XI. STATISTICAL COMMUNICATION THEORY

Prof. Y. W. Lee	H. Hemami	M. Schetzen
Prof. A. G. Bose	J. D. Holmes	H. L. Van Trees, Jr.
Prof. I. M. Jacobs	K. L. Jordan, Jr.	S. Weinreb
Prof. J. B. Wiesner	D. J. Sakrison	C. E. Wernlein, Jr.
A. D. Hause		G. D. Zames

A. WORK COMPLETED

1. NONLINEAR LEAST-SQUARE FILTERING AND FREQUENCY MODULATION

The present study has been completed by A. D. Hause. In September 1960, he submitted the results to the Department of Electrical Engineering, M.I.T., as a thesis in partial fulfillment of the requirements for the degree of Doctor of Science. The study will also be published as Technical Report 371.

Y. W. Lee

2. DISCRETE REPRESENTATION OF RANDOM SIGNALS

This study was completed and presented by K. L. Jordan, Jr. as a thesis in partial fulfillment of the requirements for the degree of Doctor of Science, Department of Electrical Engineering, M.I.T., September 1960. The study will also be published as Technical Report 378.

Y. W. Lee

3. NONLINEAR OPERATORS FOR SYSTEM ANALYSIS

This study has been completed by G. D. Zames. It was submitted as a thesis in partial fulfillment of the requirements for the degree of Doctor of Science, Department of Electrical Engineering, M.I.T., September 1960, and will also be published as Technical Report 370.

Y. W. Lee

B. A COMPARISON OF THE LAGUERRE SET WITH AN OPTIMUM SET FOR THE EXPANSION OF THE PAST OF A CERTAIN RANDOM PROCESS

As has been stated previously (1, 2), the optimum expansion of the past of a random process x(t) with respect to a weighted mean-square error involves the solution of the integral equation

$$\int_{-\infty}^{0} G(s)\, R(s,t)\, G(t)\, \phi_n(t)\, dt = \lambda_n \phi_n(s) \qquad -\infty < s \leq 0 \tag{1}$$

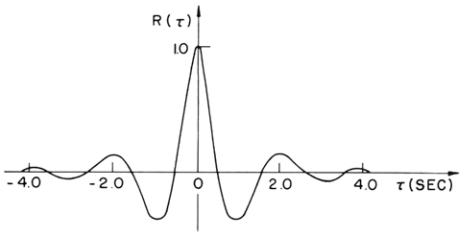


Fig. XI-1. Autocorrelation function of the process to be expanded.

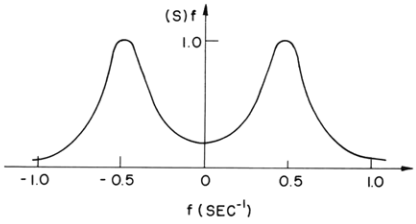


Fig. XI-2. Power density spectrum of the process to be expanded.

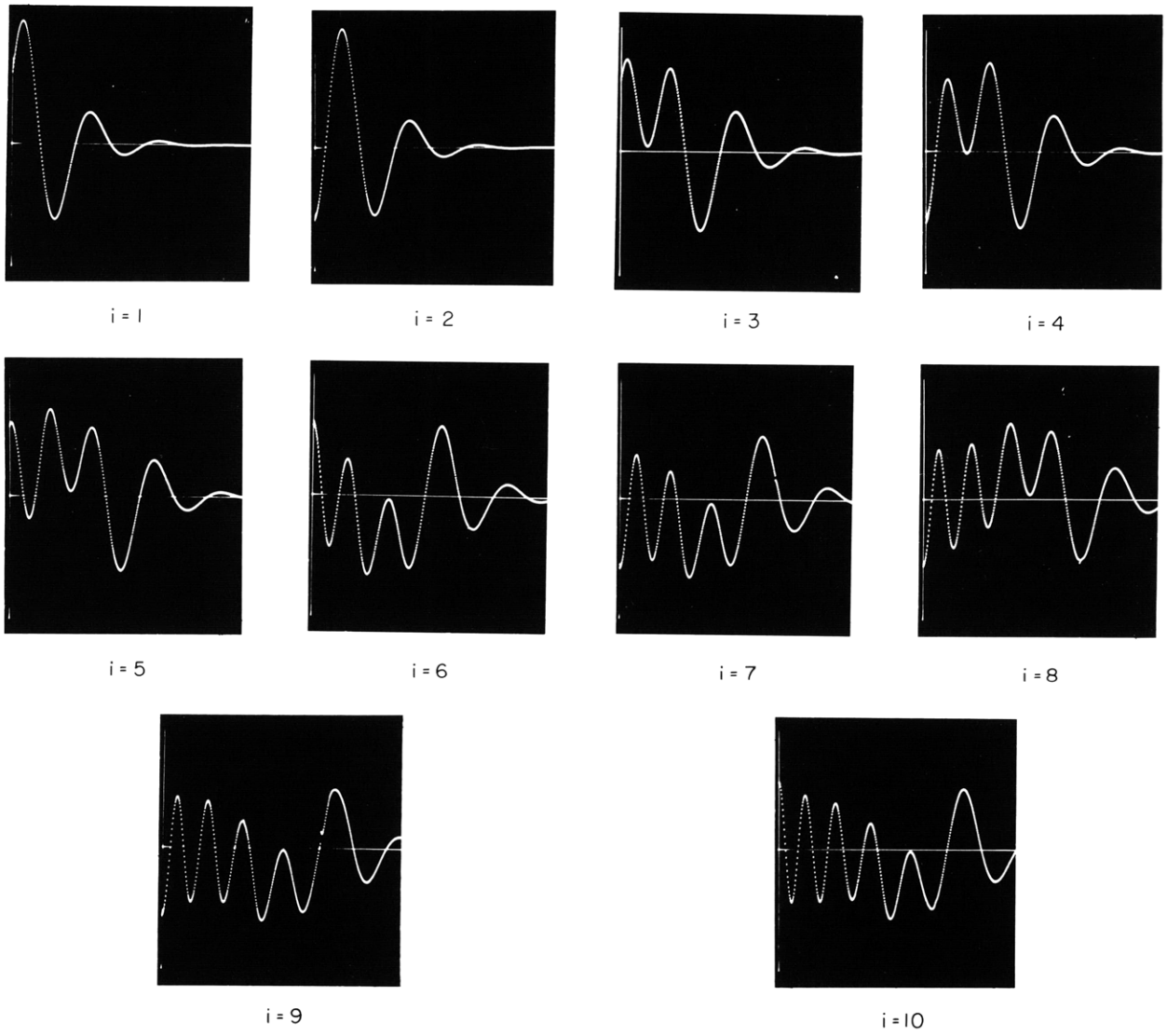


Fig. XI-3. Orthonormal set of solutions of Eq. 2.

where G(t) is the weighting function over the past of the random process x(t), and R(s, t) is the correlation function. We also note that the set $\{\phi_n(t)\}$ is the set that minimizes the mean square error of a finite-term expansion of the process

$$G(t)\, x(t) \qquad t \leq 0$$

We have developed a computer program for the solution of equations of the form of Eq. 1 on the IBM 704 computer at the Computation Center, M. I. T. This program was used on a particular example, and the resulting set of functions was used to make finite-term expansions of a sample function of the process obtained in the laboratory.

For our example we chose a stationary process with the correlation function

$$R(\tau) = \exp[-|\tau|] \cos 3\tau$$

and a weighting function G(t) = exp[t/4]. The correlation function and power density spectrum of the process are shown in Figs. XI-1 and XI-2. The process was obtained in the laboratory by passing white noise through a linear filter with the proper transfer function.

The integral equation

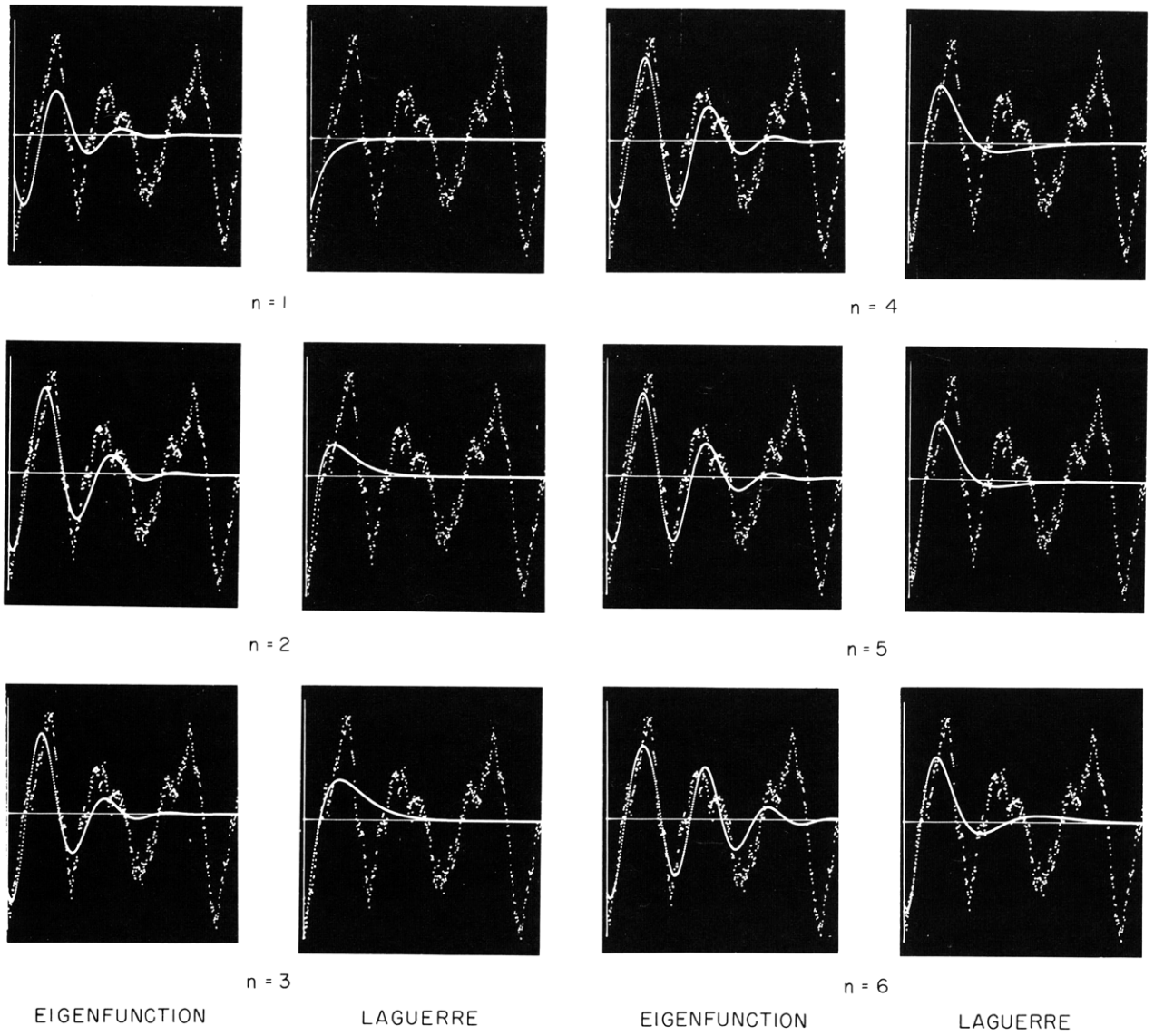
$$\int_0^\infty \exp\left[-\frac{s}{4}-\frac{t}{4}-|s-t|\right] \cos 3(s-t)\, \phi_n(t)\, dt = \lambda_n \phi_n(s) \qquad 0 \leq s < \infty \tag{2}$$

in which we have reversed the sign of the argument for convenience, was solved by means of the program. The first 10 solutions are shown in Fig. XI-3.

