XIV. PROCESSING AND TRANSMISSION OF INFORMATION*

Prof. E. Arthurs	S. Asano	T. Kailath
Prof. M. Eden	D. G. Detert	A. Paul
Prof. P. Elias	H. S. Ekre	J. C. Pennypacker
Prof. J. B. Dennis	H. A. Ernst	D. H. Pratt
Prof. R. M. Fano	R. F. Ferretti	D. L. Reich
Prof. E. M. Hofstetter	R. G. Gallager	D. Reiffen
Prof. D. A. Huffman	T. J. Goblick, Jr.	L. G. Roberts
Prof. W. W. Peterson	U. F. Gronemann	J. L. Rosenfeld
Prof. W. F. Schreiber	F. C. Hennie III	I. G. Stiglitz
Prof. C. E. Shannon	M. Horstein	N. K. H. Tam
Prof. W. M. Siebert	T. S. Huang	W. S. Wells
Prof. J. M. Wozencraft	F. Jelinek	H. P. Zeiger

A. ESTIMATING FILTERS FOR LINEAR TIME-VARIANT CHANNELS

In other work (1, 2) it has been shown that channel estimating filters play an important role in optimum receivers for linear time-variant channels. In this report we discuss these filters in greater detail, chiefly to supplement later discussions of receivers. The relation between minimum variance and maximum likelihood estimates is shown, even in the case of a "singular" channel. Some observations on the general estimation problem are also made.

1. Definition of the Problem

The situation that we shall consider is diagrammed in Fig. XIV-1. A known signal, x(t), of limited duration is transmitted through a random linear time-variant channel, A, of finite memory. The result is a waveform, z(t), which is further corrupted by additive noise, say n(t), before it is available to the receiver. Let y(t) denote the final received signal — that is, y(t) = n(t) + z(t) — and let T denote the duration of y(t).

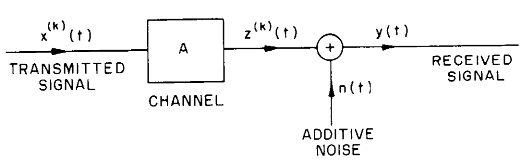

Fig. XIV-1. The channel.

Our problem is: Given y(t), and the knowledge of the statistical parameters of y(t) and n(t), we wish to derive estimates of z(t) on a minimum-variance and maximum-likelihood basis. For the minimum variance estimate we need only assume knowledge of the autocorrelation functions of y(t) and n(t); for the maximum likelihood estimate we have to assume that y(t) and n(t) are Gaussian. If we make these assumptions, we shall find that

*This work was supported in part by Purchase Order DDL B-00283 with Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology with the joint support of the U.S. Army, Navy, and Air Force under Air Force Contract AF 19(604)-5200.

the maximum likelihood estimate coincides with the minimum variance estimate, which is a well known result.

We assume that the additive noise is gaussian (but not necessarily white), that the random channel is such that its output z(t), for the given input x(t), is Gaussian, and that the mean and variance parameters of these distributions are known a priori. (The distributions themselves need not be stationary.) No further assumption is made about the structure of the channel. Scatter-multipath channels are often of this type. Our model takes account of Rayleigh fading and Rice fading (fading when a specular component is also present), with arbitrary rates of variation.

Two additional assumptions are made in order to simplify the presentation; namely, z(t) and n(t) have zero mean, and z(t) and n(t) are independent.

We shall use a discrete model for the channel and signals, as explained in reference 2. Thus for the channel model of Fig. XIV-1 we have

$$y = Ax + n = z + n \tag{1}$$

where y, x, n are column matrices that represent y(t), x(t), n(t); and A similarly represents the channel.

Furthermore, since the channel and noise are statistically independent, the probability density function of the (Gaussian) random variable y is given by

$$p(y|x) = \frac{1}{(2\pi)^{N/2}} \cdot \frac{1}{|\Phi_{yy}|^{1/2}} \exp - \frac{1}{2}\left\{y_t[\Phi_{yy}]^{-1}y\right\}$$

where N is the number of elements in the y-matrix, and $\Phi_{yy} = \Phi_{zz} + \Phi_{nn}$ is the covariance matrix of y. We shall assume that Φ_{nn} is nonsingular — an assumption that will guarantee that Φ_{yy} is also nonsingular.

2. Derivation of the Estimating Filters

We have described elsewhere (2) how the minimum variance estimate of z is to be obtained and merely quote the result here. This estimate does not require the assumption of Gaussian statistics for the noise. The other estimate that we discuss has been called the maximum likelihood estimate by Youla and Price (1). This is not the conventional designation in the statistical literature, and a more appropriate name might be "Bayes' estimate," as Davenport and Root have suggested. In this report, we shall retain the designation "maximum likelihood estimate."

a. Minimum variance estimate (2, 3)

The formula is

$$z_e = Hy = \Phi_{zz}(\Phi_{zz}+\Phi_{nn})^{-1} y \tag{2}$$

b. Maximum likelihood estimate

The maximum likelihood estimate of $\underline{z}$ is obtained by finding the $\tilde{\underline{z}}$ that maximizes the conditional probability, $p(\underline{z}|\underline{y})$, of $\underline{z}$, given $\underline{y}$. To solve for this $\tilde{\underline{z}}$ is, in general, quite difficult, but for Gaussian statistics the calculation is readily made. Therefore, here, in contrast to the practice for the minimum variance estimate, we shall specifically assume that $\underline{y}$, $\underline{z}$, and $\underline{n}$ are Gaussian. We have, then, $\underline{y} = \underline{z} + \underline{n}$, where $\underline{y}$, $\underline{z}$, $\underline{n}$ are all Gaussian with zero mean and covariance matrices $\underline{\Phi}_{yy}$, $\underline{\Phi}_{zz}$, $\underline{\Phi}_{nn}$, respectively. Furthermore, Bayes' rule gives

