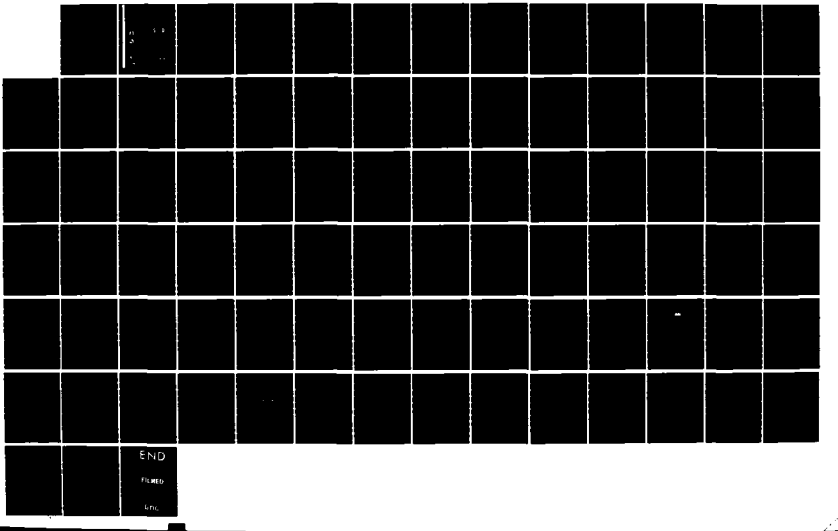
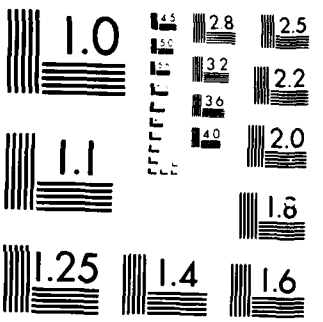
AD-A164 698 A REVIEW OF MAXIMUM ENTROPY SPECTRAL ANALYSIS AND APPLICATIONS TO FOURIER SPECTROSCOPY(U) AIR FORCE GEOPHYSICS LAB HANSCOM AFB MA E M DEWAN 03 APR 85 1/1

UNCLASSIFIED AFGL-TR-85-0091 F/G 12/1 NL





(12)

AFGL-TR-85-0091
ENVIRONMENTAL RESEARCH PAPERS, NO. 911

A Review of Maximum Entropy Spectral Analysis and Applications to Fourier Spectroscopy

EDMOND M. DEWAN

3 April 1985

OPTICAL PHYSICS DIVISION PROJECT 7670
AIR FORCE GEOPHYSICS LABORATORY
HANSCOM AFB, MA 01731

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

AD-A164698

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION: Unclassified	1b RESTRICTIVE MARKINGS: None
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution Unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S): AFGL-TR-85-0091 ERP, No. 911	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION: Air Force Geophysics Laboratory; 6b OFFICE SYMBOL (If applicable): OPA	7a NAME OF MONITORING ORGANIZATION
6c ADDRESS (City, State and ZIP Code): Hanscom AFB Massachusetts 01731	7b ADDRESS (City, State and ZIP Code)
8a NAME OF FUNDING/SPONSORING ORGANIZATION: Air Force Geophysics Laboratory; 8b OFFICE SYMBOL (If applicable): OPA	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State and ZIP Code): Hanscom AFB Massachusetts 01731	10 SOURCE OF FUNDING NOS: PROGRAM ELEMENT NO 62101F; PROJECT NO 7670; TASK NO 15; WORK UNIT NO 07
11 TITLE (Include Security Classification): A Review of Maximum Entropy Spectral Analysis and Applications to Fourier Spectroscopy	
12 PERSONAL AUTHOR(S): Dewan, Edmond M.	
13a TYPE OF REPORT: In-house; 13b TIME COVERED: FROM _____ TO _____; 14 DATE OF REPORT (Yr., Mo., Day): 1985 April 3; 15 PAGE COUNT: 86	
16 SUPPLEMENTARY NOTATION	

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB GR	Spectral analysis	Maximum Entropy
			Data analysis	Digital signal processing
			Fourier Spectroscopy	Super resolution

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

There are many situations in Air Force technology where a great increase in the resolution of power spectral densities, PSD's, would solve important problems. The problem areas range from radar, speech compression, seismic signal processing, to spectral analysis of optical signals. The maximum entropy method (MEM) can, under certain circumstances, provide an order of magnitude increase in resolution over conventional techniques. This last fact motivates the present review. The fundamental concepts and mathematical background are given and, most importantly, the peculiarities and new features introduced by the MEM technique are explained. Some signals are good candidates for this technique while others are not; thus, methods are described to distinguish between the two. In addition, some of the vulnerabilities and unanswered questions associated with MEM are mentioned.

(over)

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS ☐	21 ABSTRACT SECURITY CLASSIFICATION: Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL: Edmond M. Dewan	22b TELEPHONE NUMBER (Include Area Code): (617) 861-3092; 22c OFFICE SYMBOL: OPA

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

19. Abstract (Contd)

In particular, after the techniques to obtain the MEM-PSD are explained, an application to Fourier Spectroscopy is given to illustrate the results (as compared to conventional means). ~~Also,~~ this review employs certain features to make the MEM approach accessible to those without previous knowledge about it. These features include: (a) a short history of the PSD which describes the evolution over time of the basic concepts, (b) a discussion of the "higher principle" of maximum entropy in the context of probability theory as was originally introduced by E. T. Jaynes, (c) a simple introduction to the Z-transform and how it introduces the basic spectral line shape of the MEM-PSD as well as other peculiarities of the method which distinguish it from the conventional approach, (d) the physics behind the ~~so-called~~ "reflection coefficients" which are introduced by MEM, and (e) an extensive annotated bibliography which should make the literature available. The review concludes by discussing future possibilities regarding MEM and Fourier Spectroscopy.

Preface

I thank both G. Vanasse and N. Grossbard for many helpful suggestions and informative conversations. I also thank N. Grossbard for the programming associated with the data analysis.

Contents

Contents

Illustrations

Illustrations

Tables

A Review of Maximum Entropy Spectral Analysis and Applications to Fourier Spectroscopy

1. INTRODUCTION

This report is based on a talk that was presented on April 6, 1984, at a seminar called "MESA Workshop", sponsored by the Optical Physics Division. That seminar was organized because it has become apparent that the maximum entropy technique of spectral analysis can sometimes help in the resolution of spectra in Fourier Spectroscopy.

The importance of the Maximum Entropy Method (MEM) comes from the fact that the resulting power density spectra (PSD's) often exhibit "super-resolution"; a resolution that far exceeds that of conventional methods. This, in turn, allows one to obtain high resolution PSD's with much less data.

This report provides background information on this subject, presents some of the salient features of the technique, and illustrates its use in Fourier Spectroscopy with an experimental example. The report also reveals some of the peculiarities of the method, and shows how to distinguish which types of data represent good candidates and which represent bad candidates for its use. (There are, it turns out, cases where MEM is worse than conventional methods.) Finally, the references are annotated to guide the reader to the most useful literature. The latter contain the details that are left out here due to space limitations.

(Received for Publication 28 March 1985)

It will be assumed that the reader is acquainted with Fourier analysis, the Convolution Theorem, the Wiener-Khintchine Theorem, and the conventional approach of Blackman and Tukey. Finally, it should be mentioned that despite its length, the present report is merely a "broad brush" treatment.

The form of the one sided MFM spectrum is given by:

$$\Psi(\omega) \quad \frac{2\sigma_m^2 \, \Delta t}{\left| 1 - \sum_{k=1}^{m} a_k \, e^{i\omega(k\Delta t)} \right|^2}$$

where Δt is the sampling interval, ω is the radial frequency, which can assume any value within the Nyquist limit, and σ_m and a_k are constants to be determined from the matrix equation

$$\begin{bmatrix} \phi_0 & \phi_1 & \cdots & \phi_{m-1} & \phi_m \\ \phi_1 & \phi_0 & \cdots & \cdot & \phi_{m-1} \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \phi_{m-1} & \cdot & \cdots & \phi_0 & \phi_1 \\ \phi_m & \phi_{m-1} & \cdots & \phi_1 & \phi_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{bmatrix} = \begin{bmatrix} \sigma_m^2 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

where the ϕ_i are the autocorrelations. This is in contrast to the form obtained for the one sided spectrum from the discrete Fourier Transform approach:

$$\Psi(\omega_k) \quad \frac{2\Delta t}{N} \left| A(\omega_k) \right|^2$$

where N is the number of discrete values of the signal, and

$$A(\omega_k) \equiv \sum_{n=0}^{N-1} x_n \, e^{i\omega_k n}$$

where x_n are the discrete signal values, and ω_k is the discrete "digital radial frequency" given by $\frac{2\pi k}{N}$. The analog equivalent would be $\omega_{analog} \equiv (\omega_k / \Delta t)$. Time $t = n\Delta t$ in the continuous domain.

As can be seen, these two discrete time formulations of the power spectral density appear very different. To someone familiar only with the second form, the first form (MEM) will look strange. To give the MEM formalism a clear physical meaning, this report will first present the development of the power spectrum up to the time of Burg. We then consider Jaynes' development of the main principle behind MEM. This principle is explained in its general form and then used to derive the MEM spectrum. After that, we consider a more detailed approach that uses the "Z-transform." This transform is explained in Section 4, where it is shown that the Lorentzian line shape is due to the nature of the Z-transform, rather than depending directly on the MEM principle. In Section 5, the MEM spectrum is derived from autoregression theory, which gives the physical significance of the MEM spectrum.

The associated matrix equation is derived in Section 6, to give a physically clear picture of it. The practical methods to solve it are derived in Section 7. Another approach is through linear prediction, due to Wiener. This least squares approach is used in Section 8 to derive the matrix equation from a completely different approach. Here it is shown that least squares and MEM can lead to the same results.

In Section 9 we turn to the question of what types of data are good candidates for MEM treatment and which are not. In Section 10, we consider the so-called "Burg technique" for determining σ_m and a_i. Many advantages of the Burg technique will be pointed out. We then (Section 11) apply MEM to Fourier Spectroscopy and show some results. Finally, in Section 12, we consider the interpretation, due to Burg, of the power spectrum in terms of "reflection coefficients."

An annotated bibliography suggests further reading.

2. DEVELOPMENT OF THE THEORY

In this report, a recent and excellent historical survey of spectrum analysis by Robinson[1] is used. In addition, a very understandable introduction to time series analysis by Gottman[2] was employed.

2.1 From Pythagoras to Fourier

If one excludes calendars and time-keeping devices, then perhaps the first use of harmonic or spectral analysis was due to Pythagoras (600BC). He pointed out that a string of a musical instrument vibrated in a way that could be broken

1 Robinson, Enders A. (1982) A historical perspective of spectrum estimation, Proc. IEEE 70(No. 9):885.

2. Gottman, J.M. (1981) Time-Series Analysis, Cambridge Univ. Press.

down into "harmonics" as shown in Figure 1. He used such information for a mathematical basis of musical harmony. From this time until the 1700's, there seems to have been no other application of harmonic analysis.

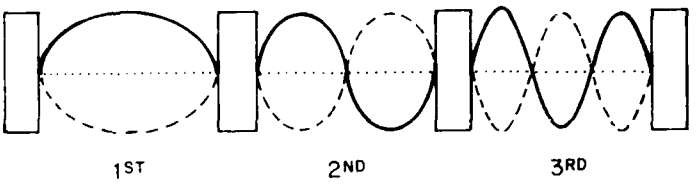


Figure 1. String Harmonics

The mathematics of spectral analysis began, *in its modern form, in the* eighteenth century. Daniel Bernoulli (1738) was the first to solve the wave equation.

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0 \quad , \tag{1}$$

where u is the vertical displacement of a string, c is the velocity of the wave, x the horizontal coordinate, and t the time. He did this by use of the now classic method of separation of variables:

$$u(x,t) = X(x)\,T(t) \quad . \tag{2}$$

The boundary conditions are given by

$$\begin{aligned} u_{x=0} &= 0 \quad , \\ u_{x=\pi} &= 0 \quad , \end{aligned} \tag{3}$$

where the end points are $x = 0, \pi$. This leads to a solution in x that employs sine (not cosine) waves and since Eq. (1) is linear, one can use superposition to get the general solution

$$u(x, t) = \sum_{k=1}^{\infty} \sin kx \{A_k \cos kct + B_k \sin kct\} \quad , \tag{4}$$

where k is an integer, due to Eq. (3).

Here, the crucial part of his result is the expression for the initial displacement of the string [t = 0 in Eq. (4)]:

$$u(x, 0) = \sum_{k=0}^{\infty} A_k \sin kx \quad , \tag{5}$$

On the left, u(x, 0) is an arbitrary function. It need not be "analytic". On the right there is an infinite sum of sines, each of which is analytic.

Thus, we have a seeming paradox. "How can a non-analytic function (with, for example, discontinuous derivatives), be represented by an infinite sum of nothing but analytic functions?" From Euler (1707–1783) and Lagrange (1736–1813) we have the well-known relation

$$A_n = \frac{2}{\pi} \int_0^{\pi} u(x, 0) \sin nx \, dx \quad , \tag{6}$$

which displays the "global" nature of A_k. It turns out that Eq. (6) is useful in clearing up the apparent paradox.

The first announcement that an almost arbitrary function could be expanded as in Eq. (5), was made in 1807 by Jean Baptiste Joseph Fourier to the French Academy. (There are constraints on the function, but analyticity is not one of them.) The distinguished mathematicians present at his talk rejected his result. Fourier himself was never able to provide a rigorous proof, although we now know he was indeed correct. The proof of Fourier's assertion made use of the "Z-transformation".[1] The MEM technique of spectral analysis (as well as a very large and increasing literature on the digital processing of data and signals) is based conceptually upon the Z-transform.

The analytic function paradox mentioned above is resolved when the Taylor series expansion (which defines analytic functions) is connected to the Fourier series expansion by means of the Z-transformation. Consider, in place of u(x, 0) above, an analytic function, f(Z), of a complex variable Z:

$$Z \equiv x + iy \equiv (re^{j\theta})^{-1}$$

Expand f(Z) in a Taylor series about the point $Z^{-1} = 0$ to obtain what we define as the Z-transform

$$f(Z) = \sum_{n=0}^{\infty} a_n Z^{-n} \quad . \tag{7}$$

If Z_0^{-1} is defined as the first singular point from the origin, then the radius of convergence extends from $Z^{-1} = 0$ to $Z^{-1} = Z_0^{-1}$, or, in other words, the region of convergence is where $|Z| > |Z_0|$. Now, letting $Z = 1 \exp(-j\theta)$ in Eq. (7), we obtain the well known complex form for the Fourier series

$$f(e^{-j\theta}) = \sum_{n=0}^{\infty} a_n [\cos n\theta + j \sin(n\theta)] \quad . \tag{8}$$

We now consider three cases to explain the paradox. In the first case, Z_0 is inside the unit circle. In this case, both f(Z) and the series in Eq. (7) are analytic functions and there is no paradox. In the second case, Z_0 is outside the unit circle. This implies that the series does not converge (as can be seen) and, again there is no paradox because the series is invalid. The third case is the one that resolves the paradox. In this case, Z_0 is on the unit circle. When this happens, the Taylor series will not converge at some or all of the points on the circle. The Taylor series defines an analytic function (differentiable to any order) inside the circle, however. On the circle (the Fourier series in θ) we have a non-analytic function. The solution to the paradox is that it is sufficient to move the singularity from the circle to the inside by a vanishingly small amount to change the given non-differentiable function of θ into one which can be differentiated to any order, that is, an analytic function. The proof of the validity of the Fourier series makes use of this fact (a limit is taken). The reader may recall that similar arguments are used in Fourier transforming the step-function. The main point here is that the Z-transform was crucial for the proof that Fourier's approach is correct.

2.2 The Periodogram as Introduced by Sir Arthur Schuster

The next major step in spectral analysis of data came with the introduction of the concept of the periodogram, $P(\omega)$, by Sir Arthur Schuster in 1898. This is defined by Eqs. (9) and (10)

$$X(\omega) \equiv \sum_{n=1}^{N} x_n e^{-jn\omega} \quad , \tag{9}$$

and

$$P(\omega) \equiv \frac{1}{N} |X(\omega)|^2 \quad , \tag{10}$$

where the x_n represent the individual values of the time series data, n is the integer giving it the (discrete) time label, $\omega \equiv 2\pi f$ where f is the frequency, and N is the total number of data points. The periodogram was conceived for the purpose of finding "hidden periodicities" in the data. Implicit in this approach is a model of the data that consists of two parts. (Here we follow Gottman.[2]) One part is purely deterministic and consists of one or more sine waves. The other part consists of white noise, e_t. For example

$$x_t = A \cos \omega t + B \sin \omega t + e_t \quad . \tag{11}$$

This approach, unfortunately, was plagued by a feature which severely limited its usefulness. As shown by theory and experiment, the variability or standard deviation of the periodogram is equal to its mean value for a given ω. In other words, the noise fluctuations were as large as the "signal". Furthermore, this feature remains, independently of the length of data record or N. (For a proof, the reader may consult Gottman[2] or Claerbout.[3]) For this reason the periodogram was considered a failure.

2.3 The Slutzky Effect and the Work of Yule

In the 1920's, the Russian scientist Slutzky made a surprising discovery. He found that a signal consisting of pure white noise, as shown in Figure 2, could give rise to a spectrum with a clear peak merely by taking a moving average prior to the spectral analysis. See Figure 3, based on Gottman.[2] In effect, he found that a moving average can introduce a periodicity, that is, "create order out of chaos". Initially, this discovery was alarming to his contemporaries because they routinely applied moving averages to their data and they did not suspect that, in so doing, they were introducing a periodic feature into their results. Once the "Slutzky Effect" became understood, it changed the way one modeled the periodicities of nature. In effect, this led to a model that consisted

3. Claerbout, Jon F. (1976) Fundamentals of Geophysical Data Processing.

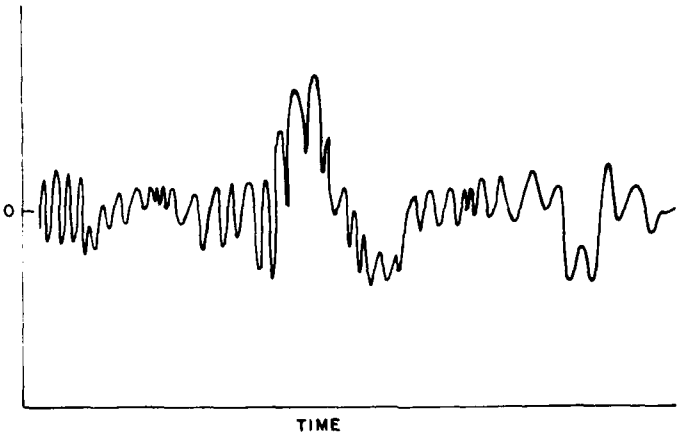


Figure 2. Pure Noise of Zero Mean and Unit Variance

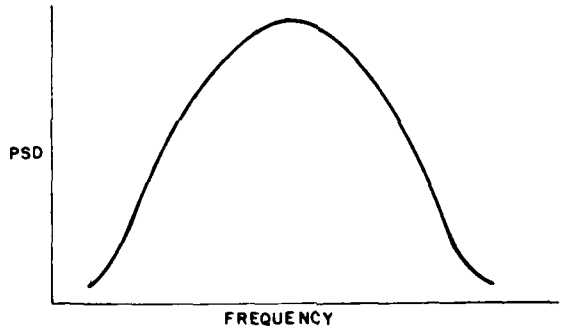


Figure 3. White Noise Spectrum of Figure 2 is Altered by 4-point Moving Average. It now has a peak due to this average

of white noise as an input to a filter, and the output (or "filtered" white noise) as "data of nature" or "signal to be analyzed". More details will be given below, but it should be emphasized that this model is exceedingly different from the one implied by Eq. (11).