The approximations of a sample function of the process over a period of time of 7.5 sec, made with the use of the solutions or eigenfunctions of the integral equation and Laguerre functions for n = 1, . . . , 10, 15, and 20 terms, are shown in Fig. XI-4. It is seen that the eigenfunctions do much better, especially in approximating the higher frequency portions, than the Laguerre functions. This is because the Laguerre functions have Fourier transforms of the form

$$\frac{1}{n!} \frac{\left(j2\pi f-\frac{1}{2}\right)^n}{\left(j2\pi f+\frac{1}{2}\right)^{n+1}} \qquad n = 0, 1, 2, \ldots$$

so that most of their energy is near the origin. It is seen from Fig. XI-2,


n = 1
n = 4
n = 2
n = 5
n = 3
n = 6
EIGENFUNCTION
LAGUERRE
EIGENFUNCTION
LAGUERRE

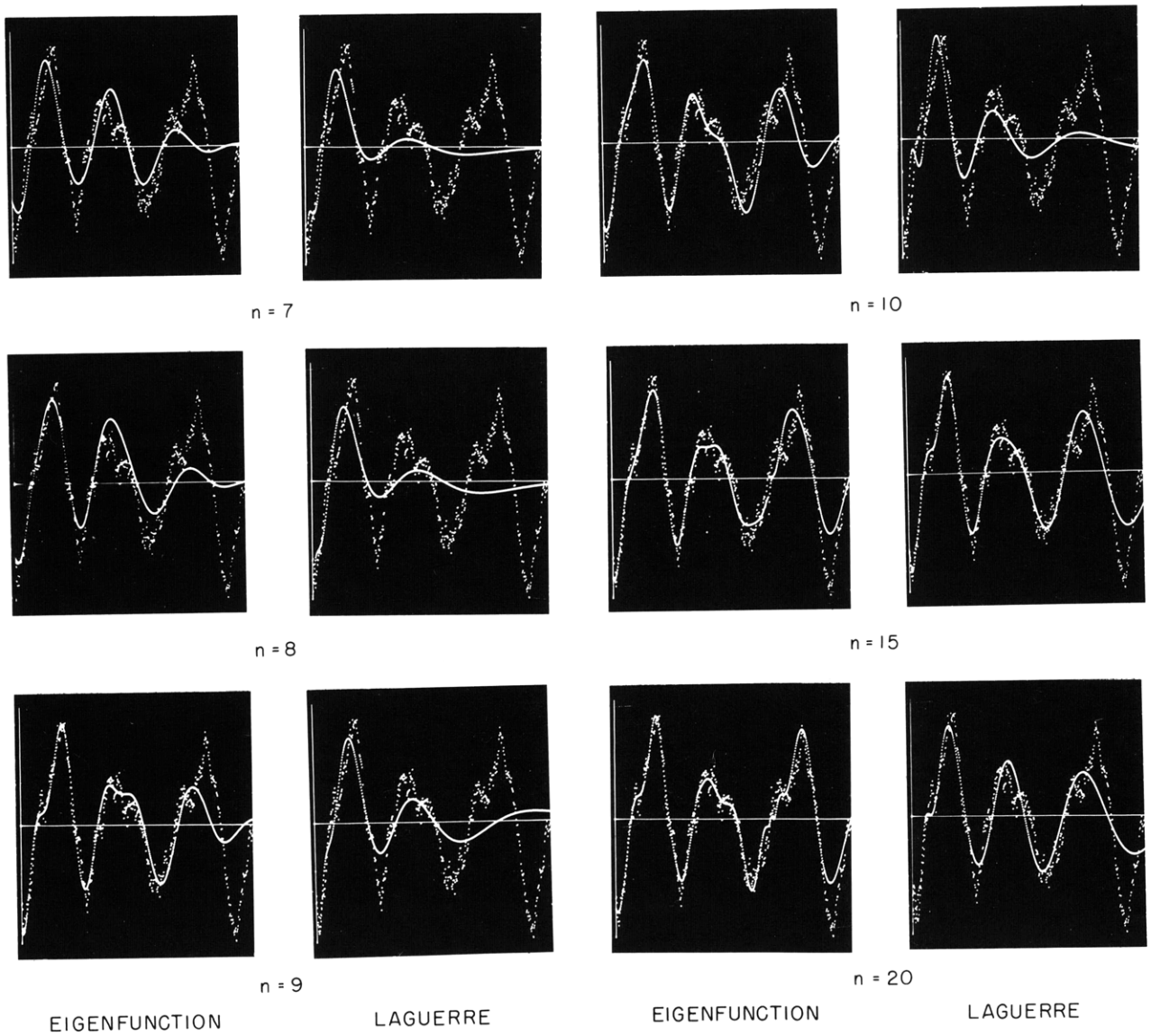


Fig. XI-4. Comparison of the expansion of a sample of the process with the use of the eigenfunctions of Eq. 2 and Laguerre functions.

however, that most of the energy in the random process is not near the origin so that the performance of the Laguerre functions is not expected to be near optimum.

K. L. Jordan, Jr.

References

1. K. L. Jordan, Jr., Characterization of random signals with an application to nonstationary linear systems, Quarterly Progress Report No. 56, Research Laboratory of Electronics, M.I.T., Jan. 15, 1960, pp. 117-125.

2. K. L. Jordan, Jr., Discrete Representations of Random Signals, Sc.D. Thesis, Department of Electrical Engineering, M.I.T., Aug. 22, 1960.

C. NOISE ANALYSIS IN MAGNETIC TAPE RECORDING

In order to examine the possibility of increasing the dynamic range in magnetic tape recording an analysis of the noise was made on an Ampex Model 351 recorder. Autocorrelation analysis was used to conveniently distinguish the various components of the recorder output noise. By analyzing the recorder output with the tape not running and with it running under different erasure conditions, the contribution to the total output noise from the tape noise, the 60-cycle noise, the remnant signal after erasure by bulk and machine-head methods, and the noise in the electronic amplifiers was determined. The results of this analysis can be found in the writer's S.M. thesis, "Noise Analysis in Magnetic Tape Recording," Department of Electrical Engineering, M.I.T., June 1960.

This analysis has shown that, consistent with good engineering design, the noise power at the machine output caused by the recording and reproducing amplifiers is nearly equal (within 2 db) to the noise power resulting from the tape itself. Thus, in order to extend the recorder's dynamic range, it is necessary to improve the amplifiers and also to increase the dynamic range of the signal on the tape. The former can be achieved by better circuit design and construction. Use of nonlinear pre-emphasis to achieve the latter will be investigated by other members of the group.

P. Piqué

D. AN ITERATIVE PROCEDURE FOR SYSTEM OPTIMIZATION

1. Introduction

We are interested in the following general problem. We are given a system, with k variable parameters $x_1, x_2, \ldots x_k$, which operates on an input v(t) and produces an output q(t). Corresponding to the input there is a desired output d(t), and we wish to minimize the performance criterion

$$M(\underline{x}) = E\{W[d(t)-q(t)]\,|\,\underline{x}\}$$

where W(e) is some non-negative function of the error, d - q.

One approach to the problem would be to use an iterative method; that is, we start with an initial parameter setting and attempt to make a sequence of adjustments which ultimately results in the optimum parameter setting. Such a procedure is useless, however, unless it can be shown that the iterative method used converges in some meaningful sense. Here stochastic approximation methods are used to show convergence. These methods are essentially gradient methods; that is, at each step we attempt to measure the direction in which $M(\underline{x})$ decreases fastest and then change the parameter setting some amount a_n in that direction. Such methods are useful only in situations in which $M(\underline{x})$ has a unique minimum. This will be the principal restriction on the classes of systems and weighting functions of the error to which the method is applicable.

Here we consider, in particular, the design of nonlinear (or linear) filters and predictors. Throughout we denote the k-tuple of parameters $x_1, x_2, \ldots x_k$ by a k-dimensional vector $\underline{x}$. The usual inner product,

$$\sum_{i=1}^{k} x_i y_i$$

is denoted $[\underline{x}, \underline{y}]$, and the k-dimensional Euclidean norm, $\|\underline{x}\|$. A unit setting of the parameter x_i and zero setting of the other k - 1 parameters will be represented by the unit vector $\underline{e}_i$, i = 1, 2, ... k.

The discussion will be carried out in terms of sampled or discrete time parameter signals and systems. A remark at the end of this report will indicate the extension of the analysis to the continuous case.

2. Physical Design Procedure

We consider the design of filters and predictors of the form shown in Fig. XI-5. The form is general in that any compact operator can be approximated arbitrarily closely by the given form if a sufficiently large number of terms is used (1, 2). It is assumed that v(m) and d(m) are the outputs of stationary ergodic sources.

The iterative adjustment is made as follows: Suppose we are at the n^{th} iteration and the parameters are at the setting $\underline{x}_n$. Let $Y_n(m) = W[d_n(m)-q_n(m)]$. We then make the 2k measurements, $Y_n^1, Y_n^2, \ldots Y_n^{2k}$, where

Y_n^1 is an observation with parameter setting $\underline{x} = \underline{x}_n + c_n\underline{e}_1$

Y_n^2 is an observation with parameter setting $\underline{x} = \underline{x}_n - c_n\underline{e}_1$

$\vdots$

Y_n^{2k-1} is an observation with parameter setting $\underline{x} = \underline{x}_n + c_n\underline{e}_k$

Y_n^{2k} is an observation with parameter setting $\underline{x} = \underline{x}_n - c_n\underline{e}_k$

We then form the vector

$$\underline{Y}_n = \left[\left(Y_n^1 - Y_n^2\right), \ldots, \left(Y_n^{2k-1} - Y_n^{2k}\right)\right] \tag{1}$$

and determine the next parameter setting in the iterative procedure by

$$\underline{x}_{n+1} = \underline{x}_n - \frac{a_n}{c_n} \underline{Y}_n \tag{2}$$

where $\{a_n\}$ and $\{c_n\}$ are sequences of positive numbers whose properties will be described later. This completes the n^{th} iteration of the procedure.

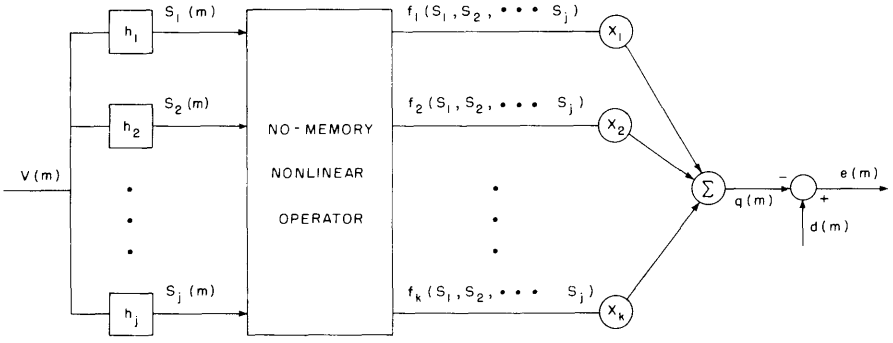


Fig. XI-5. Form of filter to be designed.