$$p(\underline{z}|\underline{y}) = \frac{p(\underline{y}|\underline{z})\, p(\underline{z})}{p(\underline{y})} = k \cdot p(\underline{y}|\underline{z})\, p(\underline{z})$$

because $p(\underline{y})$ is a constant. Now, we have

$$p(\underline{y}|\underline{z}) = p_n(\underline{y}-\underline{z}) = \frac{1}{(2\pi)^{N/2}\, |\underline{\Phi}_{nn}|^{1/2}} \exp -\frac{1}{2}\left\{(\underline{y}-\underline{z})_T \underline{\Phi}_{nn}^{-1} (\underline{y}-\underline{z})\right\} \tag{3}$$

and

$$p(\underline{z}) = \frac{1}{(2\pi)^{N/2}} \frac{1}{|\underline{\Phi}_{zz}|^{1/2}} \exp -\frac{1}{2}\, \underline{z}_T \underline{\Phi}_{zz}^{-1} \underline{z}, \text{ assuming that } \underline{\Phi}_{zz}^{-1} \text{ exists.} \tag{4}$$

Therefore

$$p(\underline{z}|\underline{y}) = k' \exp -\frac{1}{2}\left\{(\underline{y}-\underline{z})_T \underline{\Phi}_{nn}^{-1}(\underline{y}-\underline{z}) + \underline{z}_T \underline{\Phi}_{zz}^{-1} \underline{z}\right\}$$

To obtain the maximum-likelihood estimate of $\underline{z}$, we set

$$\frac{\partial p(\underline{z}|\underline{y})}{\partial \underline{z}} = 0$$

This gives

$$\frac{\partial}{\partial \underline{z}}\left\{(\underline{y}-\underline{z})_T \underline{\Phi}_{nn}^{-1}(\underline{y}-\underline{z}) + \underline{z}_T \underline{\Phi}_{zz}^{-1} \underline{z}\right\} = 0$$

or

$$-2(\underline{y}-\tilde{\underline{z}})_T \underline{\Phi}_{nn}^{-1} + 2\tilde{\underline{z}}_T \underline{\Phi}_{zz}^{-1} = 0$$

Therefore

$$\tilde{\underline{z}}_T\left(\underline{\Phi}_{nn}^{-1} + \underline{\Phi}_{zz}^{-1}\right) = \underline{y}_T \underline{\Phi}_{nn}^{-1}$$

or

$$\begin{aligned}\tilde{\underline{z}} &= \left(\underline{\Phi}_{nn}^{-1}+\underline{\Phi}_{zz}^{-1}\right)^{-1} \underline{\Phi}_{nn}^{-1}\underline{y} \\ &= \left(\underline{I}+\underline{\Phi}_{nn}\underline{\Phi}_{zz}^{-1}\right)^{-1} \underline{y} = \underline{\Phi}_{zz}(\underline{\Phi}_{zz}+\underline{\Phi}_{nn})^{-1} \underline{y} \\ &= \tilde{\underline{H}}\,\underline{y} \end{aligned} \tag{5}$$

where $\tilde{\underline{H}} = \underline{\Phi}_{zz}(\underline{\Phi}_{zz}+\underline{\Phi}_{nn})^{-1}$.

This expression for the maximum likelihood estimating filter, $\tilde{\underline{H}}$, is the same as that for the filter derived on a minimum average mean-square-error basis. This equivalence is a characteristic of Gaussian processes and has been proved several times (4). We might also point out that for Gaussian statistics, our estimate of $\underline{z}$ is optimum for more general criteria; for instance, for any function $L(\underline{z}-\underline{z}_e)$ that fulfills the requirements $L(0) = 0$, $L(-\epsilon) = L(\epsilon)$, $L(\epsilon_2) \geq L(\epsilon_1) \geq 0$, for $\epsilon_2 \geq \epsilon_1 \geq 0$. (See ref. 5.) We may point out that in the proof above we have assumed that Φ_{zz} is nonsingular, but no such restriction is necessary for the minimum variance estimate. As we might expect, therefore, this condition can be removed (cf. sec. 4).

3. Particular Cases

We shall examine the filter H in greater detail and study the effects of making additional assumptions about the channel H. The only term in the formula for H which is

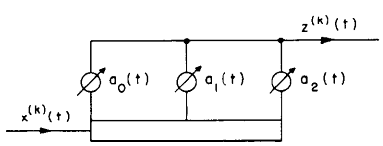

Fig. XIV-2. Simple delay-line channel.

affected thereby is $\underline{\Phi}_{zz}$. We shall calculate the precise manner in which the channel and the signal combine in this term. This will be done for four cases. (The delay-line model for the channel is shown in Fig. XIV-2.)