Shortly after the Slutzky Effect was discovered, G. Udney Yule (1927) analyzed sun spot activity. To obtain his results, he originated the idea that Schuster's model [Eq. (11)] is not of general validity. Instead of this "pure-sine-wave-plus-noise process", he suggested the "filtered noise" model in the following graphic manner.

Imagine a laboratory with a swinging pendulum in it, and suppose that an experiment is in progress in which a transducer is used to record the pendulum's angular position as a function of time. Now suppose that a group of young boys get into the laboratory and shoot peas with their pea-shooters at the pendulum. The impacts would randomly perturb both the amplitude and phase of the pendulum. The resulting time series would no longer contain a deterministic component nor would there be an exact period. Instead, the series would become unpredictable (that is, for long times into the future) and yet it would appear smooth and fairly periodic. This seemed to be the ideal model for sun spots. It became clear later that it was of general validity for the science of time series analysis.

Next, consider a discrete mathematical treatment of the "peas and pendulum" model. The pendulum can be regarded as a narrow band filter. In terms of discrete time series, the behavior of a pendulum can be modeled by a homogeneous difference equation

$$b(n) + a_1\, b(n-1) + a_2\, b(n-2) = 0 \quad , \tag{12}$$

where n is the time index, a's are constants that are given, and b(n) is the dependent variable. The solution to Eq. (12), easily found in books on difference equations, is

$$b(n) = e^{-\lambda n}\, \frac{\sin(n+1)\,\omega_0}{\sin \omega_0} \quad , \tag{13}$$

where λ and ω_0 are functions of the constants a_1 and a_2. As can be seen, Eq. (13) is the discrete version of a decaying sine wave of the type well known in elementary differential equation theory. To add the effect of a random input to Eq. (12), in this example "the peas", a term $\varepsilon(n)$ is inserted on the right hand side as in Eq. (14).

$$x(n) + a_1\, x(n-1) + a_2\, x(n-2) = \varepsilon(n) \quad , \tag{14}$$

where x(n) is now the dependent variable. The solution of Eq. (14) is given by

$$x(n) = \sum_{k=0}^{\infty} b(k)\, \varepsilon(n-k) \quad , \tag{15}$$

where b(k) is given by Eq. (13). The right hand side of Eq. (15) is the convolution of the input noise, $\epsilon(n)$, and the filter impulse response b(k), in a manner analogous to the continuous theory. Eq. (14) has become known as the AR or autoregressive model for discrete time series. In the general case any number of filter (a_i) coefficients can be involved.

Figure 4 gives the block diagram of the new time series model [redacted] from Slutzky and Yule's work. Quite independently, the mathematical [redacted] tral analysis of random processes was being developed in the [redacted] (1920–1930) by Norbert Wiener. This will be considered next.

Figure 4. Block Diagram of the New "Model" for the "Signal"

2.4 The Contribution of Norbert Wiener

In 1923, Wiener developed a model for Brownian movement. This work extended the work of Einstein (1905) and this Einstein-Wiener theory of Brownian movement is important to all theoretical studies of spectrum analysis [according to Robinson[1] who describes this work in more detail than will be done here]. In 1930, Wiener published his classic paper on Generalized Harmonic Analysis. Equation (15), and the Fourier transform version,

$$X(\omega) = B(\omega)\, E(\omega) \quad , \tag{16}$$

became his model for the random process. In Eq. (16), the capital letters stand for the Fourier transforms of the lower case quantities in Eq. (15). The convolution theorem was used [the Fourier Transform (FT) of a convolution equals the product of the FT's]. A key point in Wiener's analysis is

$$\overline{\epsilon(n)\,\epsilon(k)} = \delta_{nk} \quad , \tag{17}$$

where the bar indicates a suitable average or "expectation value".* On the left side we have, by definition, the autocorrelation of a noise process; and on the right, we have a Kronecker delta function.

Let us now consider an important observation. In the context of the signal model of Figure 4, which is used by Wiener, the usual explanation for the failure of the periodogram (added noise) is not quite valid.† The true explanation is the presence of the factor $E(\omega)$, which is a very noisy function, in Eq. (16). The FT of a rough function is not necessarily a smooth function. Indeed, Wiener showed in 1923 that the FT was rough (as Robinson's very useful survey explains).

In passing, it is of interest to note that Wiener was very interested not only in Brownian Motion, but also in a very general aspect of the mathematical characterization of noisy but semi-ordered patterns. He described in his autobiographical accounts how he was inspired by the sight of the Charles River, as seen from MIT, and wondered how one could mathematically describe the surface of the water with its random waves.

So far as spectral analysis of data is concerned, Wiener's most concrete contribution was the theorem named after him and Khintchine.‡ This states that the autocorrelation, defined by $\phi(k)$ in

$$\phi(k) \equiv \frac{1}{N} \sum_{n} x^*(n)\, x(n+k) \quad , \tag{18}$$

is related to the PSD $\Phi(\omega)$ by

$$\Phi(\omega) = \sum_{k=-p}^{P} \phi(k)\, e^{-j\omega k} \quad , \tag{19}$$

where p is the maximum lag. This implies that the key to the essentially continuous spectrum of the process in Figure 4 is in the autocorrelation. (See Y. W. Lee[4] for an elementary study of Wiener's approach.) Unfortunately, this did not lead immediately to practical data analysis. Tukey's breakthrough (to be described next) was needed for results to be achieved. As will be seen, however,

---

4. Lee, Y.W. (1960) Statistical Theory of Communication, Wiley.

*See E. Wolf, JOSA 72, No.3, (1982) (sec. 2) for an interesting contrast between the treatments of Wiener and Kolmogorov.

†See Robinson and Treitel (1980), Appendix 16-1, for a rigorous treatment.

‡See Wolf, op cit.

Wiener's work on linear prediction and the "Wiener Optimum Filter" based on least squares also has a crucial role to play in MEM spectral analysis.

2.5 The Blackman and Tukey Approach

The practical approach to the PSD (which according to Robinson[1] was mainly due to Tukey) was described in the famous book by Blackman and Tukey.[5] For further details the reader may consult that reference.

Unlike Wiener, Blackman and Tukey considered the case where the data record is finite. Using their notation, the autocorrelation, C_r, is defined (discrete case) as

$$C_r = \frac{1}{n - r} \sum_{q=0}^{q=n-r} x_q x_{q+r} \quad . \tag{20}$$

where x_q is the time series, q the time index, and r the lag. One assumes unity for the sample interval spacing in Eq. (20). The total number of data points is n. The PSD, V_r, is obtained from [in Eq. (20) x_q is <u>real</u>]:

$$V_r = \Delta\tau \cdot \left[C_0 + 2 \sum_{q=1}^{m-1} C_q \cos \left(\frac{qr\pi}{m} \right) + C_m \cos r\pi \right] \quad . \tag{21}$$

Here, $\Delta\tau$ is the sampling interval, m is the maximum lag, and the cosine arises from the fact that the original series is real. These authors then introduced ways to smooth out the PSD to get rid of the noise fluctuations that killed the usefulness of the periodogram.

They used three smoothing techniques: (a) the truncated autocorrelation, (b) "window carpentry", and (c) direct spectral smoothing. The last two techniques also help with the problem of "spectral leakage" which, as will be explained, comes from the fact that the data record is not infinitely long.

To truncate the autocorrelation, the maximum lag m in Eq. (21) is limited to be a small fraction of n, the total number of points. Typically, the maximum number of lags is taken to be 1/10 of the number of data points. The resultant smoothing of the PSD is due, as explained by Blackman and Tukey,[5] to the resulting increase in the number of degrees of freedom in their spectral model based on the chi-squared statistic (compare this with Gottman's

5. Blackman, R.B., and Tukey, J.W. (1958) <u>The Measurement of Power Spectra</u>, Dover.

discussion of the periodogram[2]). In effect, the smoothing results from the pooling of 10 elementary spectral values for each PSD result based on the 10 percent truncated autocorrelation. Their book contains confidence intervals in a table to guide one in the use of truncated autocorrelations. Further discussions of use will be found in Bendat and Piersol[6] and Otnes and Enochson.[7]

"Window carpentry" (called apodization in optics) uses a weighting factor, W_n, which is unity for zero lag and smoothly decreases to nearly zero at the maximum lag. Thus, in Eq. (21), C_q is replaced by $W_q C_q$. Blackman and Tukey showed that exactly the same smoothing effects can be obtained by means of a weighted average in the frequency domain. For example, the use of a weight given by $W_q = \frac{1}{2}(1 + \cos \frac{\pi r}{m})$, $|r| < m$, is exactly the same thing as using

$$V_r' = \frac{1}{4} V_{r-1} + \frac{1}{2} V_r + \frac{1}{4} V_{r+1} \tag{22}$$

in the frequency domain, and in either case, it is called "Hanning". (P. 36 of Blackman and Tukey[5] explains end point corrections.)

Another smoothing method, due to Welch, introduced more "degrees of freedom" in a more direct manner. In this method, the original data record was chopped into sub-sections. The PSD was obtained for each of these subsections, and then they were all averaged together [Eq. (23)], resulting in a smooth PSD.

$$P(\omega) = \frac{\sum_{i=1}^{N} P_i(\omega)}{N} \quad . \tag{23}$$

"Leakage" is another practical problem solved by Blackman and Tukey. This problem comes from the fact that a finite segment of data is, in effect, the product of an infinite data string multiplied by a rectangular pulse (equal to 1 over the available data and zero outside). From the convolution theorem it follows that this is equivalent to convolving the true spectrum with a sinc function $[(\sin \theta)/\theta]$, $\theta \equiv \omega N \Delta\tau$. Window carpentry alleviates this problem. One can use tapered weights to reduce the "side-lobes" of this spectral window by a factor of 10 (these lobes are the cause of leakage from neighboring frequencies) but, of course, there is a price. The resolution decreases, for example, by a factor of 2 (from the effect of smoothing) if one were to use a triangular window.

6. Bendat, J.S., and Piersol, A.G. (1971) Random Data: Analysis and Measurement Procedures, Wiley-Interscience.

7. Otnes, R.K., and Enochson, L. (1978) Applied Time Series Analysis, Vol. 1., Wiley.

Another approach to the leakage problem involves what Blackman and Tukey call "pre-whitening". For example, consider Figures 5 and 6. Suppose one considers the red noise shown in Figure 5. Leakage would bring about an incorrect slope from a raw analysis. If, however, the data are "first differenced", that is,

$$x_i = x_i - x_{i-1} \quad , \tag{24}$$

then the resulting spectrum, as they show, is that shown in Figure 6. This is a flat spectrum and, therefore, no leakage is possible. After this is done, the PSD is "post colored" by means of a spectral version of the inverse transform of Eq. (24).

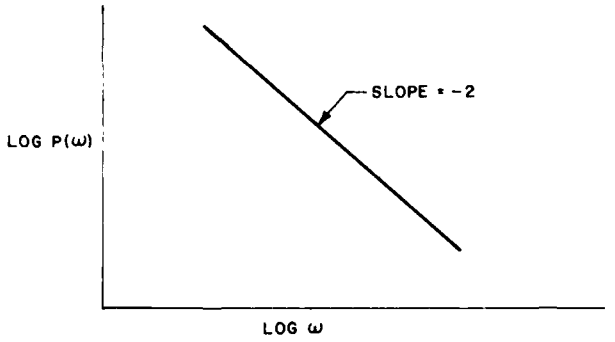


Figure 5. PSD of Original Signal Before Prewhitening

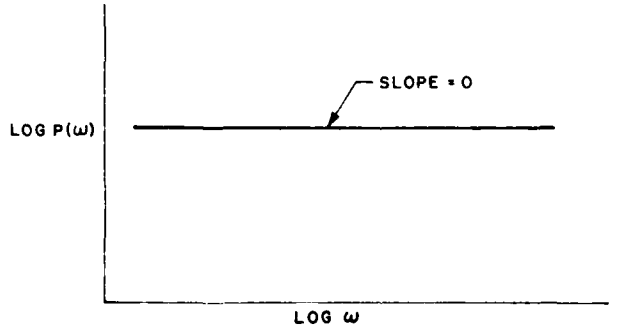


Figure 6. PSD of Signal After "First-Differencing" or Prewhitening

Prewhitening has been mentioned here because it has relevance to the MEM spectrum. The more general form of Eq. (24) is a combination of autoregression and moving averages (see Blackman and Tukey[5]). As such it represents something very analogous in concept to MEM or, as we shall see, AR analysis. In effect, prewhitening can be imagined to be an "eyeball" kind of MEM. If one could produce a perfectly flat PSD by prewhitening, the recolored spectrum would be the MEM spectrum (or, in the most general case to be explained, the ARMA spectrum). This remark will be more meaningful to the reader later on when MEM, AR, and "ARMA" are explained.

Blackman and Tukey seem to have introduced the term "aliasing" as depicted in Figure 7. (See also Goldman[8] "Information Theory".) This figure shows how the process of sampling data at intervals of $\Delta\tau$ can give rise to the illusion that a large wavelength is present when in fact it is a short one. There is an analogy* here with the effect of a stroboscope which can slow down or even stop the motion of a rotating fan. To avoid such aliasing, one filters out all high frequencies before sampling the data. The filter must remove all frequencies above the Nyquist frequency given by $(1/(2\ \Delta\tau))$ if there is to be no aliasing.

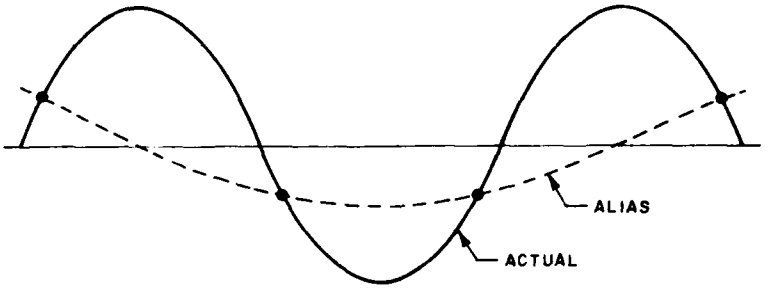


Figure 7. Actual Signal Is Given by the Full Curve. The dotted curve represents the aliased signal caused by sampling at the low frequency given by the dots

8. Goldman, S. (1953) Information Theory, Prentice-Hall, Inc.

*Another analogy has been suggested by G. Vanasse. It is the overlapping orders in a grating spectrometer. The grating samples the incoming wave front at an internal corresponding to the grating constant; it is because of this finite sampling interval that spectra of different orders are produced. If the spectral bandwidth is too large, aliasing (that is overlapping of orders) occurs. In contrast, the prism is continuous in its effect on the wavefront and no spectral orders appear.

Historically, the next breakthrough was the invention of the "Fast Fourier Transform" or FFT. This made the calculation of the PSD practical from an economic point of view and it is described in great detail in Brigham[9] as well as in many other places. Cooley and Tukey published the first algorithm for the FFT. Later on it was realized that the general idea was already in the literature; however, even today few people seem to realize that Gauss was the very first one to use it.

We have arrived at the point in this account where John Burg made his contributions, which are the MEM-PSD, and the associated methods to obtain it. In his Ph. D. thesis (Stanford), he made two very important observations which set the stage for what was to follow. First, he noted that sometimes the PSD which was obtained by the techniques described above could be negative. Theoretically such a result is complete nonsense. It is a result of the fact that the sinc function has negative sidelobes. The cure is "window carpentry", but with loss of resolution. He also noted that the PSD's obtained did not agree with the autocorrelations of the data. In his words: "These two affronts to common sense were the main reasons for the development of maximum entropy spectral analysis....".

"If one were not blinded by the mathematical elegance of the conventional approach, making assumptions as to unmeasured data and changing the data values that one knows [to be correct] would be totally unacceptable from a common sense and, hopefully, from a scientific point of view."

These are strong words, but they are well founded. Since the method he developed rests on the general principle of maximum entropy, we turn now to a discussion of the work of Jaynes, which is the foundation for the use of this principle in the present context.

3. MAXIMUM ENTROPY – THE "HIGHER PRINCIPLE"

In 1957, Edwin T. Jaynes started an unusual project in which he derived statistical mechanics on the basis of information theory. This work has grown in significance over the years. The generality is due to the fact that Jaynes has made a fundamental contribution to the concept of probability. The original approach to "probability" was due to Laplace. He regarded it as a reflection of one's state of knowledge. For example, suppose one is given a coin to flip. The question of what probability to assign to "heads" and "tails" is decided, by Laplace, by the "principle of indifference" or "the principle of insufficient reason". If

9. Brigham, E.O. (1974) The Fast Fourier Transform, Prentice Hall, Inc.

there is no known reason to prefer either "heads" or "tails", then one assigns an equal probability to them. The die is similar in this respect. If there is no known reason to suspect that it is "loaded", then one should assign an equal probability to any particular outcome of a "throw" of a die. In this case the probability would be 1/6. Unfortunately, if one did know that the die was unfair or loaded, the Laplace approach would give no answer. As will be explained below, Jaynes has shown us how to extend the Laplace principle to the case of the loaded die and even to far more general cases (one particular one being the PSD).

Laplace's subjective concept of probability went out of fashion when it was found to be very limited in application in the form that he gave us. It was replaced by another definition known as the "frequency" interpretation given in most books on the subject today. The generalization due to Jaynes, however, came at a time when other scientists such as Cox, Jeffreys, and Good were reconsidering the subjective approach. For an excellent history of the subject of probability, the reader should consult the paper by Jaynes given at the MIT symposium on Maximum Entropy (Levine[10]). Jaynes was drawn to this subject by the following strange observation. Shannon introduced the concept of "information" to communication engineering, and its mathematical form is essentially identical to that of entropy in the statistical mechanics of Boltzmann and Gibbs. "Why should they be the same?" is the question Jaynes asked himself. He was later to find the answer, which is that they turn out to be two examples of the application of the same principle. In deriving this answer, he generalized the Laplace method of assigning probabilities. He was also able to obtain results in non-equilibrium statistical mechanics.

To approach Jaynes' work, this report will use a simple example that illustrates the main ideas behind the maximum entropy concept as it applies to subjective probability assignment. The particular example used is one given by Jaynes himself in a lecture that he gave at Brandeis in 1963 (see Ref. 11).