If the design is being carried out on a computer, each of the 2k measurements for the n^{th} iteration may be carried out by using the data v(m), v(m-1) . . . , where m is the time at which the n^{th} iteration is initiated. If the design is being carried out on an operational system or analog thereof, the measurements must then be made by using v(m), v(m-1), . . . for Y_n^1, . . . , and v(m+2k), v(m+2k-1), . . . for Y_n^{2k}. The observation Y_n^i need not be made with just one sample of the output data but can be taken with the parameters fixed over several samples, and the average used for Y_n^i (i=1, 2, . . . 2k). Our theory will include all such variations.

We must place a restriction on the time interval between successive iterations, however. Thus, if for the n^{th} iteration the data v(m), v(m-1), . . . are used and the data v(m+s), v(m+s-1), . . . are used for the $n + 1^{th}$ iteration, convergence of the iterative procedure is guaranteed only for s sufficiently large. We also assume that at the start of the iterative procedure the linear memory units $h_1, h_2, \ldots h_j$ have been operating on the data long enough to come to equilibrium.

We shall now state some useful mathematical results and then show for what restrictions on the physical situation the mathematical results are applicable.

3. A Stochastic Approximation Theorem

We now state extensions of the work of Dupač (3) to the multidimensional case in which independence between samples is no longer assumed. At the outset, the proofs follow the lines of Dupač's proofs and, after a point, follow Dupač's almost exactly.

Consider the multidimensional regression function

$$M(\underline{x}) = \int Y \, dH(Y|\underline{x})$$

We assume that $M(\underline{x})$ has a unique minimum at $\underline{x} = \underline{\theta}$. Let $Y_{\underline{x}_n + c_n \underline{e}_i}$ denote an observation of Y with $\underline{x} = \underline{x}_n + c_n \underline{e}_i$ and let

$$\underline{Y}_n = \left[\ldots, Y_{\underline{x}_n + c_n \underline{e}_i} - Y_{\underline{x}_n - c_n \underline{e}_i}, \ldots \right]$$

$$\underline{M}_{c_n}(\underline{x}_n) = \frac{1}{c_n} \left[\ldots, M(\underline{x}_n + c_n \underline{e}_i) - M(\underline{x}_n - c_n \underline{e}_i), \ldots \right]$$

We now define our iterative procedure by the following recursion relation:

$$\underline{x}_{n+1} = \underline{x}_n + \frac{a_n}{c_n} \underline{Y}_n$$

In order to make some statements concerning the sequence $\{\underline{x}_n\}$ thus generated, we make the following assumptions:

(i) $E\left\{ \| \underline{Y}_n \|^2 | \underline{x}_n, \underline{x}_1 \right\} \leq \| E\{\underline{Y}_n | \underline{x}_n, \underline{x}_1\} \|^2 + S \qquad S < \infty$

(ii) (a) $K_o \| \underline{x} - \underline{\theta} \|^2 \leq [(\text{grad } M)(\underline{x}), -(\underline{x} - \underline{\theta})]$

(b) $\| (\text{grad } M)(\underline{x}) \| \leq K_1 \| \underline{x} - \underline{\theta} \| \qquad K_1 > K_o > 0$

(iii) (a) $E\{ \| E\{\underline{Y}_n | \underline{x}_n, \underline{x}_1\} - c_n \underline{M}_{c_n}(\underline{x}_n) \| \, | \underline{x}_1 \} = E\{ \| E\{\underline{Y}_n | \underline{x}_n, \underline{x}_n - \underline{x}_1\}$

$$- c_n \underline{M}_{c_n}(\underline{x}_n) \| \, | \underline{x}_1 \} \leq S_1 \frac{a_n}{c_n}$$

where $S_1 < \infty$, and $\underline{x}_n - \underline{x}_1 = \sum_{j=1}^{n-1} \frac{a_j}{c_j} \underline{Y}_j$.

(b) $\left| E\left\{ \| E\{\underline{Y}_n | \underline{x}_1, \underline{x}_n\} \|^2 - c_n^2 \| \underline{M}_{c_n}(\underline{x}_n) \|^2 | \underline{x}_1 \right\} \right| < S_2, \quad S_2 < \infty$

(iv) $\{a_n\}$ and $\{c_n\}$ are sequences of positive numbers satisfying

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty, \quad \sum_{n=1}^{\infty} a_n c_n < \infty, \text{ and } \sum_{n=1}^{\infty} \frac{a_n^2}{c_n^2} < \infty$$

(v) $\underline{x}$ is constrained to a bounded, closed, convex set X, but is free to be varied inside X. (A set X is convex if $\underline{x}_1 \in X$, $\underline{x}_2 \in X$ implies $\alpha\underline{x}_1 + (1-\alpha)\underline{x}_2 \in X$ for $0 \leq \alpha \leq 1$.) Assumptions (i), (ii), and (iii) need hold only for $\underline{x} \in X$. It is assumed that X is chosen sufficiently large that $\underline{\theta} \in X$.

We now make the following statements:

STATEMENT 1: Assumptions (i)-(v) imply the convergence of $\|\underline{x}_n - \underline{\theta}\|$ to zero in the mean-square sense. That is,

$$\lim_{n\to\infty} E\left\{\|\underline{x}_n - \underline{\theta}\|^2 \,|\, \underline{x}_1\right\} = 0$$

We now set $a_n = a/n^{\alpha}$, $c_n = c/n^{\gamma}$, and in order to satisfy assumption (iv) we require $3/4 < \alpha \leq 1$, $1 - \alpha < \gamma < \alpha - 1/2$. We also require $a > 1/k_o$ if $\alpha = 1$.

STATEMENT 2: Assumptions (i)-(v) and the choice $\alpha = 1$, $\gamma = 1/4$ imply

$$E\left\{\|\underline{x}_n - \underline{\theta}\|^2 \,|\, \underline{x}_1\right\} = 0(1/n^{1/2})$$

The sequence f(n) = 0(g(n)) implies that

$$\overline{\lim_{n\to\infty}} \left(|f(n)|/|g(n)|\right) < +\infty$$

Furthermore, this choice is optimum in the sense that no other choice guarantees faster convergence for all $Y(\underline{x})$ that satisfy assumptions (i)-(v); that is, for $\alpha \neq 1$, $\gamma \neq 1/4$, there exists a $Y(\underline{x})$ that satisfies assumptions (i)-(v) for which

$$E\left\{\|\underline{x}_n - \underline{\theta}\|^2 \,|\, \underline{x}_1\right\} = 0\left(\frac{1}{n^{1/2}} - \epsilon\right) \text{ for some } \epsilon > 0$$

STATEMENT 3: With the additional assumption,

$$\left|\frac{\partial^3 M(\underline{x})}{\partial x_i^3}\right| \leq Q < \infty \qquad i = 1, 2, \ldots, k \text{ and } \underline{x} \in X$$

the choice $\alpha = 1$, $\gamma = 1/6$ implies

$$E\left\{\|\underline{x}_n - \underline{\theta}\|^2 \underline{x}_1\right\} = 0\left(\frac{1}{n^{2/3}}\right)$$

and this choice is, again, optimum in the sense of statement 2.

4. Convergence of the Design Procedure

We now turn to the physical design situation that was described in section 2 and proceed to show sufficient conditions for

$$Y = W[d(m)-q_{\underline{x}}(m)]$$

to satisfy the assumptions of section 3. If this is done, then the implications of the statements of section 3 to our design procedure are obvious.

First, we consider conditions that will insure the satisfaction of assumptions (i) and (ii); then, we consider assumption (iii). To guarantee that assumptions (i) and (ii) are satisfied, we make the following restrictions on the physical situation:

(a) v(m) and d(m) are the outputs of stationary ergodic sources and are uniformly bounded in absolute magnitude for all m with probability one.

(b) $\sum_{m=0}^{\infty} |h_i(m)| < \infty$; i = 1, 2, ... k. The $f_i(m) = f_i[S_1(m), S_2(m), \ldots S_j(m)]$ are continuous in $S_1, S_2, \ldots, S_j$ for i = 1, 2, ... k.

(c) $P\left\{\left|\sum_{i=1}^{k} (x_i-\theta_i)\, f_i(m)\right| \geq D\|\underline{x}-\underline{\theta}\|\right\} \geq \epsilon$ for $\underline{x} \in X$ and D, $\epsilon > 0$. Or, in terms of one sample function, there exists an N_o with the property that for $N > N_o$, $\frac{n}{2N+1} \geq \epsilon > 0$, where n is the number of occurrences, $-N \leq m \leq N$, of

$$\left|\sum_{i=1}^{k} (x_i-\theta_i)\, f_i(m)\right| \geq D\|\underline{x}-\underline{\theta}\| \quad \underline{x} \in X$$

(d) W(e) has continuous first and second derivatives.

(e) W(e) is strictly convex; that is, there exists an E > 0 with the property that

$$W[\alpha a+(1-\alpha)b] \leq \alpha W(a) + (1-\alpha)\, W(b) - E\alpha|a-b| \quad \text{for } 0 \leq \alpha \leq 1/2$$

Restrictions (a)-(c) provide no serious limitation on our physical situation. Restriction (a) will surely be satisfied, since the output of any physical source is always uniformly bounded. Restriction (b) requires only that the memory elements $h_1, h_2, \ldots h_j$ be stable and that the f_i be continuous, as all physical transducers are. Restriction (c) only requires that all the h and all the f differ from one another in the prescribed sense. Restriction (c) is satisfied, for instance, if the probability of the sequence of k plus and minus signs $S_{gn}(f_i(m))$, $S_{gn}(f_2(m))$, ..., $S_{gn}(f_k(m))$ taking on a certain sequence of k plus and minus signs is nonzero for each of the 2^k possible outcomes. (This is a sufficient, but by no means necessary, condition for restriction (c) to be satisfied.) Restriction (d) is no practical limitation; restriction (e) is the only serious limitation on our method.

To show that assumptions (i) and (ii) follow from restrictions (a)-(e), we use the ergodicity of the sources to write

$$M(\underline{x}) = E\{W[d(m)-q(m)]\,|\,\underline{x}\}$$

$$= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{m=-N}^{N} W\left[d(m) - \sum_{i=1}^{k} x_i f_i(m)\right] \tag{3}$$

Assumption (i) now follows immediately from restriction (a) on the uniform boundedness of the sources and restriction (b) on the form of the filter.

