

IX. STATISTICAL COMMUNICATION THEORY

Prof. Y. W. Lee
A. G. Bose
J. Y. Hayase

K. L. Jordan
C. S. Lorens
A. H. Nuttall
K. H. Powers

J. Tierney
W. E. Wernikoff
H. E. White

A. IMPROVING THE PERFORMANCE OF A GIVEN FILTER

A theory for the experimental determination of optimum, time-invariant, nonlinear filters was developed in the Quarterly Progress Report of October 15, 1955, pages 43-49. The present report develops methods of applying the theory given in the previous report for the purpose of improving the performance of simple filters.

As we increase the complexity of the filter (i. e., we use more Laguerre coefficients to characterize the past and more gate functions for each coefficient) the number of parameters (A_a 's) necessary to specify the filter grows very rapidly. In particular, if we use s Laguerre coefficients and n gate functions, for each coefficient we have n^s parameters, (A_a 's) to evaluate. After evaluating a large number of parameters, we should like to have some guarantee that our filter would perform at least as well as, say, a linear filter or a simple nonlinear filter that can be designed with less effort. Two methods of obtaining this guarantee will now be described.

Let us, first of all, prove the existence of a property of our class of filters which will be used in one of the methods. We want to show that the class of filters employing s Laguerre coefficients of the past includes the class of filters that only uses any one of the s Laguerre coefficients. Since we can always renumber the Laguerre coefficients, it is sufficient to prove that the s -coefficient class includes the class that uses only the first Laguerre coefficient u_1 . The series representation for the general system of this one coefficient class is

$$y(t) = \sum_{i=1}^n a_i \phi_i(u_1) \tag{1}$$

We can now make use of the fact that the sum of the n gate functions of any one coefficient is unity in order to express the series representation, Eq. 1, in the form

$$y(t) = \sum_{i=1}^n a_i \phi_i(u_1) = \sum_{i=1}^n a_i \phi_i(u_1) \left[\sum_{j=1}^n \phi_j(u_2) \right] \left[\sum_{k=1}^n \phi_k(u_3) \right] \dots \left[\sum_{h=1}^n \phi_h(u_s) \right] \tag{2}$$

which is recognized to be a special case of the expansion (Eq. 9, Quarterly Progress Report, Oct. 15, 1955, p. 47) for the general s -coefficient system. In a similar way, it can be easily shown that the class of filters using s Laguerre coefficients includes all classes having less than s coefficients. Note that this property is independent of

(IX. STATISTICAL COMMUNICATION THEORY)

the nature of the u 's; they may be Laguerre coefficients of the past of the filter input $x(t)$ or they may be obtained from $x(t)$ by any linear or nonlinear operation.

We now return to the problem of designing a nonlinear filter which is guaranteed to perform at least as well as a given filter. Let $x(t)$ be the input and $z(t)$ the desired output time functions. Our problem is to design a filter that is equal or superior, with respect to a weighted mean-square criterion of the type discussed, to a given filter F . F may be linear or nonlinear. We augment the Laguerre coefficients with the output, u_o , of the given filter F , as shown in Fig. IX-1. Then, by the property demonstrated above, the filter whose output is expressible as

$$y(t) = \sum_{i=1}^n a_i \phi_i(u_o) \quad (3)$$

is a member of the class of filters which has s Laguerre coefficients augmented by u_o . If the number n of gate functions associated with the variable u_o is sufficiently large, then to any degree of approximation Eq. 3 represents the class of filters shown in Fig. IX-2, consisting of F cascaded with a no-storage filter. Since the transfer characteristic of the no-storage system can be linear, the latter class certainly contains the filter F . Hence the filter determined by the procedure indicated in Fig. IX-1, for any s , performs at least as well as the given filter F and, in fact, at least as well as F cascaded with any no-storage filter. Having determined the A_a 's of the desired filter by the experimental procedure indicated in Fig. IX-1, the filter synthesis proceeds as indicated in Fig. IX-3.

We now turn our attention to another method of designing filters to improve the performance of given filters. Let the output of the given filter F be $u_o(t)$ when its input is $x(t)$. Our object is to improve (with respect to a weighted mean-square criterion) the performance of F by paralleling it with a filter which will be determined. The A_a 's of the desired filter are chosen to minimize

$$\mathcal{E} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T G(t) \left\{ z(t) - \left[u_o(t) + \sum A_a \Phi(a) \right] \right\}^2 dt \quad (4)$$

which is equivalent to

$$\mathcal{E} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T G(t) \left\{ \left[z(t) - u_o(t) \right] - \sum_a A_a \Phi(a) \right\}^2 dt \quad (5)$$

Comparing Eq. 5 with Eq. 11, Quarterly Progress Report, October 15, 1955, page 47, we find that the optimum A_a 's are determined by an experimental procedure like that indicated in Fig. VII-8, Quarterly Progress Report, October 15, 1955, with $z(t)$

