Power Spectrum Identification for Adaptive Systems

KENNETH STEIGLITZ

Summary: This paper describes a method for identifying the parameters of a class of power spectra. In contrast to conventional methods of spectral analysis, the method assumes a particular form for the power spectrum and gives direct estimates of unknown parameters. Thus the method is faster than ordinary spectral analysis and can be easily implemented by a digital computer for use in an adaptive loop.

The derivation of the estimates assumes that the power spectrum has no zeros, and is based on well-known results in the theory of autoregressive schemes. Some ways of extending the results to the case where zeros are present in the spectrum are suggested. The method can also be used as a prewhitening technique in conjunction with ordinary spectral analysis.

A KEY PROBLEM in the design of self-optimizing and adaptive systems is the real-time estimation of unknown structural parameters of a stochastic signal. In particular, there is a need for methods of identifying parameters in the power spectral density of system signals. This makes possible the design of signal-adaptive and plant-adaptive controllers when signal or system parameters are slowly changing.

The usual methods for measuring power spectral density, such as described by Blackman and Tukey, do not involve any assumptions about the form of the power spectrum and do not give direct estimates of structural parameters.[1] This paper presents a simple method for estimating the parameters of a class of sampled power spectra. In contrast to conventional methods of spectral analysis, the method assumes a particular form for the power spectrum and gives direct estimates of parameters. For these reasons, the method described in this paper requires less time to obtain tolerably accurate results, and is better suited for use in an adaptive loop.

Paper 63-145, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the IEEE Winter General Meeting, New York, N. Y., January 27–February 1, 1963. Manuscript submitted November 5, 1962; made available for printing September 19, 1963.

KENNETH STEIGLITZ is with Princeton University, Princeton, N. J.

The author wishes to thank Professor S. S. L. Chang of New York University for his valuable comments during the course of this work.

This work was part of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Engineering Science at New York University. The research was sponsored partly by the National Science Foundation and partly by the Air Force Office of Scientific Research under Grant No. *AF-AFOSR-62-321*.

Nomenclature

x_n = discrete-time signal whose power spectrum is of interest
N = number of points of x_n that have been observed
$\varphi_{xx}(n)$ = autocorrelation function of signal x_n
$\Phi_{xx}(z)$ = power spectral density of x_n
f_j = mean-lagged product of lag j of the signal x_n
$D(z)$ = assumed denominator polynominal of $\Phi_{xx}(z)$; also a digital filter
p = assumed order of $D(z)$
α_i = assumed coefficient of z^{-i} in $D(z)$, $\alpha_0=1$
β^2 = assumed constant multiplier of $\Phi_{xx}(z)$
$\hat{\alpha}_i$ = estimated value of α_i
$\hat{\beta}^2$ = estimated value of β^2
y_n = signal obtained by passing x_n through digital filter $D(z)$

Statement of Problem

It is assumed that the power spectrum of a discrete-time signal $\{x_n\}_{n=-\infty}^{\infty}$ is of interest and that a finite length of this signal, say

$$x_1, x_2, \ldots, x_N$$

has been observed. The aliasing problem when the x_n are samples of a continuous signal will not be considered here. Instead, the problem will be formulated in terms of the sampled power spectrum

$$\Phi_{xx}(z)=\sum_{n=-\infty}^{\infty}\varphi_{xx}(n)z^{-n} \tag{1}$$

where

$$\varphi_{xx}(n)=Ex_ix_{i+n} \tag{2}$$

is the autocorrelation function of lag n.

Furthermore, the following assumptions will be made:

1. The number of points N of the sample is large so that end effects can be neglected.
2. The signal is normally distributed with zero mean and is stationary and ergodic so that equation 2 can represent either an ensemble or time average.
3. The signal has a power spectral density which can be closely represented by the following equation:

$$\Phi_{xx}(z)=\frac{\beta^2}{D(z)D(z^{-1})} \tag{3}$$

where

$$D(z)=1+\alpha_1z^{-1}+\alpha_2z^{-2}+\ldots+\alpha_pz^{-p} \tag{4}$$

has all of its zeros inside the unit circle in the z-plane.