To show that the upper inequality in assumption (ii) is satisfied, we use a Taylor's expansion about $\underline{x} = \underline{\theta}$ to write

$$\frac{\partial M(\underline{x})}{\partial x_j} = 0 + \sum_{i=1}^{k} (x_i - \theta_i) \left. \frac{\partial^2 M(\underline{x})}{\partial x_i \partial x_j} \right| \underline{x} = \underline{\theta} + x_i \tau_i \underline{e}_i \tag{4}$$

where $0 < \tau_i < 1$ for an arbitrary j, with $j = 1, 2, \ldots k$. Thus, the upper inequality in assumption (ii) follows with

$$k_1 = k \cdot \sup_{\substack{i,j=1,2,\ldots k \\ \underline{x} \in X}} \left| \frac{\partial^2 M(\underline{x})}{\partial x_i \partial x_j} \right|$$

if all the $\frac{\partial^2 M(\underline{x})}{\partial x_i \partial x_j}$ $i, j = 1, 2, \ldots k$ are bounded for all $\underline{x} \in X$. To show this, we consider

$$S_N(\underline{x}) = \frac{1}{2N+1} \sum_{m=-N}^{N} W\left[d(m) - \sum_{i=1}^{k} x_i f_i(m) \right]$$

and

$$S_N^{ij}(\underline{x}) = \frac{1}{2N+1} \sum_{m=-N}^{N} W''\left[d(m) - \sum_{i=1}^{k} x_i f_i(m) \right] f_i(m)\, f_j(m)$$

The continuity of W" and the uniform boundedness of the $f_i(m)$, $i = 1, 2, \ldots k$, and d(m) imply the equicontinuity of $S_N^{ij}(\underline{x})$ for $\underline{x} \in X$. The Arzela-Ascoli theorem then guarantees the existence of a uniformly convergent subsequence N_t, and for this subsequence

$$\begin{aligned} \frac{\partial^2}{\partial x_i \partial x_j} \lim_{t\to\infty} S_{N_t}(\underline{x}) &= \lim_{t\to\infty} \frac{\partial^2}{\partial x_i \partial x_j} S_{N_t}(\underline{x}) \\ &= \lim_{t\to\infty} S_{N_t}^{ij}(\underline{x}) \end{aligned} \tag{5}$$

for all $\underline{x} \in X$. However, by our assumption of ergodic sources, $\lim_{N\to\infty} S_N^{ij}(\underline{x})$ is unique with probability one, and hence

$$\frac{\partial^2}{\partial x_i \partial x_j} M(\underline{x}) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{m=-N}^{N} W''\left[d(m) - \sum_{i=1}^{k} x_i f_i(m) \right] f_i(m)\, f_j(m) \tag{6}$$

with probability one for all $\underline{x} \in X$. But the right-hand side of Eq. 6 is bounded by restrictions (a) and (d), and we have established the upper inequality in assumption (ii).

We now turn to the lower inequality in assumption (ii). Using Eq. 3, restrictions (c), (e), and the convexity of X, we have

$$M[(1-a)\underline{x}+a\underline{\theta}] \leq (1-a)\, M(\underline{x}) + aM(\underline{\theta}) - \epsilon DEa \left\| \underline{x}-\underline{\theta} \right\| \quad \text{for } 0 \leq a \leq 1/2 \tag{7}$$

or

$$-\frac{M[\underline{x}+a(\underline{\theta}-\underline{x})] - M(\underline{x})}{a} \geqslant M(\underline{x}) - M(\underline{\theta}) + \epsilon DE\left\| \underline{x}-\underline{\theta} \right\|$$

$$\geqslant \epsilon DE\left\| \underline{x}-\underline{\theta} \right\| \quad \text{for } 0 \leq a \leq 1/2 \tag{8}$$

The right-hand side of Eq. 8 is independent of a; hence, taking the limit of the left-hand side as a approaches zero, we have

$$\left[(\text{grad } M)(\underline{x}), \frac{-(\underline{x}-\underline{\theta})}{\left\| \underline{x}-\underline{\theta} \right\|}\right] \geqslant \epsilon DE\left\| \underline{x}-\underline{\theta} \right\| \tag{9}$$

as desired.

If we desire to use the results of statement 3, we need only require in addition that W'''(e) be continuous for all values of the argument which occur under restrictions (a) and (b) and the assumption that $\underline{x} \in X$. The boundedness of

$$\left| \frac{\partial^3 M(\underline{x})}{\partial x_i^3} \right| \qquad i = 1, 2, \ldots k$$

under this added restriction is shown in the same manner as was the boundedness of

$$\left| \frac{\partial^2 M(\underline{x})}{\partial x_i \partial x_j} \right|$$

Assumption (iii) is a restriction that the process Y_n depends only remotely on the distant past and is not in itself unreasonable. It is this assumption that requires that some time interval S be left between iterations. In a future report sufficient conditions on the process v(t), d(t) which guarantee the satisfaction of assumption (iii) will be developed, and a comment will be added to extend the analysis to the continuous time parameter case.

Grateful acknowledgment is made for many helpful discussions with Dr. A. V. Balakrishnan and for the helpful cooperation of Space Technology Laboratories, Inc., Los Angeles, California.

D. J. Sakrison

(References on following page)

References

1. N. Wiener, Nonlinear Problems in Random Theory (Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1958).

2. M. Brilliant, Theory of the Analysis of Nonlinear Systems, Technical Report 345, Research Laboratory of Electronics, M.I.T., March 3, 1958.

3. V. Dupač, On the Kiefer-Wolfowitz Approximation Method, Report 22G-008, Lincoln Laboratory, M.I.T., 24 Feb. 1960. (Translation of a paper in the Czech mathematical journal Casopis pro Pestovani Matematiky, Vol. 82, pp. 47-75, 1957.)

E. OPTIMUM SYNTHESIS OF A GAUSSIAN PROCESS FROM A NON-GAUSSIAN PROCESS

The transformation of non-Gaussian random processes into Gaussian processes was discussed in Quarterly Progress Report No. 58 (pages 165-170). In that report the question arose as to the extent of the class of random processes that could be so transformed with zero error. In this report we shall show that a sufficient condition for a random process to be a member of this class is that it be a stationary ergodic process and that it have a finite mean-square value, and that samples taken τ seconds apart become statistically independent as the separation, τ, becomes infinite. We shall refer to this class of functions as class G.

Wiener has shown (1) that any function, F, of our class G can be expanded uniquely in terms of his fundamental orthogonal functionals as

$$F(a) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{\nu=0}^{N} A_\nu G_\nu[K_\nu, a] \tag{1}$$

in which a distinguishes one member of the Gaussian ergodic ensemble from another. In terms of Wiener's nonlinear network (2), which is schematically represented in Fig. XI-6, we can write Eq. 1 as

$$F(a) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{\nu=0}^{N} A_\nu y_\nu(a) \tag{2}$$

in which $y_\nu(a)$ is the output of the nonlinear network represented by $G_\nu[K_\nu]$ as a result of the input $g(a)$ which is the a^{th} member of the Gaussian ensemble.

We shall first show that if $F(a)$ is only a single term of Eq. 1,

$$F(a) = y_m(a) \tag{3}$$

then there is a nonlinear network N, of the type described by Wiener (2), whose output is Gaussian when the input is F. To show this, we first note that Eq. 3 represents a

many-to-one mapping because many members of the α ensemble will produce the same output. Thus an inverse mapping from the F ensemble back to the whole α ensemble does not exist. However, we shall show that a mapping to a Gaussian ensemble which is a subset of the α ensemble does exist. To do this, denote the various outputs by F_β. Then we may order the α ensemble in a set of subsets such that the S_β subset of α is

$$S_\beta = \{\alpha: G_n[K_n, \alpha] = y_n(\alpha) \text{ for } n \neq m,\ G_m[K_m, \alpha] = F_\beta\} \tag{4}$$

That is, we gather together all of the members of the α ensemble that produce the same output F_β and call it the set S_β. We now form a new set, T_γ, by picking for each value of β, one member from S_β. This can be done in many ways. Thus, to each value of γ there corresponds a set, T_γ, each of whose members results in a different output, F_β. Also, different values of γ correspond to sets that differ from each other in at least one member. Now, each output, F_β, occurs with probability zero. Thus each set, S_β, has a measure zero relative to α; thus we can show that for each $\epsilon > 0$, there is an ordering

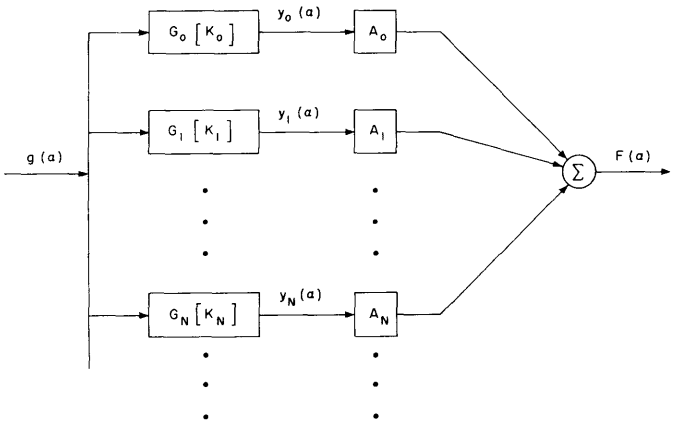


Fig. XI-6. Nonlinear network.

of the α ensemble so that for any set, T_γ, at least one member of the set is contained in every ϵ-neighborhood of the ordered set of α's. This effectively means that each set T_γ is dense in α. Now, since the α ensemble is ergodic, almost all dense subsets of α are also Gaussian ergodic ensembles. Thus, at least one set for some value of γ is a Gaussian ergodic ensemble. Denote each member of this ensemble δ. Then, from Eq. 3, we have

$$F(\delta) = y_m(\delta) \tag{5}$$

This equation now represents a one-to-one mapping and thus a right inverse exists. That is, there exists a network N for which

$$N[F(\delta)] = g(\delta) \tag{6}$$

To prove our statement for the general case, in which

$$F(\alpha) = \sum_{i=1}^{N} A_{m_i} y_{m_i}(\alpha) \tag{7}$$

we choose the set S_β to be

$$S_\beta = \left\{ \alpha : G_n[K_n, \alpha] = y_n(\alpha) \text{ for } n \neq m_i, \quad \sum_{i=1}^{N} A_{m_i} G_{m_i}\left[K_{m_i}, \alpha\right] = F_\beta \right\} \tag{8}$$

The proof is the same as before.

We have thus shown that any function of our class G may be transformed into a Gaussian process with zero error. We now note that the required network, N, is stable because its output has a finite mean-square value for an input that also has a finite mean-square value. Also, $F(\delta, t)$ and $F(\delta, t+\tau)$ become statistically independent as $\tau \to \infty$. The same is true for $g(\delta, t)$ and $g(\delta, t+\tau)$. Thus the output of N is asymptotically independent of the remote past input. Wiener has shown that all such networks can be expanded uniquely in terms of his fundamental orthogonal functionals. This is the form of the network shown in Fig. XI-6 and the form of the network that we used in the transformation of non-Gaussian processes into Gaussian processes (3). Thus, for any function of the class G, the procedure discussed previously searches out a set, δ, which we now know exists. Consequently, the transformation can be done with zero error.

It is interesting to note that although a Wiener network does not always possess an inverse in the sense that the output converges in the mean to a desired output, it always possesses an inverse in the sense that the output converges in probability to a desired output.

M. Schetzen

References

1. N. Wiener, Nonlinear Problems in Random Theory (Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1958), p. 41.

2. Ibid., Lectures 10 and 11.

3. M. Schetzen, Optimum synthesis of a Gaussian process from a non-Gaussian process, Quarterly Progress Report No. 58, Research Laboratory of Electronics, M.I.T., July 15, 1960, pp. 165-170.

