} operating on the signal $s_j(t)$.

IX. LITERATURE CITED

1. Ayres, F., Jr. Theory and problems of matrices. New York, N.Y., Schaum Publishing Co. 1962.

2. Bendat, J. S. Optimum time-variable filtering for non-stationary random processes. Technical Report NAI-54-771. (Northrop Aircraft, Inc., Hawthorne, Calif.) 1954.

3. Bendat, J. S. Principles and applications of random noise theory. New York, N.Y., John Wiley and Sons, Inc. 1958.

4. Brown, R. G. Linear filtering and prediction notes. Unpublished mimeographed notes for Electrical Engineering Course 675A. Ames, Iowa, Department of Electrical Engineering, Iowa State University of Science and Technology. ca. 1964.

5. Brown, R. G. and J. W. Nilsson. Introduction to linear systems analysis. New York, N.Y., John Wiley and Sons, Inc. 1962.

6. Dolph, C. L. and M. A. Woodbury. On the relation between Green's functions and covariances of certain stochastic processes and its application to unbiased linear prediction. American Mathematical Society Transactions 72: 519-550. 1952.

7. Fulks, W. Advanced calculus. New York, N.Y., John Wiley and Sons, Inc. 1961.

8. Kalman, R. E. A new approach to linear filtering and prediction problems. American Society of Mechanical Engineers Transactions (Journal of Basic Engineering) 82D: 35-45. 1960.

9. Wiener, N. The interpolation, extrapolation and smoothing of stationary time series. New York, N.Y., John Wiley and Sons. Inc. 1949.

10. Wong, E. and J. B. Thomas. On the multidimensional prediction and filtering problem and the factorization of spectral matrices. Franklin Institute Journal 272: 87-99. 1961.

11. Zadeh, L. Optimum filters for the detection of signals in noise. Institute of Radio Engineers Proceedings 40: 1223-1231. 1952.

X. ACKNOWLEDGMENTS

The author wishes to thank Dr. R. G. Brown, his major professor, for his helpful suggestions and comments in the selection and investigation of this topic. Credit is also due the National Science Foundation, whose support made this work possible.