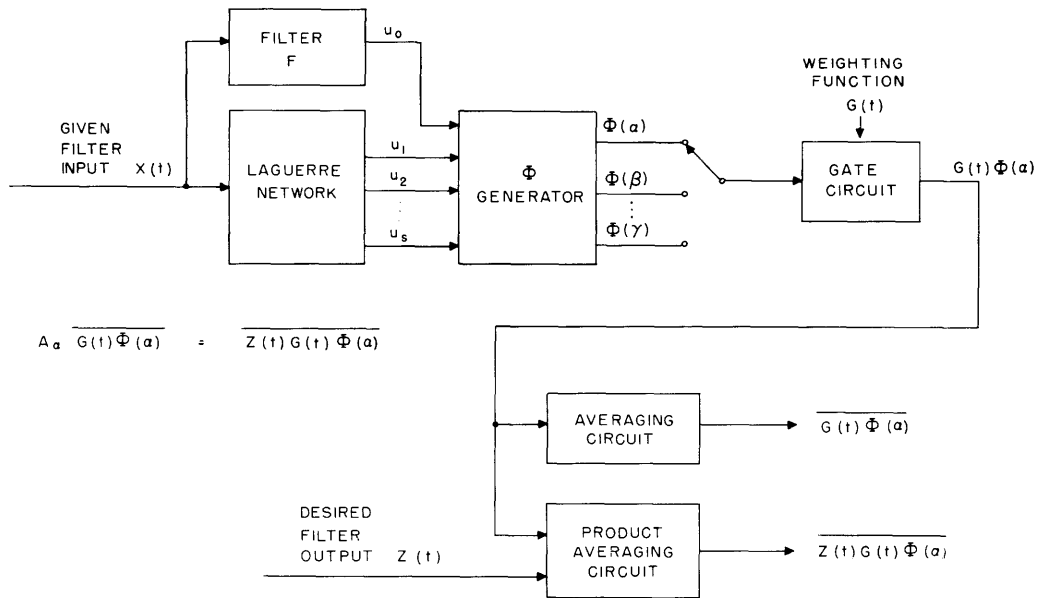


Fig. IX-1. Augmenting the Laguerre coefficients with the output of a given filter F.

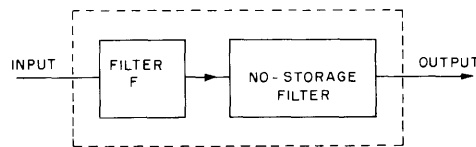


Fig. IX-2. The class of filters consisting of F cascaded with a no-storage filter.

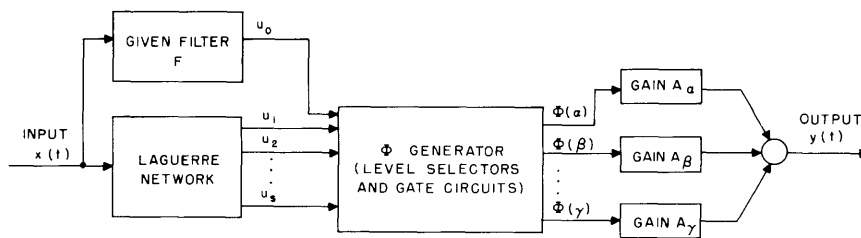


Fig. IX-3. Synthesis of a filter to improve the performance of a given filter F.

(IX. STATISTICAL COMMUNICATION THEORY)

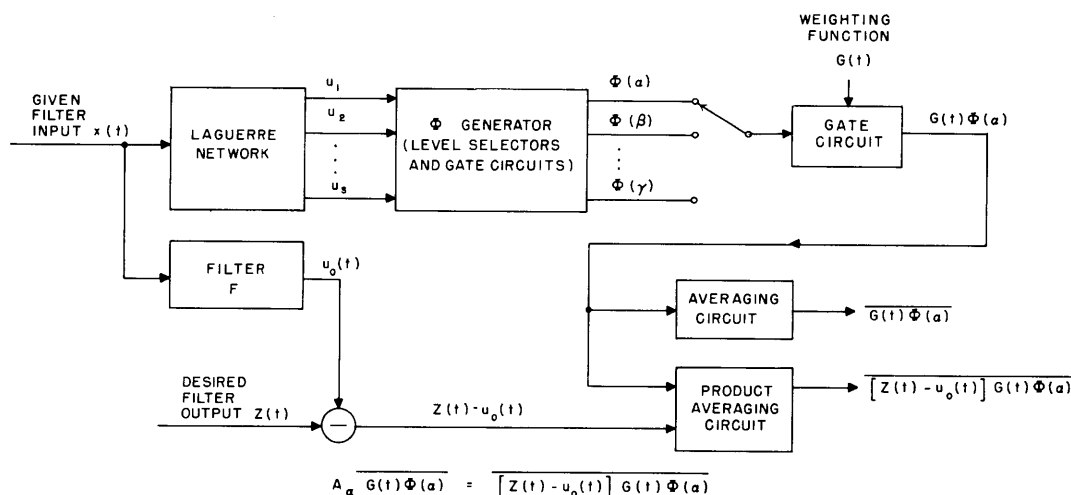


Fig. IX-4. Apparatus for determining a filter to improve the performance of F by a parallel connection.

replaced by $z(t) - u_o(t)$. The latter quantity is easily generated by applying $x(t)$ to the given filter F and subtracting its output from $z(t)$, as shown in Fig. IX-4. The parallel combination of F and the filter determined as we have just described will always perform at least as well as F, since that filter which has no transmission from input to output is a member of the class of filters considered in our theory. In other words, the solution in which all the A_α 's are equal to zero is a possible solution of Eqs. 4 and 5.

The second method for improving the performance of given filters offers the advantage of not having gate functions associated with the output time function of the given filter; therefore improvements can be made on the performance of F by very simple systems involving as few as one Laguerre coefficient. The first method does require a gate function expansion of the output of the given filter F but it has the advantage of ensuring that the performance of the over-all filter will always be at least as good as the performance of F cascaded with any no-storage system. In either method, the resultant over-all filter approaches the most general filter (of the class considered here) as the number of Laguerre coefficients is increased.

Since, in practice, the number of Laguerre coefficients will never be increased to very large numbers, it is unnecessary – sometimes it is even undesirable – in application of the theory, to confine our attention to a complete set of functions on the past of the input, such as Laguerre functions. For example, suppose we are dealing with binary signals. Then it is appropriate to replace the Laguerre network by a delay line with taps so spaced that consecutive symbols appear at adjacent taps. There are many choices we can make for our operators on the past of the input. Each choice of a finite number of operators on the past of the input implies a restriction to a class of filters

from which the test procedure, similar to that indicated in Fig. VII-8, Quarterly Progress Report, October 15, 1955, will pick the optimum for our particular problem.

A. G. Bose

B. ON THE PROBLEM OF MULTIPLE LINEAR PREDICTION

The problem of optimum linear prediction of a single discrete stationary random sequence consists in determining that linear operation to be applied to the past history of the sequence in order to minimize the mean-square difference between some particular future element of the sequence and the resultant of the linear operation. It will be shown that this problem has a unique solution that can be found by a perfectly straightforward procedure.