Case 1. Channel consists of a single path. In this case, a typical input-output relation might be

$$\begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} a_{00} & & \\ & a_{01} & \\ & & a_{02} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} + \tag{6}$$

To compute Φ_{zz} we shall use the following device. We can rewrite Eq. 6 as

$$\begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} x_0 & & \bigcirc \\ & x_1 & \\ \bigcirc & & x_2 \end{bmatrix} \begin{bmatrix} a_{00} \\ a_{01} \\ a_{02} \end{bmatrix} = \underline{X}a, \text{ for example.} \tag{7}$$

Then $\underline{\Phi}_{zz} = \overline{zz_T}$, where $\underline{z}\underline{z}_T$ is the direct, or Kronecker, product of $\underline{z}$ and $\underline{z}_T$, and the bar stands for an ensemble average over the random channel, A. Thus

$$\underline{\Phi}_{zz} = \underline{X}\overline{\underline{a}\underline{a}_T}\underline{X}_T = \underline{X}\underline{\Phi}_{AA}\underline{X}_T \tag{8}$$

This procedure yields

$$\underline{\Phi}_{zz} = \begin{bmatrix} x_0 & & \bigcirc \\ & x_1 & \\ \bigcirc & & x_2 \end{bmatrix} \begin{bmatrix} \overline{a_{00}^2} & \overline{a_{00}a_{01}} & \overline{a_{00}a_{02}} \\ \overline{a_{01}a_{00}} & \overline{a_{01}^2} & \overline{a_{01}a_{02}} \\ \overline{a_{02}a_{00}} & \overline{a_{02}a_{01}} & \overline{a_{02}^2} \end{bmatrix} \begin{bmatrix} x_0 & & \bigcirc \\ & x_1 & \\ \bigcirc & & x_2 \end{bmatrix}$$

$$= \begin{bmatrix} \overline{a_{00}^2}\, x_0^2 & \overline{a_{00}a_{01}}\, x_0x_1 & \overline{a_{00}a_{02}}\, x_0x_2 \\ \overline{a_{01}a_{00}}\, x_1x_0 & \overline{a_{01}^2}\, x_1^2 & \overline{a_{01}a_{02}}\, x_1x_2 \\ \overline{a_{02}a_{00}}\, x_2x_0 & \overline{a_{02}a_{01}}\, x_2x_1 & \overline{a_{02}^2}\, x_2^2 \end{bmatrix} \tag{9}$$

This equation can be directly verified by computation.

Case 2. Channel consists of a single time-invariant path. Case 1 should reduce to case 2 if we assume that $a_{00} = a_{01} = a_{02} = a_0$, say. This assumption would give

$$\underline{\Phi}_{zz} = \overline{a_0^2} \begin{bmatrix} x_0^2 & x_0x_1 & x_0x_2 \\ x_1x_0 & x_1^2 & x_1x_2 \\ x_2x_0 & x_2x_1 & x_2^2 \end{bmatrix} = \underline{X}\underline{\Phi}_{AA}\underline{X}_T$$

$$= \begin{bmatrix} x_0 & & \bigcirc \\ & x_1 & \\ \bigcirc & & x_2 \end{bmatrix} \begin{bmatrix} \overline{a_0^2} & \overline{a_0^2} & \overline{a_0^2} \\ \overline{a_0^2} & \overline{a_0^2} & \overline{a_0^2} \\ \overline{a_0^2} & \overline{a_0^2} & \overline{a_0^2} \end{bmatrix} \begin{bmatrix} x_0 & & \bigcirc \\ & x_1 & \\ \bigcirc & & x_2 \end{bmatrix} \tag{10}$$

This equation, too, can be directly verified. Notice, however, that in this case $\underline{\Phi}_{zz}$ can be written

$$\underline{\Phi}_{zz} = \overline{a_0^2} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} x_0 x_1 x_2 \end{bmatrix} \tag{11}$$

and therefore $\underline{\Phi}_{zz}$ is of rank 1. Thus, plainly, $\underline{\Phi}_{zz}^{-1}$ does not exist – which means that we cannot write the probability density function, Eq. 4, for $\underline{z}$, in this case. This is an example of a "singular normal distribution" (6). The reason for this peculiar behavior of $\underline{\Phi}_{zz}$ is that although we have three sample values, z_o, z_1, z_2, they do not constitute a three-dimensional random process. In fact, the randomness of $\underline{z}$ results from the random channel A, and for this we have only one random value, a_o. Therefore the three-dimensional distribution of $\underline{z}$ may be regarded as being concentrated along a particular straight line in three-dimensional space.

For the same reason, $\underline{\Phi}_{AA}$ as shown is clearly singular. In general, it is difficult to recognize when, in fact, $\underline{\Phi}_{zz}$ and $\underline{\Phi}_{AA}$ are singular. Thus, depending on the elements of $\underline{a}$, the $\underline{\Phi}_{zz}$ and $\underline{\Phi}_{AA}$ of Eqs. 8 and 9 might also be singular. Furthermore, perhaps if we used a different spacing for our discrete approximation we might now get a singular matrix. To talk about the rank of the matrix, as Cramér (6) does, is not much help to us because in our formulation the rank will depend on the size of the matrix. Fortunately, however, the situation is not so gloomy. Singular cases are rare, as shown by a theorem of Good (7). For a stationary channel, he shows that if the absolutely continuous portion of the integrated spectrum is not identically zero, the corresponding covariance matrices of any order (due to differences in sampling intervals) are always nonsingular. Thus singular channels correspond physically to pure frequency translating (Doppler) channels. The random time-invariant channel, as studied by Turin (8), for example, is a singular channel, and we shall usually discuss only this case.

However, whether or not $\underline{\Phi}_{zz}$ is singular, we can always write it in the form

$$\underline{\Phi}_{zz} = \underline{X}' \underline{\Phi}'_{AA} \underline{X}'_T \tag{12}$$

where $\underline{\Phi}'_{AA}$ is nonsingular, and the $\underline{X}$ are matrices that depend only on x(t) and are not necessarily square. For example, in the time-invariant case we can write

$$\underline{z} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} [a_0] = \underline{X}' a_0 \tag{13}$$

$$\underline{\Phi}_{zz} = \underline{X}'\,\underline{\Phi}'_{AA}\,\underline{X}'_T$$

where $\underline{\Phi}'_{AA} = \left[\overline{a_o^2}\right]$ and is nonsingular.