F. AN OPTIMUM METHOD FOR SIGNAL TRANSMISSION

1. Transmission with Irreducible Minimum Mean-Square Error

A problem in the statistical theory of communication is the design of optimum systems for the transmission and reception of messages through noisy channels. In this report, we shall describe a design of optimum systems for a class of messages for which the mean-square error is an irreducible minimum when the noise in the channel is Gaussian. The class of messages, which was discussed in Section XI-E, is the class that can be transformed into Gaussian random waves. We shall refer to this class as class G.

The general problem may be described by means of Fig. XI-7. The message, $f_m(t)$, is to be transmitted through a channel in which noise, $f_n(t)$, is added. The network N_2 is the optimum filter for the received signal $f_r(t)$, so that the output $f_o(t)$ approximates the message with minimum mean-square error, $\overline{e^2(t)} = \overline{[f_o(t)-f_m(t)]^2}$. In order that this error be an irreducible minimum, the message is optimally distorted by network N_1 to yield the signal, $f_s(t)$, for transmission through the channel. The problem is to determine the optimum combination of networks N_1 and N_2 for which the mean-square error, $\overline{e^2(t)}$, is an irreducible minimum.

We first note that, for any given network N_1, we can determine the optimum nonlinear filter, N_2, of the class described by Wiener (1) and shown in Fig. XI-8. It is determined by sequentially adjusting the amplifier gains, A_i, until the measured mean-square error is a minimum. We are assured that this procedure will always converge

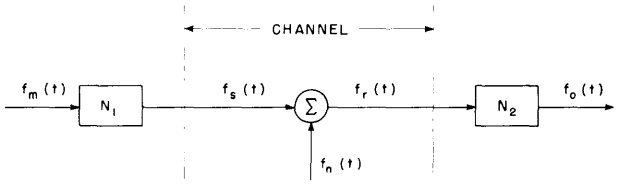


Fig. XI-7. An optimum transmitting-receiving system.

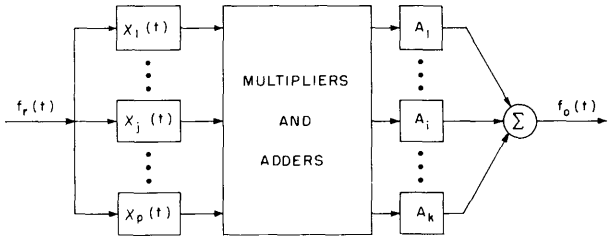


Fig. XI-8. Nonlinear network.

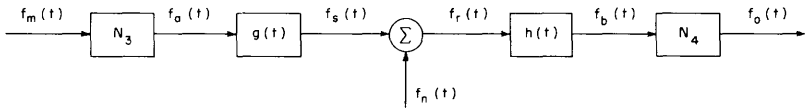


Fig. XI-9. Pertaining to the determination of the optimum system of Fig. XI-7.

since it has been shown that, for each order of nonlinearity, the set of possible outputs for various amplifier gains is a convex set (2). Thus the mean-square error is a minimum for a unique setting of the amplifier gains. However, in order to obtain the irreducible minimum mean-square error, the optimum distortion network, N_1, must also be determined. We shall now describe a procedure for determining the optimum network, N_1, if the noise in the channel is a Gaussian random variable and the message is a member of class G.

Consider the system depicted in Fig. XI-9. If the input, $f_m(t)$, is a member of class G, we can determine a nonlinear network, N_3, of the form depicted in Fig. XI-8, whose output, $f_a(t)$, is a Gaussian random variable (2). Furthermore, a right inverse exists. Thus, except for a possible delay, there exists a network N_4, also of the form depicted in Fig. XI-8, with the property that the network N_3 followed by N_4 is equivalent to a direct connection. We shall find that we do not need to determine the network N_4. We merely need to be assured of its existence to guarantee that our procedure is optimum. Let us assume for the moment that networks N_3 and N_4 have been determined and are placed as shown in Fig. XI-9. The problem is then to determine the optimum distorting network for $f_a(t)$ and receiving network for $f_b(t)$ for which the mean-square error, $\overline{e_1^2(t)} = \overline{[f_b(t)-f_a(t)]^2}$, is an irreducible minimum. However, since $f_a(t)$ and $f_b(t)$ are Gaussian random variables, it can be shown that the optimum system is linear and no nonlinear network will yield a lower mean-square error.

We thus have reduced the problem to the determination of two linear networks, h(t) and g(t). The optimum network, h(t), is given by the solution of the Wiener-Hopf equation. For such a network, the irremovable error (3) is given by

$$\overline{e_{irr}^2(t)} = \int_{-\infty}^{\infty} \frac{\Phi_{rr}(\omega)\,\Phi_{aa}(\omega) - |\Phi_{ra}(\omega)|^2}{\Phi_{rr}(\omega)}\,d\omega \tag{1}$$

in which $\Phi_{ij}(\omega)$ is the Fourier transform of the correlation function, $\phi_{ij}(\tau)$.

$$\phi_{ij}(\tau) = \overline{f_i(t)\,f_j(t+\tau)} \tag{2}$$

However, for our system,

$$f_r(t) = f_n(t) + \int_0^\infty g(x)\, f_a(t-x)\, dx \tag{3}$$

For simplicity, we shall assume $f_a(t)$ and $f_n(t)$ to be uncorrelated. Then

$$\begin{aligned} \Phi_{rr}(\omega) &= \Phi_{nn}(\omega) + |G(\omega)|^2\, \Phi_{aa}(\omega) \\ |\Phi_{ra}(\omega)|^2 &= |G(\omega)|^2\, \Phi_{aa}(\omega) \end{aligned} \tag{4}$$

and Eq. 1 becomes

$$\overline{e_{irr}^2(t)} = \int_{-\infty}^\infty \frac{\Phi_{nn}(\omega)\, \Phi_{aa}(\omega)}{\Phi_{nn}(\omega) + |G(\omega)|^2\, \Phi_{aa}(\omega)}\, d\omega \tag{5}$$

We now impose the constraint that the mean-square value of the transmitted wave, $f_s(t)$, be P. Then

$$\int_{-\infty}^\infty |G(\omega)|^2\, \Phi_{aa}(\omega)\, d\omega = P \tag{6}$$

The transfer function $G(\omega)$ that minimizes Eq. 5, subject to this constraint, is

$$|G_{opt}(\omega)|^2 = \begin{cases} \frac{1}{\beta}\left[\frac{\Phi_{nn}(\omega)}{\Phi_{aa}(\omega)}\right]^{1/2} - \frac{\Phi_{nn}(\omega)}{\Phi_{aa}(\omega)} & \frac{\Phi_{aa}(\omega)}{\Phi_{nn}(\omega)} \geqslant \beta^2 \\ 0 & \frac{\Phi_{aa}(\omega)}{\Phi_{nn}(\omega)} < \beta^2 \end{cases} \tag{7}$$

and

$$P = \frac{1}{\beta} \int_{\omega \in W} [\Phi_{nn}(\omega)\Phi_{aa}(\omega)]^{1/2}\, d\omega - \int_{\omega \in W} \Phi_{nn}(\omega)\, d\omega \tag{8}$$

in which the integration is carried out over the set W which is the set of frequencies for which $\frac{\Phi_{aa}(\omega)}{\Phi_{nn}(\omega)} \geqslant \beta^2$. The minimum irremovable error is then

$$E^2 = \min\left\{\overline{e_{irr}^2(t)}\right\} = \beta \int_{\omega \in W} [\Phi_{nn}(\omega)\Phi_{aa}(\omega)]^{1/2}\, d\omega + \int_{\omega \notin W} \Phi_{aa}(\omega)\, d\omega \tag{9}$$

in which the second integral is carried out over those frequencies that are not members of W. We note that the optimum solution may require $G_{opt}(\omega)$ to be zero over a range of frequencies. This is not physically realizable. For such cases, the optimum transfer

function is that which is equal to $|G_{opt}(\omega)|$ at those frequencies for which $|G_{opt}(\omega)| \neq 0$ and is as small as possible at those frequencies for which $|G_{opt}(\omega)| = 0$.

For example, let $\Phi_{aa}(\omega) = 1/(1+\omega^2)$, $\Phi_{nn}(\omega) = 1$ and choose $\beta = 10^{-3}$. Then the set of frequencies transmitted, W, is the set for which

$$\beta^2 = 10^{-6} \leq \frac{\Phi_{aa}(\omega)}{\Phi_{nn}(\omega)} = \frac{1}{1+\omega^2}$$

or

$$\omega < 10^3 \tag{10}$$

Then, from Eq. 8, the mean-square value of the transmitted wave is

$$P = 10^3 \int_{-10^3}^{10^3} \frac{1}{(1+\omega^2)^{1/2}}\, d\omega - \int_{-10^3}^{10^3} d\omega$$

$$= 2 \times 10^3 [\sinh^{-1} 10^3 - 1]$$

$$= 13.2 \times 10^3 \tag{11}$$

From Eq. 7, $|G_{opt}(\omega)|^2$ is given by

$$|G_{opt}(\omega)|^2 = \begin{cases} 10^3[1+\omega^2]^{1/2} - [1+\omega^2] & \omega < 10^3 \\ 0 & \omega \geq 10^3 \end{cases} \tag{12}$$

Figure XI-10 is a graph of $|G_{opt}(\omega)|$. The minimum irremovable error, E^2, is obtained from Eq. 9:

$$E^2 = 2 \times 10^{-3} \int_0^{10^3} \frac{1}{(1+\omega^2)^{1/2}}\, d\omega + 2 \int_{10^3}^{\infty} \frac{1}{1+\omega^2}\, d\omega$$

$$= 2 \times 10^{-3} [1 + \sinh^{-1} 10^3]$$

$$= 17.2 \times 10^{-3} \tag{13}$$

If we had used an amplifier of gain, $|G(\omega)| = A$, then for the same mean-square value of the transmitted wave, we would make $A^2 = \frac{P}{\pi}$. From Eq. 5, the irremovable error, $\overline{e_{irr}^2(t)}$, for this case is

$$\overline{e_{irr}^2(t)} = \int_{-\infty}^{\infty} \frac{\frac{1}{1+\omega^2}}{1+\frac{P}{\pi}\frac{1}{1+\omega^2}}\, d\omega$$

$$= \frac{\pi}{\left(1+\frac{P}{\pi}\right)^{1/2}}$$

$$= 48.4 \times 10^{-3} \qquad (14)$$

if we use the value of P from Eq. 11. Thus, we have reduced the irremovable mean-square error by a factor of 2.71 by the use of $|G_{opt}(\omega)|$ instead of an amplifier with constant gain. It should be noted that, for our solution, the phase of $G_{opt}(\omega)$ is arbitrary. Even so, in practice, $G_{opt}(\omega)$ should be chosen as a minimum-phase network in order to reduce the time delay in filtering the received signal, $f_r(t)$.

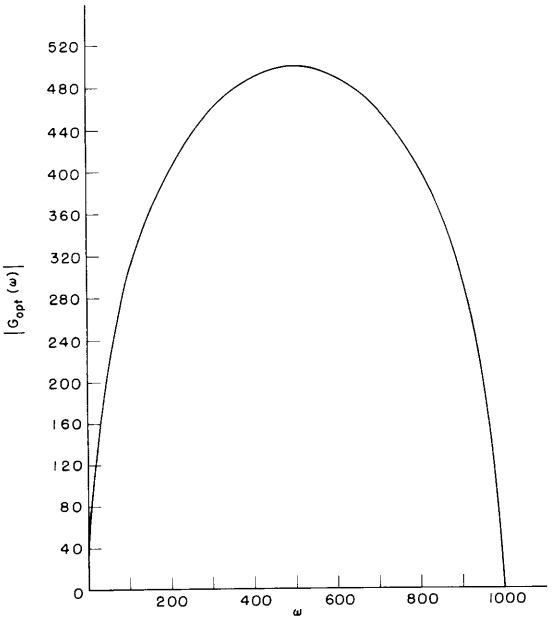


Fig. XI-10. Graph of $|G_{opt}(\omega)|$ versus ω for the example given.

