ON THE ACCURACY OF FREQUENCY DETERMINATION BY AN AUTOREGRESSIVE SPECTRAL ESTIMATOR*

G. KOVÁCS

Konkoly Observatory, Hungary

Abstract. The accuracy of frequency determination by a least squares technique for an autoregressive spectral estimator is studied and compared with the Fourier method. Using numerical tests the probability distribution function of the peak location is calculated. The autoregressive filter order is optimized in the sense of minimum variance of the peak location. Simple sinusoidal signals with additive Gaussian noise are considered and the effect of other components is only indicated. Generally, a filter order between 1/3 and 1/2 of the total data number and a not very dense data sampling, gives the most stable spectrum. The results are numerical.

1. Introduction

Since the introduction of the maximum entropy method (MEM) of spectral analysis, it has become widely used and studied in various fields dealing with problems of data processing. Though Burg's recursive scheme (Burg, 1975) has proved to be far superior to conventional methods, especially for short data records, there are two limitations in practice:

(1) Splitting and shifting of spectral peaks (Chen and Stegen, 1974; Fougere, 1977).

(2) Difficulty in obtaining general analytical expressions for its statistical appearance since MEM is a nonlinear spectral estimator.

The anomalies listed first do not occur in the nonlinear method as proposed by Fougere (1977) nor in the least squares (LS) spectral estimate suggested by Ulrych and Clayton (1976). The advantage of the LS estimate over that of Burg has been demonstrated by Ulrych and Clayton (1976) and Swingler (1979). Therefore, we limit our considerations here to the LS method.

In the case of large filter order and data number the statistical properties of the MEM spectral estimator have been determined and compared with those of the Fourier method by Kromer (1970) and Berk (1974). Confidence limits for a MEM spectral estimator using a one-way (predicting in one direction only) autoregressive (AR) model have been described by Reid (1979).

No work is available relating to the statistical properties of the general two-way LS spectral estimator. Because the frequency structure of the light variation of a variable star is the most important quantity for comparison with theory, our considerations are focused only on the accuracy of the frequency determination when the LS spectral estimator as proposed by Ulrych and Clayton (1976) is used.

* Proceedings of the 66th IAU Colloquium: *Problems in Solar and Stellar Oscillations*, held at the Crimean Astrophysical Observatory, U.S.S.R., 1–5 September, 1981.

Solar Physics **82** (1983) 123–128. 0038–0938/83/0821–0123$00.90.

2. Tests by Artificial Data

First the role of sampling time in the stabilization of the spectrum is studied. Using a test signal of the form

$$x_j = \sum_{k=1}^{3} A_k \sin(P\Delta t f_k(j-1)) + B_k \cos(P\Delta t f_k(j-1)),$$

$$A_1 = 1, \quad B_1 = 0, \quad A_2 = 0.8, \quad B_2 = -0.5, \quad A_3 = 0.2, \quad B_3 = 0.1,$$

$$P = 6.283\,185,$$

$$f_1 = 0.07 \text{ Hz}, \quad f_2 = 0.09 \text{ Hz}, \quad f_3 = 0.12 \text{ Hz}, \quad j = 1, 2, \ldots, 45,$$

$$\Delta t = 0.5 \text{ s},$$

LS spectra were calculated with different sampling rates. Once the filter order (M) and an integer number (L) of Δt is chosen, L spectra were calculated using data corresponding to the sequences: $j = 1, L+1, 2L+1, \ldots$; $j = 2, L+2, 2L+2, \ldots$; $j = L, 2L, 3L, \ldots$. Taking the average of these spectra the plot shown in Figure 1 was obtained. Numbers on the curves denote the various M, L combinations. The spectra were normalized to unity at the highest peak in each case. It is clearly seen that a denser data sampling makes the spectra unstable and it is possible to choose a proper M, L combination to deduce the exact frequencies.

A fairly crucial point in estimating MEM spectra is the determination of the order of the AR process. Several objective criteria may be used for the proper choice of the optimal value of filter order. For harmonic processes with noise, these criteria underestimate the AR order and an empirical rule can be adopted. Percy (1977) has found that the best values of M are such that $M\Delta t$ is of the same order of magnitude as the period present. Ulrych and Ooe (1979) have adopted the rule which constrains M to $N/3 \leq M \leq N/2$ (N is the number of data points).

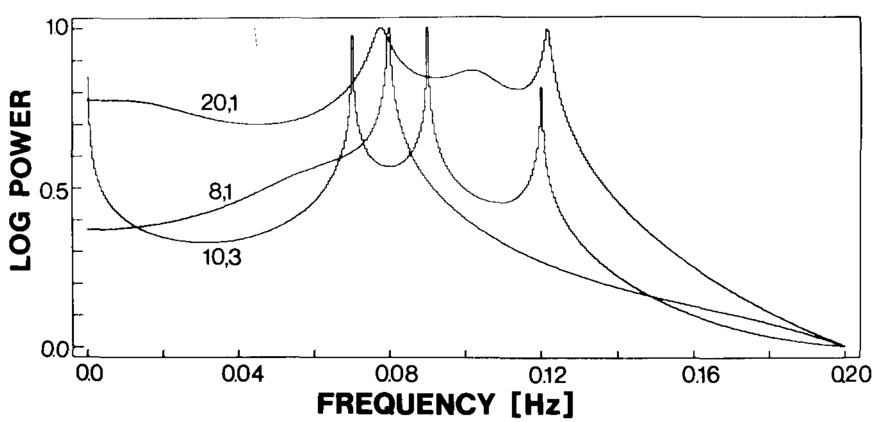

Fig. 1. LS AR spectra of a test signal with three sinusoidal components of different frequencies. The various filter order and averaging numbers are indicated near each curve (for details, see text).

https://doi.org/10.1017/S0252921100095464 Published online by Cambridge University Press

As regards the stability of the frequencies obtained by MEM, a natural criterion for the optimal filter order is that choice of M at which the variance of the peak location is minimal. Because it is almost impossible to follow this problem analytically, numerical tests are required.