However, recall that in deriving the maximum likelihood filter, we used $\underline{\Phi}_{zz}^{-1}$. Thus the proof given in section 2b is not valid in this case and will have to be modified. (Cf. discussion in sec. 4 and at the end of sec. 3.) We shall find the form of Eq. 12 in which we use a nonsingular $\underline{\Phi}'_{AA}$ useful for this purpose.

Case 3. General time-invariant channel. For simplicity, we shall consider only low-order matrices and vectors, but the relation that we shall derive is general, and the method of obtaining it in higher-order cases is the same as in the simple example considered.

Thus we might have

$$\begin{bmatrix} z_0 \\ z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} a_{00} & 0 \\ a_{11} & a_{00} \\ a_{22} & a_{11} \\ 0 & a_{22} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$$

which may be rewritten

$$\begin{bmatrix} z_0 \\ z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} x_0 & 0 & 0 \\ x_1 & x_0 & 0 \\ 0 & x_1 & x_0 \\ 0 & 0 & x_1 \end{bmatrix} \begin{bmatrix} a_{00} \\ a_{11} \\ a_{22} \end{bmatrix} = \underline{X}\underline{a}, \text{ for example,} \tag{14}$$

and we can prove that $\underline{\Phi}_{zz} = \underline{X}\,\underline{\Phi}_{AA}\,\underline{X}_T$, where

$$\underline{\Phi}_{AA} = \begin{bmatrix} \overline{a_{00}^2} & \overline{a_{00}a_{11}} & \overline{a_{00}a_{22}} \\ \overline{a_{00}a_{11}} & \overline{a_{11}^2} & \overline{a_{11}a_{22}} \\ \overline{a_{00}a_{22}} & \overline{a_{11}a_{22}} & \overline{a_{22}^2} \end{bmatrix} \tag{15}$$

This matrix is nonsingular unless either $a_{11} = a_{oo}$, or $a_{22} = a_{oo}$, or both. The most interesting case is the nonsingular one, and therefore we shall assume this situation.

Case 4. General time-variant nonsingular channel. We shall again study a simple example:

$$\begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} a_{00} & 0 \\ a_{11} & a_{01} \\ 0 & a_{12} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$$

which may be rewritten

$$\begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} x_0 & 0 & 0 & 0 \\ 0 & x_1 & x_0 & 0 \\ 0 & 0 & 0 & x_1 \end{bmatrix} \begin{bmatrix} a_{00} \\ a_{01} \\ a_{11} \end{bmatrix} = \underline{\underline{X}}\underline{a}, \text{ for example.} \tag{16}$$

Then $\underline{\Phi}_{zz} = \underline{\underline{X}}\underline{\Phi}_{AA}\underline{\underline{X}}_T$, where

$$\underline{\Phi}_{AA} = \begin{bmatrix} \overline{a_{00}^2} & \overline{a_{00}a_{01}} & \overline{a_{00}a_{11}} & \overline{a_{00}a_{12}} \\ \overline{a_{00}a_{01}} & \overline{a_{01}^2} & \overline{a_{01}a_{11}} & \overline{a_{01}a_{12}} \\ \overline{a_{11}a_{00}} & \overline{a_{11}a_{01}} & \overline{a_{11}^2} & \overline{a_{11}a_{12}} \\ \overline{a_{12}a_{00}} & \overline{a_{12}a_{01}} & \overline{a_{11}a_{12}} & \overline{a_{12}^2} \end{bmatrix}$$

$$= \left[\begin{array}{c:c} \underline{\Phi}_{00} & \underline{\Phi}_{01} \\ \hdashline \underline{\Phi}_{10} & \underline{\Phi}_{11} \end{array}\right], \text{ for example.}$$

Here, $\underline{\Phi}_{00}$ is the covariance matrix for the values a_{00}, a_{01}, ..., assumed by the top $a_0(t)$. Notice that in all of these four cases the arrangements of the sample values of the $\underline{a}$ vector could be arbitrary; this would have led to different forms for the $\underline{X}$ and $\underline{\Phi}_{AA}$ matrices. The forms that we have given appear to be the most convenient ones. However, some of the other arrangements are also of interest. We shall merely give one such illustration for case 4. Thus Eq. 16 could also be written

$$\begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} x_0 & 0 & 0 & 0 \\ 0 & x_0 & x_1 & 0 \\ 0 & 0 & 0 & x_1 \end{bmatrix} \begin{bmatrix} a_{00} \\ a_{11} \\ a_{01} \\ a_{12} \end{bmatrix} = \underline{\underline{X}}\underline{a}, \text{ for example,}$$

and now

$$\underline{\Phi}_{AA} = \begin{bmatrix} \overline{a_{00}^2} & \overline{a_{00}a_{11}} & \overline{a_{00}a_{01}} & \overline{a_{00}a_{12}} \\ \overline{a_{11}a_{00}} & \overline{a_{11}^2} & \overline{a_{11}a_{01}} & \overline{a_{11}a_{12}} \\ \overline{a_{01}a_{00}} & \overline{a_{01}a_{11}} & \overline{a_{01}^2} & \overline{a_{01}a_{12}} \\ \overline{a_{12}a_{00}} & \overline{a_{12}a_{11}} & \overline{a_{12}a_{01}} & \overline{a_{12}^2} \end{bmatrix}$$

Thus in these four cases we have obtained a more explicit formula for $\underline{\Phi}_{zz}$:

$$\underline{\Phi}_{zz} = \underline{X}\underline{\Phi}_{AA}\underline{X}_T \tag{17}$$

where the $\underline{X}$ and the $\underline{\Phi}_{AA}$ are to be appropriately chosen for the case under consideration. However, we have found that in case 2, the matrix $\underline{\Phi}_{zz}$ is singular. (Also notice that the $\underline{\Phi}_{AA}$ for case 3 cannot be obtained from that of case 4 by making suitable a-values equal. This is for the same reasons as for case 2.)