The optimum transmitting-receiving system is thus designed by first determining network N_3 by the method previously described (2); $|G_{opt}(\omega)|$ is given by Eq. 7. The tandem connection of N_3 and g(t) is then the optimum distorting network, N_1, of Fig. XI-7. We now observe that it is not necessary to determine h(t) and N_4 in order to obtain the optimum receiving system. Rather, we may determine the optimum network, N_2, of Fig. XI-7 directly by the method discussed in this report. However, since $f_r(t)$ is a

Gaussian random variable, the amplifier gains A_i will not interact and only one adjustment of each amplifier gain will be necessary.

2. Transmission with Irreducible Minimum Probability of Error

We now shall show that by slightly modifying the optimum system described in section 1, we can transmit signals through Gaussian channels with an irreducible minimum probability of error. We shall limit our discussion to binary signals.

The method is based upon the fact that the output of a filter, which is within the minimum mean-square error of some desired output, is the conditional mean. Thus, from Fig. XI-7, if we let x be the amplitude of the message, then the output of the optimum filter, N_2, at time t_1 is

$$f_o(t_1) = \int_{-\infty}^{\infty} xP[x/f_r(t);t\leq t_1]\,dx \tag{15}$$

in which $P[x/f_r(t);t\leq t_1]$ is the probability that the message's amplitude was x, given the complete past of the received signal, $f_r(t)$. If the message is a binary signal that assumes the values +1 and -1, then

$$f_o(t_1) = P[1/f_r(t);t\leq t_1] - P[-1/f_r(t);t\leq t_1] \tag{16}$$

We note that $f_o(t_1) > 0$ only if $P[1/f_r(t);t\leq t_1] > P[-1/f_r(t);t\leq t_1]$ and vice versa. Thus we can form a signal which, at each instant of time, is the message with a minimum probability of error by having the optimum network N_2 of Fig. XI-7 followed by a saturating amplifier whose output, $f_c(t)$, is +1 if $f_o(t) > 0$, and -1 if $f_o(t) < 0$. We now note that since

$$1 = P[1/f_r(t);t\leq t_1] + P[-1/f_r(t);t\leq t_1] \tag{17}$$

we may write Eq. 16 as

$$f_o(t_1) = 2P[1/f_r(t);t\leq t_1] - 1 \tag{18}$$

Now, the mean-square error has been made an irreducible minimum by optimally distorting the message with network N_1. In terms of Eq. 18, this means that if the amplitude of the message is +1, then $P[1/f_r(t);t\leq t_1]$ is as close to +1 as is theoretically possible. Also, if the amplitude of the message is -1, then $P[1/f_r(t);t\leq t_1]$ is as close to zero as is theoretically possible. This implies that, at each instant of time, $f_c(t)$ is the message with an irreducible minimum probability of error. An example of a message for which this type of transmission would be desirable is infinitely clipped speech. If, however, the message is a sequence of binary digits that are positive or negative

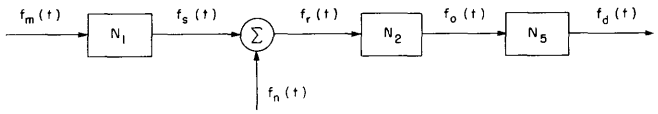


Fig. XI-11. An optimum system for binary digits.

pulses of duration T, the system that we have just described would not be the desired optimum. For such a signal, we desire the output also to be pulses of duration T that are +1 or -1 with an irreducible minimum probability of error. To accomplish this, we note that

$$\int_{t_1-T}^{t_1} P[x/f_r(t);t\leq t_1]\,dt_1$$

is the probability that the message pulse, which is known to exist in the interval (t_1-T, t_1), had the amplitude x, given the complete past of the received signal. From Eq. 16, we have

$$I_1 = \int_{t_1-T}^{t_1} f_o(t_1)\,dt_1 = \int_{t_1-T}^{t_1} P[1/f_r(t);t\leq t_1]\,dt_1 - \int_{t_1-T}^{t_1} P[-1/f_r(t);t\leq t_1]\,dt_1 \qquad (19)$$

We thus note that $I_1 > 0$ only if the probability that the binary digit was positive is greater than the probability that it was negative, given the complete past of the received signal. Thus we can transmit binary digits with an irreducible minimum probability of error with the system of Fig. XI-11. The networks N_1 and N_2 are the optimum networks described in section 1 of this report. The network N_5 has an output, $f_d(t)$, which is a sequence of binary digits. Each digit has an amplitude that is +1 or -1; the sign depends on whether

$$I_n = \int_{t_n-T}^{t_n} f_o(t)\,dt$$

is greater than or less than zero. The times $[t_n-T]$ are the known starting times of the binary digits.

M. Schetzen

References

1. N. Wiener, Nonlinear Problems in Random Theory (Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1958).

2. M. Schetzen, Optimum synthesis of a Gaussian process from a non-Gaussian process, Quarterly Progress Report No. 58, Research Laboratory of Electronics, M.I.T., July 15, 1960, pp. 165-170.

3. Y. W. Lee, The Statistical Theory of Communication (John Wiley and Sons, Inc., New York, 1960).

G. OPTIMUM COMPENSATION FOR NONLINEAR CONTROL SYSTEMS

1. Nonlinear Compensation Problem

In feedback control systems the output is provided by a physical device capable of meeting the power requirements of the control system. To optimize the performance of the system, the input to these fixed elements must be altered in some manner. The presence of these fixed elements in the system provides the fundamental distinction between compensation problems and filter problems. For linear systems with random inputs, the relation between the two problems is well known.

For nonlinear systems, the relation between the two problems is still not clear. Physical systems are most easily described by differential equations. Classical nonlinear theory has been essentially a study of the solutions and properties of nonlinear differential equations. However, a general statistical optimization technique that makes use only of differential equations appears difficult to develop. The modern functional expansion techniques for treating statistical problems outlined by Wiener (1) provide a basis for solving the nonlinear optimization problem.

In this report a method of solving the nonlinear compensation problem by effectively combining the use of nonlinear differential equations and functional expansion techniques is outlined.

The general configuration of a feedback control system is shown in Fig. XI-12. Here, r(t) is a sample function from a stationary random process. The desired output of the over-all system is a function $y_d(t)$ which is related to either r(t) or the signal part of r(t) in some known manner. If there were no fixed elements, we would have the nonlinear filter problem illustrated in Fig. XI-13.

One method of solving this filter problem would be to specify the input-output relation of the filter to be of the form

$$y_b(t) = \int_0^t K_1(t-\tau_1)\, r(\tau_1)\, d\tau_1 + \int_0^t \int_0^t K_2(t-\tau_1, t-\tau_2)\, r(\tau_1)\, r(\tau_2)\, d\tau_1 d\tau_2 + \cdots \qquad (1)$$

and then find a series of kernels $K_1, K_2, K_3, \ldots$ so that the expected value of $[y(t)-y_b(t)]^2$ is minimized.

Now, if we could find a realizable system C_a or C_b with the property that $y_a(t) = y_b(t)$, then the original compensation problem would be solved.

The class of compensation problems to be considered may now be more clearly defined. It is assumed that the fixed elements may be described by a nonlinear differential equation of the form

$$P_1(x, x', x'', \ldots x^{(r)}) = P_2(y, y', y'', \ldots y^{(s)}) \qquad (2)$$

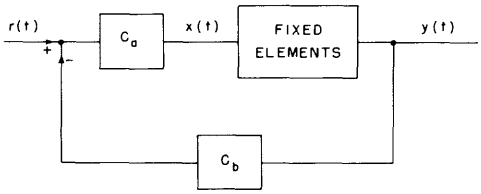


Fig. XI-12. General control system configuration.

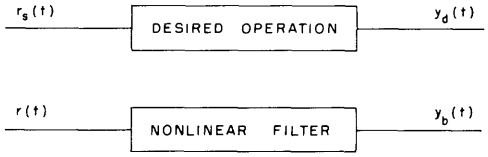


Fig. XI-13. Optimum filter.

where x is the input to the fixed elements, y is the output of the fixed elements, and P represents a function which is a polynomial. Typical examples of the occurrence of equations of this type are

$$ay'' + by' + cy + dy^3 = x \tag{3}$$

which represents motion with nonlinear restoring force;

$$ay'' + b(y')^2 + cy = x \tag{4}$$

which represents motion with nonlinear damping; and

$$ay'' + by' + cy = P(x) \tag{5}$$

which represents the general category of systems that are a cascade of a nonlinear no-memory polynomial device and a linear memory device.

More general polynomials such as

$$ay''y + b(y')^3 + cy^2 = dx + ex''x' + fx^3 \tag{6}$$

are included in this category.

Before outlining the method of solution of the stated compensation problem, related work should be reviewed. Recent work in the theory of optimum nonlinear systems uses the functional power-series representation conceived by Wiener (1). This is a generalization of the convolution integral. In this method, the output y is expressed as:

$$y(t) = \int_{-\infty}^{\infty} K_1(t-\tau)\, x(\tau)\, d\tau + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_2(t-\tau_1, t-\tau_2)\, x(\tau_1)\, x(\tau_2)\, d\tau_1 d\tau_2$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_3(t-\tau_1, t-\tau_2, t-\tau_3)\, x(\tau_1)\, x(\tau_2)\, x(\tau_3)\, d\tau_1 d\tau_2 d\tau_3 + \cdots \tag{7}$$

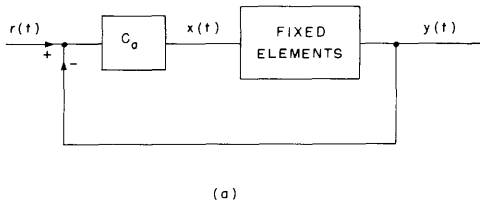
Brilliant (2) developed a theory of analysis for nonlinear systems of this form. He developed methods for cascading, adding, and multiplying systems of this type. George (3) also used the functional representation. Working primarily with nonlinear systems that are composed of nonlinear no-memory devices and linear memory devices,

he developed an algebra for combining systems. He extended Brilliant's use of multidimensional transforms and developed many properties that we shall find useful. Although some results on Gaussian white-noise inputs were included, George worked principally with deterministic inputs. In control systems, the division of a nonlinear physical device into a linear memory part and a nonlinear no-memory part is often unwieldy and almost always artificial. By working directly with the differential equation of the physical system, there is a greater possibility of attaching the correct physical significance to one's final answer. Zames (4, 5) has dealt with the problem of inversion of functionals and feedback around functionals. He has developed an algorithm for determining the resultant kernels. The part of his work pertaining to control systems was deterministic.

