LA-6895-MS
Informal Report

C.3

UC-32

Issued: August 1977

Signal Processing by P Log P Maximum Entropy Analysis

J. E. Brolley

los alamos
scientific laboratory
of the University of California
LOS ALAMOS, NEW MEXICO 87545

UNITED STATES
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION
CONTRACT W-7405-ENG. 36

Printed in the United States of America. Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161
Price: Printed Copy $3.50 Microfiche $3.00

SIGNAL PROCESSING BY P LOG P MAXIMUM ENTROPY ANALYSIS

by

J. E. Brolley

ABSTRACT

P Log P maximum entropy analysis has been applied to a bank of Butterworth filters. Resolution enhancement was obtained. Comparison is done with the prewhitening autocorrelation method.

I. INTRODUCTION

Signal processing in the deterministic sense has been a viable bourgeoning activity since the time of Fourier. The question that has generally been asked is what is the spectral decomposition? Not, what is the most probable spectral decomposition? The deterministic point of view leads to artificial assumptions and patchwork when finite data sequences are to be analyzed by the classical methods.

In 1967 Burg[1] broke the chains of classicism and adopted a position that signal analysts might have discerned years earlier if they had followed the initial period of development of quantum mechanics. His basic tenets are as follows. If S_i is the spectral power of a finite sequence, the best choice of S is that for which the entropy H is a maximum

$$H \sim \Sigma \ell n S_i = \text{maximum}. \tag{1}$$

Moreover, this operation must be carried through, subject to the condition that it be independent of all of the unknown autocorrelations. Had the data sequence been infinite, then there would be no unknown autocorrelations. If Φ_j denotes the unknown autocorrelations resulting from a finite data sequence, the maximization shall be independent of Φ_j.

$$\frac{\partial H}{\partial \Phi_j} = 0. \tag{2}$$

Pursuit of the philosophy expressed by Eqs. (1) and (2) leads to analytical realizations for computer manipulation. Applications to geophysics and other fields have been numerous and fruitful. Although results obtained from short data sequences have been quite useful, Percy[2] and others noted that some frequency distortion occurs. The reason for this has not so far been identified.

It is clear that Burg's approach is pointing in the right direction, namely, try to construct a theory that does not make assumptions concerning quantities that cannot be measured. However, it appears possible to apply probability theory in a different manner.[3]

In what follows, the discussion will be focused on digital operations. However, it will be clear that the procedure applies to analogue systems as well as noncircular functions.

II. METHOD

A variety of digital filters have been developed for decomposing data sequences.[4,5,6] Low-pass, high-pass, and band-pass filters of various kinds are available. Spectral decomposition of a data sequence by passing it through a bank of band-pass filters has been practiced for a long time. For many applications the filters must be broadly tuned and overlapping. In the case of very weak signals such broadbanding may have to be carried to extremes in order to discern any meaningful output from the filter bank. Resolution has been lost and consequently the probability of successfully identifying a signal of known spectral characteristics is diminished.

It is possible to apply probability theory to improve the interpretation of the filter output and to take into account noise generated in the filter channels. This can be done by adapting the arguments of Jaynes[7] and Frieden.[8]

A bank of Butterworth[4,5,6] filters was constructed to perform the initial signal processing. Although other types of filters could be used, these were employed because of their lack of ripple and simple analytical representation.

They are characterized by an essentially flat top with power gain of unity and smooth roll-off. The half power points are designated ω_1 and ω_2, and the center frequency (angular) by ω. Such a filter bank is schematically depicted in Fig. 1, these being equally spaced, overlapping channels extending from ω_{min} to ω_{max}. Each channel has a power gain curve given by

$$G = \frac{1}{1 + \left(\dfrac{\tan^2 \frac{\omega T}{2} - \omega_1{}' \omega_2{}'}{\omega_c{}' \tan \frac{\omega T}{2}} \right)^{2N}} \tag{3}$$

$$\omega_1{}' = \tan \frac{\omega_1 T}{2} \tag{4}$$

$$\omega_2{}' = \tan \frac{\omega_2 T}{2} \tag{5}$$

$$\omega_c{}' = \omega_2{}' - \omega_1{}' \tag{6}$$

T = time interval between data points
ω = input frequency.

