TABLE I

Covariance matrix for letter A

1.034	1.281	0.351	−0.293	0.098	0.301	0.141	1.336
	1.967	0.664	−0.219	0.259	0.556	0.276	2.094
		7.138	1.192	2.726	1.116	0.678	2.097
			2.269	1.367	0.146	0.201	−0.308
				5.727	1.280	0.933	2.107
					2.941	1.949	2.197
						1.577	1.229
							6.606

Mean vector for letter A

7.825	6.750	5.835	8.525	6.615	7.065	7.865	4.435

Covariance matrix letter for B

4.792	4.417	4.244	2.406	1.798	0.790	0.785	2.993
	5.074	4.636	2.798	1.824	0.639	0.644	2.799
		5.428	3.224	2.111	0.903	1.131	2.943
			5.287	3.006	1.326	1.897	2.648
				3.574	2.229	2.471	1.915
					4.008	2.405	1.106
						4.507	1.727
							3.972

Mean vector for letter B

5.760	5.715	5.705	4.150	6.225	6.960	6.750	3.910

TABLE II
COMPARISON OF RESULTS OF FEATURE ORDERING PROCEDURES

	True divergence					Marill–Green divergence order				
	Feature ordering	Error % A	B	Avg.	Divergence	Feature ordering	Error % A	B	Avg.	Divergence
1	4	9.8	15.3	12.5	6.5	4	9.8	15.3	12.5	6.5
2	14	5.3	14.8	10.0	11.9	14	5.3	14.8	10.0	11.9
3	124	3.6	7.5	5.5	19.0	124	3.6	7.5	5.5	19.0
4	1234	2.7	4.2	3.4	23.6	1247	2.9	7.7	5.3	19.9
5	12345	2.3	4.1	3.2	26.0	12467	1.8	5.9	3.8	25.2
6	123467	1.6	4.4	3.0	29.8	123467	1.6	4.4	3.0	29.8
7	1234567	0.9	2.6	1.7	33.1	1234567	0.9	2.6	1.7	33.1
8	12345678	1.9	1.7	1.8	36.1	12345678	1.9	1.7	1.8	36.1

	Maximum divergence linear discriminant order					Results for both minimum expected error and approximation to maximum divergence linear discriminant functions				
	Feature ordering	Error % A	B	Avg.	Divergence	Feature ordering	Error % A	B	Avg.	Divergence
1	1	11.4	31.5	21.4	3.9	1	11.4	31.5	21.8	3.9
2	12	6.3	20.4	13.3	8.6	14	5.3	14.8	10.0	11.9
3	127	6.3	18.1	12.2	9.8	124	3.6	7.5	5.5	19.0
4	1247	2.9	7.7	5.3	19.9	1245	3.7	6.0	5.3	21.8
5	12467	1.8	5.9	3.8	25.2	12345	2.3	4.1	3.2	26.0
6	124567	2.0	4.2	3.1	28.6	123457	1.9	3.0	2.4	28.0
7	1245678	2.0	2.7	2.3	31.6	1234567	0.9	2.6	1.7	33.1
8	12345678	1.9	1.7	1.8	36.1	12345678	1.9	1.7	1.8	36.1

Linear discriminants
- maximum divergence = (0.77, −0.43, −0.02, 0.25, −0.552, −0.21, 0.31, 0.038)
- approximation to maximum divergence = (0.675, −0.471, −0.154, 0.508, −0.177, −0.035, 0.081, −0.012)
- minimum error = (0.709, −0.490, −0.115, 0.459, −0.135, −0.054, 0.099, 0.025)

3) *Approximation to maximum divergence linear discriminant function:* For this test the covariance matrices given in Table I were averaged and (2) and (3) solved. These equations are shown in Kullback[2] to determine the linear discriminant maximizing the divergence when the covariance matrices are equal.

$$J = \delta'\Sigma^{-1}\delta \tag{2}$$

$$\gamma = \Sigma^{-1}\delta \tag{3}$$

where Σ is the average covariance matrix, δ is the difference in the means of the two classes, and γ is the vector of coefficients of the linear discriminant. This procedure is by far the simplest to implement, since it requires only the averaging of the covariance matrices and two relatively simple matrix multiplications. Nevertheless, quite reasonable results were obtained.