In our derivation of the maximum likelihood estimator we used the quantity $\underline{\Phi}_{zz}^{-1}$. This is invalid for case 2. But the derivation of the minimum variance estimator did not involve $\underline{\Phi}_{zz}$ and it gives an identical result. Moreover, the final formula for the estimator does not involve $\underline{\Phi}_{zz}^{-1}$. Therefore we would expect that the formula for $\underline{H}$ is correct in all cases, but the derivation for the maximum likelihood criterion should be modified. This will be done in section 5. For the present, by using Eq. 17 we can write

$$\underline{H} = \underline{X}\underline{\Phi}_{AA}\underline{X}_T(\underline{\Phi}_{nn}+\underline{X}\underline{\Phi}_{AA}\underline{X}_T)^{-1} \tag{18}$$

4. Rederivation of Maximum Likelihood Estimating Filter

We can write, for all cases, $\underline{y} = \underline{A}\underline{x} + \underline{n} = \underline{X}\underline{a} + \underline{n} = \underline{z} + \underline{n}$, with the $\underline{X}$ and $\underline{a}$ suitably defined. In this proof, as opposed to the one in section 3, we shall first try for a maximum likelihood estimate of $\underline{a}$ and use this to get the estimate for $\underline{z}$. Therefore we consider

$$p(\underline{a}|\underline{y}) = \frac{p(\underline{y}|\underline{X}\underline{a})\ p(\underline{a}|\underline{X})\ p(\underline{X})}{p(\underline{y})} = k \cdot p(\underline{y}|\underline{X}\underline{a})\ p(\underline{X}|\underline{a})\ p(\underline{a})$$

Now

$$p(\underline{y}|\underline{X}\underline{a}) = p_n(\underline{y}-\underline{X}\underline{a}) = \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{|\underline{\Phi}_{nn}|^{1/2}} \exp - \frac{1}{2}\left\{(\underline{y}-\underline{X}\underline{a})_T\underline{\Phi}_{nn}^{-1}(\underline{y}-\underline{X}\underline{a})\right\}$$

where $p(\underline{X})$ = constant; $p(\underline{a}|\underline{X}) = p(\underline{a})$ because the channel and the transmitted signal are assumed to be independent; and

$$p(\underline{a}) = \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{|\underline{\Phi}_{AA}|^{1/2}} \cdot \exp - \frac{1}{2}\left\{\underline{a}_T\underline{\Phi}_{AA}^{-1}\underline{a}\right\} \tag{19}$$

and this is defined because $\underline{\Phi}_{AA}$ is chosen to be nonsingular. This is the crux of the proof.

$$p(\underline{a}|\underline{y}) = k' \cdot \exp - \frac{1}{2}\left\{(\underline{y}-\underline{X}\underline{a})_T\underline{\Phi}_{nn}^{-1}(\underline{y}-\underline{X}\underline{a})+\underline{a}_T\underline{\Phi}_{AA}^{-1}\underline{a}\right\}$$

where k' is a constant; that is, it is independent of $\underline{a}$. For a maximum likelihood estimate of $\underline{a}$, we set

$$\frac{\partial p(\underline{a}|\underline{y})}{\partial \underline{a}} = 0$$

that is,

$$\frac{\partial}{\partial \underline{a}}\left\{(\underline{y}-\underline{X}\underline{a})_T\underline{\Phi}_{AA}^{-1}(\underline{y}-\underline{X}\underline{a})+\underline{a}_T\underline{\Phi}_{AA}^{-1}\underline{a}\right\} = 0$$

or

$$-2(\underline{y}-\underline{X}\underline{a})_T\ \Phi_{nn}^{-1}X + 2\underline{\tilde{a}}_T\underline{\Phi}_{AA}^{-1} = 0$$

which yields

$$\underline{\tilde{z}} = \underline{X}\underline{\tilde{a}} = \underline{X}\left(\underline{\Phi}_{AA}^{-1}+\underline{X}_T\underline{\Phi}_{nn}^{-1}\underline{X}\right)^{-1}\ \underline{X}_T\underline{\Phi}_{nn}^{-1}\underline{y} \tag{20}$$

This equation gives the estimate of $\underline{z}$ on a maximum likelihood basis. (We have shown a different expression for this estimate in sec. 4a.)

Thus on a minimum-variance basis we have

$$\underline{z}_e = \underline{H}\,\underline{y} = \underline{\Phi}_{zz}(\underline{\Phi}_{zz}+\underline{\Phi}_{nn})^{-1} = \underline{X}\underline{\Phi}_{AA}\underline{X}_T(\underline{\Phi}_{nn}+\underline{X}\underline{\Phi}_{AA}\underline{X}_T)^{-1}$$

Therefore we should have the identity

$$\underline{H} = \underline{X}\underline{\Phi}_{AA}\underline{X}_T(\underline{\Phi}_{nn}+\underline{X}\underline{\Phi}_{AA}\underline{X}_T)^{-1} = \underline{X}\left(\underline{\Phi}_{AA}^{-1}+\underline{X}_T\underline{\Phi}_{nn}^{-1}\underline{X}\right)^{-1}\ \underline{X}_T\underline{\Phi}_{nn}^{-1} \tag{21}$$

This can be verified purely on a matrix basis.