Multiple prediction is concerned not with a single sequence, but rather with a finite set of n stationarily correlated sequences, whose elements can be regarded as the coordinates of an n -dimensional vector sequence. The problem of prediction of the future values of the n sequences consists in determining that matrix operation which, when applied to the past history of the vector sequence, minimizes the mean-square errors of prediction of each coordinate of the vector. It will be shown that the solution to this latter problem does not share the uniqueness property enjoyed by the single prediction problem. For this reason, the general multiple-prediction problem has not yet been completely solved in closed form, although a scalar series solution suitable for machine computation has been given by Wiener and Rankin (1).

We shall review the single prediction problem and formulate the multiple one, in the hope that additional interest will be stimulated in this field of research. For simplicity, however, we shall restrict our consideration to prediction one-element ahead.

1. The Wold-Kolmogorov Decomposition of a Single Sequence

Let $\{f_i\}$ ($i = \dots, -1, 0, 1, \dots$) be a discrete stationary regular sequence (2) with spectral function $\Lambda_{ff}(z)$. Let $\{a_i\}$ ($i = 0, 1, 2, \dots$) be a set of coefficients and then consider the error sequence

$$\xi_k = f_k - \sum_{i=0}^{\infty} a_i f_{k-1-i} \quad (1)$$

It has been shown in (2) that those elements of the set $\{a_i\}$ which minimize the mean-square error ξ_k^2 are given by

$$a_i = \frac{1}{2\pi i} \oint \frac{1}{z} \left[1 - \frac{\lambda_{ff}(0)}{\lambda_{ff}(z)} \right] \frac{dz}{z^{i+1}} \quad (2)$$

(IX. STATISTICAL COMMUNICATION THEORY)

where $\lambda_{ff}(z)$ is that Szegö factor of $\Lambda_{ff}(z)$ which is analytic and nonvanishing inside the unit circle. If we examine the crosscorrelation coefficient between the error and the sequence f , we find that

$$\begin{aligned} \overline{\xi_k f_{k+m}} &= \overline{f_k f_{k+m}} - \sum_{i=0}^{\infty} a_i \overline{f_{k-1-i} f_{k+m}} \\ &= \frac{1}{2\pi i} \oint \Lambda_{ff}(z) \left[1 - \frac{1}{z} A\left(\frac{1}{z}\right) \right] \frac{dz}{z^{m+1}} \\ &= \frac{\lambda_{ff}(0)}{2\pi i} \oint \lambda_{ff}(z) \frac{dz}{z^{m+1}} \end{aligned} \quad (3)$$

However, since $\lambda_{ff}(z)$ is analytic in $|z| < 1$, it follows that the integral vanishes for negative values of m . We say that ξ is orthogonal to the past of f in the sense that $\overline{\xi_k f_{k+m}}$ is zero when $m < 0$. Since the past elements of ξ are linear combinations of the past elements of f , it is clear that ξ is orthogonal to its own past and is a purely random sequence.

If we write Eq. 1 in the form

$$f_k = \sum_{i=0}^{\infty} a_i f_{k-1-i} + \xi_k \quad (4)$$

we see that each element of the sequence f is expressed as the sum of two parts: the first is a linear combination of the past elements of f , while the second is orthogonal to those past elements. This is the so-called Wold-Kolmogorov decomposition (3) of the random sequence f . The sequence $\{\xi_i\}$ may be regarded as the "innovation" or new information contained in each element of the sequence $\{f_i\}$.

Denoting by U that transformation which converts each element of a sequence into the next preceding one, that is

$$U f_k = f_{k-1} \quad (5)$$

we may express Eq. 1 in the operational form

$$\xi_k = B(U) f_k \quad (6)$$

where $B(U)$ is a power-series expansion in positive powers of U . The coefficient β_i of the i^{th} power of U becomes from Eq. 1, $\beta_0 = 1$, $\beta_i = -a_{i-1}$ ($i > 0$), or from Eq. 2,

$$\beta_i = \frac{1}{2\pi i} \oint \frac{\lambda_{ff}(0)}{\lambda_{ff}(z)} \frac{dz}{z^{i+1}} \quad (7)$$

The function $B(z)$ is thus the power-series expansion about the origin of the function $\lambda_{ff}(0)/\lambda_{ff}(z)$ which is analytic and nonvanishing inside $|z| < 1$. It follows that the inverse $B^{-1}(z)$ has the same property, hence the elements of the sequence f can be expressed in terms of the past and present elements of the innovation sequence.

$$f_k = B^{-1}(U) \xi_k \quad (8)$$

It may be shown quite readily from Eq. 1 that

$$\Lambda_{\xi\xi}(z) = B(z) B\left(\frac{1}{z}\right) \Lambda_{ff}(z) \quad (9)$$

However, since the sequence $\{\xi_i\}$ is purely random, we have $\Lambda_{\xi\xi}(z) \equiv \overline{\xi_k^2} = \lambda_{ff}^2(0)$.

The Wold-Kolmogorov decomposition of a regular sequence amounts, in effect, to the determination of a function $B(z)$, analytic and nonvanishing inside the unit circle, satisfying the conditions

$$B(0) = 1$$

and

$$B(z) B\left(\frac{1}{z}\right) = \frac{\lambda_{ff}^2(0)}{\Lambda_{ff}(z)} \quad (10)$$

The Szegő factorization theorem may be employed to deduce uniquely that

$$B(z) = \frac{\lambda_{ff}(0)}{\lambda_{ff}(z)} \quad (11)$$

This result constitutes the solution of the single-prediction problem.

2. The Vector Sequence

Let the set $\{f_{1i}, f_{2i}, \dots, f_{ni}\}$ represent n stationarily correlated random sequences with correlation coefficients

$$R_m^{(ij)} = \overline{f_{ik} f_{j, k+m}} \quad (i, j = 1, 2, \dots, n) \quad (12)$$

and spectral functions

$$\Lambda_{ij}(e^{i\theta}) = \sum_{m=-\infty}^{\infty} R_m^{(ij)} e^{im\theta} \quad (13)$$

These spectral functions form the elements of an $n \times n$ Hermitian matrix $\mathcal{A}(e^{i\theta})$ whose determinant

$$\Delta(z) = |\mathcal{A}(z)| \quad (14)$$

(IX. STATISTICAL COMMUNICATION THEORY)

is real, nonnegative, and even on the unit circle.