4) *Maximum divergence linear discriminant function:* For this test no averaging was performed on the covariance matrices. Equations (4) and (5) which are derived in Kullback were solved iteratively and the features ordered as to the magnitude of their respective coefficients.

$$\Sigma_1\gamma - \lambda\Sigma_2\gamma = \delta \tag{4}$$

where

$$\lambda = \frac{\gamma'\Sigma_2\gamma[(\gamma'\Sigma_2\gamma)^2 - (\gamma'\Sigma_1\gamma)^2 + (\gamma'\delta)^2(\gamma'\Sigma_1\gamma)]}{\gamma'\Sigma_1\gamma[(\gamma'\Sigma_2\gamma)^2 - (\gamma'\Sigma_1\gamma)^2 - (\gamma'\delta)^2(\gamma'\Sigma_2\gamma)]}. \tag{5}$$

The need to solve these equations made this procedure very difficult to implement. A large number of solutions exist, indicating that the divergence was not a unimodal function of the vector of coefficients γ. The various possible solutions were found by initiating the iterations with different values of λ, and the divergence for each solution was calculated. In this way the solution producing the largest divergence was obtained.

5) *Minimum expected error linear discriminant function:* For this test (6) and (7) given in Kullback were solved iteratively, and the linear discriminant yielding the minimum expected error was thereby obtained.

$$\lambda = \frac{q_\alpha\gamma'\Sigma_1\gamma}{q_\beta\gamma'\Sigma_2\gamma} \tag{6}$$

$$[\Sigma_1 + \lambda\Sigma_2]\gamma = \delta \tag{7}$$

where

$$q_\beta = \frac{\gamma'(\mu_1 - \mu_2) - \gamma'\Sigma_1\gamma q_\alpha}{\gamma'\Sigma_2\gamma}.$$

The quantities q_α and q_β are such that $\alpha = \mathfrak{F}(q_\alpha)$ and $\beta = \mathfrak{F}(q_\beta)$ where $\mathfrak{F}$ is the standardized normal distribution function, and α and β are errors of the first and second kind, respectively. To obtain the minimum total expected error, the linear discriminant which minimizes β for any fixed α was found for various values of α. The one producing the minimum total error rate was then the desired optimum linear discriminant. The feature selection was again performed by ordering the magnitude of the respective coefficients. It should be noted that the feature ordering resulting from this procedure was identical to that obtained from Procedure 3. This was due to the fact that the linear discriminant obtained in each case was very nearly the same.

G. P. CARDILLO[3]
K. S. FU
School of Elec. Engrg.
Purdue University
Lafayette, Ind.

[3] Now with General Research Corp., Santa Barbara Calif.

[2] S. Kullback, *Information Theory and Statistics*. New York: Wiley, 1963.

On Power Spectrum Identification Methods

In a recent paper by Tretter and Steiglitz,[1] a method for identifying power spectral density functions was presented. The main problem discussed in that paper was how to achieve parameter identification when the spectral density functions contain zeros. A search technique was suggested to carry out the solution. The purpose of this correspondence is to point out that the same problem has also been studied in a recent paper by the author.[2] However, the results are completely different. It is felt that a comparison of the two results, along with the classical results by Whittle,[5] might be of

Manuscript received July 24, 1967.

some interest to the readers. For the convenience of discussion, the author's results will be briefly introduced using the same notations as in Tretter and Steiglitz.