N is the order of the filter and determines how far the flat top extends and how steep the roll-off is. In general it is not practical to exceed $N = 10$ in

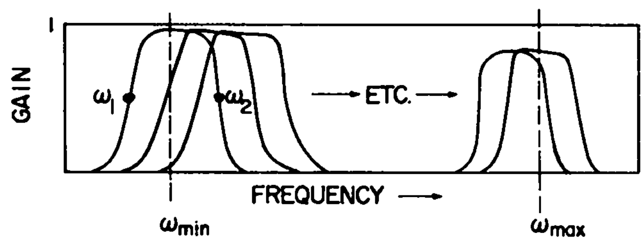

Fig. 1. Ensemble of overlapping Butterworth filters having equal bandwidths (not a necessary restriction).

ordinary computer operations because too much precision will be required. In the examples which follow, N = 8. The computer realization of Eq. (3) is a set of recursion relations.

With the filter bank of Fig. 1 in hand it is then reasonable to ask what is the most probable distribution of power sources in the original signal which would produce the observed output. As in Burg's real world philosophy a finite data sequence is assumed for input to the filter. It is further assumed there are M filters in the bank and that the original power sources are distributed amongst J equally spaced frequency intervals, and that J > M and the two frequency spaces precisely overlap. The power out of every filter element m is then

$$P_m = \sum_{j=1}^{J} E_j S(m,j) + n_m \; . \tag{7}$$

E_j is the power in the j^{th} interval of J space, S is the Green's function of the problem, and n_m is the noise in the m^{th} channel of M space. Digital band-pass filters have properties reminiscent of analogue filter circuits having R, L, and C. It takes time for them to come to equilibrium output and time to drain energy after the input signal has been turned off. The present calculation is structured to conserve energy; the filters are required to run until they have emptied. In what follows, positivity of P_m is desired. Since n_m may have negative values in some problems, a bias B and a new noise matrix N are introduced by letting

$$N_m = n_m + B \qquad N_m \geqslant 0,\ B \geqslant 0. \tag{8}$$

B is adjusted to guarantee $N_m \geqslant 0$. To a sufficient approximation it is assumed that

$$P_0 = \sum_{m=1}^{M} Pm \tag{9}$$

respresents the total input power to the filter bank. The Green's function S is the normalized overlap integral of the filter gain function between any two elements m and j of the frequency spaces. The probability arguments of Frieden and Jaynes are now applied to produce the extremum relation

$$-\sum_{j=1}^{J} \hat{E}_j \ln \hat{E}_j - \rho \sum_{m=1}^{M} \hat{N}_m \ln \hat{N}_m$$

$$-\sum_{m=1}^{M} \lambda_m \left[\sum_{j=1}^{J} \hat{E}_j S(m,j) + \hat{N}_m - B - P_m \right]$$

$$- \mu \left[\sum_{j=1}^{J} E_j - P_0 \right] = \text{maximum} \quad . \tag{10}$$

^ denotes a restored value. λ_m and μ are Lagrangian multipliers to be evaluated, and ρ is a free parameter which expresses an estimate of the relative importance and knowledge of the noise.[8] The restored quantities are then given by

$$\hat{E}_j = \text{EXP} \left[- 1 - \mu - \sum_{m=1}^{M} \lambda_m S(m,j) \right] \tag{11}$$

$$\hat{N}_m = \text{EXP} \left[- 1 - \mu - \frac{\lambda_m}{\rho} \right] . \tag{12}$$

λ_m and μ may be determined from the M + 1 equations

$$0 = F_m = e^{-1-\mu} \sum_{j=1}^{J} S(m,j) \, \text{EXP} \left[- \sum_{m'=1}^{M} \lambda_{m'} S(m',j) \right]$$

$$+ \text{EXP} \left[- 1 - \frac{\lambda_m}{\rho} \right] - B - P_m \tag{13}$$

$$0 = F_{nq} = e^{-1-\mu} \sum_{j=1}^{J} \text{EXP} \left[- \sum_{m=1}^{M} \lambda_m S(m,j) \right] - P_0 \; . \tag{14}$$