The Wold-Kolmogorov decomposition in the multiple-sequence case consists of finding the set of coefficients $\{\eta_{ijk}\}$ and $\{\gamma_{ijk}\}$, with $\eta_{ij0} = \gamma_{ij0} = \delta_{ij}$, for which

$$f_{im} = \sum_{j=1}^n \sum_{k=0}^{\infty} \eta_{ijk} \xi_{j,m-k} \quad (15)$$

$$\xi_{im} = \sum_{j=1}^n \sum_{k=0}^{\infty} \gamma_{ijk} f_{j,m-k} \quad (16)$$

$$\overline{\xi_{ik} \xi_{jm}} = \delta_{km} \delta_{ij} \sigma_j^2 \quad (17)$$

The variance σ_j^2 is the minimum mean-square error for prediction of the sequence $\{f_{ji}\}$ one-element ahead and δ_{ij} is the Kroneker delta. The element ξ_{ik} is the part of f_{ik} that is orthogonal to the pasts of all the sequences $\{f_{1i}, f_{2i}, \dots, f_{ni}\}$. If we let $f_k = (f_{1k}, f_{2k}, \dots, f_{nk})$ and $\xi_k = (\xi_{1k}, \xi_{2k}, \dots, \xi_{nk})$ represent the present elements of n -dimensional vector random sequences, and let the transformation U be defined as in Eq. 5, we may express Eqs. 15 and 16 in the form

$$f_k = \mathcal{H}(U) \xi_k \quad (18)$$

$$\xi_k = \mathcal{H}^{-1}(U) f_k \quad (19)$$

where $\mathcal{H}(U)$ is a matrix of operators whose elements are defined by

$$H_{ij}(z) = \sum_{k=0}^{\infty} \eta_{ijk} z^k \quad (20)$$

In order that $\mathcal{H}(U)$, as well as its inverse, contain only positive powers of U , its determinant $|\mathcal{H}(z)|$ must be analytic and nonvanishing inside the unit circle. The constraint on the coefficient η_{ik0} imposes the additional condition $|\mathcal{H}(0)| = 1$. A straightforward calculation from Eqs. 15, 16, and 17 yields

$$\Lambda_{ij}(z) = \sum_{r=1}^n H_{jr}(z) H_{ir}\left(\frac{1}{z}\right) \sigma_r^2 \quad (21)$$

which, expressed in matrix form, becomes

$$\mathcal{A}(z) = \mathcal{H}\left(\frac{1}{z}\right) \mathcal{S} \tilde{\mathcal{H}}(z) \quad (22)$$

where $\tilde{\mathcal{H}}$ is the transpose of \mathcal{H} , and \mathcal{S} is a diagonal matrix whose r^{th} element is

σ_r^2 . If we denote by

$$\sigma^2 = |\mathcal{S}| = \prod_{r=1}^n \sigma_r^2 \quad (23)$$

the determinant of the diagonal matrix, we obtain from Eq. 22 the determinantal equation

$$|\mathcal{H}(z)| \left| \mathcal{H}\left(\frac{1}{z}\right) \right| \sigma^2 = \Delta(z) \quad (24)$$

By the method of Szegő, we set $\Delta(z) = \delta(z) \delta(1/z)$, where $\delta(z)$ is analytic and nonvanishing in $|z| < 1$. From the analyticity requirements of the determinants of $\mathcal{H}(z)$ and its inverse, and from the condition $|\mathcal{H}(0)| = 1$, we obtain

$$|\mathcal{H}(z)| = \frac{\delta(z)}{\delta(0)} \quad (25)$$

and

$$\sigma^2 = \delta^2(0) = \exp \frac{1}{2\pi i} \oint \log \Delta(z) \frac{dz}{z} \quad (26)$$

Since σ^2 is the product of the prediction variances of each coordinate of the vector process, the vanishing of σ^2 implies the perfect prediction of at least one member of the set of random sequences from the past history of all the members of the set. Thus a necessary condition that the vector process be regular of rank n is that the integral

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |\log |\Delta(e^{i\theta})|| d\theta$$

be finite (4).

If we denote by $\mathcal{G}(z)$ the inverse of the matrix $\mathcal{H}(z)$, we can write Eq. 22 as

$$\mathcal{G}\left(\frac{1}{z}\right) \mathcal{A}(z) \tilde{\mathcal{G}}(z) = \mathcal{S} \quad (27)$$

thus illustrating another interpretation of the multiple-prediction problem. The Wold-Kolmogorov decomposition consists in the determination of that matrix $\mathcal{G}(z)$ by which the spectral matrix is pre- and post-multiplied in accordance with Eq. 27 in order that the resultant be a diagonal matrix with constants along the diagonal. The matrix $\mathcal{G}(z)$ is constrained by the conditions that its elements be analytic in $|z| < 1$ and that its determinant be nonvanishing there as well. We shall show that even these constraints are not sufficient to specify $\mathcal{G}(z)$ uniquely.

Let us suppose that, in a given case, a certain matrix $\mathcal{G}(z)$ is known to satisfy the required conditions. Let that matrix produce from the vector sequence $\{f_i\}$ an

(IX. STATISTICAL COMMUNICATION THEORY)

innovation sequence $\{\xi_i\}$ in accordance with Eq. 19. Letting \mathcal{A} be a matrix of n^2 numbers a_{ij} satisfying

$$\sum_{j=1}^n a_{ij} a_{kj} \sigma_j^2 = \delta_{ik} \sigma_k^2 \quad (28)$$

we consider the vector sequence $\{\zeta_i\}$ formed by the operation

$$\zeta_k = \mathcal{A} \xi_k \quad (29)$$

whose coordinates are given by

$$\zeta_{ik} = \sum_{m=1}^n a_{im} \xi_{mk} \quad (30)$$

This operation represents a unitary transformation of the present value of the innovation vector. It follows directly from Eqs. 28 and 30 that the correlation coefficients of the sequence $\{\zeta_i\}$ become

$$\overline{\zeta_{ik} \zeta_{jm}} = \delta_{km} \delta_{ij} \sigma_j^2 \quad (31)$$

On comparing Eq. 31 with Eq. 17, we see that the vector sequence $\{\zeta_i\}$ is a perfectly good innovation vector that has identical correlation coefficients with $\{\xi_i\}$. If the matrix $\mathcal{G}(z)$ satisfies Eq. 27, so also does the product matrix $\mathcal{A}\mathcal{G}(z)$. Thus, in the multidimensional process, the Wold-Kolmogorov decomposition (hence the solution to the multiple-prediction problem) is unique only up to a unitary transformation.