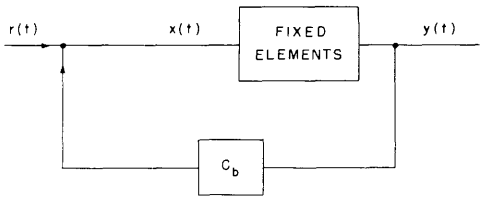
As in linear systems, it seems that application of Wiener's concepts to filter problems has proceeded more rapidly than the application to control system problems. By combining the advantages of two different analytical techniques, it is hoped that a reasonably practical, general method of solution to a large class of nonlinear control problems may be found.

2. Method of Approach and Preliminary Results

The general configuration of a feedback control system is shown in Fig. XI-12. Let us consider two important special cases of this general configuration. For $C_b = 1$, we have the "series" compensation problem (Fig. XI-14a). For $C_a = 1$, we have the



(a)



(b)

Fig. XI-14. (a) Series compensation. (b) Feedback compensation.

"feedback" compensation problem (Fig. XI-14b). In many cases the proper choice of configuration will be obvious. Presumably, there are cases in which a convergent

solution could be obtained by using both C_a and C_b when neither one alone would be satisfactory. These problems are subjects for future research.

Recall that the basic problem was, first, to find a set of kernels $K_1, K_2, K_3, \ldots K_n$ that specify the optimum nonlinear filter. For nonlinear operations on Gaussian inputs, an analytic solution for these kernels may be obtained by using Chesler's (6) techniques. For an arbitrary input, it appears that numerical methods will be required.

The basic equations describing the system in Fig. XI-14 are:

$$P_1(x, x', x'', \ldots x^{(r)}) = P_2(y, y', y'', \ldots y^{(s)}) \tag{8}$$

$$x(t) = \int_0^t C_1(t-\tau)\, e(\tau)\, d\tau + \int_0^t \int_0^t C_2(t-\tau_1, t-\tau_2)\, e(\tau_1)\, e(\tau_2)\, d\tau_1 d\tau_2 + \ldots \tag{9}$$

$$e(t) = r(t) - y(t) \tag{10}$$

$$y(t) = \int_0^t K_1(t-\tau)\, r(\tau)\, d\tau + \int_0^t \int_0^t K_2(t-\tau_1, t-\tau_2)\, r(\tau_1)\, r(\tau_2)\, d\tau_1 d\tau_2 + \ldots \tag{11}$$

For definiteness, consider fixed elements described by

$$y'' + ay' + by = x + dx^3 \tag{12}$$

or

$$L[y] = x + dx^3 \tag{13}$$

A straightforward approach would be to substitute Eqs. 9, 10, and 11 in Eq. 8. This gives an equation of the form

$$\begin{aligned}
&L\left[\int_0^t K_1(t-\tau)\, r(\tau)\, d\tau + \int_0^t \int_0^t K_2(t-\tau_1, t-\tau_2)\, r(\tau_1)\, r(\tau_2)\, d\tau_1 d\tau_2 + \ldots\right] \\
&= \left\{\int_0^t C_1(t-\tau)\left[r(\tau) - \int_0^\tau K_1(\tau-\epsilon)\, r(\epsilon)\, d\epsilon - \int_0^\tau \int_0^\tau K_2(\tau-\epsilon_1, \tau-\epsilon_2)\, r(\epsilon_1)\, r(\epsilon_2)\, d\epsilon_1 d\epsilon_2 + \ldots\right]\right. \\
&\quad + \int_0^t \int_0^t d\tau_1 d\tau_2 C_2(t-\tau_1, t-\tau_2)\left[r(\tau_1) - \int_0^{\tau_1} K_1(\tau_1-\epsilon)\, r(\epsilon)\, d\epsilon - \ldots\right]\left[r(\tau_2) - \int_0^{\tau_2} K_1(\tau_2-\epsilon)\, r(\epsilon)\, d\epsilon + \ldots\right] \\
&\quad + \iiint_0^t d\tau_1 d\tau_2 d\tau_3 C_3(t-\tau_1, t-\tau_2, t-\tau_3)\left[r(\tau_1) - \int_0^{\tau_1} K_1(\tau_1-\epsilon)\, r(\epsilon)\, d\epsilon + \ldots\right] \\
&\quad \times \left[r(\tau_2) - \int_0^{\tau_2} K_1(\tau_2-\epsilon)\, r(\epsilon)\, d\epsilon + \ldots\right]\left[r(\tau_3) - \int_0^{\tau_3} K_1(\tau_3-\epsilon)\, r(\epsilon)\, d\epsilon + \ldots\right] + \ldots\Big\} \\
&\quad + d\left\{\left[\int_0^t C_1(t-\tau)\left[r(\tau) - \int_0^\tau K_1(\tau-\epsilon)\, d\epsilon + \ldots\right] d\tau\right]^3 + 3\left[\int_0^t d\tau C_1(t-\tau)\left[r(\tau) - \int_0^\tau K_1(\tau-\epsilon)\, r(\epsilon)\, d\epsilon + \ldots\right]\right]^2\right. \\
&\quad \times \left[\int_0^t d\tau_1 d\tau_2 C_2(t-\tau_1, t-\tau_2)\left[r(\tau_1) - \int_0^{\tau_1} K_1(\tau_1-\epsilon)\, r(\epsilon)\, d\epsilon - \ldots\right]\right. \\
&\quad \left.\left.\times \left[r(\tau_2) - \int_0^{\tau_2} K_1(\tau_2-\epsilon)\, r(\epsilon)\, d\epsilon + \ldots\right]\right] + \ldots\right\}
\end{aligned} \tag{14}$$

By equating terms of the same order in r(t), a set of equations can be obtained that could be solved successively, not simultaneously, for the various C_m. The easiest approach to solving these equations is through multidimensional transform theory (3). Even with the use of transforms, a straightforward approach turns out to be too tedious to be valuable as a general method. An algorithm has been developed that gives the desired result in a much simpler fashion.

(a) Algorithm for Determining Series Compensator Kernels

The input-output relation of the fixed elements is described by a nonlinear differential equation. First, the terms are classified according to the order of their nonlinearity. Thus x^3, $x'(x'')^2$, $xx'x''$ are all third-order input terms. There is a basic expansion for all n^{th}-order terms. The algorithm developed enables us to write an equation for C_m in the transform domain by considering the expansion of the n^{th}-order terms directly. Consider the equation

$$ay^2y' + by^3 + cy^2 + d(y')^2 + L(y) = L(x) + ex^3 + fx'(x'')^2 + g\,x\,x'\,x'' \tag{15}$$

In order to solve for the m^{th}-order compensator kernel C_m, we can write an equation of the form

$$Q_m^3 + Q_m^2 + Q_m^1 = P_m^1 + P_m^3 \tag{16}$$

where Q_m^3 represents the contribution of all third-order output terms, P_m^1 represents the contribution of all linear input terms, and the other terms have similar definitions. The transform of the unknown kernel appears in P_m^1. All other quantities are known, so that the explicit answer follows immediately.

We shall see that the basic expansion for all n^{th}-order terms is a combination of different partitions. The difference between terms within a given order is taken into account by use of a characteristic coefficient that modifies the basic expansion.

Thus the problem is to find an efficient method for constructing the terms, P_m^n and Q_m^n.

(i) Construction of P_m^n; Contribution of an n^{th}-order nonlinearity of the input to an m^{th}-order series compensator kernel

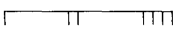
P_m^n consists of a sum of m-n+1 terms. Thus, we may write

$$P_m^n = \sum_{i=n}^{m} P_m^n(i) \tag{17}$$

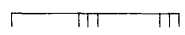
and determine each $P_m^n(i)$. $P_m^n(i)$ is constructed in three steps. First, form all partitions of m objects into n cells. For example, the partitions for m = 7 and n = 3 are:

(I) 

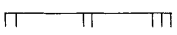
p_{115}

(II) 

p_{124}

(III) 

p_{133}

(IV) 

p_{223}

These are the basic or major partitions. Physically, each object represents a different frequency-domain variable. Second, consider all combinations of n compensator kernels C_j, where $\Sigma j = i$. For $i = 5$ and $n = 3$, the possible combinations are C_1, C_2, C_2 and C_1, C_1, C_3. Compare each set of kernels with each major partition in turn. Looking at p_{115} and C_1, C_2, C_2 first, we see that they could not occur together. C_2 represents a kernel that is a function of at least two variables, $C_2(S_1, S_2)$. But two of the partitions in p_{115} have only one variable in them. Thus, we cannot associate C_1, C_2, C_2 and p_{115}. Examples of correct associations are:

(Ia)

C_1	C_1	C_3

(IIa)

C_1	C_2	C_2

Third, look at the ways in which the optimum filter kernels K_β can combine with the compensator kernels C_j. Here, the necessary restriction is that $\Sigma \beta = m$ and that the total number of kernels $K_\beta = i$. For partition Ia, we see that suitable combinations are:

C_1	C_1	C_3		
K_1	K_1	K_1	K_1	K_3

or

C_1	C_1	C_3		
K_1	K_1	K_1	K_2	K_2

These represent the transformed terms:

$$[C_1(S_1)(1-K_1(S_1))][C_1(S_2)(1-K_1(S_2))][C_3(S_3,S_4,S_5+S_6+S_7)(1-K_1(S_3))(1-K_1(S_4))K_3(S_5,S_6,S_7)]$$

and

$$[C_1(S_1)(1-K_1(S_1))][C_1(S_2)(1-K_1(S_2))][C_3(S_3, S_4+S_5, S_6+S_7)(1-K_1(S_3))K_2(S_4, S_5)K_2(S_6, S_7)]$$

To complete $P_7^3(5)$, consider the remaining partitions and the kernels associated with them. In tabular form, the complete process and results are:

$P_7^3(5)$:

$p_{115} \rightarrow$

C_1	C_1	C_3		
K_1	K_1	K_1	K_1	K_3
C_1	C_1	C_3		
K_1	K_1	K_1	K_2	K_2

$\rightarrow C_1(S_1)\, C_1(S_2)\, C_3(S_3, S_4, S_5+S_6+S_7)$
$\times \prod_{i=1}^{4} [1-K_1(S_i)]\, K_3(S_5, S_6, S_7)$

$\rightarrow C_1(S_1)\, C_1(S_2)\, C_3(S_3, S_4+S_5, S_6+S_7)$
$\times \prod_{i=1}^{3} [1-K_1(S_i)]\, K_2(S_4, S_5)\, K_2(S_6, S_7)$

$p_{124} \rightarrow$

C_1	C_1	C_3
K_1	K_2	$K_1\, K_1\, K_2$
C_1	C_2	C_2
K_1	$K_1\, K_1$	$K_1\, K_3$
C_1	C_2	C_2
K_1	$K_1\, K_1$	$K_2\, K_2$

$\rightarrow C_1(S_1)\, C_1(S_2+S_3)\, C_3(S_4, S_5, S_6+S_7)$
$\times [1-K_1(S_1)]\, K_2(S_2, S_3)[1-K_1(S_4)][1-K_1(S_5)]\, K_2(S_6, S_7)$

$\rightarrow C_1(S_1)\, C_2(S_2, S_3)\, C_2(S_4, S_5+S_6+S_7)$
$\times \prod_{i=1}^{4} [1-K_1(S_i)]\, K_3(S_5, S_6, S_7)$

