MAXIMUM ENTROPY SPECTRAL ANALYSIS

by

John Parker Burg

* * * * * * * * * * * * * *

ABSTRACT

The usual digital method of obtaining a power spectrum estimate from a measured autocorrelation function makes the assumption that the correlation function is zero at all lags for which no estimate is available and uses some treatment of the estimated lags to reduce the effect of truncation of the autocorrelation function. The method discussed in this paper instead retains all of the estimated lags without modification and uses a non-zero estimate for the lags not directly estimated. The particular estimation principle used is that the spectral estimate must be the most random or have the maximum entropy of any power spectrum which is consistent with the measured data.

This new analysis technique gives a much higher resolution spectral estimate than is obtained by conventional techniques with a very little increase in computing time. Comparisons will illustrate the relative performance.

MAXIMUM ENTROPY SPECTRAL ANALYSIS

In the digital measurement of the power spectrum of a stationary, band-limited time function, the two basic theoretical equations are those defining the autocorrelation function and the power spectrum. (See Slide 1.) These definitions involve infinite summations and thus must be modified for practical applications. This paper presents an unconventional method of power spectrum estimation from a finite number of autocorrelation lag values.

The conventional manner of treating the lack of data beyond the last available lag value of the autocorrelation function is to introduce the concept of a weighting function. There are many such weighting functions (Slide 2 shows the Hanning and Bartlett ones), but all have the common feature of multiplying the unmeasured lag values by zero. The particular weightings applied to the known lag values governs the shape of the corresponding frequency window which when convolved with the true power spectrum gives the estimated spectrum. The different approach used in this paper retains all of the known lag values without modification and estimates the values of the unknown autocorrelation lags.

In extending the known portion of the autocorrelation function, the basic autocorrelation function theorem shown in Slide 3 is used. For this N+1 by N+1 toeplitz matrix to be semi-positive definite, its determinant has to be non-negative. Thus, if the values of $\varphi(\tau)$, ($\tau = 0$ to N-1), are known, then $\varphi(N)$ must have a value which makes the determinant non-negative. The determinant is a quadratic function in $\varphi(N)$. Thus there are two values of $\varphi(N)$ which make the determinant equal to zero. These two values and all the values in between are the theoretically allowable ones. At times, the range of possible values is quite narrow so that one can predict the next value of the auto-correlation function with considerable accuracy. Such a case is shown in Slide 4 where knowing the first 20 values of the autocorrelation function allows one to state that $\varphi(20)$ lies in the range of 0.98901 ± 0.11099.

In selecting the predicted value of $\varphi(N)$, the most natural choice would be the midpoint of the allowable range. Then, using this value, one could then determine the allowable range of $\varphi(N+1)$ and again use the midpoint as its estimated value. This procedure could be continued on to extend the known autocorrelation function indefinitely in a mathematically acceptable fashion. Using this method, Slide 5 shows the estimated values of $\varphi(20)$ through $\varphi(40)$ for the example shown in Slide 4. As one can see, this extension method is much closer to the actual autocorrelation values than the zero extension commonly used.

Let us now consider the information theory concept of entropy. Surprisingly enough, this concept which relates to randomness or uncertainty, is strongly connected to the particular correlation extension method just discussed. Slide 6 presents the mathematical expression involved in defining the entropy of a Gaussian stationary time series. This expression is usually used in connection with the channel capacity of a communication system, where maximum entropy corresponds to maximum information transmission rate. A reasonable question is one of asking what power spectrum has the maximum entropy, knowing the first N values of its autocorrelation value. The solution to this problem turns out to be fairly simple and is given in Slide 7. The matrix equation may be recognized as that of designing a N+1 point prediction error filter $(1, \Gamma_2, \ldots, \Gamma_N, \Gamma_{N+1})$ with the mean square error being given by P_{N+1}. The top equation for P(f) can be seen to be P_{N+1}/W divided by the power response of the prediction error filter. W is of course the fold-over frequency of the band-limited time series which has a sampling period of $\Delta t = \frac{1}{2W}$.

Slide 8 illustrates the time series, X_n, being operated on by the optimum N point prediction filter, with output $\hat{X}_n$. The output, ϵ_n, of the N+1 point prediction error filter is just the difference between the true value of X_n and the predicted value $\hat{X}_n$. Since the prediction filter is optimum in the mean square error sense, the average product of ϵ_n with any of the N previous values of X_n is zero.