K. H. Powers

References

1. N. Wiener and B. Rankin, Multiple prediction, *Ann. Math. Stat.* (to be published).
2. K. H. Powers, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Jan. 15, 1956, p.80.
3. A. N. Kolmogorov, *Bull. State Univ., Moscow, Ser. Math.* 2 (1941).
4. V. Zasuhin, *Compt. rend. Acad. Sci. U.S.S.R.* 33 (1941).

C. STATIONARY STATE DISTRIBUTION OF DISCRETE MARKOV SYSTEMS

For some time, a number of people have felt that it should be possible to obtain the properties of a Markov system directly from the graph that represents it. R. W. Sittler has shown how to obtain the generating function of the transitional probabilities (1).

(IX. STATISTICAL COMMUNICATION THEORY)

This report describes three methods for obtaining the stationary state distribution for an aperiodic recurrent Markov system directly from representative graph. Flow graphs similar to the Markov graph are used to solve the linear set of equations that is involved in the calculations.

Discrete Markov systems are defined by a discrete set of states (s_1, s_2, \dots, s_m) with an associated matrix of conditional probabilities $P = p_{jk}$ of going from state s_j to state s_k . These conditional probabilities from one state to another are restricted to being functions of only the two states s_j and s_k .

The graph representing this system is a collection of nodes representing the states (s_1, s_2, \dots, s_m) and branches representing the conditional transitions between the states. Branches representing zero probability are usually omitted. Along with the initial starting probabilities, the system's statistical properties are completely determined in the sense that they exist.

For the Markov systems composed of recurrent states, this report gives three methods for obtaining the stationary state probabilities. These methods are based on the well-known equations:

$$\sum_{r=1}^m P(s_r) p_{rk} - P(s_k) = 0 \quad (1)$$

$$\sum_{k=1}^m P(s_k) = 1 \quad (2)$$

$$\sum_{k=1}^m p_{rk} = 1 \quad r = 1, 2, \dots, m \quad (3)$$

where $P(s_k)$ is the stationary state probability, and p_{rk} is the conditional probability.

In each of the methods the flow graph that was used to perform the calculations corresponds exactly to the Markov graph except for minor changes in structure.

In the first method, $m - 1$ equations of Eq. 1 are solved with the use of a flow graph in terms of one state probability $P(s_j)$ in order to obtain the set of relations

$$P(s_k) = T_k \cdot P(s_j) \quad k = 1, 2, \dots, m \quad (4)$$

With the help of Eq. 2, the desired solution is obtained:

$$P(s_k) = \frac{T_k}{\sum_{k=1}^m T_k} \quad (5)$$

(IX. STATISTICAL COMMUNICATION THEORY)

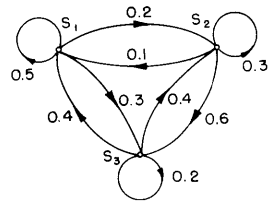


Fig. IX-5. Markov diagram.

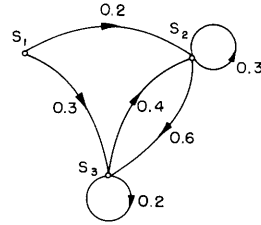


Fig. IX-6. First method for the stationary state distribution.

The second method uses $m - 1$ equations of Eq. 1, and Eq. 2, in flow-graph form. The state probabilities are obtained directly from the flow graph.

The third method consists essentially of calculating the diagonal cofactors Δ_k of the transition matrix minus the unit matrix and uses Eq. 6 to calculate the desired probabilities.

$$P(s_k) = \frac{\Delta_k}{\sum_{k=1}^m \Delta_k} \quad (6)$$

Calculation of the cofactor Δ_k is performed directly from the Markov graph using either Mason's rule (2) or the author's rule of nonintersecting loops (3). As an example, consider finding the stationary state probabilities of the Markov system of Fig. IX-5.

In using the first method, all branches converging on a single state node s_j are removed so that the $m - 1$ equations of Eq. 1 are represented as in Fig. IX-6. The equivalence relations T_k from s_j to s_k are then calculated for $k \equiv 1, 2, \dots, m$. Stationary state probabilities result from the application of Eq. 5. In this example,

$$T_1, T_2, T_3 : 1, \frac{7}{8}, \frac{33}{32}$$

Thus the stationary state probabilities are

$$P(s_1), P(s_2), P(s_3) : \frac{32}{93}, \frac{28}{93}, \frac{33}{93}$$

The second method uses the $m - 1$ equations of Eq. 1, and Eq. 2, in flow-graph form. All branches converging on a single state node s_j are first removed so that the Markov diagram represents Eq. 1. Equation 2 is represented by attaching a branch with a coefficient of -1 from each state node to state node s_j and driving state node s_j by a unit source, as is done in Fig. IX-7. The value of each state node, as determined by flow-graph manipulation, is the state probability. From Fig. IX-7 the following calculations are made.

$$P(s_1), P(s_2), P(s_3) : \frac{32}{93}, \frac{28}{93}, \frac{33}{93}$$

(IX. STATISTICAL COMMUNICATION THEORY)

The third method is based on properties of singular determinants that have an eigenvalue equal to +1. Normally, the solution to Eq. 1 is the cofactor of a row multiplied by an arbitrary constant of the matrix (P - I).

$$P(s_1), P(s_2), \dots, P(s_m) : A_{1j}C, A_{2j}C, \dots, A_{mj}C$$

However, in the calculation it is found that the row cofactors are equal to the diagonal cofactors. This is a result of Eq. 3.