3.1 The Brandeis Die

Jaynes' example is that of the loaded die. Over the years it has become famous as the "Brandeis die". It is described as follows. Suppose we first consider a fair die. ($P_i = 1/6$.) Let each face be numbered according to the number of spots, so that the average number of spots per throw would be

---

10. Levine, R. D., and Tribus, M. (1979) The Maximum Entropy Formalism, MIT Press, Cambridge, Mass.

11. Rosenkrantz, R. D., (Ed.) (1983) E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics, Reidel Pub. Co., Boston, MA.

$$\sum_{i=1}^{6} (i)\ P_i = 3.5 \tag{25}$$

Of course we have, by definition,

$$\sum_{i=1}^{6} P_i = 1 \tag{26}$$

which says that the sum of all the probabilities add up to unity or certainty. But now consider a loaded die. In particular, suppose that all one knows is that the average number of spots is 4.5, that is

$$\sum_{i=1}^{6} (i)\ P_i = 4.5 \tag{27}$$

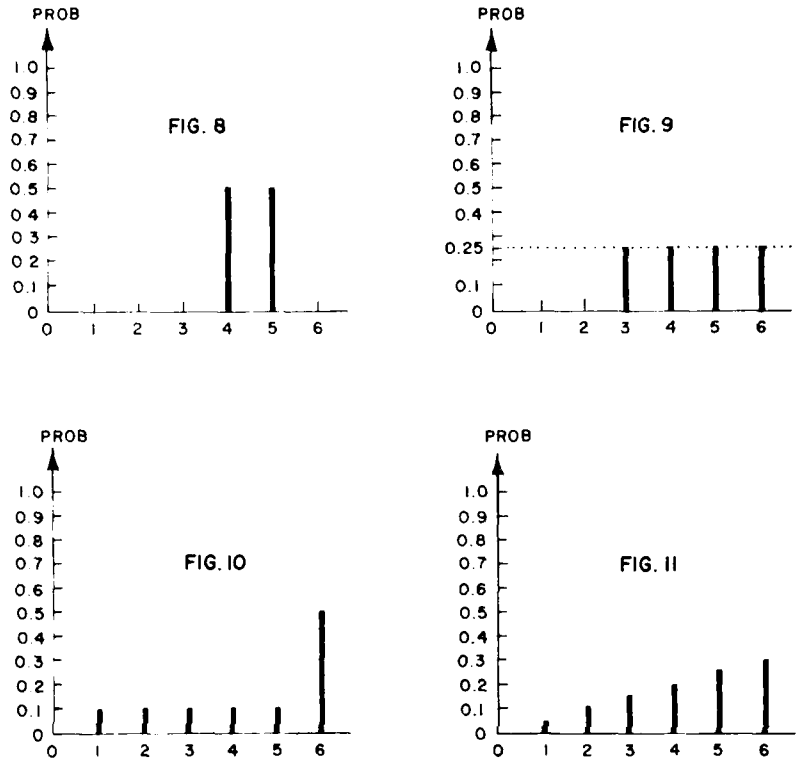
The question is, what probability should be assigned to P_i given the constraints Eqs. (26) and (27)? The Laplace type of approach does not help very much. Certainly it gives no numerical result. Let us explore some possibilities. Figure 8, for example, gives the correct answer, for the average, but it is simply not reasonable. Figure 9 is more honest, but it still jumps to conclusions. Figure 10 is even worse because it is uneven without reason. Figure 11 is the best one up to this point. It is based upon:

$$P_j = \frac{(12j - 7)}{210} \tag{28}$$

Unfortunately, one cannot employ a linear relation and obtain an average greater than 5 without introducing a negative probability. This shows that the general curve must not be linear.

The above sequence illustrates that, to be as objective as possible, one must avoid letting P_i reflect more information than is actually given. In effect, we want P_i to reflect only the known constraints that have been imposed upon it, and then be as uncommittal as possible with regard to anything else. How then does one do this? The clue is given by Shannon's definition of information, S_I, given by

$$S_I = -\sum_i P_i \ln P_i \tag{29}$$



Figures 8–11. Various Assignments of Probabilities for Die Throws. Figure 11 is the most fair of the four, yet it too is not general enough. Text gives explanation

Thus, the prescription of Jaynes for the case at hand is to maximize S in Eq. (29) with the constraints of Eqs. (26) and (27).

If this procedure is compared to the one which is used in statistical mechanics, one will see the connection with probabilities based on the frequency interpretation. The statistical mechanical approach is based on permutations and combinations of states under such constraints as: (a) there is a constant number of particles, (c) there is a given average energy, and so on. Applying this to a die, one would consider the number of ways of getting N_1 throws of single spots, N_2 double spots, and so on, up to N_6 throws of six spots out of a total of N throws. The Bernoulli Formula (so famous in statistical mechanics) is, for this particular case:

$$W = \frac{N!}{N_1!N_2!\ldots N_6!} \tag{30}$$

When N and N_i are all sufficiently large one can use the well known Stirling Approximation for the factorial. This gives us

$$\lim_{N\to\infty} \frac{1}{N} \ln W \cong -\sum_i (N_i/N) \ln (N_i/N) \tag{31}$$

Here the quantities (N_i/N) represent "frequencies". Now, if one maximizes the right side of Eq. (31), one is assigning frequencies that occur the greatest number of ways. This is exactly the method for finding the frequency probability distributions. But this is also exactly what is done by maximizing S in Eq. (29). Thus, it can be seen that the approach of Jaynes is consistent with the frequency interpretation. It is more general in that it can be used even in a case where only one "throw" is involved. It is also much simpler, and can be considered to be a short cut for calculating "greatest number of ways" types of probabilities.

In the following, the Jaynes formalism will be derived and it will be applied to the Brandeis Die problem. As an introduction, however, consider Figure 12, which represents a summary of the formalism. The constraints consist of any number of average values that are given, plus the fact that the probabilities must sum to unity. By maximizing entropy, it turns out that the probability will always have the form of an exponential. Another thing to notice is that this formalism is identical to that found in statistical mechanics, where the quantity Z is called the "partition function". Figure 12 is in essence, the basis of the Jaynes program. It also underlies the MEM technique for the PSD, as will be seen.

3.2 Derivation of Jaynes' Formalism

Jaynes' procedure was to maximize S in Eq. (29) with the constraint conditions

$$\sum_i P_i f_r(\theta_i) = \langle f_r(\theta_i)\rangle \qquad (r = 1, \ldots, n) \tag{32}$$

and

$$\sum_i P_i = 1 = \sum_i f_0(\theta_i) P_i \tag{33}$$

where $f_0 \equiv 1$ and $\langle f_0\rangle = 1$. Note that we used $f_0 = 1$ to make the notation uniform. To maximize entropy (or minimize information) under our constraints, we must

PROBLEM:

TO FIND P_i FOR "i" MUTUALLY EXCLUSIVE AND EXHAUSTIVE OUTCOMES.

i IS AN INTEGER.

GIVEN:

1 – $1 - \sum_i P_i = 1$

2 – AVERAGES OF f_r

$$\sum_i P_i f_r(\theta_i) = \langle f_r(\theta_i) \rangle \quad (1 \le r \le k)$$

$\theta_i \equiv$ STATE OF NATURE

THE MEM ANSWER:

$$P_i = \exp\left[-\lambda_o - \lambda_1 f_1(\theta_i) - \ldots - \lambda_k f_k(\theta_i)\right]$$

WHERE

$$\lambda_o = \ln Z \qquad Z \equiv \sum_i \exp\left[-\lambda_1 f_1 \cdots - \lambda_k f_k\right]$$

$$\langle f_r(\theta_i) \rangle = -\frac{\partial Z}{\partial \lambda_r}$$

Figure 12. The Jaynes Formalism

use Lagrange multipliers. [See Page[12].] Thus, the variation is taken and set equal to zero:

$$\delta[S_I - \mu_0\langle f_0\rangle - \lambda_1\langle f_1\rangle - \ldots] = 0 \tag{34}$$

The μ_0 and λ_i constants are the Lagrange multipliers. Inserting for S and the $\langle f_i\rangle$'s from Eqs. (29), (32), and (33)

$$\delta \sum_j [P_j \ln P_j + \mu_0 P_j + \lambda_1 f_1(\theta_j) P_j + \ldots.] = 0 \tag{35}$$

The variation is taken with respect to P_j. For this reason each square bracket for a given j must vanish. Consider the jth term. We have

12. Page, L. (1935) Introduction to Theoretical Physics, Van Nostrand Co., N.Y., p. 311.

$$\delta P_j \left[\ln P_j + \left(\frac{P_j}{P_j} + \mu_0 \right) + \lambda f_1(\theta_j) + \ldots \right] = 0 \tag{36}$$

If we now call $\lambda_0 \equiv 1 + \mu_0$, we obtain

$$P_j = \exp [- \lambda_0 - \lambda_1 f_1(\theta_j) - \ldots] \tag{37}$$

which gives us the form of the probability and explains why it must be exponential.

The next question is, how does one obtain λ_0? Equation (37) can be rewritten

$$P_j = e^{-\lambda_0} e^{-\lambda_1 f_1(\theta_j)} e^{-\lambda_2 f_2(\theta_j)} \ldots \quad , \tag{38}$$

and Eq. (33) can be written as

$$\Sigma P_j = 1 = e^{-\lambda_0} \sum_j e^{-\lambda_1 f_1(\theta_j)} e^{-\lambda_2 f_2(\theta_j)} \quad \ldots \tag{39}$$

$$\equiv e^{-\lambda_0} Z \quad ,$$

where

$$Z \equiv \sum_j e^{-\lambda_1 f_1(\theta_j)} e^{-\lambda_2 f_2(\theta_j)} \ldots . \tag{40}$$

Hence we have

$$e^{-\lambda_0} = \frac{1}{Z} \tag{41}$$

or

$$\lambda_0 = \ln Z \tag{42}$$

for λ_0.

To obtain λ_1 (and λ_j in general) we first consider

$$\langle f_1 \rangle = \sum_j f_1(\theta_j) P_j = \sum_j f_1(\theta_j)\, e^{-\lambda_0}\, e^{-\lambda_1 f_1(\theta_j)} \ldots .$$

$$\langle f_1 \rangle = \frac{1}{Z} \sum f_1(\theta_j)\, e^{-\lambda_1 f_1(\theta_j)} \ldots . \tag{43}$$

from Eq. (41), hence

$$\langle f_1(\theta_j) \rangle = \frac{-\partial[\ln Z]}{\partial \lambda_1} \tag{44}$$

that is, one can obtain Eq. (43) from Eq. (44) and therefore they are the same.

At this point, Figure 12 has been explained. Considering the number of uses it has found, it is quite simple and straightforward. (See the Jaynes' references in Rosenkrantz[11] for an indication of the scope of the applications.)

3.3 Solution of the Brandeis Die Problem

It is very instructive to work out one case in complete numerical detail. We will solve the Brandeis die problem.

As was shown, P_j is given by

$$P_j = \frac{1}{Z}\, e^{-\lambda_1 f(\theta_j)} = \frac{1}{Z}\, e^{-(j)\lambda_1} \tag{45}$$

where $f(\theta_j) \equiv j$ = the number of spots on a face. The Z is given by

$$Z = \sum_j e^{-\lambda_1 f_1(\theta_j)} = \sum_{j=1}^{6} e^{-j\lambda_1} \tag{46}$$

To simplify calculations define $x \equiv e^{-\lambda_1}$. To reduce the six member sum in Eq. (46) to a simpler expression, we consider

$$\frac{1}{1-x} = 1 + x + x^2 + \ldots \quad , \tag{47}$$

which can be obtained from the geometric series expansion. Equation (47) does not terminate. We wish to obtain the finite sum given by $\sum_{i=1}^{m} x^i$. First consider

$$\frac{1 - x^{m+1}}{1 - x} = 1 + x + \ldots + x^m \tag{48}$$

This is the same as Eq. (47) but the term $x^{m+1}/(1 - x)$ is subtracted from it. To get rid of the number 1 on the right side, we subtract unity from both sides:

$$Z = \frac{1 - x^{m+1}}{1 - x} - 1 \quad (\text{where } m = 6)$$
$$= \frac{x(1 - x^6)}{(1 - x)} \tag{49}$$

The expression for Z is now manageable.

We are given the fact that the average number of spots is 4.5 instead of 3.5. The latter would be the result of a fair die. Thus the formalism shows that, to obtain λ_1, we may use

$$\frac{-\partial(\ln Z)}{\partial \lambda_1} = 4.5 = \frac{1 - 7x^6 + 6x^7}{(1 - x)(1 - x^6)} \tag{50}$$

and a hand calculator can be used to solve Eq. (50) numerically. A more convenient form for Eq. (50) is

$$3x^7 - 5x^6 + 9x - 7 = 0 \tag{51}$$

The solution is $x = 1.449 = e^{-\lambda_1}$, and $\lambda_1 = -0.371$, $Z = 26.66$, and since

$$\left(P_i = \frac{x^i}{Z}\right) \tag{52}$$

$$(P_1, \ldots P_6) = (0.054,\ 0.079,\ 0.114,\ 0.165,\ 0.239,\ 0.347) \tag{53}$$

It should be mentioned that when one makes the transition from the discrete case to the continuous one, the definition of entropy is no longer that given above. Jaynes pointed this out and has shown how to solve the difficulty (Shannon, it seems, did not discuss this difficulty). The solution involves continuous transformation groups. The interested reader may consult Jaynes for further information.

The work of Jaynes remains a matter of some controversy at this time. Prof. Abner Shimoney et al[13] use a rather elaborate mathematical treatment which, they assert, proves that the maximum entropy formalism is not universally correct. In fact, they claim that it works only for a very highly limited set of problems. Their argument seems, to this writer, mainly directed to the applications in statistical mechanics. A mathematical proof that MEM procedures, as used to obtain PSD, are rigorously valid seems to be needed. On the other hand, one can hardly dispute the success of MEM in PSD analysis; and, there is the likelihood that the mathematical objections will be resolved in the future. Furthermore, it will be shown that the manner in which MEM is employed in this report allows it to be replaced by a "least square-error" approach, which in turn leads to exactly the same answers. In this sense one knows that there is little to fear regarding the controversial aspects of the general theory of MEM analysis in the context of the applications to be discussed.*

3.4 The MEM Spectrum

Initially, one is given the autocorrelations, which will be labeled ϕ_k ($-m \leq k \leq +m$) where m is the maximum lag. The Wiener-Khintchine Theorem relates the PSD, $\Phi(\omega)$ to the ϕ_k by

$$\phi_k = \int_{-\omega_N}^{+\omega_N} \Phi(\omega)\, e^{i\omega k \Delta t}\, d\omega \tag{54}$$

where $\omega_N \equiv 2\pi f_N = \pi/\Delta t$ is the Nyquist frequency and Δt is the sampling interval. Eq. (54) gives us our "constraints", and the next step is to maximize the entropy. This entropy, or rather the "entropy rate", is related to the PSD by, essentially,

$$S \equiv \int_{-\omega_N}^{+\omega_N} \ln \Phi(\omega)\, d\omega \tag{55}$$

---

13. Shimony, A., and Dias, P. M. (1981) A critique of Jaynes' maximum entropy principle, Advances in Applied Math. 2, 172-211.

*A discussion of Maximum Entropy as a special case of Maximum Likelihood will be found in Ch. 8 of Deconvolution by P. A. Jansson. This chapter was written by B. Roy Frieden. Perhaps this will limit the objections of Shimony et al.

Derivations of this equation are in Burgs' thesis[17] and in Haykin.[14] These derivations assume that the P_i associated with $\Phi(\omega)$ are "multivariate Gaussian". Jaynes, however, shows that his formalism,[15] (Fig. 12) leads directly to

$$P(y) = \frac{1}{Z} e^{\left[-\Sigma \lambda_k \phi_k(y)\right]} \tag{56}$$

where y is a vector representing the time series. Thus MEM already implies that P(y) is multivariate Gaussian. In other words, it was not really necessary for Burg to assume a "Gaussian random process". Equation (56) is a Gaussian, because the definition of ϕ_k makes the exponent quadratic.

We now perform the variation to get the MEM PSD. Lagrange Multipliers will be used, just as in the previous case.

$$\delta \int_{-\omega_n}^{\omega_n} \left[\ln \Phi(\omega) - \sum_{k=-m}^{m} \lambda_k \left\{ \Phi(\omega)\, e^{-i\omega\Delta t_k} - \phi_k \right\} \right] d\omega \ . \tag{57}$$

The variation is taken over $\Phi(\omega)$. The result is

$$\int_{-\omega_n}^{\omega_n} \left[\frac{1}{\Phi(\omega)} - \sum_{k=-m}^{m} \lambda_k e^{i\omega\Delta t_k} \right] \delta\Phi(\omega) = 0 \ . \tag{58}$$

Again, since each δ (ω) for each ω is independent, each square bracket must equal zero, and this gives

$$\Phi(\omega) = \frac{1}{\sum_{k=-m}^{m} \lambda_k e^{i(\omega\Delta t_k)}} \ . \tag{59}*$$

which is the MEM spectrum.* At this point the values of the λ_i are unknown. To obtain them, one must resort to a technique called Levinson Recursion, which

14. Haykin, S. (Ed.) (1979) Nonlinear Methods of Spectral Analysis, Springer-Verlag.

15. Jaynes, Edwin T. (1982) On the rationale of maximum-entropy methods, Proc. IEEE 70(No. 9):939. The issue is entitled Spectral Estimation.

*The coefficients here are related to the a's given in the MEM equation of the introduction through an autocorrelation.

depends on Toeplitz matrixes. These will be explained below; however, it is first necessary to explain the "Z transformation."

4. THE Z-TRANSFORMATION

The following will provide the reader with the minimum background needed for understanding the remainder of this report. The books by Claerbout, Oppenheim and Schafer, Bracewell, and Robinson and Treitel are suggested for further reading. The text below will, for the most part, follow the treatment of Claerbout. Since the Z-transform forms the basis of the MEM approach to the PSD, it follows that it also explains the basic shape of a spectral line obtained from MEM. This shape is the "Lorentzian" distribution. Since the latter corresponds to the shape found in atomic and molecular radiation, the coincidence appeared to Jaynes as magical. Of course, there is no magic. This result is an inherent feature of the Z-transform formalism.

The Lorentzian shape of a spectral line results, as we shall see, from the location of a "pole near the unit circle" (which relates to damped sinusoidal oscillations) as will be explained. This Lorentzian feature also explains another very important aspect of MEM spectra. Unlike the spectra obtained by conventional methods where the height of the spectral line is proportional to the power, the MEM approach (see Lacoss, [16] and Burg's Ph. D. thesis[17]) results in a line that must be integrated to obtain a quantity proportional to power. Another interesting distinction is that, in conventional spectra, the resolution of the spectrum is determined by the length of the data record whereas in the MEM-PSD formalism it is determined by pole position.

4.1 Definition of the Z-Transform

Consider a function of time sampled at discrete intervals, Δt, which we shall for the sake of convenience take as unity. (See Figure 13.) The Z transform of this time function is obtained by multiplying each ordinate, a_i, of the time series by successively higher powers of Z as in

$$F(Z) = \sum_n a_n Z^n \tag{60}$$

16. Lacoss, R. T. (1971) Data Adaptive Spectral Analysis Methods in Ref. 7, (p. 134), of Main References.

17. Burg, John P. (1975) Maximum Entropy Spectral Analysis, Ph. D. Dissertation, Dept of Geophysics, Stanford University.

where in Figure 13 we have $F(t_n) = a_n$. Another convention consists of using negative powers of Z for the expansion, as mentioned earlier when Fourier's contribution was discussed. The one that will be used here follows that given by Laplace. The discrete Fourier Transform is obtained from Eq. (60) by making the substitution $Z = e^{i\omega \Delta t} = e^{i\omega}$ (since $\Delta t = 1$ here). In some conventions the negative exponent is used in the substitution. (Robinson and Treitel, for example, use this and it will be employed in later sections.)