Assumption 3—that the unknown spectrum has no zeros—is rather restrictive. The possibility of relaxing this assumption will be discussed later.

The problem, then, is to estimate the parameters $\alpha_1, \alpha_2, \ldots, \alpha_p$, and β^2, given N observed points of the signal.

Most Likely Estimates

The problem as just stated is equivalent to a well-known problem in mathematical statistics: that of estimating the coefficients of an autoregressive scheme. The solution to this latter problem can be found in the literature; a good discussion is given by Hannan, for example.[2] The solution given here will be essentially the same as Hannan's, except for the fact that the argument will be in terms of power spectra.

Define a new signal y_i by passing x_i through the digital filter $D(z)$. That is, put

$$y_i=x_i+\alpha_1x_{i-1}+\alpha_2x_{i-2}+\ldots+\alpha_px_{i-p} \tag{5}$$

or, in z-transform notation,

$$Y(z)=D(z)X(z) \tag{6}$$

The stochastic variable y_t is normally distributed. Furthermore, its power spectral density is

$$\Phi_{yy}(z)=D(z)D(z^{-1})\Phi_{xx}(z)=\beta^2$$

so that signal y_t is Gaussian-distributed white noise with a mean square value β^2. The joint probability density function of the observed sample $(y_1, y_2, \ldots, y_N)$ is then

$$p(y_1, y_2, \ldots, y_N)=\frac{1}{(2\pi)^{N/2}\beta^N}\exp\left(-\frac{1}{2\beta^2}\sum_{t=1}^{N} y_t^2\right) \tag{7}$$

The maximum likelihood estimates of the unknown parameters, denoted by $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_p$, and $\hat{\beta}^2$, are obtained by maximizing this probability. Thus, the following set of equations must be solved:

$$\frac{\partial \log p(y_1,y_2,\ldots,y_N)}{\partial \alpha_j}=0,\ j=1,2,\ldots,p \tag{8}$$

and

$$\frac{\partial \log p(y_1,y_2,\ldots,y_N)}{\partial \beta^2}=0 \tag{9}$$

When equation 5 is substituted in equation 7 and the indicated operations are carried out, the most likely estimated result is

$$\begin{bmatrix}\hat{\alpha}_1\\ \hat{\alpha}_2\\ \hat{\alpha}_3\\ \cdot\\ \cdot\\ \hat{\alpha}_p\end{bmatrix}=-\begin{bmatrix} f_0 & f_1 & f_2 & \cdots & f_{p-1}\\ f_1 & f_0 & f_1 & \cdots & f_{p-2}\\ f_2 & f_1 & f_0 & \cdots & f_{p-3}\\ \cdot & & & \cdot & \cdot\\ \cdot & & & \cdot & \cdot\\ f_{p-1} & \cdots & & & f_0\end{bmatrix}^{-1}\begin{bmatrix}f_1\\ f_2\\ f_3\\ \cdot\\ \cdot\\ f_p\end{bmatrix} \tag{10}$$

and

$$\hat{\beta}^2=\sum_{i,j=0}^{p}\hat{\alpha}_i\hat{\alpha}_j f_{|i-j|}=f_0+\hat{\alpha}_1 f_1+\hat{\alpha}_2 f_2+\ldots+\hat{\alpha}_p f_p \tag{11}$$

where the f_j are the mean lagged products

$$f_j=\frac{1}{N-j}\sum_{t=1}^{N-j} x_t x_{t+j},\quad j\geq 0 \tag{12}$$

The mean lagged products are unbiased estimates of the autocorrelation function $\varphi_{xx}(j)$ and represent the bulk of the computations involved in almost any spectral measurement problem.