Let us consider the possibility that the average product of ϵ_n with all previous values of X_n is zero. In this case, the N point prediction filter is also the optimum infinitely long prediction filter. This would also mean that the predictability of X_n would not be increased by using more of the past of X_n. This possibility does in fact correspond to the maximum entropy assumption since an increase in the predictability of X_n would mean a decrease in its uncertainty or entropy. Continuing on, we can note that in this case where $\overline{\epsilon_n X_{n-s}} = 0$ for $S > 0$, we can note then ϵ_{n-1} is simply a linear combination of X_{n-s} for $S > 0$ and thus $\overline{\epsilon_{n-1} \epsilon_n} = 0$. In fact, ϵ_n is uncorrelated with all other points of the error trace and thus the output of the prediction error filter has a white power spectrum of total power P_{N+1} or of a density level of P_{N+1}/W. Slide 9 shows the simple relation between input and output power spectra. This viewpoint makes the equation for P(f) quite understandable.

Slide 10 shows a comparison of the maximum entropy technique with a conventional spectral estimation method using the Bartlett weighting function. The true spectrum is the same as referred to in Slides 4 and 5. The conventional estimate is very poor because zero is a very poor extension of the autocorrelation function whereas the maximum entropy extension was reasonably accurate. Besides having very high resolution, the new method has much lower "side lobes". This term is put into quotes since, for the maximum entropy technique, side lobe analysis has no meaning because the autocorrelation function is now infinitely long.

Slide 11 shows a maximum entropy spectrum and a Bartlett spectrum of some ambient seismic noise recorded on a horizontal seismometer at the Wichita Mountains Seismological Observatory. The autocorrelation function used had 100 lag values. Note that most of the peaks in the M.E. spectrum are also seen at a subdued level in the Bartlett spectrum. This is reasonable since if one convolves the M.E. spectrum with the Bartlett frequency window, one would obtain exactly the Bartlett spectrum. This statement is obvious if one considers the equivalent operation on the autocorrelation function.

Slide 12 is presented to show that from a theoretical viewpoint, the maximum entropy concept could be used as a different limiting process in defining a power spectrum.

THE AUTOCORRELATION FUNCTION OF A SAMPLED TIME SERIES, X_n, IS DEFINED AS

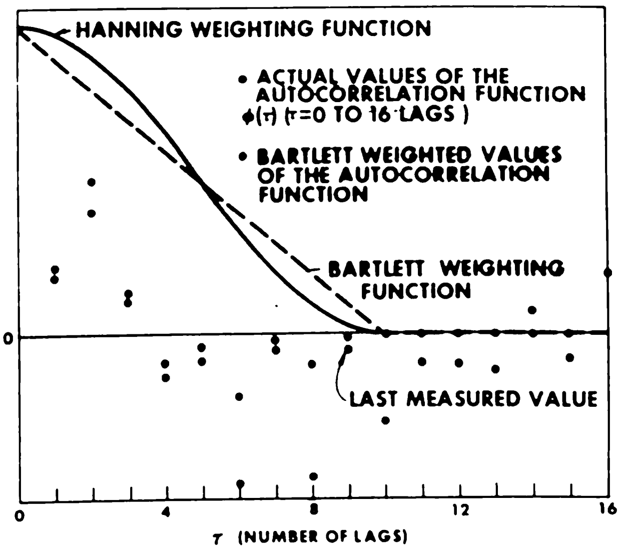
$$\phi(\tau) = \lim_{T \to \infty} \frac{1}{2T+1} \sum_{n=-T}^{+T} X_n X_{n+\tau} \quad (-\infty < \tau < \infty).$$

THE POWER SPECTRUM IS THEN DEFINED AS

$$P(f) = \frac{1}{W} \sum_{\tau=-\infty}^{+\infty} \phi(\tau) \cos(2\pi f \tau \Delta t), \quad (0 \le f \le W = \frac{1}{2\Delta t}),$$

WHERE Δt IS THE SAMPLE PERIOD OF THE TIME SERIES.