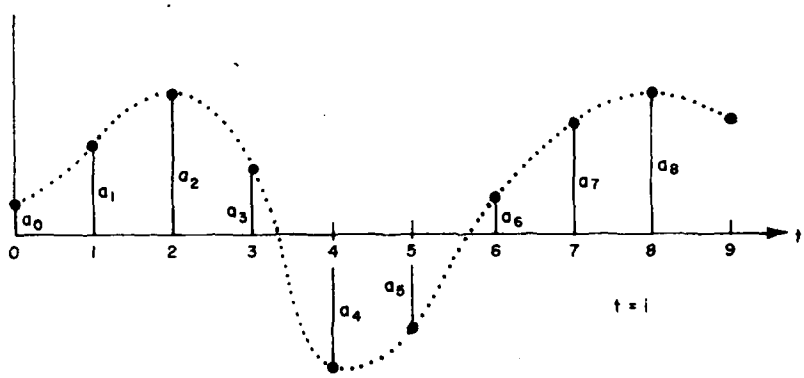


Figure 13. Signal Sampled at Discrete Times. $\Delta t = 1$, thus times are given by integers

The fact that the coefficients of each power of Z in the expansion [Eq. (60)] give the Inverse Fourier Transformation (when $Z = e^{i\omega}$) comes as a surprise to most people. This is because in ordinary Fourier Analysis, the inverse is obtained by calculating an integral which can be difficult to evaluate. In contrast, in the present case one merely needs to "point to a coefficient". Why does the formalism operate in such a simple manner?

To examine the formalism, insert $Z = e^{i\omega}$ in Eq. (60) and call $F(z = e^{i\omega}) \equiv A(\omega)$

$$A(\omega) = \sum_n a_n (e^{i\omega})^n \tag{61}$$

The inverse is defined as (from Fourier Transform theory)

$$a_t = \frac{1}{2\pi} \int_{-\infty}^{\infty} A(\omega)\, e^{-i\omega t}\, d\omega \tag{62}$$

Next we employ a key observation. It is

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} Z^n\, d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{in\omega}\, d\omega = \begin{cases} 1\ , & \text{if } n = 0 \\ 0\ , & \text{if } n \neq 0 \end{cases} \tag{63}$$

Notice that we now limit our integral to the unit circle. This is because we are considering discretized functions with implied Nyquist limits. The function simply repeats outside of these limits. The inverse, a_t, then becomes

$$a_t \equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum a_n \left(e^{i\omega}\right)^n e^{-i\omega t}\, d\omega = a_t \tag{64}$$

Thus, this shows us one of the reasons why the Z transform brings about unexpected simplicities.

$A(\omega)$ in Eq. (61) can be viewed in an alternative manner which gives it an interpretation familiar to people in the field of optics. If $A(\omega)$ were constructed by successively adding terms in Eq. (61) it would take on more "structure" as each time a sample is added. Such a build up of a spectrum is a synthesis method used in some optical instruments, according to G. Vanasse, and the general concept is discussed at the end of the first chapter in Bracewell.[18]

4.2 The Z-Transform Convolution Theorem

We wish to show that, in analogy to the Fourier Transform theory of linear systems (see Lee[4] or Hsu[19]), there is a convolution theorem for the Z-transformation: "The product of two Z-transforms is equal to the Z-transform of the convolution". Figure 14 shows a block diagram for a filter, B, with an input x and output y. Let B(Z) be the Z-transform of the impulse response of the filter

18. Bracewell, R. (1965) The Fourier Transform and Its Applications, McGraw Hill, Chapter 1.

19. Hsu, H. P. (1970) Fourier Analysis, Simon and Schuster (Tech Outlines).

Figure 14. Input X Enters Filter B, and Emerges as Output Y

and let Y(Z) and X(Z) be the Z-transforms of output and input respectively. Then, by analogy to the continuous case, we have

$$Y(Z) = \sum_{n=0}^{\infty} x_n Z^n \cdot \sum_{n=0}^{N} b_n Z^n = \sum_{n=0}^{\infty} y_n Z^n \tag{65}$$

Consider a simple case where the filter's maximum length, N, is given as N = 2 (3 terms). Then

$$\begin{aligned} Y(Z) &= (x_0 Z^0 + x_1 Z^1 + x_2 Z^2 + \ldots)(b_0 Z^0 + b_1 Z^1 + b_2 Z^2 \\ &= x_0 b_0 Z^0 + (x_1 b_0 + x_0 b_1) Z^1 + (x_2 b_0 + x_1 b_1 + x_0 b_2) Z^2 \\ &\quad + (x_3 b_0 + x_2 b_1 + x_1 b_2) Z^3 + \ldots = \sum_{n=0}^{\infty} y_n Z^n \end{aligned} \tag{66}$$

Now identifying the coefficients of powers of Z we obtain

$$y_k = \sum_{n=0}^{2} x_{k-n} b_n \tag{67(a)}$$

or in general

$$y_k = \sum_{n=0}^{N} x_{k-n} b_n \tag{67(b)}$$

The product of the coefficients of powers of Z gives one a convolution in the coefficients. This is an interesting fact. The reason is that, for each power of Z, for y_n on the right, there must be a sum of products on the left all of which involve that same exponent, n, of Z. For this to be true, each left member of the product must have powers that decrease as the right member's power of Z increases so that their sum of powers is equal to n. This is another reason for the Z-transform's strange simplicity.

4.3 The Wiener-Khintchine Theorem

The reader may recall that this theorem states that the Fourier Transform of the autocorrelation is the PSD, and vice versa (see Lee[4].) We now examine the Z-transform version of this since it is part of the MEM procedure.

Consider B(Z) above. We wish to obtain the Z-transform version of a PSD. The PSD is $|B(\omega)|^2$ thus, using $Z = e^{i\omega}$

$$|B(\omega)|^2 \equiv B(\omega)\overline{B}(\omega) = \sum_{n=0}^{\infty} b_n \left(e^{i\omega}\right)^n \cdot \sum_{n=0} \overline{b}_n \left(e^{-i\omega}\right)^{n'} = B(Z)\overline{B}(Z^{-1}) \qquad (68)$$

Thus $B(Z)\overline{B}(1/Z)$ is the Z-transform version of the PSD. Notice that here we have been allowing b_i to be complex for generality. From this point on b_i will be considered real. In this case, the PSD becomes $B(Z)B(1/Z)$. We define the autocorrelation, r_n, as

$$r^k \equiv \sum_n b_n b_{n+k} \qquad (69)$$

Next one multiplies out $B(Z)B(1/Z)$ above to obtain

$$B(Z)B(1/Z) = \sum_{-N}^{N} r^n Z_n = \sum_{n=0}^{N} r^n \left(Z_n + \frac{1}{Z_n}\right) \qquad (70a)$$

To see why this is true, consider again the three term filter. Then

$$B(Z)B(1/Z) = (b_0 + b_1 Z^1 + b_2 Z^2)(b_0 + b_1 Z^{-1} + b_2 Z^{-2})$$

$$= \frac{b_2 b_0}{Z^2} + \frac{b_1 b_0 + b_2 b_1}{Z} + (b_0 b_0 + b_1 b_1 + b_2 b_2) Z^0 \qquad (70b)$$

$$+ (b_0 b_1 + b_1 b_2) Z^1 + (b_0 b_2) Z^2$$

In Eq. (70) one substitutes $Z = e^{i\omega}$ and makes use of the trigonometric identity

$$2 \cos \theta = e^{i\theta} + e^{-i\theta} \quad . \qquad (71)$$

Thus

$$B(Z)B(1/Z) = \sum_{n=0}^{N} r_n [2 \cos(n\omega)] \qquad (72)$$

This is the Wiener Khintchine relation for a real time series [compare the discussion of the Blackman-Tukey approach where Eq. (72) is used to get the PSD]. Again, there is a surprising simplicity here.

4.4 The Z-Transform of $e^{-i\omega_0 t}$

This discussion is designed to serve two purposes: (1) to give one a "feel" for the Z-transform, and (2) give the essential explanation of the MEM Lorentzian line shape. The left side of

$$1 + e^{-i\omega_0} Z + e^{-2i\omega_0} Z^2 + e^{-3i\omega_0} Z^3 + \ldots = \frac{1}{1 - Z e^{-i\omega_0}} \qquad (73)$$

represents the Z-transform of $e^{-i\omega_0 t}$ which is turned "on" at $t = 0$ and is sampled at intervals of $\Delta t = 1$. (Such a sequence is called a "causal" one, since, when the concept is applied to a filter's impulse response, the filter that results operates only on past, and not future, inputs. See Robinson and Treitel.[20] The right hand side is obtained from the Bernoulli or geometric expansion used in reverse (see

20. Robinson, E. A., and Treitel, S. (1980) Geophysical Signal Analysis.

the die problem). The root in the denominator, Z_0, is given by $Z_0 = e^{i\omega_0}$. This root is on the unit circle and therefore, the expansion does not converge. The reader should compare this to the discussion of the Fourier series and the Z-transform above. We avoid division by zero by moving Z_0 an infinitesimal distance ε outside of the unit circle $\left[Z_0 = (1 + \varepsilon) \text{ times } e^{i\omega_0}\right]$. This converts Eq. (73) into

$$B(Z) = \frac{1}{1 - \left(\frac{Z}{Z_0}\right)} = 1 + \sum_{n=1}^{\infty} \left(\frac{Z}{Z_0}\right)^n = 1 + \sum_{n=1}^{\infty} Z^n \frac{e^{-in\omega_0}}{(1 + \varepsilon)^n} \quad . \tag{74}$$

(Note, this is essentially the same as moving the singularity to a point just inside of the unit circle in the Fourier series discussion. A convention, however, was switched here.) To obtain the PSD

$$\overline{B}(1/Z)\, B(Z) \equiv \text{PSD} \quad , \tag{75}$$

we define, for ease in calculation,

$$A(Z) \equiv \frac{1}{B(Z)} = \left(1 - \frac{Z}{Z_0}\right) \tag{76}$$

From Eq. (76) we see

$$A(Z) = 1 - \frac{e^{i(\omega-\omega_0)}}{(1 + \varepsilon)} \quad , \tag{77}$$

where $Z = e^{i\omega}$, and

$$\overline{A}(1/Z) = 1 - \frac{e^{-i(\omega-\omega_0)}}{(1 + \varepsilon)} = \left(1 - \frac{1}{Z\,\overline{Z}_0}\right) \tag{78}$$

The inverse PSD is then

$$A(Z)\,\overline{A}(1/Z) = \left[1 - \frac{e^{i(\omega-\omega_0)}}{(1 + \varepsilon)}\right]\left[1 - \frac{e^{-i(\omega-\omega_0)}}{(1 + \varepsilon)}\right] \quad . \tag{79}$$

Expanding, completing the square, and using $2\cos\theta = e^{i\theta} + e^{-i\theta}$,

$$A(Z)\,A(1/Z) = \left[1 - \frac{1}{(1+\varepsilon)}\right]^2 + \frac{2}{(1+\varepsilon)}\left[1 - \cos(\omega-\omega_0)\right] \quad . \tag{80}$$

The last parentheses can be rewritten, by using a trigonometric identity, as $2\sin^2\frac{(\omega-\omega_0)}{2}$, which in turn, is approximately $\frac{(\omega-\omega_0)^2}{2}$ for small argument. Using Eq. (76), we get the PSD from

$$B(Z)\,B(1/Z) = \frac{1}{A(Z)\,\overline{A}(1/Z)} = \frac{1}{\left[\frac{\varepsilon}{1+\varepsilon}\right]^2 + \frac{4(1+\varepsilon)}{(1+\varepsilon)^2}\left(\frac{\omega-\omega_0}{2}\right)^2} \quad . \tag{81}$$

Thus,

$$B(Z)\,\overline{B}(1/Z) = \frac{(1+\varepsilon)^2}{\varepsilon^2 + 4(1+\varepsilon)\left(\frac{\omega-\omega_0}{2}\right)^2} \sim \frac{1}{\varepsilon^2 + (\omega-\omega_0)^2} \quad , \tag{82}$$

where ε is taken as very small for the right side of the approximation sign. This is the Lorentzian shape as promised.

5. A COMPARISON BETWEEN THE AR AND MEM SPECTRA

There are two purposes for this comparison. The first is to give the reader an intuitive interpretation of the MEM PSD given in Eq. (59). The form of the latter is always strange to someone who is familiar with the conventional approach but not with the MEM formalism. The second purpose is to provide background information to be used below in the discussion of the calculation of the unknown coefficients, λ_k, or their equivalent, in the MEM PSD. In the following part of this section, the approaches of Robinson and Treitel[20] pp. (401-405), and pp. (444-446), as well as Robinson[21] pp. 235-241 will be employed.

First, recall the discussion of Yule's work on auto-regressive (AR) analysis. From Eq. (14), (using y instead of x and switching to subscripts)

21. Robinson, E.A. (1967) Statistical Communication and Detection, Hafner.

$$y_t + \sum_{n=1}^{m} a_n y_{t-n} = \varepsilon_t \tag{83}$$

represented the AR process and ε_t represented noise with mean equal to zero, and variance equal to σ^2. The Z transformation of Eq. (83) is

$$Y(Z)\left(1 + \sum_{1}^{m} a_n Z^n\right) = \varepsilon(Z) \quad . \tag{84}$$

or, with $a_0 \equiv 1$,

$$Y(Z) = \frac{\varepsilon(Z)}{\sum_{m=0}^{m} a_n Z^n} \tag{85}$$

or

$$Y(Z) = \frac{\varepsilon(Z)}{A(Z)} \quad . \tag{86}$$

The PSD is obtained from $Y(Z)\overline{Y}(1/Z)$ and $Z = e^{-i\omega}$, (where we have changed our convention)*

$$|Y(\omega)|^2 = \frac{|\varepsilon(Z)|^2}{|A_m(Z)|^2} = \frac{\sigma^2}{|A_m(Z)|^2} = \frac{\sigma^2}{|A_m(\omega)|^2} \tag{87}$$

($\Delta t \equiv 1$ here) where use was made of Eq. (17) and the fact that the variance of $\varepsilon(Z)$ is σ^2. Equation (87) is the basic AR PSD. The MEM PSD, rewritten here for comparison, is

*The change from $z = e^{i\omega}$ to $z = e^{-i\omega}$ merely changes the sign of the complex exponential in the definition of the Fourier Transform and its inverse. There is no fixed convention in the literature.

$$|\Phi(\omega)|^2 = \frac{1}{\sum_{k=-m}^{m} \lambda_k e^{-i\omega(k\Delta t)}} \qquad (88)$$

Notice the similarity in form between Eq. (88) and the expansion of Eq. (87) [see Eq. (85)]. It will be explained below that they are identical. In any case, the derivation of Eq. (87) given here, gives another, perhaps more intuitively satisfying, derivation of this form.

6. THE TOEPLITZ EQUATIONS FOR THE MEM PSD "COEFFICIENTS"

The constants, λ_k in Eqs. (87) or (88) remain to be determined. The discussions given below and in the next section (Section 7) are devoted exclusively to this subject. Section 6 derives the main equations (Toeplitz) which relate the coefficients to the autocorrelations of the data, and Section 7 describes the ingenious technique for the solution of these equations.

6.1 Derivation of the Toeplitz Equations

According to a theorem by Fejér (which will be discussed more in Section 6.2 and which is discussed at length in Robinson,[21] p. 235, one can write the right hand side of Eq. (88) (in Z-transform form) as

$$\Phi(Z) = \frac{\sigma_m^2}{A_m(Z)\,A_m(1/Z)} \quad , \qquad (89)^*$$

where

$$A_m(Z) \equiv 1 + a_1 Z + \ldots + a_m Z^m \quad . \qquad (90)$$

and the coefficients, a_i, are presumed to be real. Here, $A_m(Z)$ is called the prediction error (P. E.) filter. This is a reasonable term to use since, in Eqs. (84) and (85), $\varepsilon(Z)$ is the prediction error. Further explanation will be given below when Wiener's theory of linear prediction is discussed. Returning to Eq. (89) and the Fejér Theorem, the latter shows that $A_m(Z)$ and $A_m(1/Z)$ can

*In the treatment the convention $\Delta t = 1$ is employed. If it had not been, the numerator would have been $\sigma_m^2 \Delta t$.

put into a form which is called "factored". "Factored" means that one can arrange all the roots of $A_m(Z)$ and $A_m(1/Z)$ in a manner that $A_m(Z)$ has all of its roots outside of the unit circle and $A_m(1/Z)$ has all of its roots inside. This will be discussed more in Section 6.2. (The reader will recall the discussion of the single MEM spectral line and how important the position of a root can be for stability.) Unfortunately it is not possible to give a comprehensive discussion of the concept of "minimum phase" in this report. (For this the reader may consult Robinson and Silvia,[22] Robinson and Treitel[20] and Claerbout.[3]) Nevertheless, it should be mentioned that, when the "factorization" is carried out, $A_m(Z)$ is called "minimum phase" or "minimum delay" (the two terms are synonymous) and, in contrast, $A_m(1/Z)$ is called "maximum delay" or "maximum phase".

The next step in the derivation consists of multiplying both sides of Eq. (89) by A(Z) giving

$$\Phi(Z)\, A_m(Z) = \frac{\sigma_m^2}{A_m(1/Z)} \quad . \tag{91}$$

The left hand side is the product of the Z transforms of Φ and A. It was shown above that this is identical to the Z transform of the convolution of Φ and A. Now consider the right hand side of Eq. (91). We have a constant σ_m^2 divided by $A(1/Z)$, which is a polynomial with all of its zeroes inside of the unit circle. As will be shown in Section 6.2, $[A(1/Z)]^{-1}$ is stable in the expansion of powers of Z^{-n}. The next step is to use the Wiener-Khintchine Theorem in its Z-Transform form

$$\Phi(Z) = \sum_{-\infty}^{\infty} \phi_n Z^n \tag{92}$$

where ϕ_n are the autocorrelations. First, we expand the inverse of A(1/Z) on the right hand side of Eq. (91) to obtain

$$\Phi(Z)\, A(Z) = \sigma_m^2 [1 + (\text{terms in neg. powers of } Z)] \tag{93}$$

and then use Eq. (92) to replace $\Phi(Z)$ on the left side of Eq. (93). Next we equate the coefficients of like powers of Z on both side of Eq. (93) remembering that

22. Robinson, E.A., and Silvia, M.T. (1978) Digital Signal Processing and Time Series Analysis, Holden Day, Inc.

$\phi_n \equiv \phi_{-n}$ for real series. It will not be done here, but it is simple to carry out. The result is m + 1 simultaneous linear equations that can be written in matrix form as

$$\begin{bmatrix} \phi_0 & \phi_1 & \cdots & \phi_{m-1} & \phi_m \\ \phi_1 & \phi_0 & \cdots & & \phi_{m-1} \\ \cdot & \phi_1 & \cdots & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ & & & & \phi_1 \\ \phi_m & \phi_{m-1} & \cdots & \phi_1 & \phi_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ \\ a_m \end{bmatrix} = \begin{bmatrix} \sigma_m^2 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \\ 0 \end{bmatrix} \tag{94}$$

Equation (94) is the fundamental equation that gives the values of a_i and σ_m^2 in the MEM PSD [Eq. (88)]. The matrix on the left side is called a Toeplitz matrix. A Toeplitz matrix has matrix elements t_{ij} that are a function of (i - j). This relationship, which makes all the elements with a constant difference between the indexes identical, gives a Toeplitz matrix a characteristic "banded" appearance, because all terms in the upper-left-to-lower-right diagonals are equal.