In summary, then, the following computations are performed:

1. From the N sample points of the signal, the mean lagged products $f_0, f_1, \ldots, f_p$ are calculated in accordance with equation 12.
2. The $p\times p$ matrix $(f_{|i-j|};\ i,j=1, \ldots, p)$ is formed and inverted.
3. $\hat{\alpha}_j$ $(j=1, \ldots, p)$ are calculated from equation 10.
4. $\hat{\beta}^2$ is calculated from equation 11.

These computational steps are shown diagrammatically in Fig. 1.

Variability of Estimates

If this identification method is to be used in an adaptive loop, some knowledge is required about the accuracy of the estimates for a given N. It can be shown[2] that the vector $\hat{\alpha}-\alpha$ defined by

$$\hat{\alpha}-\alpha=\begin{bmatrix}\hat{\alpha}_1-\alpha_1\\ \hat{\alpha}_2-\alpha_2\\ \cdot\\ \cdot\\ \cdot\\ \hat{\alpha}_p-\alpha_p\end{bmatrix}$$

is asymptotically normally distributed with zero mean and covariance matrix

$$\frac{\beta^2}{N-p}\begin{bmatrix}\varphi_0 & \varphi_1 & \varphi_2 & \cdots & \varphi_{p-1}\\ \varphi_1 & \varphi_0 & \varphi_1 & \cdots & \varphi_{p-2}\\ \varphi_2 & \varphi_1 & \varphi_0 & \cdots & \varphi_{p-3}\\ \cdot & & & \cdot & \cdot\\ \cdot & & & \cdot & \cdot\\ \cdot & & & \cdot & \cdot\\ \varphi_{p-1} & \cdots & & & \varphi_0\end{bmatrix}^{-1} \tag{13}$$

This can be estimated conveniently by

$$\frac{\hat{\beta}^2}{N-p}\begin{bmatrix}f_0 & f_1 & f_2 & \cdots & f_{p-1}\\ f_1 & f_0 & f_1 & \cdots & f_{p-2}\\ f_2 & f_1 & f_0 & \cdots & f_{p-3}\\ \cdot & & & \cdot & \cdot\\ \cdot & & & \cdot & \cdot\\ \cdot & & & \cdot & \cdot\\ f_{p-1} & \cdots & & & f_0\end{bmatrix}^{-1} \tag{14}$$

which does not use any quantities which have not already been calculated.

The distribution of the estimate $\hat{\beta}^2$ is difficult to calculate since it is a more complicated function of the f_j's. It is easy, however, to derive distribution of

$$\frac{1}{N}\sum_{t=1}^{N} y_t^2=\sum_{i,j=0}^{p}\alpha_i\alpha_j f_{|i-j|} \tag{15}$$

and this will give some (optimistic) indication of the variability of $\hat{\beta}^2$. With this in mind consider the random variable

$$\sum_{t=1}^{N}(y_t/\beta)^2$$

This is the sum of squares of independent, normally distributed random variables whose means are zero and whose variances are one. This random variable is χ^2-distributed with N degrees of freedom. Cramér[3] shows that with increasing N the χ^2-distribution becomes asymptotically normal with mean N and standard deviation $\sqrt{2N}$. Therefore, the random variable of equation 15 is asymptotically normally distributed with mean β^2 and standard deviation $\sqrt{2/N}\beta^2$ and hence $\sqrt{2/N}\hat{\beta}^2$ can be used as a low estimate of the standard deviation of $\hat{\beta}^2$.

Extension to Spectra With Zeros

As mentioned previously, the assumption that the unknown spectrum does not have any zeros is rather restrictive. It would, therefore, be desirable to extend this method so that it is applicable to more general forms. Unfortunately there seems to be an essential difficulty in doing so, and the solutions to the more general problems become very involved and are not suited for real-time computation. There are some general situations where something can be done, however, and these will now be discussed.

Suppose that the signal of interest has an unknown power spectrum of the form

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

and that the locations of the zeros are known (at least approximately). Then the signal can be prefiltered by a digital filter $1/N(z)$ (or an equivalent analog filter). The resultant signal will then be of the requisite form and the method described in this paper can be used to determine the pole locations and β^2.