The proof is not too difficult, and we shall leave it as an interesting exercise for the reader. Notice, however, that except in case 1, the matrix X is not square. Another method of proof is to make a formal series expansion – the Neumann series – of the terms $(\underline{\Phi}_{nn}+\underline{X}\underline{\Phi}_{AA}\underline{X}_T)^{-1}$ and $\left(\underline{\Phi}_{AA}^{-1}+\underline{X}_T\underline{\Phi}_{nn}^{-1}\underline{X}_T\right)^{-1}$ on each side of Eq. 21, and verify the identity term by term. We remark that the direct matrix proof of Eq. 21 enables us to prove in another way the equivalence of minimum variance and maximum likelihood estimators.

5. Solution of the Equations for the Estimating Filter

The equations that we have obtained for the estimating filter H – for example,

$$\underline{H} = \underline{\Phi}_{zz}(\underline{\Phi}_{zz}+\underline{\Phi}_{nn})^{-1}$$

$$\underline{H} = \underline{X}\left(\underline{\Phi}_{AA}^{-1}+\underline{X}_T\underline{\Phi}_{nn}^{-1}\underline{X}\right)^{-1} \underline{X}_T\underline{\Phi}_{nn}^{-1}$$

are discrete analogs of the Wiener-Hopf equation. This equation is, in general, difficult to solve, and explicit answers for the continuous case have only been found in a few special cases by a variety of methods that are used by Kac and Siegert, Zadeh and Miller, Slepian, Youla, Price, Middleton, and others. We have found solutions in some cases on a discrete basis. It is important to stress, however, that the significance of the preceding formulas is chiefly that they give us functional forms for ideal receivers for general channels. An explicit solution of the equations for H would, no doubt, be valuable, but the solution is only valid under the strict assumptions of our model. Actual situations differ inevitably from the model and here the functional forms of the receivers for our model serve us better by helping to make extensions and extrapolations to real situations. In this connection, approximate solutions for H are useful and one method of obtaining them is by the use of Neumann's series. We shall illustrate this by considering

$$\underline{H} = \underline{\Phi}_{zz}(\underline{\Phi}_{zz}+N_o\underline{I})^{-1} = \frac{1}{N_o}\,\underline{\Phi}_{zz}\left(\underline{I}+\frac{\underline{\Phi}_{zz}}{N_o}\right)^{-1}$$

$$= \frac{1}{N_o}\,\underline{\Phi}_{zz}\left(I - \frac{\underline{\Phi}_{zz}}{N_o} + \frac{\underline{\Phi}_{zz}^2}{N_o^2} - \dots\right) = \frac{1}{N_o}\,\underline{\Phi}_{zz} - \dots + \dots \qquad (22)$$

This infinite geometric series, or Neumann series, converges for sufficiently large N_o. (A more detailed discussion of the convergence properties can be found in books on integral equations.) In many cases an approximate representation of $\underline{H}$ by $\underline{\Phi}_{zz}/N_o$ is useful and instructive.

6. Concluding Remarks

We have given a discussion of estimates for random channels, directed particularly toward aiding future discussions of optimum receivers. We shall now add a few remarks on the estimation problem.

We have deduced several forms of estimates for the quantity z(t). Since we know the transmitted signal x(t), we can calculate what the channel A actually was by "de-convolution." Equation 25 gives us this channel estimate. However, our objective might be to make measurements on a general (time-variant) channel to enable us to synthesize a model that will (with suitable delays) respond to inputs in the same fashion as the original channel would respond. For time-variant channels, it appears that even in the absence of noise, there are situations in which it is impossible to determine the actual time-variant channel by input-output measurements only (9). Some methods have been discussed in reference 9 for making such determinations when they are possible. R. G. Gallagher (10) has used some of these ideas in measurements on time-variant telephone lines.

A general study of channel estimation was recently made by Levin (11). Since he also gives an extensive bibliography, we shall not discuss this problem further, except for a special case studied by Turin (8).

For the noisy channel case, Turin was the first to show that crosscorrelation was the best measurement technique for a time-invariant channel. This follows also from our Eq. 20, when we assume a single time-invariant path. In this case, $\underline{\Phi}_{AA}^{-1} = 1/\sigma^2$, say, $\underline{\Phi}_{nn}^{-1} = 1/N_o$, and $\underline{X}$ is a column vector; and $\underline{X}_T\underline{X} = 2E$, where E is the energy in the transmitted signal x(t). Therefore

$$\underline{\tilde{a}} = \underline{H}'_T\underline{X}_T\underline{\Phi}_{nn}^{-1}\underline{y} = \left(\frac{1}{\sigma^2} + \frac{2E}{N_o}\right)^{-1} \frac{1}{N_o}\underline{X}_T\underline{y}$$

and $\underline{X}_T\underline{y}$ is a crosscorrelation. Our matrix approach to this problem also reveals the possibility of generalizing this result to more complicated situations – with many time-invariant paths, time-variant channels, and so forth.

We note that this work applies directly to the problems of estimating and detecting random and gaussian signals in gaussian noise. This kind of situation is encountered in radio astronomy.

A few remarks should be made about notation and analysis. The matrix approach is much more convenient for obtaining general block-diagram solutions for the estimators and receivers than the integral-equation method. To derive the equations for a general channel as we have just done, using integral equations only, is rather complicated. However, when it comes to solving the matrix equations in particular cases, it is convenient to pass (formally) to the limit and obtain our results in continuous form. The advantage is that much work has been done on (linear) integral equations, while matrix equations have not been as extensively investigated. In addition to the matrix method as we have used it, and the integral-equation method, as used, for example, by Price (1), we can use Grenander's method of observable coordinates (12, 13). This method is probably the most rigorous, but the least intuitive or physical, of the three methods. It is useful in detection theory because it points out clearly the occurrence and existence of "singular" cases, when perfect detection is possible (12).