Because our autocorrelations are even functions, the matrix representation will be a symmetric matrix, so that this matrix not only has the "banded" appearance characteristic of a Toeplitz matrix, but is also symmetric. Another way to see it is to realize that the bottom row is the reverse of the top row, and the second from the bottom is the reverse of the second from the top, and so on. A Toeplitz matrix gives rise to the "Levinson Recursion" to be described in Section 7.

There are two names given in the literature for Eq. (94). On one hand they (except for the top one) are called the "normal equations". As will be shown, Eq. (94) can be derived from a least squares approach to prediction; and "least squares" always has its "normal equations" as is well known. The second name for Eq. (94) is the "Yule-Walker equations" after the two authors who derived it for their work in AR.

One other comment should be added regarding the right hand side of Eq. (94) which consists of nothing but zeroes with the exception of the first member which is σ_m^2. This structure arose because there were no positive powers of Z on the right side of Eq. (93). This is an unusual situation to a person schooled in the continuous formalism. However, there is a familiar analogue of this, the treatment of causal and anti-causal filters. (For example, see Guillemin.[23]) There,

23. Guillemin, E. A. (1949) The Mathematics of Circuit Analysis, Wiley and Sons; (1963) Theory of Linear Physical Systems, Wiley and Sons.

the analogue to the above treatment is accomplished by means of contour integrations. These contours either contain singularities or they do not. When they do not, the integrals vanish, and vice versa. This underlies for example, the theory of Hilbert Transforms for causal functions. We have here yet another simplification due to the Z-transform. (The interested reader should also consult Lee's[4] treatment of spectral factorization in the continuous domain. Robinson & Treitel[20] also discuss spectral factorization.

6.2 Comments on Factorization and Minimum Phase

The Fejér Theorem was used to get the essential result [Eq. (94)]. The reader may want to know a little more of what lies behind it. The key elements will be described and references will be provided for the full treatment. The first important fact is that the filter represented by $A_m(Z)$ is causal. It starts at $t = 0$ and goes into the future. During convolution, it therefore acts exclusively upon the past inputs. Let us now factor it in the algebraic sense [not in the sense of "factoring" the quantity $A(Z)\,A(1/Z)$ used in the previous section]. This gives us

$$A(Z) \equiv \sum_{n=0}^{N} a_n Z^n = (b_0 + b_1 Z)(c_0 + c_1 Z)(d_0 + d_1 Z) \ldots \quad . \tag{95}$$

To understand a little more about "minimum phase" and "maximum phase", let us take the simplest case of Eq. (95) where only one root is involved. The generalization from this case will not introduce any new concepts. The first question that arises is, When is $[A(Z)]^{-1}$ stable? Using the simple case

$$A(Z) = (b_0 + b_1 Z) \quad , \tag{96}$$

we obtain for the inverse

$$\frac{1}{A(Z)} = \frac{1}{b_0 + b_1 Z} = \frac{1}{b_0(1 - \alpha Z)} = \frac{1}{b_0}(1 + \alpha Z + (\alpha Z) \quad + \ldots) \quad , \tag{97a}$$

where the root $Z_0 = \frac{1}{\alpha}$ and $\alpha \equiv (-b_1/b_0)$. Thus it is obvious that an expansion in positive powers of Z will be stable whenever $\alpha < 1$, which means Z_0 must be outside of the unit circle and, by a definition given above, A(Z) is minimum phase. "Invertability" and "minimum delay" or "minimum phase" are identical in their meaning (p. 157, Robinson and Silvia[22]).

The next question is, when is $[A^*(1/Z)]^{-1}$ stable in an expansion of negative powers of Z? The asterisk denotes that we take the complex conjugate. The reason that we have switched to the general, complex case is that all treatments of the Fejér Theorem (for example, Robinson[21]) use this form. Here one has

$$\frac{1}{A^*(1/Z)} = \frac{1}{\left(b_0^* + b_1^* \frac{1}{Z}\right)} = \frac{1}{b_0^*\left(1 - \frac{1}{\alpha' Z}\right)} = \frac{1}{b_0^*}\left(1 + (\alpha' Z)^{-1} + (\alpha' Z)^{-2} + \ldots\right) \quad . \tag{97b}$$

The root $Z_0^* = 1/\alpha'$, $\alpha' \equiv (-b_0^*/b_1^*)$; and, the expansion converges only when $\alpha > 1$ or Z_0^* is now <u>inside</u> of the unit circle. The generalization of this was used to derive Eq. (94).

The "factorization" of A(Z) A*(1/Z) reduces in essence to the situation shown in Figure 15. Here the two roots Z_i and $1/Z_i^*$ are shown, one inside and one outside of the unit circle. In the general case [Eq. (95)], one can arrange to have all the roots of $|A(Z)|^2$ into two categories, namely, those which are inside and those which are outside of the unit circle. In general, if one is given only $|A(Z)|^2$, then there would be numerous ways of assigning the roots of the component parts, A(Z), and A*(1/Z). Thus, $|A(Z)|^2$ would remain the same no matter how these roots are assigned. On the other hand, <u>making</u> A(Z) minimum delay and A*(1/Z) maximum delay represents a unique assignment with desirable stability properties. The reader who wishes to study the general treatment may consult pp. 235-238 of Robinson[21]; however, the above gives the main ideas for concepts involved.

The concepts of minimum phase and spectral factorization have many implications in data analysis. Chapter 3 of Claerbout[3] explains these concepts in six different ways. The above represents merely one of them. It is also discussed in Section 8.3 of Robinson and Silvia[22] as well as in Ch. 3 and Ch. 7 of their book. In Ch. 7, the so-called Payley-Wiener condition is described in the clearest manner known to the present author. This condition is not only a key concept behind the minimum phase property but it is a condition which, when violated, leads to a "deterministic" time series. This last point is mentioned for a very specific reason here (despite space limitations). The conceptual foundations behind the MEM PSD <u>rule out</u> deterministic signals. The reader can understand this when it is pointed out that the model for the signal is the model consisting of filtered white noise (or the "peas and the pendulum model"). A signal containing, for example, a sine wave (along with everything else such as filtered white noise) will violate our model. It will violate the Payley-Wiener condition. A MEM spectrum of such a signal could, and should, give us nonsense.

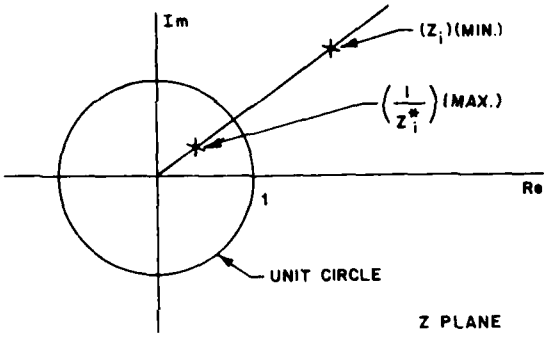


Figure 15. Illustration of Factorization of Roots Z_i Into Minimum Phase and Maximum Phase Pairs

In view of this, then, it is somewhat surprising that many of the "tests" of the MEM PSD method to be found in the literature use sine waves. If the MEM method is discredited by such tests, then, one may perhaps regard such tests as invalid, and it seems that they should be ignored.* *After all, why should this technique be required to work under the very conditions that violate its assumptions?

7. LEVINSON RECURSION (AS MODIFIED BY ROBINSON AND WIGGINS)

This technique is used to solve the Toeplitz Equations, (94). It was first developed by Levinson in order to solve Eq. (94) when it appeared in linear prediction, *as described in Wieners' book on time series.*[24] Levinson's work is reprinted in that book in the form of appendixes. Robinson, in his IEEE review article, has pointed out that these papers of Levinson are masterpieces of explanation and that they made it possible for the engineering community to employ Wiener's work on time series.

This technique of recursion is a basic part of all computer programs for calculating MEM PSD's. It depends, as was mentioned previously, on the Toeplitz symmetry; and the purpose is to obtain σ_m^2 and the a_i in Eq. (94). In

*However, this issue is not resolved at this time. The fact that any sine wave is, in practice, finite may make the Paley-Wiener condition not relevant. On the other hand, Gottman says that the deterministic signals must first be removed in order that there be a valid application of these methods.

24. Wiener, N. (1949) Extrapolation, Interpolation, and Smoothing of Stationary Time Series, Cambridge, MA, Technology Press of the MIT.

order to explain the procedure, we shall commence with a 3×3 version of Eq. (94) and then show how to proceed to the 4×4 case. Consider

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 \\ \phi_1 & \phi_0 & \phi_1 \\ \phi_2 & \phi_1 & \phi_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1^{(2)} \\ a_2^{(2)} \end{bmatrix} = \begin{bmatrix} P_2 \\ 0 \\ 0 \end{bmatrix} . \tag{98}$$

Here the σ_2^2 is replaced by P_2 for convenience. The superscript 2 on the a_i identify them as being associated with a PE filter with a_2 as the highest coefficient.

One now takes advantage of the symmetry (the bottom row is the reverse of the top row of the matrix, and so on). To extend Eq. (98) to a 4×4 matrix, we write

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \phi_3 \\ \phi_1 & \phi_0 & \phi_1 & \phi_2 \\ \phi_2 & \phi_1 & \phi_0 & \phi_1 \\ \phi_3 & \phi_2 & \phi_1 & \phi_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1^{(2)} \\ a_2^{(2)} \\ 0 \end{bmatrix} = \begin{bmatrix} P_2 \\ 0 \\ 0 \\ \Delta_3 \end{bmatrix} , \tag{99}$$

where Δ_3 is to be regarded as a discrepancy. From Eq. (99)

$$\Delta_3 \equiv \sum_{i=0}^{2} \phi_{3-i} a_i^{(2)} , \tag{100}$$

where $a_o^{(2)} \equiv 1$. Now consider the 4th order filter in

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \phi_3 \\ \phi_1 & \phi_0 & \phi_1 & \phi_2 \\ \phi_2 & \phi_1 & \phi_0 & \phi_1 \\ \phi_3 & \phi_2 & \phi_1 & \phi_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1^{(3)} \\ a_2^{(3)} \\ a_3^{(3)} \end{bmatrix} = \begin{bmatrix} P_3 \\ 0 \\ 0 \\ 0 \end{bmatrix} . \tag{101}$$

Next, we use the "trick", based on the symmetry (which allows one to reverse the vectors without reversing the matrix) to rewrite Eq. (99) as

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \phi_3 \\ \phi_1 & \phi_0 & \phi_1 & \phi_2 \\ \phi_2 & \phi_1 & \phi_0 & \phi_1 \\ \phi_3 & \phi_2 & \phi_1 & \phi_0 \end{bmatrix} \begin{bmatrix} 0 \\ a_2^{(2)} \\ a_1^{(2)} \\ 1 \end{bmatrix} = \begin{bmatrix} \Delta_3 \\ 0 \\ 0 \\ P_2 \end{bmatrix} . \tag{102}$$

As can be seen, it is the Toeplitz symmetry which insures that Eq. (102) and Eq. (99) are exactly the same equation. Now let us add to Eq. (99) a certain fraction of Eq. (102), given by C_3, and try to reproduce Eq. (101).

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \phi_3 \\ \phi_1 & \phi_0 & \phi_1 & \phi_2 \\ \phi_2 & \phi_1 & \phi_0 & \phi_1 \\ \phi_3 & \phi_2 & \phi_1 & \phi_0 \end{bmatrix} \left\{ \begin{bmatrix} 1 \\ a_1^{(2)} \\ a_2^{(2)} \\ 0 \end{bmatrix} + C_3 \begin{bmatrix} 0 \\ a_2^{(2)} \\ a_1^{(2)} \\ 1 \end{bmatrix} \right\}$$

$$= \left\{ \begin{bmatrix} P_2 \\ 0 \\ 0 \\ \Delta_3 \end{bmatrix} + C_3 \begin{bmatrix} \Delta_3 \\ 0 \\ 0 \\ P_2 \end{bmatrix} \right\} \rightarrow \begin{bmatrix} P_3 \\ 0 \\ 0 \\ 0 \end{bmatrix} . \tag{103}$$

Our unknowns are P_3 and C_3 on the right side. To make the bottom element vanish, we have,

$$\Delta_3 + C_3 P_2 = 0 , \tag{104}$$

or

$$C_3 = \frac{-\Delta_3}{P_2} . \tag{105}$$

To obtain agreement for the top element on the right,

$$P_3 = P_2 + C_3 \Delta_3 . \tag{106}$$

$$P_3 = P_2 \left[1 - \left(\frac{\Delta_3}{P_2} \right)^2 \right] = P_2 \left[1 - |C_3|^2 \right] \quad , \tag{107}$$

where Eq. (105) was used. It must be recalled that Δ_3 is defined by Eq. (100), that is,

$$\Delta_3 \equiv \sum_{j=0}^{2} \phi_{3-i} \, a_i^{(2)} \quad . \tag{108}$$

Similarly, using Eq. (101) in connection with Eq. (103) (left side)

$$\begin{bmatrix} 1 \\ a_1^{(3)} \\ a_2^{(3)} \\ a_3^{(3)} \end{bmatrix} = \left\{ \begin{bmatrix} 1 \\ a_1^{(2)} \\ a_2^{(2)} \\ 0 \end{bmatrix} + C_3 \begin{bmatrix} 0 \\ a_2^{(2)} \\ a_1^{(2)} \\ 1 \end{bmatrix} \right\} , \tag{109}$$

or

$$a_s^{(3)} = a_s^{(2)} + C_3 \, a_{3-s}^{(2)} \quad , \qquad (s = 0, 1, 2, 3) \quad , \quad a_0^{(i)} \equiv 1 \tag{110}$$

The above recursion from the 3×3 to the 4×4 case can be generalized in an obvious way to

$$C_N = \frac{-\Delta_N}{P_{N-1}} \quad ,$$

$$\Delta_N = \sum_{n=0}^{N-1} a_n^{(N-1)} \, \phi_{N-n} \quad , \tag{111}$$

$$P_N = P_{N-1}(1 - |C_N|^2) \quad ,$$

$$a_s^{(N)} = a_s^{(N-1)} + C_N \, a_{N-s}^{(N-1)} \quad , \quad (s = 0, 1, \ldots, N) \quad , \text{ and } \quad a_N^{(N)} = C_N \quad , \quad a_N^{N-1} = 0$$

The recursion is initiated by (using a 1 by 1 matrix and the general formulas)

$$
\begin{aligned}
P_0 &= \phi_0 \quad , \\
\Delta_1 &= \phi_1 a_0^{(0)} = \phi_1 \quad , \\
C_1 &= \frac{-\Delta_1}{P_0} = \frac{-\phi_1}{\phi_0} \quad , \\
P_1 &= P_0(1 - |C_1|^2) \quad , \\
a_1^{(1)} &= a_0^{(0)} C_1 = C_1 = \frac{-\phi_1}{\phi_0} \quad , \\
a_0^{(0)} &= a_0^{(1)} = 1 \quad .
\end{aligned}
\tag{112}
$$

A treatment is given in Haykin[14] (Section 2.4) that parallels this treatment and which the reader may consult. The quantities C_N play a crucial role in what is to follow below. They are called "reflection coefficients".

8. WIENER PREDICTION AND LEAST SQUARES

The following treatment follows Robinson and Treitel.[20] The original work is in Wiener's "Time Series",[24] and explanations of the continuous version are given in Lee,[4] Laning and Battin[25] and in Bendat.[26] The present treatment will serve a number of purposes: (1) it will give an alternative way to understand Eq. (94), that is, the Toeplitz Equations, (2) it will explain better why $A_m(Z)$ is called the "prediction error" (PE) filter, (3) it will illustrate why the MEM PE filter whitens the spectrum until, at some order of $A_m(Z)$, the PSD is flat, (4) by showing that both MEM and least squares lead to Eq. (94) one can convince oneself that the objections of Shimony et al to MEM do not apply to the MEM PSD as we know it, and (5) it will provide basic information needed in order to understand the so-called "Burg Technique" to be explained in Section 10.

---

25. Laning, J.H., and Battin, R.H. (1956) Random Process in Automatic Control, McGraw-Hill.

26. Bendat, J.S. (1958) Principles and Applications of Random Noise Theory, Wiley & Sons.

Figure 16 shows a block diagram that describes the problem to be solved. The input, x_t, is given. We wish to obtain predicted values of x_t at a time α later, or, in other words, we want $x_{t+\alpha} \equiv z_t$. To accomplish this, one has a filter described in terms of its weighting coefficients, f_t. This filter is to be designed so that its output will be as close as possible to z_t.

Figure 16. Prediction Filter

To design the filter, Wiener chose to use a "least square error" criterion, and this is the basis of the entire procedure of the present section. The mean square error is given by J, defined by

$$J \equiv \overline{(z_t - y_t)^2} \tag{113}$$

One now applies the convolution theorem to describe the effect of the filter in Figure 16.

$$y_t = x_t * f_t \equiv \sum_{\tau=0}^{m} f_\tau \, x_{t-\tau} \quad . \tag{114}$$

where the asterisk indicates convolution. We will use Eq. (114) in Eq. (113) for y_t in the following. To obtain the optimum values of f_i, one minimizes J by setting the derivative of J with respect to each f_i equal to zero:

$$\frac{\partial J}{\partial f_i} = 0 \quad . \tag{115}$$

Before carrying this out, we define the cross correlation, $\phi_{xy}(\tau)$, by

$$\phi_{xy}(\tau) \equiv \overline{x_{t+\tau} \, y_t} \quad . \tag{116}$$

If Eq. (115)[with Eq. (114) and Eq. (113) substituted for y_t and J] is carried out with i = 1, and if one makes use of the definition [Eq. (116)] for $\phi_{xz}(\tau)$, one obtains

$$\sum_{\tau=0}^{m} f_\tau \, \phi_{xx}(1 - \tau) = \phi_{zx}(1) \quad . \tag{117}$$

Here we used

$$\phi_{zx}(1) = \overline{z_t x_{t-1}} \quad \text{and} \quad \phi_{xx}(1 - \tau) = \overline{x_{t-\tau} x_{t-1}} \quad . \tag{118}$$

For the general case of $\partial J / \partial f_i = 0$ one obtains

$$\sum_{\tau=0}^{m} f_\tau \, \phi_{xx}(j - \tau) = \phi_{zx}(j) \quad , \qquad (j = 0, 1, \ldots, m) \quad . \tag{119}$$

This equation is the discrete version of the Wiener-Hopf Equation (compare with Lee[4] and Haykin[14]) and it constitutes the "normal equations" of the least square formulation here.

Eq. (119) is also used to give a more convenient expression for J, namely

$$J = \phi_{zz}(0) - \sum_{\tau=0}^{m} f_\tau \, \phi_{zx}(\tau) \quad . \tag{120}$$

[Details of the algebraic manipulations left out above will be found in Robinson and Treitel,[20] p. 147.]

We now specialize to the one step predictor, or

$$z_t = x_{t+1} \quad . \tag{121}$$

In other words, we now wish to predict x_t one unit of time in the future, that is, x_t is the input and we want x_{t+1} as the output of the filter. For this case we have

$$\phi_{zx}(j) \equiv \phi_{xx}(j + 1) \quad , \tag{122}$$

and the normal equations, Eqs. (117), become

$$\sum_{\tau=0}^{m} f_\tau \, \phi_{xx}(j - \tau) - \phi_{xx}(j + 1) = 0 \quad . \tag{123}$$

Thus Eq. (123) is the Wiener-Hopf one step predictor equation which, when solved, gives the optimum coefficients, f_i.