1



2

THE BASIC AUTOCORRELATION FUNCTION THEOREM

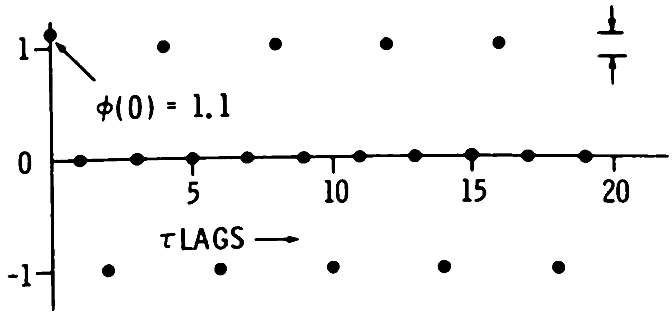
$\phi(0), \phi(1), \ldots, \phi(N)$ ARE THE FIRST N+1 VALUES OF AN AUTOCORRELATION FUNCTION IF, AND ONLY IF, THE N+1 BY N+1 TOEPLITZ MATRIX,

$$\begin{bmatrix} \phi(0) & \phi(1) & \cdots & \phi(N) \\ \phi(1) & \phi(0) & & \\ \vdots & & \ddots & \\ & & \phi(0) & \phi(1) \\ \phi(N) & & \phi(1) & \phi(0) \end{bmatrix}$$

IS SEMI-POSITIVE DEFINITE.

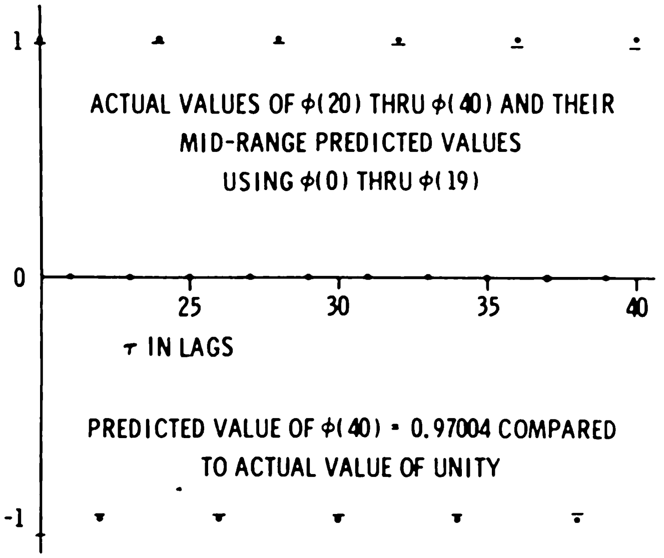
3

AUTOCORRELATION OF A UNIT POWER SINUSOID OF FREQUENCY W/2 WITH 10% RANDOM NOISE



THE VALUES OF $\phi(0)$ THRU $\phi(19)$ REQUIRE THAT $\phi(20)$ LIE IN THE RANGE OF 0.98901 ± 0.11099

4



5

THE ENTROPY OF A GAUSSIAN BAND-LIMITED TIME SERIES WITH POWER SPECTRUM P(f) IS PROPORTIONAL TO

$$\int_0^W \log P(f)\, df \, .$$

UNDER THE CONSTRAINT THAT P(f) AGREES WITH THE N+1 MEASURED VALUES OF THE AUTOCORRELATION FUNCTION, i.e.,

$$\int_0^W P(f) \cos(2\pi f \tau \Delta t)\, df = \phi(\tau), \quad (\tau = 0 \text{ to } N),$$

WHAT P(f) HAS THE MAXIMUM ENTROPY?

6

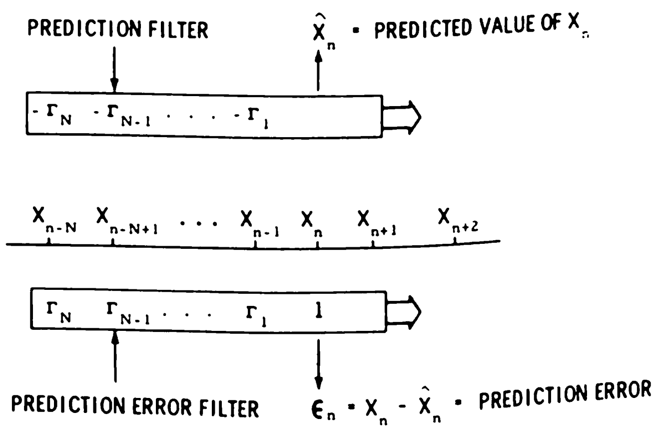
THE MAXIMUM ENTROPY SPECTRUM IS GIVEN BY

$$P(f) = \frac{P_{N+1}/W}{\left|1 + \sum_{n=1}^{N} \Gamma_{n+1} e^{-i2\pi f n \Delta t}\right|^2} ,$$