T. Kailath

References

1. R. Price, Notes on ideal receivers for scatter channels, Group Report 34-39, Lincoln Laboratory, M.I.T., May 1955.

2. T. Kailath, Correlation detection of signals perturbed by a random channel, Trans. IRE, Professional Group on Information Theory, Special Issue on Matched Filter Detection, IT-6, pp. 361-366, June 1960.

3. B. Friedland, Theory of Time-Varying Sampled-Data Systems, Technical Report T-19/8, Electronics Research Laboratories, Columbia University, New York, April 1957.

4. G. W. Preston, The equivalence of optimum transducers and sufficient and most efficient statistics, J. Appl. Phys. 24, 841 (1953).

5. S. Sherman, Non mean-square error criteria, Trans. IRE, IT-4, pp. 125-126 (1958).

6. H. Cramér, Mathematical Methods of Statistics (Princeton University Press, Princeton, N.J., 1946).

7. I. J. Good and K. Caj. Good, A paradox concerning rate of information, Information and Control, pp. 113-126, May 1958; pp. 195-197, June 1959.

8. G. L. Turin, Communication through Noisy, Random-Multipath Channels, Technical Report 116, Lincoln Laboratory, M.I.T., May 1956.

9. T. Kailath, Sampling Models for Linear Time-Variant Filters, Technical Report 352, Research Laboratory of Electronics, M.I.T., May 24, 1959.

10. R. G. Gallagher, Measurement of Time-Varying Telephone Characteristics, Group Report 25G-0003, Lincoln Laboratory, M.I.T., September 1959.

11. M. J. Levin, Estimation of the Characteristics of Linear Systems in the Presence of Noise, Sc.D. Thesis, Electrical Engineering Department, Columbia University, New York, N.Y., July 1959; also see Trans. IRE, CT-7, pp. 50-56, March 1960.

12. U. Grenander, Stochastic Processes and Statistical Inference, Arkiv. Math. 2, 195 (1950).

13. W. B. Davenport, Jr. and W. L. Root, An Introduction to the Theory of Random Signals and Noise (McGraw-Hill Book Company, Inc., New York, 1958).

B. EXTRACTION OF INFORMATION FROM A PAIR OF PERIODICALLY VARYING RANDOM WAVEFORMS*

This research was completed and was presented to the Department of Electrical Engineering, M.I.T., in partial fulfillment of the requirements for the degrees of Bachelor of Science and Master of Science.

I. G. Stiglitz

C. OPTIMUM DIVERSITY COMBINERS

Recent work on optimum multipath receivers (1, 2) has led to a simple solution of the following general problem in diversity combination:

A signal x(t) is transmitted through a set of random linear time-variant channels $A_1, A_2, \ldots, A_r$, and is further corrupted by additive noises $n_1, n_2, \ldots, n_r$ (as shown in Fig. XIV-3) before being available to the receiver as signals $y_1(t), y_2(t), \ldots, y_r(t)$.

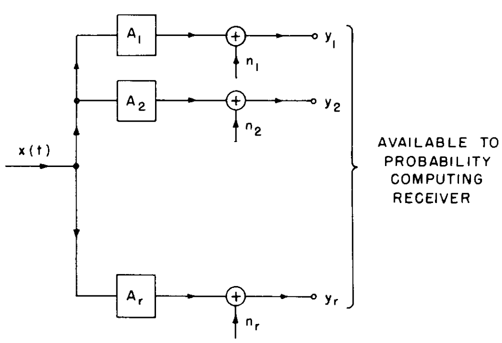

Fig. XIV-3. The diversity channel.

The receiver then has to compute the a posteriori probability $p(x(t)\,|\,y_1(t), y_2(t), \ldots, y_r(t))$.

The noises n_i are assumed to have Gaussian statistics that are known to the receiver. The noises need not be mutually independent, but they are assumed to be independent of the statistics of the channels, A_i.

For the channels A_i we can assume that (a) the statistics are Gaussian and known or (b) the statistics are arbitrary, but the first- and second-order statistics (the means and the covariances) are known. For case (b), we have obtained a solution that is valid only for weak (or threshold) signals.

The solution depends on the following result which we derived in reference 2. (It can also be simply deduced by using some of the results of Sec. XIV-A.) For a single

*This work was supported in part by the Office of Naval Research under Contract Nonr-1841(57).

channel, A_1, we can write in matrix notation (ref. 2 and Eqs. 7, 13, 14, 16 of Sec. XIV-A):

$$y_1 = Xa_1 + n_1 = Xa_{1r} + X\bar{a}_1 + n_1 \tag{1}$$

where $\underline{y}_1$ and $\underline{n}_1$ are column vectors comprised of sample values of the received $y_1(t)$ and $n_1(t)$ waveforms, $\underline{a}_1$ is a column vector whose elements depend on the values of the channel impulse response $a_1(t,\tau)$, and $\underline{X}$ is a matrix – not usually one-dimensional – with elements dependent on the signal x(t). We write $\underline{a}_{1r}$ and $\bar{\underline{a}}_1$ to denote the random and the mean components of $\underline{a}_1$. Then, for the covariances we have

$$\underline{\Phi}_{y_1y_1} = \underline{X}\underline{\Phi}_{a_1a_1}\underline{X}_t + \underline{\Phi}_{n_1n_1} \tag{2}$$

where

$$\underline{\Phi}_{y_1y_1} = \overline{\underline{y}_1\underline{y}_{1t}}, \quad \underline{\Phi}_{a_1a_1} = \overline{(\underline{a}_1-\bar{\underline{a}}_1)(\underline{a}_1-\bar{\underline{a}}_1)_t}, \quad \underline{\Phi}_{n_1n_1} = \overline{\underline{n}_1\underline{n}_{1t}} \tag{3}$$

in which the subscript t denotes the transposition of the matrix.