III. RESULTS

The first example consists of passing a simple sine wave of period 5 s and duration 512 through a bank of 30 broadbanded filters with resonant frequencies extending from 0.18 s^{-1} to 0.22 s^{-1}. The full width at half maximum

$(\omega_2 - \omega_1)$ was adjusted to be 2.5 times the difference between adjacent resonant frequencies. The filter output is shown in Fig. 2. The restored frequency space was then divided into 65 equal intervals and maximum entropy analysis applied. B = 0.05 and ρ = 50 were somewhat arbitrarily chosen. In Fig. 3 the result of the calculation is exhibited. It will be observed that the peak has been considerably sharpened and that the correct frequency has been restored. No attempt was made in this example to create the sharpest possible peak. It has also been shown with synthetic examples that for closely beating waves observed over short intervals that this method will often reveal the existence of two waves when the Fourier periodogram indicates only one.

As a second example we consider the data sequence of 2048 points displayed in Fig. 4. This sequence was kindly supplied by A. Cox and S. Hodson of LASL. It is known to possess many different spectral components. Only the first three dominant components will be considered here. The time interval is 0.028 days. The Fourier periodogram of the first three dominant components is shown in Fig. 5. After appropriate transformation of the frequency scale, the first three peaks are found to occur at approximately 5.9, 4.5, and 3.4 days. The same data were then passed through a bank of 30 filters; again the bandwidth was chosen as 2.5 times the difference of adjacent frequency intervals. The maximum entropy analysis yielded the result shown in Fig. 6. Again B = 0.05 and ρ = 50.

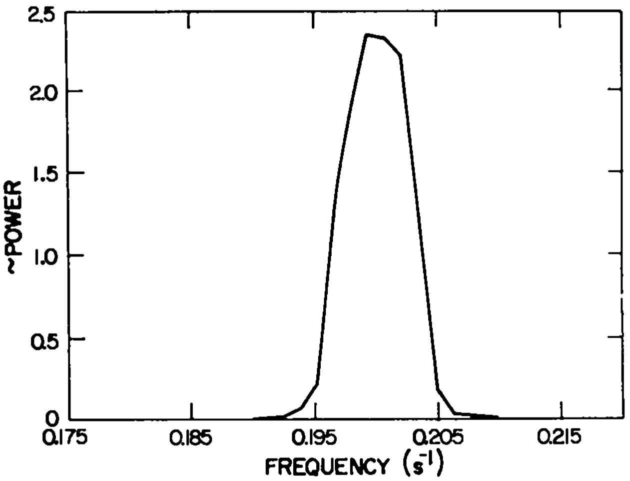

Fig. 2. Output of a broadbanded overlapping Butterworth filter bank (30 components) for a sine wave input of 5-s period.

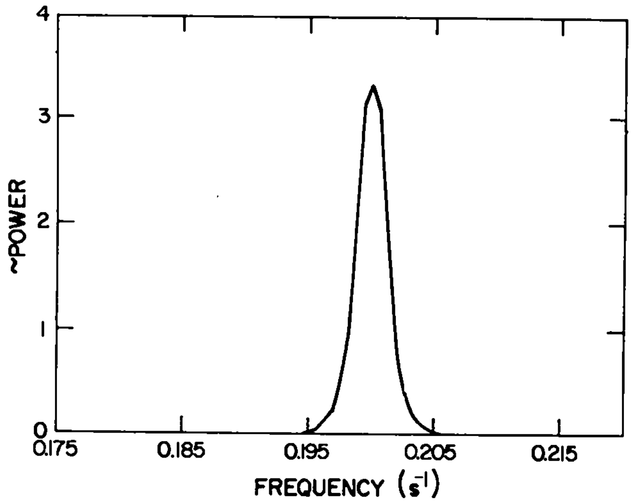

Fig. 3. Maximum entropy restoration of signal shown in Fig. 2. The restored frequency space has been divided into 65 intervals.