The discrete random process x which has a spectral density function

$$\phi_{xx}(z) = \beta^2 \frac{N(z)N(z^{-1})}{D(z)D^{-1})}$$

where

$$N(z) = \sum_{n=0}^{K} a_n z^{-n}, \qquad D(z) = \sum_{n=0}^{L} b_n z^{-n},$$

and $a_0 = b_0 = 1$, can be generated by the system

$$\sum_{n=0}^{L} b_n x(n) = \beta^2 \sum_{n=0}^{k} a_n \omega(n) \tag{1}$$

where $\omega(n)$ is a white random process of unity constant power spectrum. Now autocorrelate both sides of (1)

$$\begin{aligned} \sum_{j=0}^{L} b_j \left[\sum_{i=0}^{L} b_i R_{xx}(k+j-i) \right] &= \beta^2 \sum_{i=0}^{K-k} a_i a_{i+k}, \qquad 0 \leq k \leq K \\ &= 0, \qquad\qquad k > K \end{aligned} \tag{2}$$

where $R_{xx}(k)$ is the autocorrelation function of x. Equation (2) implies that for $k > K$

$$\sum_{i=0}^{K} b_i R_{xx}(k+j-i) = 0, \qquad j = 0, 1, \cdots, L. \tag{3}$$

Take the first L equations of (3) for $k = K+1$, and replacing $R_{xx}(i)$ by the mean logged products f_i, the solutions for b_n can be easily written as

$$b = -F^{-1}f \tag{4}$$

where b is column vector $\{b_i; i = 1, \cdots, L\}$, F is the matrix $\{f_{|i-j+K|}; i, j = 1, \cdots, L\}$, and f is the column vector $\{f_{i+K}; i = 1, 2, \cdots, L\}$. Now pass the signal x through the filter $D(z) = \sum_{n=0}^{L} b_n z^{-n}$ in which the b_n are known; the output signal y is a moving-average-scheme time series the spectrum of which is an all-zero function. Using (2) and the method by Wold,[3] the solutions for a_n are identified from the relationship[2]

$$\prod_{i=1}^{K} (1 - u_i z^{-1}) = \sum_{i=0}^{K} a_n z^{-n} \tag{5}$$

where u_i is a root of $N(z)$ and is given by one of the two values (one is the inverse of the other) of

$$\frac{r}{2} \pm \sqrt{\frac{r^2}{4} - 1} \tag{6}$$

r being a root of

$$\sum_{i=0}^{K} R_{yy}(K-i) r^{K-i} = 0$$

in which $R_{yy}(k)$, the autocorrelation function of y, is defined by the left-hand side of (2). Finally the solution for β^2 is

$$\beta^2 = \frac{R_{yy}(0)}{\sum_{n=0}^{K} a_n^2}. \tag{7}$$

It is interesting to see that when the order of $N(z)$ is zero, (4) and (7) become identical to the all-pole spectrum estimates.[1],[4] However, the above results are equally applicable to non-Gaussian signals.

A few comments on all the methods are in order:

1) The author's method shows that it is possible to obtain parameter estimates analytically and less complex. Hence computational approach appears undesirable and unnecessary.

2) The condition that the roots of $N(z)$ be situated inside the unit circle is required by Tretter and Steiglitz's method but not by the author's method. An appropriate value (greater than one or less than one) can be chosen from (6) for each u_i on prior ground. Therefore, the method is more general.

3) Since the residual function R given by Tretter and Steiglitz is a highly nonlinear function of the parameters a_n and b_n, the surface of R is usually multimodal. Therefore, it is quite difficult to see that the procedure of first minimizing R with respect to a_n and then calculating b_n would always lead to the true minimum value of R. It appears that the initial conditions of b_n in the search procedure would, in general, affect the results. Unless convergence of the proposed computational solution can be guaranteed, a higher-dimensional $(K+L)$ multimodal search would be necessary to minimize R. This, of course, will be quite a difficult computational job.