The cofactor and minor are related by Eq. 7.

$$A_{ij} = (-1)^{i-j} M_{ij} \quad (7)$$

where M_{ij} is the matrix (P - I) with the i^{th} column and j^{th} row removed. In the minor M_{ij} , the i^{th} row is reconstructed as the sum of all the rows in the minor. The new coefficient of the i^{th} row and r^{th} column will then be

$$\sum_{\substack{k=1 \\ k \neq j}}^m (p_{rk} - \delta_{rk}) \quad \begin{array}{l} r = 1, 2, \dots, m \\ r \neq i \end{array}$$

which is equal to

$$-(p_{rj} - \delta_{rj})$$

Changing the sign of the i^{th} row and permuting the i^{th} row into the j^{th} row produces the result:

$$A_{ij} = (-1)^{i-j} M_{ij} = M_{ii} = A_{ii} \quad (8)$$

The cofactor A_{ii} of the matrix (P - I) is then related to the nonintersecting determinant of the node s_i by the relation

$$A_{ii} = (-1)^{m-1} \Delta_i \quad (9)$$

Methods for calculating Δ_i are described in references 2, 3, and 4. Subsequent use of Eq. 2 then gives the desired result of Eq. 6.

From Fig. IX-5 the cofactors are determined,

$$\Delta_1, \Delta_2, \Delta_3 : 0.32, 0.28, 0.33$$

Applying Eq. 6 to these results then gives the state probabilities:

$$P(s_1), P(s_2), P(s_3) : \frac{32}{93}, \frac{28}{93}, \frac{33}{93}$$

(IX. STATISTICAL COMMUNICATION THEORY)

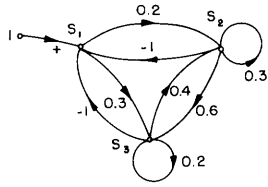


Fig. IX-7. Second method for the stationary state distribution.

A comparison of the three methods shows that each method has its advantages. The first method simplifies the diagram. The second requires only one calculation to obtain a single stationary probability. The third requires no modification of the diagram. Of course, each method also has its disadvantages. The first and third require essentially a solution for the whole

set of probabilities before any one can be found. The second method is practical only when a great number of the states have a finite probability of moving to one state. It is hard to say that one method is better than another without specifying the system under study. Each method has its own particular field of application wherein it works best.

There is a fourth method which involves the mean duration of a recurrent event. Calculations of the fourth method correspond to those of the first method.

C. S. Lorens

References

1. R. W. Sittler, Sc.D. Thesis, Department of Electrical Engineering, M.I.T., 1955.
2. S. J. Mason, Feedback theory – further properties of signal flow graphs, Technical Report 303, Research Laboratory of Electronics, M.I.T. (to be published).
3. C. S. Lorens, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Jan. 15, 1956, p. 97.

D. BASIC THEORY OF FLOW GRAPHS – INVERSION

Inversion plays an important part in the setting up and reduction of flow graphs. This report presents three essentially different methods of inverting the dependency in a flow graph. The first two methods are general; the third is limited to linear graphs.

Flow graphs represent a means of making positive statements and interrelating these statements with functional dependency. The statement

$$g(x) - ay + h(z) = z \tag{1}$$

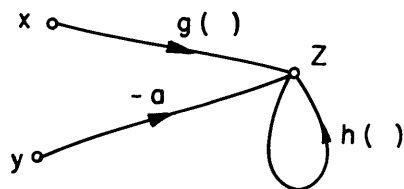


Fig. IX-8. Positive statement.

is represented in Fig. IX-8.

Inversion changes the independent variable to a dependent variable and vice versa. It is essentially a process of turning around and it is so represented in manipulating the flow graph.

Inversion has meaning only for two types of paths: a path from a strictly independent variable to a strictly dependent variable, and a path that forms a loop.

1. First Inversion Method

Consider the general node "z" shown in Fig. IX-9a, where we are interested in inverting the path from y_1 through z to x_1 . It is assumed that the inverse exists. If it did not, the inversion could not be performed. Figure IX-9a is the representation of the equations

$$\begin{aligned} f_1(y_1) + f_2(y_2) + h(z) &= z \\ g_1(z) &= x_1 \\ g_2(z) &= x_2 \end{aligned} \tag{2}$$

One method of inverting Eqs. 2 gives

$$\begin{aligned} f^{-1}(u) &= y_1 \\ z - h(z) - f_2(y_2) &= u \\ g_1^{-1}(x_1) &= z \\ g_2(z) &= x_2 \end{aligned} \tag{3}$$

The set of Eqs. 3 is represented in Figs. IX-9b and c.

Thus the general inversion rule is formulated:

1. Invert the direction of the path, splitting all nodes along the way and inserting the inverse function on the inverted branches.
2. At each split node attach all of the outgoing branches (except for the inverted branch) to the first node that represents the original variable. Attach all of the incoming branches (except for the inverted branch) to the second node, multiplying each function by -1.

The first rule simplifies in the special case of either one branch, leaving a node or one branch entering a node. These degenerate conditions with their inverse are shown in Fig. IX-10. Note that only in the case of "one branch entering the node before or after inversion" does the inverted node still remain the former variable.

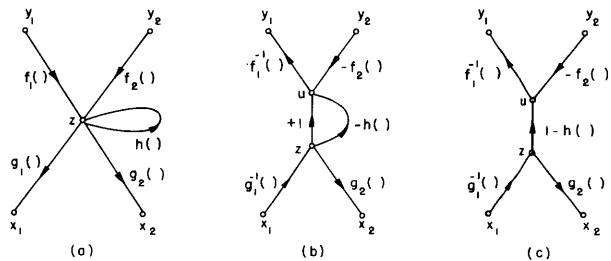


Fig. IX-9. (a) A general node "z". (b) First inversion method. (c) First inversion method.

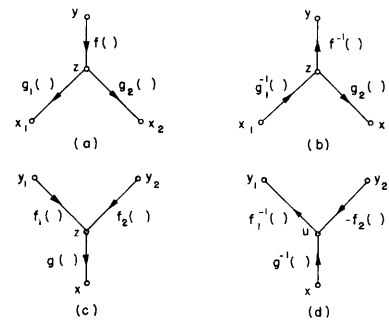


Fig. IX-10. Degenerate cases of the first inversion method. (a) First degenerate condition; (b) inverse of Fig. IX-10a; (c) second degenerate condition; (d) inverse of Fig. IX-10c.