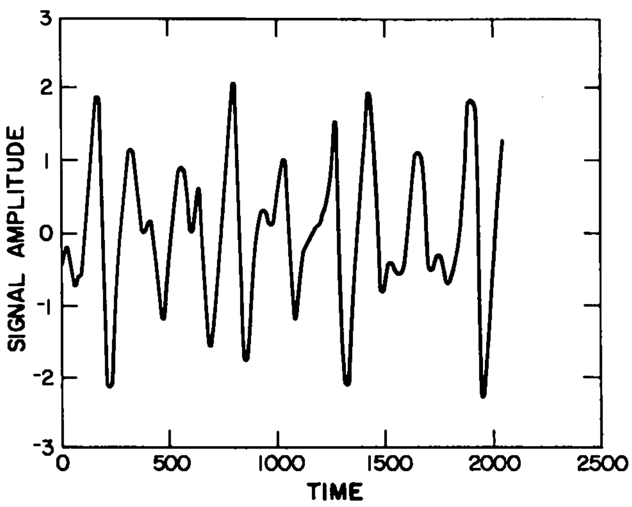

Fig. 4. Data sequence of Cox and Hodson. Special time units.

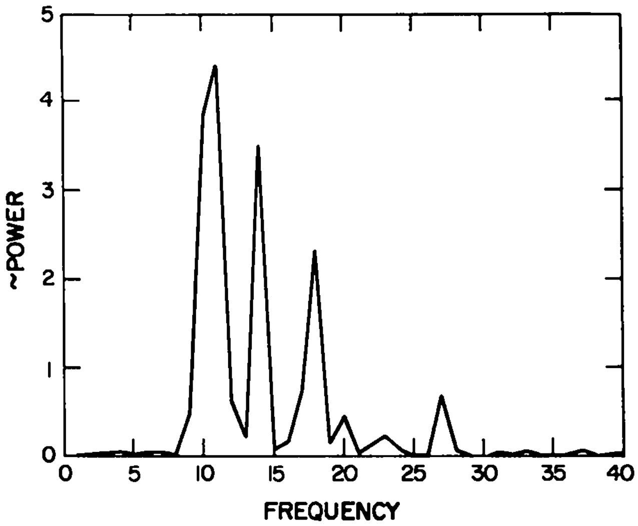

Fig. 5. Fourier periodogram of the data shown in Fig. 4. Only the first few dominant components are displayed. Special frequency units.

Attention will now be focused on the peak indicated as being approximately 5.9 days by Fourier analysis. Cox and Hodson have applied the prewhitening autocorrelation method[9] to the data sequence and have determined the first three dominant periods to be 6.0425, 4.4007, and 3.4304 days. However, the maximum entropy analysis indicates that the 6-day peak may be split.

After a suitable transformation of the frequency scale in Fig. 7 it is found that the components of the "6-day" period are ~ 6.1 and ~ 6.7 days. It is clearly interesting to examine this further. A nonlinear optimization code to fit either 3 or 4 circular functions to the 2048-point data sequence was prepared. This code was driven by program OPAC,[10] and could search over amplitudes, phases, and periods. Search proceeded alternately over amplitude plus phase and then period. With a fit of three functions a goodness of fit measure $\Phi = 51$ was obtained and periods of 6.032, 4.4004, and 3.4419 days were obtained. A search over a fit of four functions gave $\Phi = 53$. These two values of Φ are essentially indistinguishable. However, this search returned 6.5728 and 6.1538 days for the split "6-day period." It was further shown that with the phases obtained, this split period mocked up the 6.032-day period when searched over with one circular function.