At this point we shall change the notation to the "prediction error" notation for reasons which will become apparent. This is done as follows

$$\begin{aligned} a_0 &= 1 \quad , \\ a_1 &= -f_0 \quad , \\ a_2 &= -f_1 \quad , \qquad J \equiv J_{m+1} \equiv \sigma_N^2 \quad . \\ &\cdot \\ &\cdot \\ &\cdot \\ a_N &= -f_N \quad , \end{aligned} \tag{124}$$

where $N \equiv (m + 1)$, that is, we now have an Nth order PE filter.

In terms of this notation, Eqs. (123) become

$$\sum_{\tau=0}^{m} (-a_{\tau+1} \, \phi_{xx}(j - \tau) - a_0 \, \phi_{xx}(j + 1) = 0 \quad . \tag{125}$$

We now define $\mu \equiv \tau + 1$ and multiply Eq. (125) through by (-1) obtaining

$$\sum_{\mu=0}^{N} a_\mu \, \phi_{xx}(j + 1 - \mu) = 0 \quad , \qquad (j = 0, 1, \ldots N\text{-}1) \quad . \tag{126}$$

In the new notation we have Eq. (120) in the form

$$\sigma_N^2 = \sum_{\mu=0}^{N} a_\mu \, \phi_{xx}(\mu) \tag{127}$$

If one combines Eq. (127) with Eq. (126) one obtains the previous Eqs. (94):

$$\begin{bmatrix} \phi_{xx}(0) & \cdots & \phi_{xx}(N) \\ \phi_{xx}(1) & & \phi_{xx}(N-1) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \phi_{xx}(N) & & \phi_{xx}(0) \end{bmatrix} \begin{bmatrix} a_0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_N \end{bmatrix} = \begin{bmatrix} \sigma_N^2 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} . \tag{128}$$

Previously we used $\phi_{xx}(j) \equiv \phi_j$.

This, Eq. (128), is a very interesting result because we now see that it can be obtained from least squares analysis without any help from the "higher principle" of maximum entropy. This implies that MEM and least squares have a deep connection. To this writer's knowledge, this connection has not yet received an explanation.

At this point, let us reconsider the question of why we have been calling the a_i in Eq. (124) the PE filter as contrasted to the Wiener prediction filter. The latter is given by (one step)

$$y_t \equiv \hat{x}_{t+1} = \sum_{k=1} f_k \, x_{t-k+1} \quad . \tag{129}$$

This is modeled in Figure 17a by the part of the system within the dotted lines. In contrast, there is the prediction error filter. This is given by

$$e_t = x_t - \hat{x}_t = x_t - \sum_{k=1}^{m} f_k x_{t-k} \equiv \sum_{j=0}^{m} a_j x_{t-j} \quad . \tag{130}$$

This is represented by the whole of Figure 17a or its re-representation in Figure 17b. Equation (130) explains the change of notation given in Eq. (124). The output of the PE filter is the error, and when the error is white, the filter has embodied all the predictable parts of the input signal. In effect, the predictor filter represents the inverse of the filter which is used in the signal model as depicted in Figure 4.

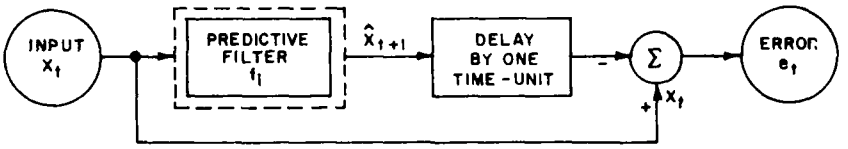


Figure 17a. Wiener Filter, f_i

Figure 17b. P. E. Filter, a_i

Lee[4] (p. 410) gave a very useful physical description of the Wiener prediction filter. (Note that the general form of Wiener's Optimum Filter did more than predict. *It also did the best job possible, in a sense, to get rid of noise.*) Figure 18 gives this description. At the top of the figure is given the "model" for the signal (or data). It consists of a random pulse generator which serves as input to a filter or "pulse shape network" (they are the same thing). Note the impulse response drawn on the filter block. Up to this point we have what was in Figure 4, but with more detail. Next, we consider the Wiener-prediction filter. The signal first enters the pulse shape inverter, or the inverse of the filter used in the signal model. The effect of this first filter is to take signal inputted to the inverter and convert it back into the original pulse noise of the signal model (that is, the output of the top left block). These pulses are then fed to a new filter or pulse shape network that differs from the one of the signal model in only one respect. The initial part of the impulse response is missing, and the amount missing corresponds to the time elapsed in the predictive time interval which we have been taking as one time step. Notice that if one were to try to predict too far into the future, the impulse response might be too small to be of any value. Eventually it would die, and the predictor would predict the value zero.

The above physical description seems to be quite helpful. The reader should consult Lee[4] to see the proof that the above is correct. What is most remarkable, in the opinion of the present writer, is the fact that such an ingenious "device" was automatically "invented" by "least squares" mathematics. As we have seen, maximum entropy accomplishes the same thing.

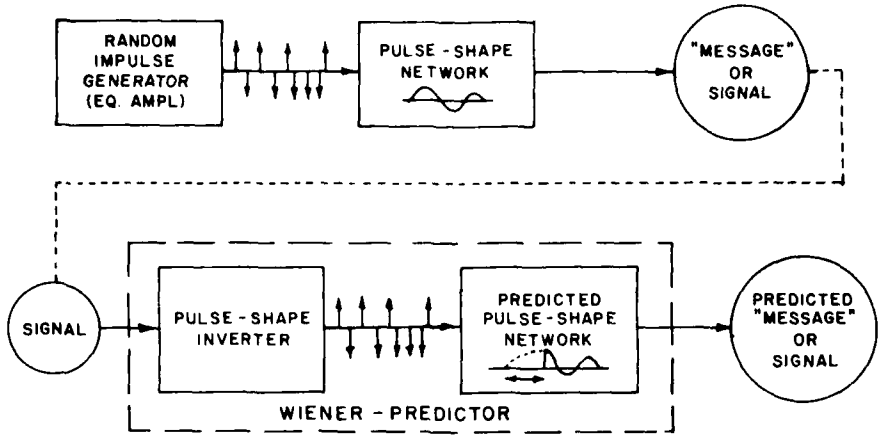


Figure 18. Y.W. Lee's Diagram of the Prediction Filter. Note that the prediction is zero for times in the future that exceed the duration of the impulse response

9. GOOD AND BAD CANDIDATES FOR MEM

Not all data lend themselves favorably to MEM-PSD analysis. There are three main situations which make a specific set of data a bad candidate: (1) The MEM signal model, that is, the AR model, does not correspond to the data source. (As will be shown, there are two other models.) (2) The data are too noisy. (MEM-PSD analysis is especially vulnerable to noise.) And (3), the autocorrelation has decayed to the point where no statistically reasonable extension of it will be above the noise level. Case (2), as will be shown, is closely related to (1). Case (3) is true because, as Burg has shown, one way to understand how MEM "works" is to think of it as a way to extend the autocorrelation from the given values to all the values that lie outside the data range. In other words, the linear predictor implied in the MEM analysis is, in effect, used to extend the autocorrelation from $-\infty$ to $+\infty$. If such prediction yields nothing new, then nothing will be gained by MEM analysis over the conventional approach.

9.1 The Various Signal Models

In the following we shall follow Robinson and Treitel,[20] Ch 16.

9.1.1 THE AR MODEL

As we have seen, where x_t is input and y_t is output, the model is given by

$$y_t = x_t - \sum_{i=1}^{N} a_i\, y_{t-i} \quad . \tag{131}$$

The model uses white noise as input, x_t. In the PE form, it is

$$y_t + \sum_{i=1}^{N} a_i\, y_{t-i} = \sum_{i=0}^{N} a_i\, y_{t-1} = x_t \quad , \tag{132}$$

where $a_0 = 1$. Taking the Z-transform of Eq. (132) yields

$$Y(Z)\, A(Z) = X(Z) \quad , \tag{133}$$

where

$$A(Z) \equiv \sum_{i=0}^{N} a_i\, Z^i \quad . \tag{134}$$

We used the convolution theorem to get Eq. (133). Now divide both sides of Eq. (133) by A(Z),

$$Y(Z) = \frac{X(Z)}{A(Z)} \quad . \tag{135}$$

From this one can get the PSD version, along the lines of our previous discussions, by means of $|A(Z)|^2 \equiv A(Z)A^*(1/Z)$ and so forth, and arrive at

$$|Y(Z)|^2 = \frac{|X(Z)|^2}{|A(Z)|^2} \xrightarrow{(Z = e^{-i\omega})} \frac{|X(\omega)|^2}{|A(\omega)|^2} \quad . \tag{136}$$

The above is somewhat in the category of "review" material and we have seen that the poles [or zeroes of A(z)] give us the "lines" in the spectra. In this connection the reader should recall the discussion of the basic "MEM line" and its Lorentzian property.

9.1.2 THE MA MODEL

Let us now consider a new model. This is related to the previous discussion of the Slutzky Effect. It is called the moving average (MA) model which is given by

$$y_t = \sum_{n=0}^{N} b_n x_{t-n} \tag{137}$$

Here it is the input values that are weight averaged by the filter. (AR corresponds to "feedback" and MA to "feedforward" if one wishes to use engineering terminology). As before, one takes the Z-transform of Eq. (137):

$$Y(Z) = B(Z)\, X(Z) \quad , \tag{138}$$

and obtains, going through the same manipulations as above:

$$|Y(\omega)|^2 = |B(\omega)|^2 \, |X(\omega)|^2 \quad . \tag{139}$$

Here it is the shape of $|B(\omega)|^2$ that gives the spectrum its form since $|X(\omega)|^2$ is "white" or constant. The Blackman and Tukey approach is "MA".

9.1.3 THE ARMA MODEL

This model is a combination of the above two models. In this case the model is given by

$$y_t = \sum_{n=0}^{N} b_n x_{t-n} - \sum_{i=1}^{M} a_i y_{t-i} \quad . \tag{140}$$

and, for obvious reasons,

$$|Y(\omega)|^2 = \frac{|B(\omega)|^2}{|A(\omega)|^2} \, |X(\omega)|^2 \quad . \tag{141}$$

Makhoul's tutorial[27] calls this a "pole-zero model" because $|B(\omega)|^2$ can supply zeroes, and the zeroes of $|A(\omega)|^2$ give poles since A is in the denominator. He also discusses the AR and MA models.

---

27. Makhoul, J., "Linear Prediction: A Tutorial Review", Ref. 7 of Main Ref. p. 99.

At this point, Robinson and Treitel make a very important observation. First they made computer simulations based on the AR model, the MA model, and the ARMA model. Then they obtained the PSD's for the three cases based on the known signal models. Next they used procedures to obtain the PSD's from the data themselves. The procedures were based, in each case, on the three different models, that is, MEM for AR, Blackman and Tukey for MA, and a procedure known as "Box Jenkins" for ARMA. Figure 19 shows their results. In each case where the simulation model matched the method of analysis, the match is perfect or very good. In the other cases, the results could be invalid. In particular, it should be pointed out that AR gave a narrow line whereas the MA treatment gave the correct shape in the case where the signal was generated by a MA process. Thus, we see that AR could give a false "high resolution" in this case.

MEM works best when the AR model fits the data. Fortunately, there is some relief from this problem. A theorem has been proven in the literature which says that all the methods converge to the same PSD result provided that sufficiently high order is used. It may therefore be much more practical to use MEM for certain types of ARMA case, since the latter is relatively difficult to carry out.

9.2 Noise

As Kay and Marple[28] have pointed out most clearly in their excellent review article, noise is very effective in corrupting MEM spectra because it has the effect of changing the signal model. To see this, let y_t be a noise corrupted AR process called x_t. Then

$$y_n = x_n + w_n \quad , \tag{142}$$

where w_n represents the white noise of variance σ_w^2 and is uncorrelated. The PSD of y, called $P_y(Z)$ in Z transform notation, is given by

$$P_y(Z) = \frac{\sigma^2 \Delta t}{A(Z)\, A^*(1/Z)} + \sigma_w^2 \Delta t \quad . \tag{143}$$

This expression should be clear from the previous discussions. The Δt here simply gives one a different dimensional form that is sometimes useful, and it is the form Kay and Marple used. This can be rewritten as

28. Kay, S. M., and Marple, S. L. (1981) Spectrum analysis—a modern perspective, Proc. IEEE 69, pp. 1380-1419.

$$P_y(Z) = \frac{\left\{\sigma^2 + \sigma_w^2 \left(A(Z)\ A^*(1/Z)\right)\right\} \Delta t}{A(Z)\ A^*(1/Z)} \quad . \tag{144}$$

This result is very informative because it tells us that, merely by adding noise, an AR process can be converted into an ARMA process. Note that it now has both poles and zeroes!

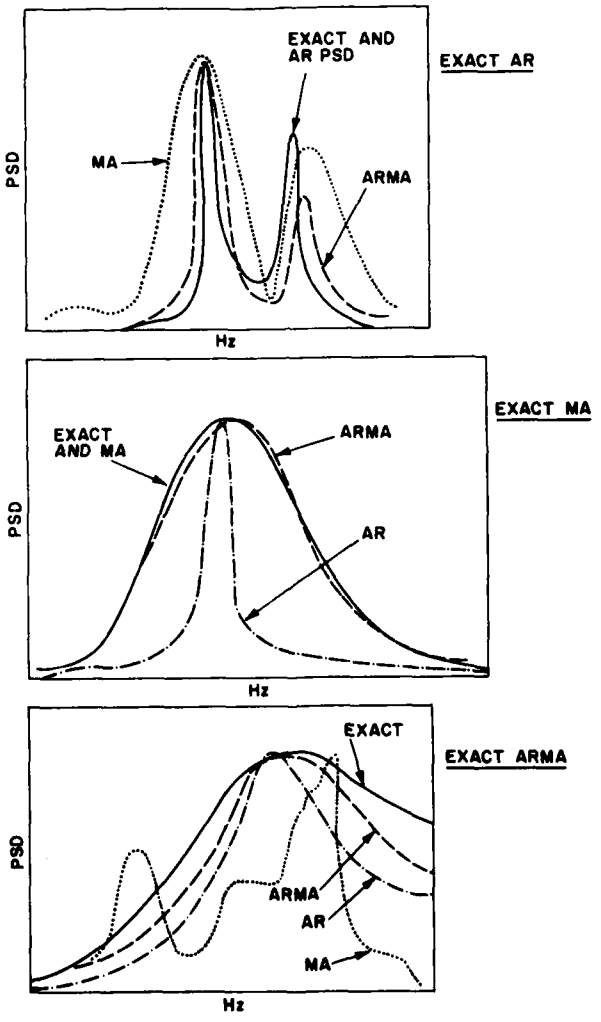


Figure 19. Synthesis and Analysis of AR, MA, and ARMA Signals (After Robinson and Treitel[18])

Figure 20 (taken from Kay and Marple) shows the result of an experiment which involves two values of added noise. When the Signal to noise ratio is high, the lines are very clearly resolved (left side). Lowering the signal to noise ratio (SNR) from 25 db to 5 db washes out all the good resolution. One concludes that "noise is very bad for MEM spectral analysis."

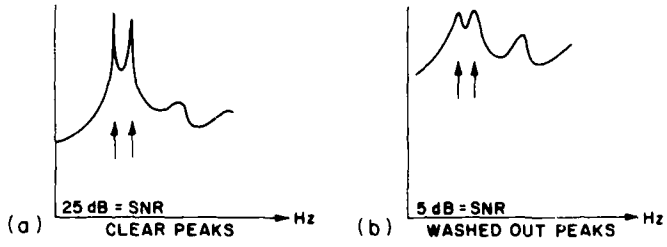


Figure 20. Effects of Noise on a MEM Spectrum (after Kay and Marple[28])

At this point it is appropriate to quote two passages from Jaynes[15] on this subject. He says that MEM is "... a precision, high-performance machine ... and can deliver that high performance only when fed high quality fuel". By this he means that one must feed it exact autocorrelations, or rather they must be as noise free as possible. A little noise can push the poles further away from the unit circle. He goes on to say that MEM "fails to take noise into account, a factor that orthodox methods do deal with usefully, and sometimes ever optimally". He then called for a "full Bayesian Solution" and went on to point the way for one to proceed.

Abels[29] in a review article mentioned an interesting idea. He suggested maximizing entropy

$$\int_{-\infty}^{\infty} \ln P(\omega)\, d\omega \quad , \tag{145}$$

29. Abels, J.G. (1974) Maximum entropy spectral analysis, Reprinted in Childers.[30]

30. Childers, D.G. (1978) Modern Spectrum Analysis, IEEE Press.

but with an alteration of the constraint to the following form

$$\frac{1}{N} \sum_{n} \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega)\, e^{-i\omega n}\, d\omega - \phi_n \right|^2 = \sigma^2 \quad . \tag{146}$$

The difference between this form and the conventional one of MEM is that here one has σ^2 instead of zero and the σ^2 is taken to represent the uncertainty introduced by the noise. The above makes use of Abel's continuous notation, however, the reader is reminded that (a) implementation requires a discrete formulation and (b) Eq. (145) is not valid in the continuous case as was pointed out by Jaynes and as has already been mentioned.

Of course, if one must cope with noise, one can always increase the order of A(Z) to compensate for the change in the signal model brought about by that noise.

9.3 The Limiting Cases Test

Finally, we consider case (3) which was first suggested to the present writer by George Vanasse. Figure 21 illustrates this case. Qualitatively the argument goes as follows. If one is given a "short" autocorrelation in the sense that it looks as if it would "go on" in a similar fashion, then it would appear from Figure 21a that one has a good candidate for MEM (forgetting for the moment the problems already discussed in cases (1) and (2). In contrast, suppose one has a case such as that given in Figure 21b. The latter represents an autocorrelation that slowly "dies" and which, so far as one can know, looks as if it has permanently decayed. As has been mentioned, MEM linearly predicts the autocorrelation to cover all values of time. What sort of prediction would one expect in the predicted extensions of this autocorrelation? From the physics of the linear prediction process as shown in Figure 18, it is clear that the predictor would predict zeroes. In other words MEM cannot give something for nothing.