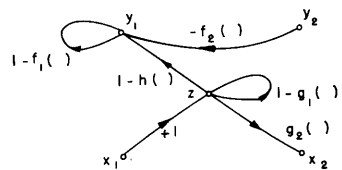


Fig. IX-11. Second inversion method.

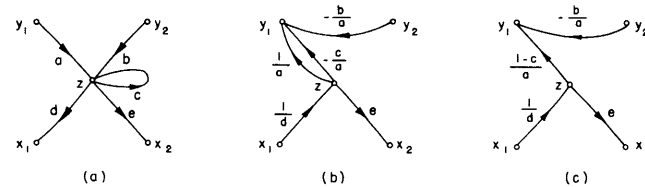


Fig. IX-12. (a) A linear node "z". (b) Third inversion method. (c) Third inversion method.

2. Second Inversion Method

The second general method involves another type of inversion that superficially avoids the inverse functions of the inverted branches. The method is quite important in linear graphs wherein it avoids the introduction of fractions.

Consider again the set of Eqs. 2 and their graphical representation, Fig. IX-9a. An inversion is indicated by Eqs. 4, which are represented in Fig. IX-11.

$$\begin{aligned}
 y_1 - f_1(y_1) - f_2(y_2) + z - h(z) &= y_1 \\
 z - g_1(z) + x_1 &= z \\
 g_2(z) &= x_2
 \end{aligned}
 \tag{4}$$

From this type of inversion we are able to formulate the following general rule:

1. Take the nose of each branch which is to be inverted and place it with its tail, changing the branch function to 1 minus the former branch function.
2. Take the nose of all incoming branches along with the nose of the inverted branch, changing the sign of the incoming branch functions. The former self-loop function is changed to 1 minus the former function.

It should be noted that all nodes have a self-loop. Some of them are zero.

3. Third Inversion Method

If we restrict ourselves to linear portions of a graph, a third type of inversion is possible. We need the linear property so that superposition will be valid. This method was originally developed by S. J. Mason and can be found in reference 1.

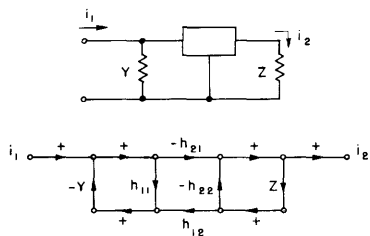


Fig. IX-13. Loaded transistor and flow graph.

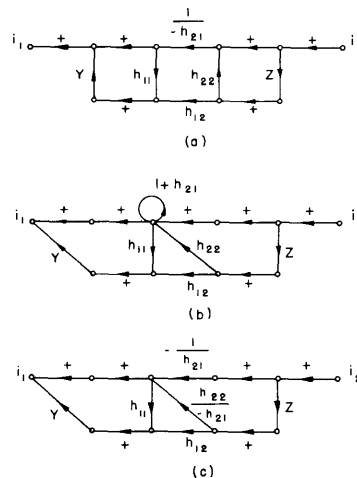


Fig. IX-14. (a) First inversion method.
 (b) Second inversion method.
 (c) Third inversion method.

(IX. STATISTICAL COMMUNICATION THEORY)

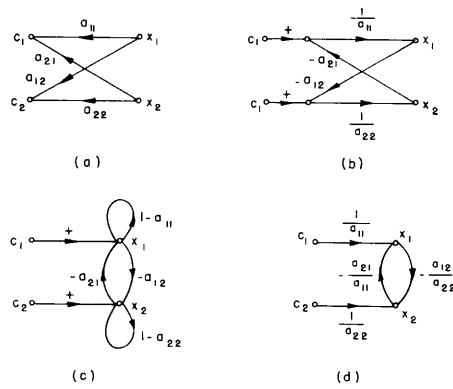


Fig. IX-15. (a) Second-order set of equations. (b) First inversion method. (c) Second inversion method. (d) Third inversion method.

Consider the linear graph of Fig. IX-12a, which represents Eqs. 5.

$$ay_1 + by_2 + cz = z$$

$$dz = x_1 \tag{5}$$

$$ez = x_2$$

A method of inversion of the path y_1 to x_1 through z is represented by Eqs. 6 (see Figs. IX-12b and c):

$$\frac{1}{a} z - \frac{c}{a} z - \frac{b}{a} y_2 = y_1$$

$$\frac{1}{d} x_1 = z \tag{6}$$

$$ez = x_2$$

Thus we are led to the following rule for the inversion of linear graphs:

1. Invert the path, changing the branch function to its reciprocal.
2. Move the nose of each incoming branch along with the nose of the inverted branch, multiplying the branch function by minus the value of the inverted branch function.

4. Comparison

As a means of comparison, consider inverting the main forward path of the flow graph which represents the loaded transistor, Fig. IX-13. The three inversion methods are illustrated in Figs. IX-14a, b, and c.

(IX. STATISTICAL COMMUNICATION THEORY)

Another example of inversion is the inversion of the two main paths ($x_1 C_1$) and ($x_2 C_2$) of Fig. IX-15a, which represents a second-order set of linear equations. For comparison, the three methods are shown in Figs. IX-15b, c, and d.

The first method has the advantage of leaving the form unchanged. The second method seems more complicated than the other two but it has the advantage of not being involved with inverse functions or fractions. The third method retains the original set of variables.

Inversion represents a convenient method of manipulating the dependency of a flow graph. Its two main uses are: it is an aid in reducing a graph to the desired relations and it is an aid in setting up the graph.

C. S. Lorens

References

1. S. J. Mason, Proc. IRE 41, 1144 (Sept. 1953).

E. DIGITAL PROBABILITY DENSITY ANALYZER

A digital probability density analyzer has been devised for operation in conjunction with the M. I. T. digital correlator. This combination enables the measurement of the first-order probability density of a voltage signal, $x(t)$, simultaneously with the correlation function. In addition, means have been incorporated in the analyzer for extending the method to second-order densities.