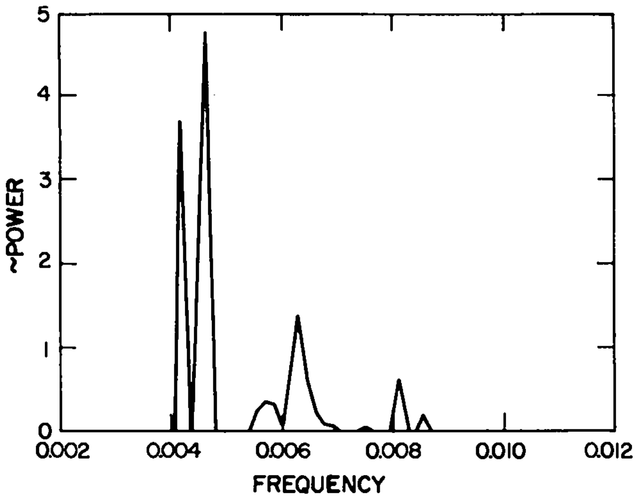

Fig. 6. Maximum entropy analysis of the region displayed in Fig. 5. Special frequency units.

IV. CONCLUSIONS

Maximum entropy analysis has been applied to a digital filter bank. The restored signal was of the correct frequency and much sharper than the filter output. In a second example the potential for revealing doublets, which were indicated as being simple lines in other methods of analysis, was demonstrated. Moreover it is clear that the procedure described here applies to radio frequency and other analogue filter banks as well as to noncircular functions.

Further study is required of the various aspects of this problem. Some areas may be noted. The theory is nonlinear and therefore sensitive to scaling of the input signal power. Optimum normalization of the input should be studied. Very likely there exist improved methods for solving the nonlinear system of equations. These should be examined and applied. The signal-to-noise ratio problem should be studied in detail. Concomitantly the role of ρ and B must be examined in detail.

Apart from the main stream problems just enumerated, the desirability of constructing other filter banks such as the Chebychev should be reviewed. Application to noncircular functions should be made.

REFERENCES

1. J. P. Burg, "Maximum Entropy Spectral Analysis," 37th Annual Meeting, Society of Exploration Geologists, Oklahoma City, October 1967, unpublished.

2. J. R. Percy, "Observational Studies of Multiperiodicity in Pulsating Variable Stars, Using Maximum Entropy Spectral Analysis," in Proc. Solar and Stellar Pulsation Conference, Los Alamos, New Mexico, August 3-5, 1976, Los Alamos Scientific Laboratory report LA-6544-C (October 1976) p. 61.

3. J. E. Brolley, "P Log P Maximum Entropy Analysis of Time Series," Bull. A.A.S., 9, 287 (1977).

4. L. R. Rabiner and B. Gold, Theory and Application of Digital Signal Processing, (Prentice-Hall, New York, 1975).

5. S. D. Stearns, Digital Signal Analysis, (Hayden Book Co., Rochelle Park, NJ, 1975).

6. A. V. Oppenheim and R. W. Schafer, Digital Signal Processing, (Prentice-Hall, New York, 1975).

7. E. T. Jaynes, "Prior Probabilities," IEEE Trans. Syst. Sci. Cybern. SSC-4, 227 (1968).

8. B. R. Frieden, "Restoring with Maximum Likelihood and Maximum Entropy," J. Opt. Soc. Am., 62, 511 (1972).

9. J. Lafler and T. D. Kinman, "An RR Lyrae Star Survey with the Lick 20-Inch Astrograph. II The Calculation of RR Lyrae Periods by Electronic Computer," Astro. J. Suppl., 11, 216 (1965).

10. Documented by the Los Alamos Computer Center Facility Program Library: E401, 15 July 1975.

Maximum Entropy Generators for Energy-Based Models. Rithesh Kumar1â˜— Sherjil Ozair1 Anirudh Goyal1 Aaron Courville1,2 Yoshua Bengio1,2,3. 1Mila, UniversitÃ© de MontrÃ©al 2Canadian Institute for Advanced Research (CIFAR). 3 Institute for Data Valorization (IVADO). Abstract. Maximum likelihood estimation of energy-based models is a challenging problem due to the intractability of the log-likelihood gradient. In this work, we propose learning both the energy function and an amortized approximate sampling mecha-nism using a neural generator network, which provides an efcient approximation of the log-