4) It is noted that a similar nonlinear minimization problem also arose in a method proposed by Whittle.[5] His method calls for the minimization of a function

$$\sum C_k R_{xx}(k) = \min \tag{8}$$

where C_k is the coefficient of z^k in the Laurent expansion of the function

$$\frac{D(z)D(z^{-1})}{N(z)N(z^{-1})} = \sum C_k z^k. \tag{9}$$

Clearly, C_k is a nonlinear function of a_n and b_n. Note that the roots of $N(z)$ also have to be less than unity in magnitudes for (9) to be a convergent series. When the order of $N(z)$ becomes zero, (8) yields the same all-pole results given by (4).

Tien C. Hsia
Dept. of Elec. Engrg.
University of California
Davis, Calif. 95616

References

[1] S. A. Tretter and K. Steiglitz, "Power-spectrum identification in terms of rational models," *IEEE Trans. Automatic Control (Short Papers)*, vol. AC-12, pp. 185–188, April 1967.
[2] T. C. Hsia and D. Landgrebe, "On a method for estimating power spectra," *IEEE Trans. Instrumentation and Measurement*, vol. IM-16, pp. 255–257, September 1967.
[3] H. Wold, *A Study in the Analysis of Stationary Time Series*. Stockholm: Alxiqvist and Wikell, 1954.
[4] K. Steiglitz, "Power spectrum identification for adaptive systems," *IEEE Trans. Applications and Industry*, vol. 83, pp. 195–197, May 1964.
[5] Wold,[3] see P. Whittle, Appendix 2.

Authors' Reply[1]

Although our method is more complex computationally, the resulting parameters are optimum in the maximum likelihood sense for Gaussian signals. Therefore, under relatively weak conditions on the spectral density, these estimates are consistent, asymptotically unbiased, and asymptotically minimum variance. It has been pointed out by Whittle,[5] p. 213, that the variances of the coefficients of $N(z)$ estimated according to the maximum likelihood or equivalently minimum residual criterion can be significantly smaller than that of the coefficients estimated using Wold's method. Hsia and Landgrebe have provided no measure of the precision of their estimates.

With respect to Comment 2 of Hsia it should be observed that the reciprocals of the roots of $N(z)$ must be roots of $N(z^{-1})$. Consequently there is no loss of generality in choosing those roots lying inside the unit circle for $N(z)$.

We agree that minimizing a nonlinear function of several viariables is not easy and that care must be taken to insure finding an absolute rather than a local minimum. This was not found to be a major problem in simulations. The required computations were relatively easily performed by an IBM 7094 computer.

With respect to Comment 4 of Hsia it should be observed that Whittle's criterion is equivalent to the minimum residual criterion. If a finite record of a discrete stochastic process $x(k)$ with spectral density $\Phi_{xx}(z)$ is passed through a filter $D(z)/N(z)$ resulting in the output $y(k)$, the average square value of $y(k)$ is

$$\begin{aligned} \frac{1}{M}\sum_{k=1}^{M} y^2(k) &= \frac{1}{2\pi j M} \oint \frac{D(z)D(z^{-1})}{N(z)N(z^{-1})} X(z)X(z^{-1}) \frac{dz}{z} \\ &= \sum_{k=-\infty}^{\infty} c_k R_{xx}(k) \end{aligned}$$

where

$$\frac{D(z)D(z^{-1})}{N(z)N(z^{-1})} = \sum_{k=-\infty}^{\infty} c_k z^{-k}$$

$X(z)$ is the z transform of the finite sample of $x(k)$, and where end effects have been neglected.

S. A. Tretter
Dept. of Elec. Engrg.
University of Maryland
College Park, Md. 20740
Kenneth Steiglitz
Dept. of Elec. Engrg.
Princeton University
Princeton, N. J. 08540

[1] Manuscript received October 18, 1967.

Spectral Density of the Output of Aperiodic Samplers

The purpose of this correspondence is to present a simple method of evaluating the spectral density of the output of an aperiodic sampler or gate with stationary random inputs. It is assumed that the sampling

Manuscript received May 4, 1967; revised July 3, 1967.