$\rightarrow C_1(S_1)\, C_2(S_2, S_3)\, C_2(S_4+S_5, S_6+S_7)$
$\times \prod_{i=1}^{3} [1-K_1(S_i)]\, K_2(S_4, S_5)\, K_2(S_6, S_7)$

$p_{133} \rightarrow$

C_1	C_1	C_3
K_1	K_3	$K_1\, K_1\, K_1$
C_1	C_2	C_2
K_1	$K_1\, K_2$	$K_1\, K_2$

$\rightarrow C_1(S_1)\, C_1(S_2+S_3+S_4)\, C_3(S_5, S_6, S_7)$
$\times [1-K_1(S_1)]\, K_3(S_2, S_3, S_4) \prod_{i=5}^{7} [1-K_1(S_i)]$

$\rightarrow C_1(S_1)\, C_2(S_2, S_3+S_4)\, C_2(S_5, S_6+S_7)$
$\times \prod_{i=1, 2, 5} [1-K_1(S_i)]\, K_2(S_3, S_4)\, K_2(S_6, S_7)$

$p_{223} \rightarrow$

C_1	C_2	C_2
K_2	$K_1 K_1$	$K_1 K_2$
C_2	C_2	C_1
$K_1 K_1$	$K_1 K_1$	K_3
C_1	C_1	C_3
K_2	K_2	$K_1 K_1 K_1$

$$\rightarrow C_1(S_1+S_2)\, C_2(S_3, S_4)\, C_2(S_5, S_6+S_7) \times K_2(S_1, S_2) \prod_{i=3,4,5} [1-K_1(S_i)]\, K_2(S_6, S_7)$$

$$\rightarrow C_2(S_1, S_2)\, C_2(S_3, S_4)\, C_1(S_5+S_6+S_7) \times \prod_{i=1}^{4} |1-K_1(S_i)|\, K_3(S_5, S_6, S_7)$$

$$\rightarrow C_1(S_1+S_2)\, C_1(S_3+S_4)\, C_3(S_5, S_6, S_7) \times K_2(S_1, S_2)\, K_2(S_3, S_4) \prod_{i=5}^{7} |1-K_1(S_i)|$$

There are three quantities that remain to be specified: the sign of each p, the numerical coefficient of each p, and the characteristic coefficient associated with each p.

(a) The sign of each p equals $(-1)^r$, where r is defined as the number of kernels K_β with a subscript β other than one.

(b) The numerical coefficient is determined by two multiplicative factors. The first factor is equal to the number of distinct arrangements that are possible by changing the ordering within the various subpartitions. (A subpartition is the partition within a specific kernel.) The second factor comes from the rearrangement of the various kernels in all possible distinct ways. This factor is simply equal to the number of permutations of n objects. Kernels with the same subscript and same number of variables are considered identical in this permutation.

(c) The characteristic coefficient is a function only of the argument of p and not of the particular kernel it represents. For example, the characteristic coefficient of $x'' x' x$ pertaining to p_{115} is

$$\frac{1}{3}\left[S_1^2 S_2^1 \cdot \left(\sum_{i=3}^{7} S_i\right)^0 + S_2^2 \cdot \left(\sum_{i=3}^{7} S_i\right)^1 \cdot S_1^0 + \left(\sum_{i=3}^{7} S_1\right)^2 S_1^1 S_2^0\right]$$

which reduces to

$$S_1^2 S_2 + S_2^2 \left(\sum_{i=3}^{7} S_i\right) + \left(\sum_{i=3}^{7} S_i\right)^2 S_1$$

This comes from the general rule of writing the variables as a product of a summation of S's. The number of S's in each sum is equal to the argument of the p function. The exponent to which each sum is raised is equal to the order of the respective derivative. The terms are then permuted. For a term containing no

derivatives, the characteristic coefficient is simply 1.

As an example, consider a typical term of $P_7^3(5)$.

$$p''_{124} = \begin{array}{|c|c|c|} \hline \multicolumn{3}{|c|}{\text{┌ ╥ ╥╥╥}} \\ \hline C_1 & C_2 & C_2 \\ \hline K_1 & K_1\ K_1 & K_1\ K_3 \\ \hline \end{array} \quad C_1(S_1)\ C_2(S_2, S_3)\ C_2(S_4, S_5+S_6+S_7) \times \prod_{i=1}^{4} [1-K_1(S_i)]\ K_3(S_5, S_6, S_7)$$

(a) r = 1, so the sign is $(-1)^1 = -1$.

(b) The first factor of the numerical coefficient is 2. This comes from rearranging the subpartition corresponding to $C_2(S_4, S_5+S_6+S_7)$. The second factor is 3.

(c) The characteristic coefficient is 1 because there are no derivatives. The complete term is

$$-6\ C_1(S_1)\ C_2(S_2, S_3)\ C_2(S_4, S_5+S_6+S_7) \prod_{i=1}^{4} [1-K_1(S_i)]\ K_3(S_5, S_6, S_7)$$

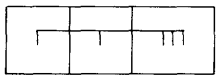
As one would expect, the compensator term is rather complicated. However, when we look at the examples and means of synthesis, we shall see that the kernels are built up by combinations of smaller, less formidable systems.

(ii) Construction of Q_m^n; Contribution of an n^{th}-order nonlinearity of the output to an m^{th}-order series compensator kernel

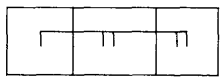
Nonlinear output terms are subject to similar, although simpler, expansion rules. By looking at the original equations, we see that

$$y(t)^n = \left\{ \int_0^t K_1(t-\tau)\ r(\tau)\ d\tau + \int_0^t \int_0^t K_2(t-\tau_1, t-\tau_2)\ r(\tau_1)\ r(\tau_2)\ d\tau_1 d\tau_2 + \ldots \right\}^n$$

Therefore, to find Q_m^n, we construct all possible partitions of m objects into n cells. For example, let n = 3 and m = 5. Possible partitions are:



q_{113}



q_{122}

which represent $K_1(S_1)\,K_1(S_2)\,K_3(S_3, S_4, S_5)$ and $K_1(S_1)\,K_2(S_2, S_3)\,K_2(S_4, S_5)$, respectively.

The sign of each term is positive. The numerical coefficient is equal to the number of different ways in which the partition can be ordered. The characteristic coefficient

is determined in exactly the same manner as it was for an input term.

Note that in order to find the functional series representation of any device when the input-output relation is a differential equation, we apply the Q_m^n algorithm and get the answer directly.

We now have all of the parts of the algorithm that are necessary for writing equations for the series compensator kernels by inspection. In the concluding part of this report (to be published in Quarterly Progress Report No. 60, January 15, 1961) examples demonstrating this procedure and a similar procedure for feedback compensator kernels will be given.

H. L. Van Trees, Jr.

References

1. N. Wiener, Nonlinear Problems in Random Theory (Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1958).

2. M. B. Brilliant, Theory of the Analysis of Nonlinear Systems, Technical Report 345, Research Laboratory of Electronics, M.I.T., March 3, 1958.

3. D. A. George, Continuous Nonlinear Systems, Technical Report 355, Research Laboratory of Electronics, M.I.T., July 22, 1959.

4. G. D. Zames, Nonlinear operators – cascading, inversion, and feedback, Quarterly Progress Report No. 53, Research Laboratory of Electronics, M.I.T., April 15, 1959, pp. 93-108.

5. G. D. Zames, Realizability of nonlinear filters and feedback systems, Quarterly Progress Report No. 56, Research Laboratory of Electronics, M.I.T., Jan. 15, 1960, pp. 137-143.

6. D. A. Chesler, Nonlinear Systems with Gaussian Inputs, Sc.D. Thesis, Department of Electrical Engineering, M.I.T., Jan. 11, 1960.

H. A PROBLEM IN RADIO ASTRONOMY*

A project has been initiated which will have as its goal the measurement of the galactic deuterium to hydrogen abundance ratio. This figure is of importance to the astrophysicist in the study of element synthesis in the universe. This measurement will be based upon comparison of the intensity of the 21-cm spectral line of atomic hydrogen with a similar line of atomic deuterium at 91.6 cm. The hydrogen line has been the subject of intensive observation since its detection, in 1951, by Ewen and Purcell. However, the deuterium line will be approximately a thousand times weaker and its presence has eluded four attempts at detection (1, 2, 3, 4). The conclusion drawn from these investigations is that the galactic deuterium to hydrogen abundance ratio is less than 1/4000,

*This research was supported in part by National Science Foundation under Grant G-13904.

which is just short of the probable terrestrial value of 1/6600. A further discussion of the radio astronomical aspects of the problem is given in a paper by the present writer (5).

The problem is essentially that of detecting a 0.025 per cent, 3-kc wide, dip in the noise power spectrum received by the antenna. An averaging time of approximately 4 weeks is needed to reduce statistical fluctuations to a sufficiently small value. Conventional analog radiometers are limited by lack of stability over the long period of time required for the measurement.

Our approach to this problem will be to use real-time digital autocorrelation of the incoming signal and then computer Fourier transformation to obtain the spectral density. The advantage of this method lies in the high degree of accuracy that can be obtained. The radiometer and correlator are now undergoing development.

The digital correlator is unique in that only one-bit samples of the incoming time function are used. That is, each sample is coded into a "1" if the polarity of the time function is greater than zero, or into a "0" if it is less than zero. For Gaussian time functions, the normalized one-bit autocorrelation function, $\rho_y(\tau)$, can be corrected to give the normalized true autocorrelation function, $\rho_x(\tau)$, by a simple relation derived by Van Vleck (6),

$$\rho_x(\tau) = \sin\left[\frac{\pi}{2}\rho_y(\tau)\right]$$

Our work (7) indicates that the rms fluctuation of the measured one-bit autocorrelation function is approximately $\pi/2$ times the rms fluctuation of an autocorrelation function derived from many-bit samples. This loss in rms fluctuation is far outweighed by the accuracy and speed with which the autocorrelation can be carried out. The one-bit correlator that is under construction will be capable of correlating a required 4 weeks of 30-kc information at 3-kc resolution in real time. The same operation with 36-bit words on the IBM 704 computer would require 16 years of computer time.

S. Weinreb

References

1. G. G. Getmanzev, K. S. Stankevitch, and V. S. Troitsky, Detection of monochromatic radio emission of deuterium from the center of the galaxy on a wavelength of 91.6 cm., Doklady Akad. Nauk S. S. S. R. 103, 783-786 (1955).

2. G. J. Stanley and R. Price, An investigation of monochromatic radio emission of deuterium from the galaxy, Nature 177, 1221-1222 (1956).

3. R. L. Adgie and J. S. Hey, Intensity of the radio line of galactic deuterium, Nature 179, 370 (1957).

(References continued on following page)

4. R. L. Adgie, An attempt to detect the 327 Mc. line of galactic deuterium, Paris Symposium on Radio Astronomy (Stanford University Press, Stanford University, Calif., 1958), pp. 352-354.

5. S. Weinreb, The Detection of UHF Radiation from Deuterium in Interstellar Space, S. B. Thesis, Department of Electrical Engineering, M. I. T., June 1958.

6. J. H. Van Vleck, The Spectrum of Clipped Noise, Report No. 51, Radio Research Laboratory, Harvard University, July 1943.

7. S. Weinreb, Autocorrelation Functions of Clipped Gaussian Processes, Internal Memorandum, Research Laboratory of Electronics, M. I. T., July 29, 1960.