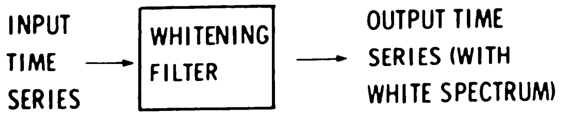
WHERE P_{N+1} AND THE Γ_{n+1} ARE OBTAINED FROM THE MATRIX EQUATION

$$\begin{bmatrix} \phi(0) & \phi(1) & \cdots & \phi(N) \\ \phi(1) & \phi(0) & & \\ \vdots & & \ddots & \\ & & \phi(0) & \phi(1) \\ \phi(N) & & \phi(1) & \phi(0) \end{bmatrix} \begin{Bmatrix} 1 \\ \Gamma_2 \\ \vdots \\ \Gamma_N \\ \Gamma_{N+1} \end{Bmatrix} = \begin{Bmatrix} P_{N+1} \\ 0 \\ \vdots \\ 0 \\ 0 \end{Bmatrix} .$$

7



8

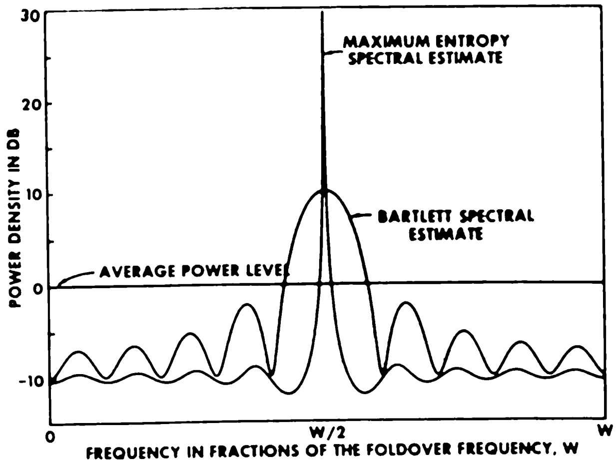


$$\text{INPUT POWER SPECTRUM} = \frac{\text{OUTPUT POWER SPECTRUM}}{\text{POWER RESPONSE OF FILTER}}$$

$$P(f) = \frac{P_{N+1}/W}{\left|1 + \sum_{n=1}^{N} \Gamma_{n+1}\, e^{-i2\pi f n \Delta t}\right|^2}$$

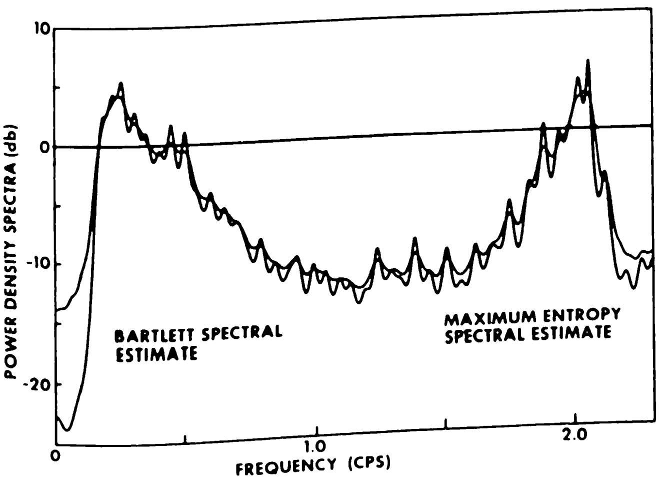
9

ESTIMATES OF A UNIT POWER SINUSOID OF FREQUENCY W/2 WITH 10% WHITE RANDOM NOISE ADDED



10

WMO AMBIENT SEISMIC NOISE RECORDED BY A HORIZONTAL SEISMOMETER



11

THE MAXIMUM ENTROPY SPECTRUM AS AN ALTERNATE DEFINITION OF A POWER-DENSITY SPECTRUM

$$P(f) = \lim_{N\to\infty} \frac{1}{W} \sum_{\tau=-\infty}^{+\infty} \phi_N(\tau) \cos(2\pi f \tau \Delta t), \quad \text{WHERE}$$

	WIENER	MAXIMUM ENTROPY
For $\tau = 0$ to N	$\phi_N(\tau) = \phi(\tau)$	$\phi_N(\tau) = \phi(\tau)$
For $\tau > N$	$\phi_N(\tau) = 0$	$\phi_N(\tau)$ = MAXIMUM ENTROPY EXTENSION

AS $N \to \infty$, THESE TWO DEFINITIONS ARE THE SAME.

12