Now, assuming Gaussian statistics, the receiver that essentially calculates $p(\underline{x}|\underline{y})$ (certain "bias" terms are omitted here) has to compute the quantity

$$\Lambda = 2(\underline{X}\bar{\underline{a}}_1)_t\, \underline{\Phi}^{-1}_{n_1n_1}\underline{y}_1 + (\underline{y}_1-\underline{X}\bar{\underline{a}}_1)_t\, \underline{\Phi}^{-1}_{n_1n_1}\underline{X}_1\underline{\Phi}_{a_1a_1}\left(\underline{I} + \underline{X}_t\underline{\Phi}^{-1}_{n_1n_1}\underline{X}\underline{\Phi}_{a_1a_1}\right)^{-1}\underline{X}_{1t}$$
$$\cdot\ \underline{\Phi}^{-1}_{n_1n_1}(\underline{y}_1-\underline{X}\bar{\underline{a}}_1) \tag{4}$$

This is true, provided that $\underline{\Phi}_{n_1n_1}$ exists; no assumptions as to the nonsingularity of $\underline{X}$ or $\underline{\Phi}_{a_1a_1}$ are required. In certain cases – for example, nonselective slow fading – $\Phi_{a_1a_1}$, as defined, is singular (2, 3). With a slight modification of the definitions of $\underline{X}$ and $\underline{a}$, we can define the covariance so that $\underline{\Phi}_{AA}$ is always nonsingular (2, 3). Then we can write

$$\Lambda = 2(\underline{X}\bar{\underline{a}}_1)_t\, \underline{\Phi}^{-1}_{n_1n_1}\underline{y} + (\underline{y}_1-\underline{X}\bar{\underline{a}}_1)_t\, \underline{\Phi}^{-1}_{n_1n_1}\underline{X}_1\left(\underline{\Phi}^{-1}_{a_1a_1} + \underline{X}_t\underline{\Phi}^{-1}_{n_1n_1}\underline{X}\right)^{-1}\underline{X}_{1t}\underline{\Phi}^{-1}_{n_1n_1}(\underline{y}_1-\underline{X}\bar{\underline{a}}_1) \tag{5}$$

Physical interpretations of the action of this receiver – for example, an estimator-correlator feature – are discussed in references 1 and 2.

To adapt this solution to the general diversity problem we write the set of equations (cf. Fig. XIV-3):

$$\left.\begin{aligned}\underline{y}_1 &= \underline{X}\underline{a}_1 + \underline{n}_1\\ \underline{y}_2 &= \underline{X}\underline{a}_2 + \underline{n}_2\\ &\vdots\\ \underline{y}_r &= \underline{X}\underline{a}_r + \underline{n}_r\end{aligned}\right\} \tag{6}$$

as a single matrix equation

$$\underline{y} = \underline{X}\underline{a} + \underline{n} \tag{7}$$

where $\underline{y}$, $\underline{a}$, $\underline{n}$ are (super) column vectors composed of the separate column vectors $\underline{y}_1$, $\underline{y}_2, \ldots, \underline{y}_r, \underline{a}_1, \underline{a}_2, \ldots, \underline{a}_r, \underline{n}_1, \underline{n}_2, \ldots, \underline{n}_r$. Thus, for example, $\underline{y}_t = \lfloor \underline{y}_{1t} \vdots \underline{y}_{2t} \vdots \cdots \vdots \underline{y}_{rt} \rfloor$. The covariances are defined for y, a, n as in Eq. 3,

$$\underline{\Phi}_{yy} = \overline{\underline{y}\underline{y}_t} = \underline{X}\underline{\Phi}_{aa}\underline{X}_t + \underline{\Phi}_{nn} \tag{8}$$

Now the optimum probability computer is of the same form as Eq. 4 or Eq. 5, except that the subscript 1 is dropped.

If we do not assume Gaussian statistics for the A_i, we have the approximate solution for weak signals:

$$\Lambda = 2(\underline{X}\underline{\bar{a}})_t\, y + (\underline{y}-\underline{X}\underline{\bar{a}})_t\, \underline{\Phi}_{nn}^{-1}\underline{X}\underline{\Phi}_{aa}\underline{X}_t\underline{\Phi}_{nn}^{-1}(\underline{y}-\underline{X}\underline{\bar{a}}) \tag{9}$$

This solution is often a useful approximation even in the case of Gaussian statistics.

We have left the solution in discrete form, but the continuous analogs, obtained formally by letting the sampling grid become infinitely dense, are easily deduced.

Thus almost in one fell swoop, the solution of the diversity problem for correlated gaussian noises and arbitrary (gaussian) multipath channels in each diversity branch, is obtained. It is readily seen that the combination rules of Kahn (4), Brennan (5), Law (6), Pierce (7), and others are included in the formulation given above. More details, and explicit solutions for some cases, will be described later.

T. Kailath

References

1. T. Kailath, Correlation detection of signals perturbed by a random channel, Trans. IRE, IT-6, No. 3, 361-366 (June 1960).

2. T. Kailath, Optimum receivers for randomly varying channels, Proc. Fourth London Symposium on Information Theory, August 1960, London (to be published).

3. See Section XIV-A of this report.

4. L. Kahn, Ratio squarer, Proc. IRE 42, 1704 (November 1954).

5. D. G. Brennan, On the maximum signal-to-noise ratio realizable from several noisy signals, Proc. IRE 43, 1530 (October 1955).

6. H. B. Law, The detectability of fading radiotelegraph signals in noise, Proc. IEE, Part B, 104, 130-140 (March 1957).

7. J. N. Pierce, Theoretical diversity improvement in frequency-shift keying, Proc. IRE 46, 903-910 (May 1958).