As another example, assume that the signal of interest, x_n, is the sum of two independent signals, one of whichh as a

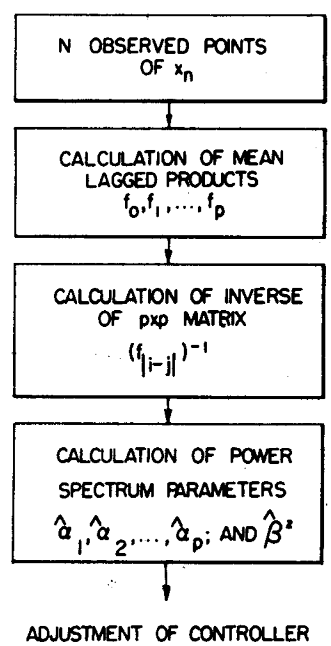

Fig. 1. Estimation of power spectrum parameters

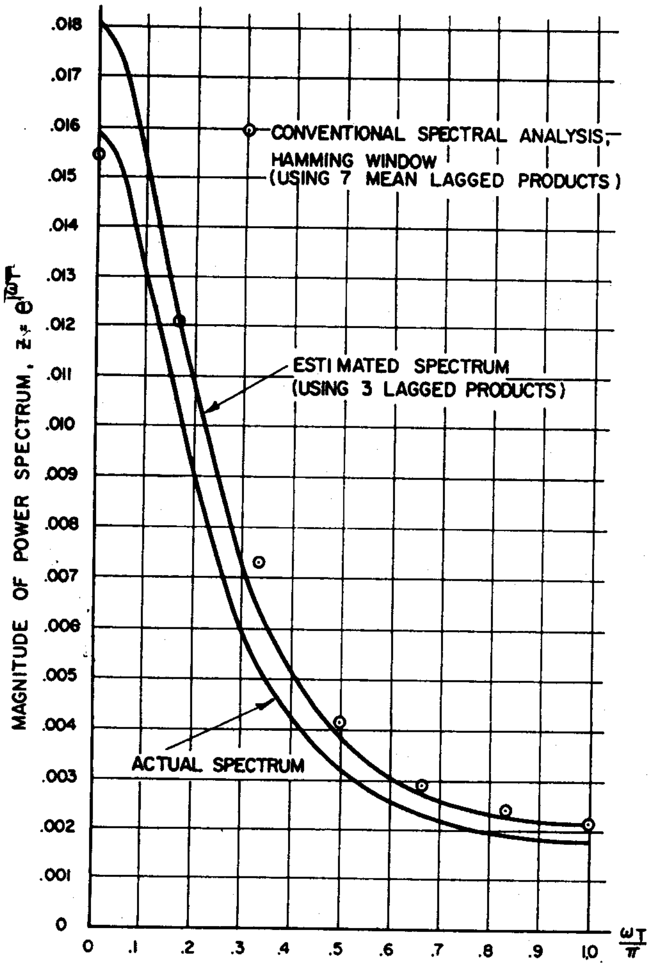

Fig. 2. Comparison of actual and estimated power spectra

known power spectrum (such as white noise of a given amplitude), and the other of which has only poles in its power spectrum. That is, assume

$$\Phi_{xx}(z) = \Phi_{xx}^*(z) + \frac{\beta^2}{D(z)D(z^{-1})}$$

The autocorrelation function of the signal is, then, the sum of known and unknown components:

$$\varphi_{xx}(n) = \varphi_{xx}^*(n) + \varphi_{xx}'(n)$$

Subtracting known components from the computed f_n yields mean lagged products

$$f_0 - \varphi_{xx}^*(0), f_1 - \varphi_{xx}^*(1), \ldots, f_p - \varphi_{xx}^*(p)$$

can that be used to estimate $D(z)$ and β^2.