We conclude this section with Figure 22 which shows what happens when one performs various types of analysis. "MEM via Burg Algorithm" will be described in the next section. "Yule-Walker" or YW refers to obtaining the coefficients from the autocorrelations in the manner described above. The figure is from Kay and Marple.[28]

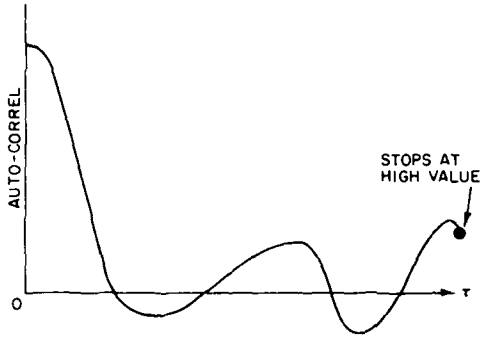


Figure 21a. An Autocorrelation That Is a Good MEM Candidate

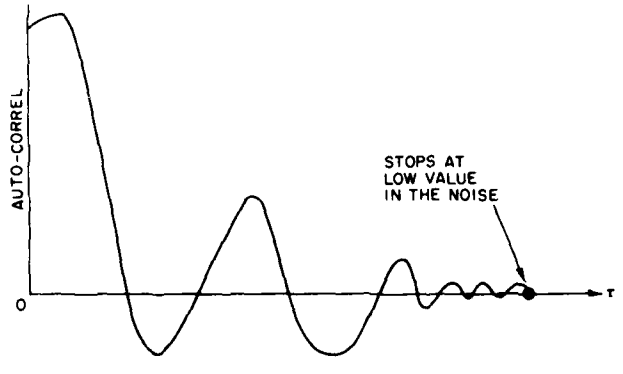


Figure 21b. An Autocorrelation That Is a Bad MEM Candidate

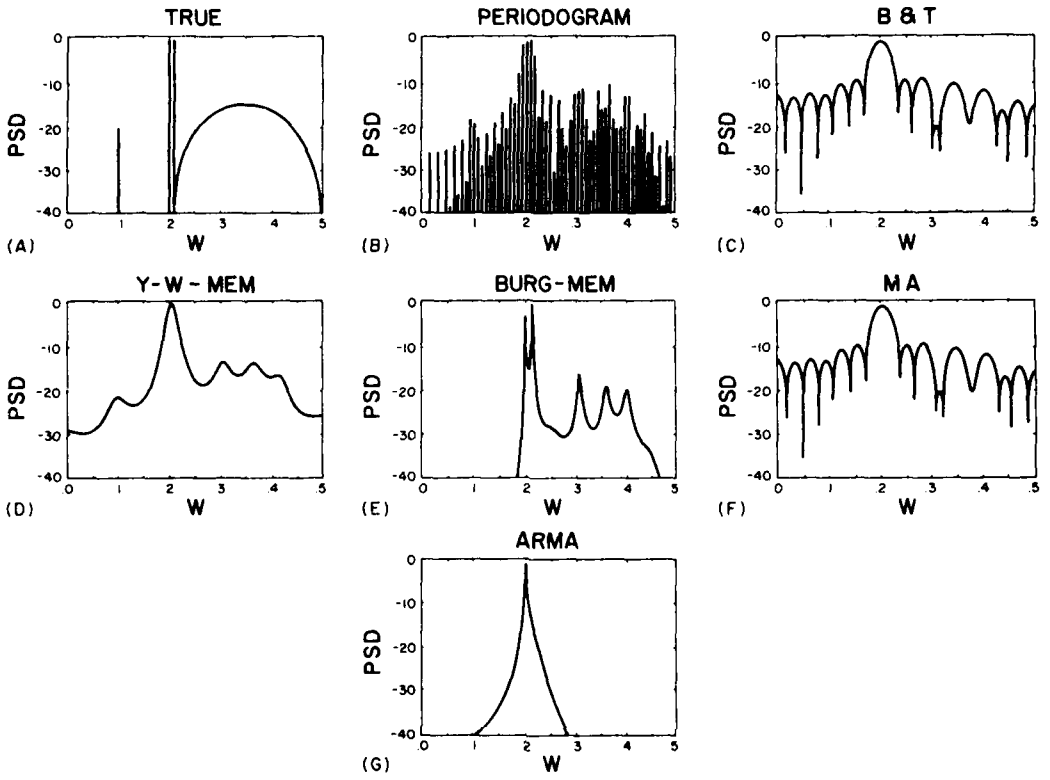


Figure 22. Various PSDs and the True PSD After Kay and Marple

10. THE BURG TECHNIQUE

The MEM method for obtaining PSD's was of course invented by Burg. However, he is also the inventor of the "Burg Technique" which is used to obtain the a_i coefficients. It is important to make a distinction between these two contributions of Burg to avoid confusion.

Burg was very interested in the fact that the series of reflection coefficients (together with the zero lag autocorrelation) represents a new way to embody the "second order statistics" of a time series. He thought that it could be more fundamental than the usual methods, namely the PSD and the autocorrelation. In addition, his choice of the word "reflection coefficient" for the C_N's rested on the fact that, in the analysis of seismological signals where acoustic waves pass through many layers of earth, the resulting C_N's are indeed the reflection coefficients of these layers. In fact, the "Burg technique" can be regarded as a "wave

propagation model" where the forward and backward predictors play the roles of upward and downward going waves (Claerbout, [3] p. 136).

Our motives for presenting the "Burg technique" are (1) it is the one most often used in MEM-PSD analysis, and (2) we used it to obtain our experimental results to be described in Section 11. It should be mentioned that, in the context of Fourier spectroscopy, the Yule-Walker method, which relies on the autocorrelations themselves, would seem to be the ideal method at first sight. This is of course due to the fact that the autocorrelations themselves are the raw data of the instrument. As one will see in Section 11, there are important exceptions to this rule. In the following we shall use Burg's own description of his method, but there are many other treatments to be found in the literature (such as Robinson and Treitel, [20] Claerbout[3] and Haykin[14]).

Before we consider the mathematical details (which can be omitted by the reader not interested) it is of interest to list, qualitatively, the most important features of this method.

1. The Burg technique obtains the C_N's (as well as the a_i's) directly from raw time series data.

2. This method completely avoids the problem which arises when one calculates an autocorrelation from a finite piece of data. The reader will recall that the autocorrelation is calculated from

$$C_r = \frac{1}{N - r} \sum_{q=0}^{q=N-r} x_q \, x_{q+r} \tag{147}$$

if one follows Blackman and Tukey. It is clear from this that when r gets close to N, the overlap becomes small and serious difficulties can be expected. This problem is particularly acute when one has a short data sample or a collection of such samples. This feature implies that, when one has a number of short pieces of data, then the MEM-PSD method, together with the Burg technique, is the best available route to the PSD.

3. This method incorporates the Levinson Recursion in its machinery.

4. The mean square error of both the forward predictor and backward predictor is minimized, or rather it is the average of both that is minimized.

5. The stability of the PE filter is guaranteed. This is not true for the Yule-Walker case.

6. Given raw data, this method would provide a much higher resolution for the PSD than would the Yule-Walker (YW) case.

We now consider the details. The best method to describe this recursive procedure is to consider the case where one has at the start the N-1th PE filter

$$A_{N-1}(Z) = 1 + a_1^{(N-1)} Z + a_2^{(N-1)} Z^2 + \ldots + a_{N-1}^{(N-1)} Z^{N-1} \tag{148}$$

and where one seeks $A_N(Z)$. We know from Levinson Recursion that

$$A_N(Z) = 1 + \left(a_1^{(N-1)} + C_N a_{N-1}^{(N-1)}\right) Z + \ldots + \left(a_{N-1}^{(N-1)} + C_N a_1^{(N-1)}\right) Z^{N+1} + C_N Z^N \quad . \tag{149}$$

where C_N is still unknown. It can be obtained from the data in the following way.

Suppose that the data consist of a number of short samples, for example $(x_0, x_2, \ldots x_{10})$, $(x_{100}, x_{101}, \ldots, x_{110})$ and so on; and let us assume that one has a total of M sets of these N+1 tuplets (where N = 10 for the two examples). The key calculation is to minimize the mean square error ε^2, given by

$$\overline{\varepsilon^2} = \overline{[A(Z)X(Z)]^2} \quad . \tag{150}$$

The value of this quantity over all the pieces of data is given by (m refers to the m^{th} (N+1) - tuplet, $(x_1, m, x_2, m, \ldots, x_{N+1}, m)$ where m = 1 to M.)

$$\overline{\varepsilon^2} = \sum_{m=1}^{M} W_m \left| \left[C_N x_{1,m} + \left(a_{N-1}^{(N-1)} + C_N a_1^{(N-1)}\right) x_{2,m} + \ldots \right.\right.$$
$$\left.\left. + \left(a_1^{(N-1)} + C_N a_{N-1}^{(N-1)}\right) x_{N,m} + x_{N-1,m} \right] \right|^2 \quad , \tag{151}$$

where, in Eq. (150), we used the convolution theorem and

$$\sum_{m=1}^{M} W_m = 1, \; W_m \geq 0 \quad .$$

We now define

$$e_m \equiv a_{N-1} x_{2,m} + \ldots + a_1 x_{N,m} + x_{N+1,m} \quad ,$$

and

$$b_m \equiv x_{1,m} + a_1 x_{2,m} + \ldots + a_{N-1} x_{N,m} \quad . \tag{152}$$

where e_m and b_m are the forward and backward error values respectively.

Now, it can be seen that Eq. (151) can be written

$$\overline{\varepsilon^2} = \sum_{m=1}^{M} W_m |e_m + C_N b_m|^2 \quad , \tag{153}$$

but it is equally valid to estimate $\overline{\varepsilon^2}$ from

$$\overline{\varepsilon^2} = \sum_{m=1}^{M} W_m |b_m + C_N e_m|^2 \quad , \tag{154}$$

and for this reason Burg realized that a better estimator would be obtained if the average of Eqs. (153) and (154) were used in place of either one alone. Thus he used

$$\overline{\varepsilon^2} \equiv \frac{1}{2} \sum_{m=1}^{M} W_m \left[|e_m + C_N b_m|^2 + |b_m + C_N e_m|^2 \right] \quad . \tag{155}$$

To obtain C_N, one minimizes $\overline{\varepsilon^2}$ in Eq. (155) by setting its derivative with respect to C_N equal to zero. The result is, simply,

$$C_N = \frac{(-2) \sum_{m=1}^{M} W_m (b_m e_m)}{\sum_{m=1}^{M} W_m \left(e_m^2 + b_m^2 \right)} \quad . \tag{156}$$

As the references show, the denominator will always be greater than the numerator, and it is this fact which guarantees stability ($|c| \leq 1$). As Claerbout,[3] p. 136, asserts, the b_m and e_m vectors play the role of "upward and downward going waves" in the seismic model.

II. EXPERIMENTAL RESULTS

From the AFGL Fourier spectrometer, George Vanasse provided some data in the form of an autocorrelation for an experiment to test the MEM approach on such data. Neil Grossbard provided the software and suggested the approach that is used here.

Originally the autocorrelations contained 132,000 points. We analyzed 8,200 points and, as will be seen, the results were excellent. On the other hand, we did not analyze the original data, and, in that sense the test is incomplete. There is a problem in the application of MEM-PSD analysis to data which contain many lines. The 132,000 point autocorrelations represented between 3 to 4 thousand spectral lines. The computer we used had a core memory of about 10,000 numbers. But, as we have seen, to have a single complex sinuisoid $e^{i\omega_o t}$, one pole was needed; and hence, to have a real spectral line, one would need, as a lower bound, more than two reflection coefficients. This limitation on MEM is the key factor in its application to optics. At least initially, we wished to avoid the problem of an exceedingly long computer run (which would be needed if many more points than could be fit into core were employed).

To circumvent the difficulty, the original data were pre-filtered by Mark Esplin in a manner which necessitated the use of a conventional Fourier spectral analysis on the original data as well as an inverse transform. In this way we were able to generate an interferogram that corresponds to a spectrum of about 100 lines, however, the test must be considered preliminary until actual data are used. Because the results of the test were good, we plan to carry out the procedure on such actual data.

Another problem arose from the fact that the autocorrelations with which we began were not positive definite (in the sense that their Toeplitz Matrix did not satisfy that condition). To cope with this problem, we did not use the Yule-Walker technique on the autocorrelations, but instead did something suggested by N. Grossbard. We applied the "Burg Technique" directly upon the symmetrical autocorrelation. In this manner, the original problem did not give rise to instability because, as was shown, the Burg Technique guarantees stability.

The PSD was obtained using

$$P_{Burg}(\omega) = \frac{P_M}{\left| \sum_{n=0}^{N} a_n e^{-in\omega\Delta x} \right|^2} \quad , \tag{157}$$

and then calculating

$$PSD(\omega) = (P_{Burg}(\omega) \cdot N)^{1/2} \Delta x \quad . \tag{158}$$

Equation (157) was obtained using the Burg method, as mentioned, and Eq. (158) was reported as the "PSD". The square root was inserted due to the fact that if one were to use conventional analysis, the Fourier Transform of the autocorrelation of an autocorrelation would be the square of the PSD. Since this is not exactly true in our application, Eq. (158) is not really proven to be mathematically correct.* The fact that it gave good results is its only real justification. The factor Δx in Eq. (158) was inserted for dimensional reasons (recall that Δx or Δt is frequently set = 1).

The results are given in Figure 23. While the abscissa is unlabeled, it is identical to the one in Figure 24 which gives the 132,000 point result obtained by conventional means. Notice that in Figure 23, the line pairs in Figure 24 labeled 6, 16, 18 and 19 were not resolved.

Next, in the spectrum of Figure 23, a Newton-Raphson technique was used to search for the line peaks. Table 1 gives the wavenumber (cm^{-1}) locations found in this way and compares these to the locations in Figure 24. Considering the reduction from 132,000 points to 8,200 points, this result is very encouraging in spite of the previously mentioned prefiltering.

Finally, we estimated a quantity that was proportional to the power of these lines by integrating under each peak. To do so, we filled in the PSD estimates between the half-power points (or local minima) and numerically integrated. Figure 25 shows a graph of these results. The ordinate's units were arbitrary. A comparison between Figure 24 and Figure 25 shows that the two families of lines (large and small peaks) are clearly displayed.

*Probably $PSD(\omega) = (P_B(\omega)\Delta x)^{1/2}$ is more accurate.

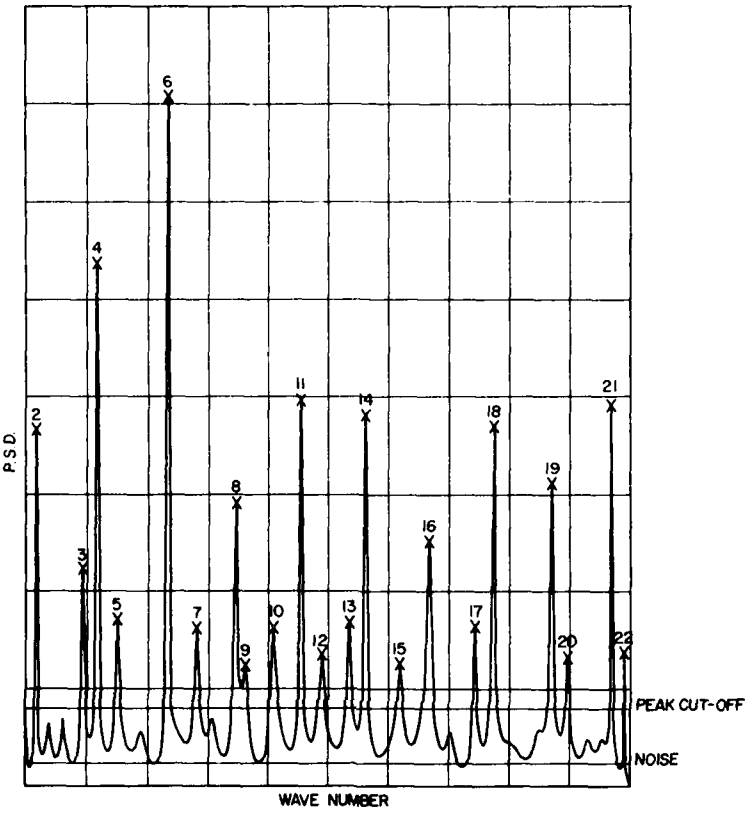


Figure 23. MEM Analysis of Filtered Data of G. Vanasse. A small fraction of the data was used.

These results show us that, there is excellent agreement between spectral line positions calculated by Blackman-Tukey with a large number of points and MEM applied to a much shorter interferogram. On the other hand, the discrepancy in line intensities requires more investigation.

Any problem with core memory is a hardware problem. If the PSD in question has too many lines (at too high a density to be filtered by analog means) then the solution is simply to wait until the hardware becomes available, or else make more efficient use of existing core capacities. On the other hand, if one has an insufficient number of points to calculate a sufficient number of PE coefficients (or reflection coefficients) in relation to the number of actual spectral lines present, then the MEM procedure will fail.

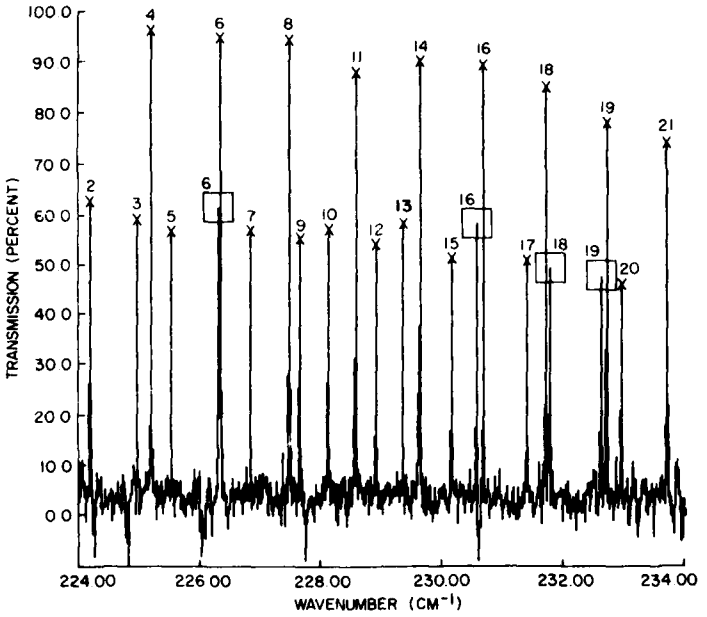


Figure 24. Original FFT Analysis on Entire Data Sample. The numbers refer to the same lines in Figures 23–25

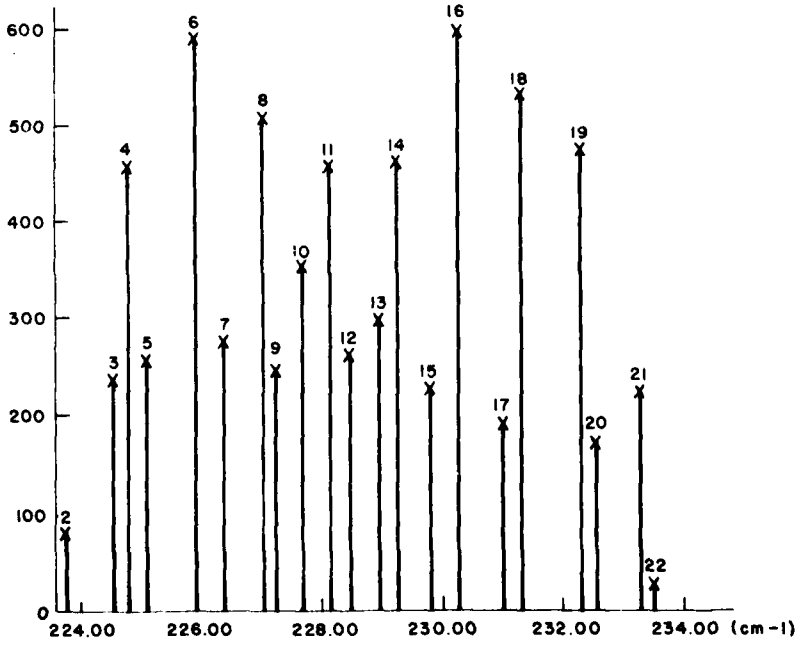


Figure 25. Graph of Integrated Lines Derived From MEM PSD of Figure 23. There is excellent agreement with the positions of the lines in Figure 24, but the intensities disagree by a large amount. Ordinate scale is arbitrary

Table 1. Comparison of Line Positions From Fourier Transform and Burg MEM

No.	Wave No. of Fourier Transform	Burg
2	224.1830	224.1773
3	224.9531	224.9395
4	225.1736	225.1774
5	225.5168	225.5104
6	226.2935	226.3258
6	226.3266	
7	226.8234	226.8076
8	227.4541	227.4449
9	227.6106	227.6134
10	228.1040	228.0849
11	228.5585	228.5522
12	228.9092	228.8937
13	229.3574	229.3486
14	229.6374	229.6270
15	230.1739	230.1893
16	230.5861	230.6805
16	230.6920	
17	231.4190	231.4182
18	231.7232	231.7356
18	231.7877	
19	232.6393	232.7046
19	232.7301	
20	232.9632	232.9711
21	233.7132	233.7046
22	234.1133	233.902

12. REFLECTION COEFFICIENTS – THE WAVE OF THE FUTURE

Historically, it was the PSD (via the prism or grating) of electromagnetic waves which eventually led to the creation of Quantum Mechanics. It was the spectral line of atomic radiation that led to the Bohr Atom. The PSD has also played an essential role in time series analysis, more so than its counterpart, the autocorrelation. The latter, via the Wiener Khintchine Theorem, contains the same "information", but it is mainly used as a way to get the PSD (as we have seen).