In the process of correlation, the correlator produces, for each sample of $x(t)$, a pair of time-modulated pulses separated by a time proportional to the amplitude of the signal sample. If time, measured from the first pulse as reference, is divided into equal intervals and the number of times that the second pulse appears in a certain interval is counted for a large number of samples, we have an approximate measure of the amplitude probability density of $x(t)$.

Error in the probability measurement is minimized by a correlator circuit (1, 2) that compensates for drift in the sections generating the time-modulated pulses from the samples of $x(t)$.

The analyzer has been designed for fifty intervals; in other words, the amplitude variation of $x(t)$ has been divided into fifty levels.

A block diagram of the analyzer is shown in Fig. IX-16. The first pulse "1" of the pulse pair initiates a gate pulse "2" which allows counter C_1 to count the interval pulses "3." The circuitry of the counter allows a positive pulse to appear at the output j during the interval between interval pulses $j - 1$ and j . Gate G_2 then opens during a specific interval, depending on what position switch S is in; if the second pulse "4" of the pair appears during that interval, then the count in counter C_2 is increased by one.

(IX. STATISTICAL COMMUNICATION THEORY)

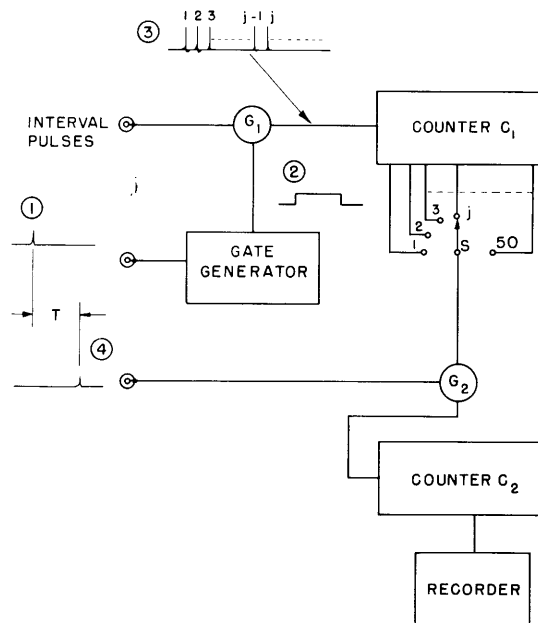


Fig. IX-16. Digital probability density analyzer.

By leaving S in the first position for a large number of samples and then consecutively stepping it through the other positions for the same number of samples, a histogram plot is obtained. This approximates the probability density of $x(t)$.

The analyzer is now under construction.

A. G. Bose, K. L. Jordan

References

1. Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., July 15, 1955, pp. 48-49.
2. Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., Jan. 15, 1956, pp. 104-105.

F. ANALOG PROBABILITY DENSITY ANALYZER

The amplitude distribution analyzer using a diode level selector, introduced by A. G. Bose in the Quarterly Progress Report of October 15, 1955, page 58, was experimentally tested in a slightly modified form.

The original diode level selector in the analyzer used two diodes in a shunt circuit to bypass the radiofrequency carrier to ground except when the input signal, $x(t)$, is within the amplitude range $E - \Delta E/2$ to $E + \Delta E/2$. A modified version of this level selector is shown in Fig. IX-17. It incorporates a series combination of two diodes

(IX. STATISTICAL COMMUNICATION THEORY)

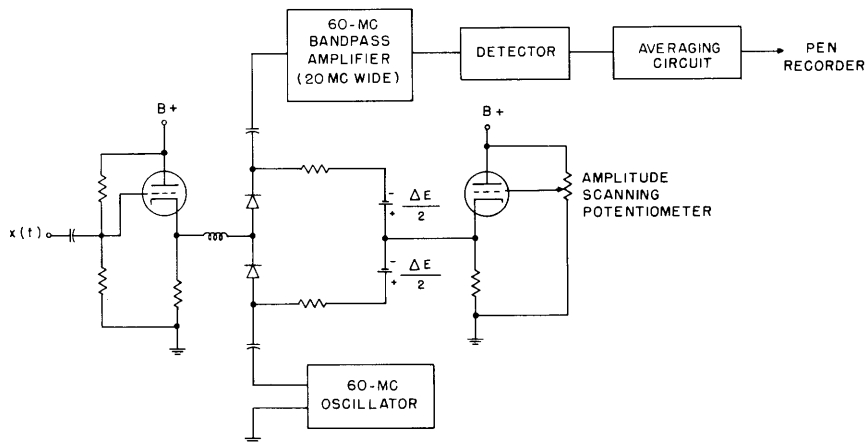


Fig. IX-17. Amplitude distribution analyzer with series diode level selector.

both of which conduct when $x(t)$ is within the amplitude channel under observation. Although the two circuits appear to furnish an equivalent means of obtaining the same result, the series circuit has produced better experimental results because it avoids the difficulty of obtaining a low 60-mc impedance to shunt the carrier frequency to ground.

The width of the amplitude slice of the series slicer circuit is dependent on the value of the diode bias potential, ΔE , and also on the voltage drop that results from diode current flowing through the impedance of the cathode-follower signal and bias sources. Cathode-follower amplifiers are used for both the signal and bias sources for the purpose of minimizing the source impedances and the resulting voltage drops in these impedances from the diode current. It is desirable to have the slice width principally determined by the battery bias, ΔE , because it is not influenced by changes of tube characteristics or supply voltages.

The pulse-recovery time of the level selector diodes seemed to be a possible limitation on the system frequency response that could not be accurately calculated. For this reason, the first experimental work was the measurement and optimization of the pulse response of the system. A rise time of $0.1 \mu\text{sec}$ was obtained from the detector output by using a 60-mc amplifier with a bandpass of 20 mc. The calculated value for the minimum output pulsewidth that results from analyzing a 32-kc sine wave through 50 amplitude levels is $0.2 \mu\text{sec}$. Since the sine wave is an extreme case of slope amplitude correlation (the maximum slope of a sine wave always occurs in the amplitude region near the axis), the analyzer should give satisfactory operation with uncorrelated signal inputs that contain frequency components somewhat larger than 32 kc.

H. E. White