Other situations suggest themselves. Some pole locations may be known in advance, for example. These poles can be removed before analysis by a digital or analog filter. Alternatively, the maximum likelihood equations 8 and 9 can be reworked.

Conclusions

The following method of self-optimizing control is proposed in this paper: A controller is designed whose optimum operation depends on the knowledge of the parameters $\alpha_1, \alpha_2, \ldots, \alpha_p$, and β^2 of the sampled power spectrum of some system signal. From a record of this signal of length N the estimates $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_p$, and $\hat{\beta}^2$ are periodically calculated by a digital computer and used to adjust the controller. In a particular application, the choice of N is a critical and difficult problem. N must be chosen large enough so that the estimates of the power spectrum parameters are accurate enough to be useful. On the other hand, N should not be so large that the system reacts to obsolete information.

The method described in this paper may also be used as a first step in a conventional spectral analysis. After $D(z)$ is estimated, the original signal can be passed through a filter $D(z)$ and subjected to further spectral analysis by conventional methods. If the form assumed for the spectrum was appropriate the output will be nearly white, and this method will amount to an "automatic" prewhitening technique which can be used in conjunction with conventional spectral analysis. (See Tukey for a discussion of prewhitening.[1,4])

Finally, it might be mentioned that the method described in this paper can be used with the adaptive information processing method described by Chang.[5]

Appendix

To demonstrate the method, a sequence of 210 independent normal random numbers was passed through the digital filter $1/(1-0.5z^{-1})$. The resultant time series then had a power spectrum

$$\frac{1/252}{(1-0.5z^{-1})(1-0.5z)}$$

Thus for this signal, assuming $p=2$,

$\alpha_1 = -0.5$

$\alpha_2 = 0.0$

$\beta^2 = 0.00397$

Mean lagged products were computed:

$f_0 = 0.00624$

$f_1 = 0.00307$

$f_2 = 0.00147$

and equation 10 used to give the estimates

$\hat{\alpha}_1 = -0.495$

$\hat{\alpha}_2 = 0.0070$

$\hat{\beta}^2 = 0.00473$

The estimated covariance matrix of the $\hat{\alpha}_j$ was calculated from equation 14:

$$\begin{bmatrix} 0.0048 & -0.0024 \\ -0.0024 & 0.0048 \end{bmatrix}$$

and it is seen that the $\hat{\alpha}_j$ are well within one standard deviation of the α_j. Optimistically estimated standard deviation of $\hat{\beta}^2$ is

$$\sqrt{2/N}\,\hat{\beta}^2 \approx (1/10)\hat{\beta}^2$$

so that 16% actual deviation is reasonable.

Fig. 2 shows plots of the actual and the estimated power spectrum. Also shown are the results of a conventional spectrum analysis using a Hamming window and seven mean lagged products.[1] Note that more than twice as many multiplications were required by the conventional method to produce similar accuracy, and that the results are not in a form that is suited for adaptive control.

References

1. The Measurement of Power Spectra, R. B. Blackman, J. W. Tukey. Dover Publications, Inc., New York, N. Y., 1959.

2. Time Series Analysis, E. J. Hannan. John Wiley and Sons, Inc., New York, N. Y., 1960.

3. Mathematical Methods of Statistics, H. Cramér. Princeton University Press, Princeton, N. J., 1946.

4. Emphasizing the Connection Between Analysis of Variance and Spectrum Analysis, J. W. Tukey. *Technometrics*, vol. 3, May 1961, pp. 191–219. Also, *Bell Telephone System Monograph 3906*, New York, N. Y.

5. Adaptive Information Processing, S. S. L. Chang. Presented at the 1962 WESCON Convention, Aug. 1962, Los Angeles, Calif.

A reprint from **APPLICATIONS AND INDUSTRY,** published by
The Institute of Electrical and Electronics Engineers, Inc.
Copyright 1964, and reprinted by permission of the copyright owner
The Institute assumes no responsibility for statements and opinions made by contributors. Printed in the United States of America

May 1964 issue