As has already been mentioned, Burg has given us yet another fundamental way to represent the "second order statistics" contained in the PSD and autocorrelation. This new representation consists of ϕ_0, the zeroth lag autocorrelation, plus the reflection coefficients, or (ϕ_0, C_1, C_2, ... C_N). The ϕ_0 corresponds to the scale of the PSD, and the C_i's determine its shape. Furthermore, Burg has suggested that this new representation may be even more fundamental than the PSD. In the opinion of the present author, this idea is worthy of consideration.

Burg shows that there is a one-to-one correspondence between the autocorrelation and (ϕ_0, C_1, ... C_N). To see this, recall the recursion relations, (111),

$$\Delta_N = \sum_{n=0}^{N-1} \phi_{N-n}\, a_n \quad . \tag{159}$$

$$C_N = \frac{-\Delta_N}{P_{N-1}} \quad , \tag{160}$$

$$P_N = P_{N-1}(1 - |C_N|^2) \quad . \tag{161}$$

Burg showed, incidentally, that whenever $|C_N| = 1$, the recursion terminates, the poles jump to the unit circle, and the resulting time series becomes deterministic. (The last is due to the fact that the prediction error P_N becomes zero as shown by Eq. (161.)

Combining Eqs. (159) and (160), we obtain

$$\phi_N = -\sum_{n=1}^{N-1} \phi_{N-n}\, a_n^{(N-1)} - C_N\, P_{N-1} \quad . \tag{162}$$

Parenthetically, it should be pointed out that Eq. (162) shows how to extend the autocorrelation; and Burg has proven that the MEM-PSD could be obtained by conventional means (Wiener-Khintchine Theorem) from these autocorrelations by using Eq. (162) to extend them to $\pm\infty$. In any case, Eq. (162) shows that it is possible to put the C_N's into one-to-one correspondence with the given autocorrelations. The recursion, it may be shown, can be run in both directions in the parameter N. More details are given in Burg's thesis.

There is a similar one-to-one correspondence between the PSD and the ϕ_0, C_1, ... C_N representation. The PSD is

$$\Phi(\omega) = \frac{P_N \Delta t}{|A(\omega)|^2} \quad , \tag{163}$$

as is now very familiar. The $A_N(Z)$ can be shown to satisfy

$$A_{N+1}(Z) = A_N(Z) + C_N Z^{N+1} A_N(1/Z) \quad . \tag{164}$$

Eq. (164) is derived from

$$A_N(Z) \equiv \sum_{n=0}^{N} a_n^{(N)} Z^n \quad , \tag{165}$$

together with the recursion of Levinson

$$\begin{aligned} a_0^{(N+1)} &= a_0^{(N)} \\ a_1^{(N+1)} &= a_1^{(N)} + C_N a_N^{(N)} \\ &\cdot \qquad\quad \cdot \\ &\cdot \qquad\quad \cdot \\ a_N^{(N+1)} &= a_N^{(N)} + C_N a_1^{(N)} \\ a_{N+1}^{(N+1)} &= C_N a_0^{(N)} \quad . \end{aligned} \tag{166}$$

(Compare Eq. (111) and see p. 166 of Robinson and Treitel[20].) Using $Z = e^{-i\omega}$, in Eq. (164), we obtain

$$A_{N+1}(\omega) = A_N(\omega) + C_N e^{-2\pi i f \Delta t N} A_N^*(\omega) \quad , \tag{167}$$

which is the frequency domain representation of Levinson Recursion. This relates the PSD to the $(\phi_0, C_1, \ldots C_N)$ through Eq. (163).

We now come to an interesting question. "When is the $\phi_0, C_1, \ldots C_N$ representation superior in some way to the PSD or to the autocorrelation?" One application is prospecting for oil by analyzing acoustic signals (due for example to an explosion) which contain information about the location of the layers under the ground. Figure 26 shows how a signal pulse can give the location of the layers but how reverberations can make the signal very complex.* The important point is that, in the present examples, the C_i's contain the physics of the situation in the most practical manner of speaking. Also, as has been mentioned, it was this particular field of research that led Burg to his "technique" as well as to his MEM-PSD.

There is another place where the reflection coefficients represent the physics behind the signal. This is in the field of speech compression. For reasons of economics it is very important to find ways to represent speech signals concisely. The signal is therefore "parameterized" and these numbers are transmitted. An earlier method consisted of sending the pitch (fundamental frequency), together with linear prediction coefficients. The sounds were regenerated at the other end. (A "voiced" vs "non-voiced" parameter was included). Later on it was found that there were many "round-off" problems that could be solved if the linear prediction coefficients were replaced by reflection coefficients. This was pointed out to the present author by Caldwell Smith of RADC. The explanation probably has a relation to the following interesting fact. Figure 27 represents an "acoustic tube" model for the vocal tract. This model plays a fundamental role in the representation of speech sounds (see Robinson and Silvia[22]). In this case, the reflection coefficients correspond to the size of the cross-sections of the "pieces of tube" in the model. Again the C_N's have a physical significance.

Let us now consider the question of the future role of the $(\phi_0, C_1, \ldots C_N)$ representation in Fourier Spectroscopy. The question reduces to the following. Since Fourier Spectroscopy is mainly used to identify molecular and atomic species, does the new representation in some way present us with a more economical way to do the job? To a fertile mind, such a question may present many ramifications. For example, would it not be interesting if a PSD could be "deconvolved" in a manner to render species identification more simple?

---

*According to Robinson et al, prospecting since the 1960's would not have been possible without data analysis. This is due to the fact that since that time, the nature of the locations of the oil were such that it was necessary to apply "deconvolution" techniques and vast amounts digital signal processing (a trillion bits a day at present) merely to find the layers.

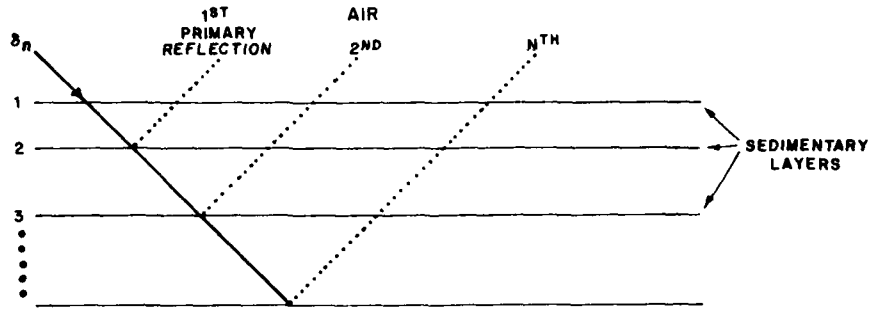


Figure 26a. First Reflection From Layers

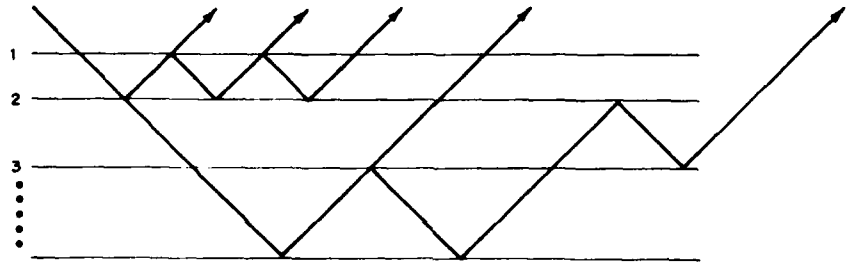


Figure 26b. Reverberation. All combinations of reflection and reverberation occur, which make the initial pulse δ result in a complicated signal

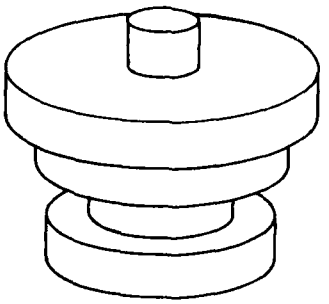


Figure 27. A Diagram That Depicts an "Acoustic Tube" Model. Models such as this one are used to analyze speech signals. The reflection coefficients are the areas of the tube sections

If it should turn out that the new representation were to present us with a way to realize significantly great simplifications, then this in turn would suggest that the reflection coefficients may have a physical significance in quantum mechanics. If anything, at least it is an interesting idea.

13. CONCLUSION

The relatively recent method of MEM-PSD analysis promises to aid the technology of Fourier Spectroscopy. Its main advantage is that under appropriate conditions, it permits a drastic reduction in the number of points of the initial interferogram without significant loss of resolution. In addition, it may lead to still further developments in spectral analysis due to the entirely new conceptual basis and philosophy which underlies the mathematics of the technique.

References (Annotated)

This listing includes key references on the subject of MEM and on the background mathematical information needed for understanding MEM. The works by Burg (Ref. 17) and Haykin (Ref. 14) can be regarded as two "textbooks" for MEM. The Sept. 1982 issue of the Proceedings of the IEE devoted to spectral estimation (including Refs. 1 and 16), Kay and Marple (Ref. 28), and Childers (Ref. 30) are the main existing review articles. The bibliographies in Burg, Kay and Marple, and Childers should cover almost all the literature until a fairly recent date. The books by Claerbout (Ref. 3), and by Robinson and Treitel (Ref. 20) may be regarded as the best available sources for the mathematical background to the subject.

1. Robinson, Enders A. (1982) A historical perspective of spectrum estimation, Proc. IEEE 70(No. 9):885.

Parts of this article were employed for the first part of the present report. It contains a large amount of information that was left unmentioned, and the reader should find it very informative.

This entire issue of the Proceedings of the IEEE, entitled "Spectral Estimation", is devoted to the subject of spectral analysis. All of the papers can be regarded as reviews of particular areas.

2. Gottman, J.M. (1981) Time-Series Analysis, Cambridge Univ. Press.

This book is an elementary approach to the subject and the mathematical level is much lower than most readers would expect; however, the subject matter, in reality, is advanced. It treats AR, MA, ARMA, the methods, and procedures. The historical discussions were used in the text above. The discussion of noise

in the periodogram is the clearest I have found. His discussion on the deterministic part of the time series is unique in that it makes it absolutely clear that "pure sine waves" are unimportant in modern spectral analysis. He shows how to detect them and eliminate them from the data before applying the new methods. If the researchers who used sine waves to test the MEM-PSD approach had read this text, they may have avoided some serious mistakes.

3. Claerbout, Jon F. (1976) Fundamentals of Geophysical Data Processing.

This book was used for the treatment of the Z-transform in the present report. It is an excellent source for mathematical background material. Ch. 3 contains an an important discussion on "spectral factorization" and Ch. 8 explains why the C_N are the actual physical "reflection coefficients" in geophysical exploration. This book is very concise, but it is also easy to comprehend. Such a combination is unique.

4. Lee, Y.W. (1960) Statistical Theory of Communication, Wiley.

This book gives the clearest available discussion on generalized harmonic analysis as created by Wiener. The explanation of Wiener's Optimum Filter in the continuous domain is explained in great detail here. In the text I used his physical picture of the Wiener Predictor. This is a very well written book so far as clarity and attention to detail is concerned. It does for the continuous domain what Robinson et al have done for us in the discrete or digital domain; it is another masterpiece of exposition.

5. Blackman, R.B., and Tukey, J.W. (1958) The Measurement of Power Spectra, Dover.

This is the "classic" book on the subject, but it is sometimes difficult to understand. For example, the reason for the "noise" in the periodogram and PSD is not really explained here, and one could get the impression that it is caused by "leakage". As Gottman's book as well as that of Claerbout explain, it is due to the statistics of the noise!

6. Bendat, J.S., and Piersol, A.G. (1971) Random Data: Analysis and Measurement Procedures, Wiley-Interscience.

This book gives the background of the conventional approach to the PSD. It also gives the detailed procedures used in that approach.

7. Otnes, R.K., and Enochson, L. (1978) Applied Time Series Analysis, Vol. 1, Wiley.

This contains many of the details on how to apply the conventional approach to the PSD.

8. Goldman, S. (1953) Information Theory, Prentice Hall, Inc.

9. Brigham, E.O. (1974) The Fast Fourier Transform, Prentice Hall, Inc.

This book is devoted to explaining the "FFT", but in doing so it gives what I found to be the best explanation of the discrete Fourier Transform to be found in the literature.

10. Levine, R.D., and Tribus, M. (1979) The Maximum Entropy Formalism, MIT Press, Cambridge, Mass.

11. Rosenkrantz, R.D., (Ed.) (1983) E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics, Reidel Publ. Co., Boston, MA.

All of the publications of E. T. Jaynes on the basic principle of applying maximum entropy to probability theory will be found in this collection with two exceptions. They are the "Mobil-Socony Lectures", which is not available, and his work on MEM-PSD analysis which is given as Ref. 15.

12. Page, L. (1935) Introduction to Theoretical Physics, Van Nostrand Co., N.Y., p. 311.

13. Shimony, A., and Dias, P.M. (1981) A Critique of Jaynes' Maximum Entropy Principle, Advances in Applied Math. 2:172-211.

This report contains the strongest criticism of Jaynes' work.

14. Haykin, S., (Ed.) (1979) Nonlinear Methods of Spectral Analysis, Springer-Verlag.

15. Jaynes, Edwin T. (1982) On the rationale of maximum-entropy methods, Proc. IEEE 70(No. 9):939. The issue is entitled Spectral Estimation.

It was most appropriate of the IEEE to include a paper by Jaynes, who, in effect, made the entire subject of MEM-PSD analysis possible. All of his other works (see Rosenkrantz[11] for a collection of them) deal with the general principle of MEM and its applications to Statistical Mechanics, Probability Theory, and Decision Theory. This review by Jaynes is the only one by him to date on the subject of MEM-PSD analysis. Although the present report does not consider the subject of determining the order of the PE filter, this problem is considered to be not completely solved. Only ad hoc solutions exist. This problem plus others are considered by Jaynes, and he points the way for future research. (Noise is another of the problems that he considers still unsolved by MEM.)

16. Lacoss, R.T. (1971) Data Adaptive Spectral Analysis Methods in Ref. 7, (p. 134), of Main References.

This reference contrasts various ways to obtain the PSD and points out that it is necessary to integrate the MEM lines in order to obtain a quantity proportional to power. Table 1 of that report contrasts MEM with two other methods in this regard.

17. Burg, John P. (1975) Maximum Entropy Spectral Analysis, Ph.D. Dissertation, Dept of Geophysics, Stanford University.

18. Bracewell, R. (1965) The Fourier Transform and Its Applications, McGraw Hill, Chapter 1.

In my opinion, this book is one of the clearest ones devoted to the continuous Fourier Transform. It also gives useful geometric interpretations.

19. Hsu, H. P. (1970) Fourier Analysis, Simon and Schuster (Tech Outlines).

This "problems and solutions" book provides a large amount of basic information in a concise manner.

20. Robinson, E. A., and Treitel, S. (1980) Geophysical Signal Analysis.

As should be evident from the text, this was the most important background information source. It is a masterpiece of exposition. The present author has learned that any textbook with E. A. Robinson as an author or co-author was very easy to follow and pleasurable to read.

21. Robinson, E. A. (1967) Statistical Communication and Detection, Hafner.

The text made use of the Fejer Theorem as treated in the present report. There is also mention in this book by Robinson of the connection between the Paley-Wiener Theorem and the deterministic signal. It is all very well written.

22. Robinson, E. A., and Silvia, M. T. (1978) Digital Signal Processing and Time Series Analysis, Holden Day, Inc.

This is a well written, clear, and comprehensive treatment. The discussion on minimum phase and the Paley Wiener Theorem are very helpful.

It is an excellent introduction since it covers all the basic material. It also gives a good discussion of applications. Their discussions on speech deconvolution and seismic deconvolution were used in the present report.

23. Guillemin, E. A. (1949) The Mathematics of Circuit Analysis, Wiley and Sons; (1963) Theory of Linear Physical Systems, Wiley and Sons.

These two books are classics and they explain many of the subtle aspects of Fourier Analysis that are not to be found elsewhere. The discussions are very clear. "Minimum phase" and the "Payley-Wiener Condition" are well explained. The discussions make it obvious that the engineering applications of this subject demand sophistication. The books both deal with the continuous domain.

24. Wiener, N. (1949) Extrapolation, Interpolation, and Smoothing of Stationary Time Series, Cambridge, MA, Technology Press of the MIT.

Originally classified secret this report was declassified and published in the open literature in 1949. The report describes the Optimum Filter of Wiener as it was first conceived. The mathematics was too advanced for the engineering community. Two appendices by Levinson, originally published elsewhere, are included. These two papers made this work available to the engineers. The Levinson Recursion, in slightly different form, first appeared here in this book (or rather in the reports used for the appendix of this book).

25. Laning, J.H., and Battin, R.H. (1956) Random Processes in Automatic Control, McGraw-Hill.

Contains an excellent treatment of the Optimum Wiener Filter and Predictor in the continuous domain.

26. Bendat, J.S. (1958) Principles and Applications of Random Noise Theory, Wiley & Sons.

This book contains a good description the Wiener Optimum Filter and Predictor.

27. Makhoul, J. (1975) Linear Prediction: A Tutorial Review", reprinted in Ref. 30, p. 99.

This is an important review paper and is well written.

28. Kay, S.M., and Marple, S.L. (1981) Spectrum Analysis—A Modern Perspective, Proc. IEEE 69, pp. 1380-1419.

This is a comprehensive review article on the present subject. The present report used one of his figures. This review by Kay and Marple goes quite far beyond MEM-PSD analysis and it considers all the other methods available. It is a veritable monograph on the general topic of PSD analysis.

29. Abels, J.G. (1974) Maximum entropy spectral analysis, Reprinted in Childers.[30]

A good and also concise introduction to the subject.

30. Childers, D.G. (1978) Modern Spectrum Analysis, IEEE Press.

This volume is a collection of reprints of what were the key papers up to the time of publication. In the text the tutorial review by Abels and one by Makhoul, as well as a paper by Lacoss have been referenced. This volume edited by Childers is essential for anyone considering doing serious work in MEM-PSD estimation.

A paper not directly referenced in the text but important in this field is: Paley, R.E.A., and Wiener, N. (1934) Fourier Transforms in the Complex Domain, American Mathematical Society.

This is the reference where their famous theorem first appeared. The explanation here is in terms of "quasi-analytic functions" in the complex domain. The "physics" is not apparent.

END

FILMED

4-86

DTIC