The logarithm of the normalized standard deviation of peak location as a function of M/N is plotted in Figure 2. Each curve has been shifted vertically somewhat, but the scale is the same, as indicated and the zero points correspond to the first points of the curves. The same hundred realizations of signals in the form *deterministic signal* + *Gaussian component* were used to calculate the variance of the peak location. The results did not change significantly on using more realizations. For the curve at the bottom, the deterministic component consisted of two sinusoids of the form

$$\sin(P(j-1)\Delta t 0.07) + 0.5 \sin(P(j-1)\Delta t 0.09), \qquad P = 6.283\,185,$$

and for the other curves one sinusoid only, i.e.

$$\sin(P(j-1)\Delta t 0.07), \qquad P = 6.283\,185.$$

The location of the peak at ~ 0.07 Hz was used to calculate the variance in the two sinusoids case. Numbers near the curves indicate the data number, the sampling time (Δt) and the standard deviation of the noise. Vertical arrows show the minima of each curve. It is seen that in the case of a signal with a single sinusoidal component the best result is expected when $\frac{1}{3} \leq M/N \leq \frac{1}{2}$. The shift of the optimal M above half of the data

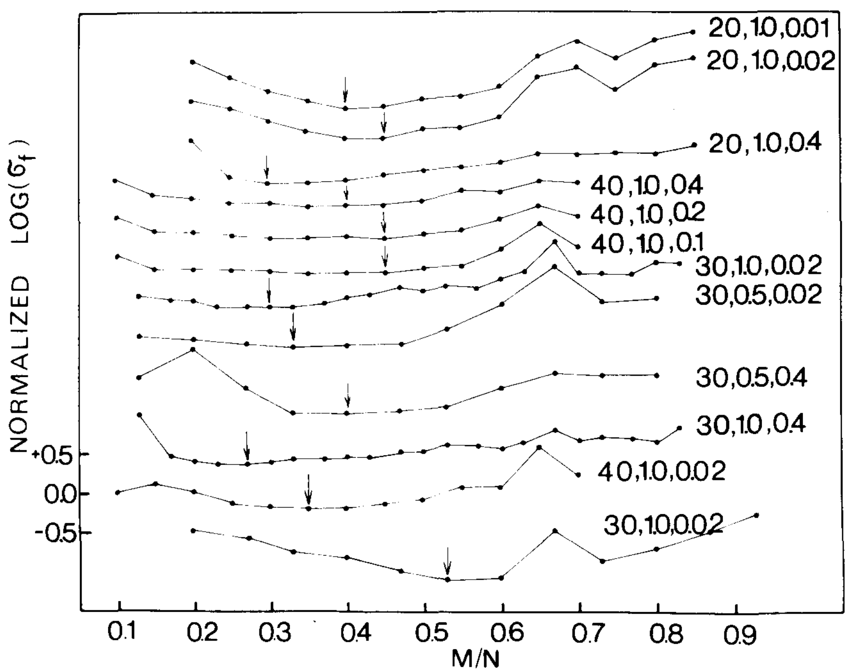

Fig. 2. Dependence of the standard deviation of peak location on the relative filter length (M/N) for different signals (for details, see text).

https://doi.org/10.1017/S0252921100095464 Published online by Cambridge University Press

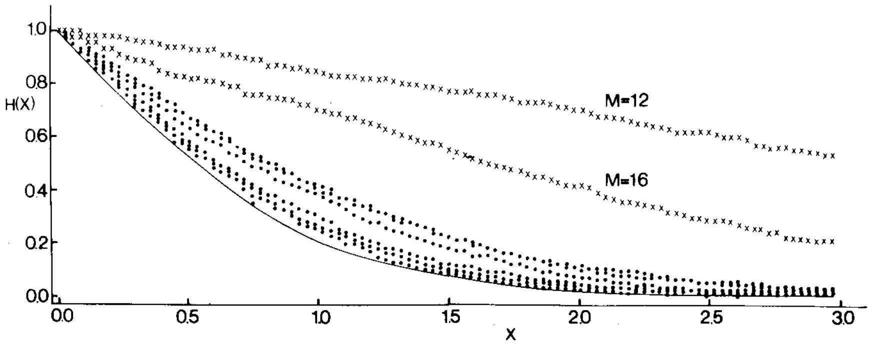

Fig. 3. Empirical probability distribution functions of the frequency shift when using the Fourier method (continuous line) and the LS AR spectral estimator (other symbols). $H(x)$ means the probability that the value of $\sqrt{N}\, N\Delta t\, |\Delta f|\, A/\sigma$ exceeds the value of x (for details see text).

number in the two sinusoids case is obvious. We conclude that Ulrych and Ooe's (1979) empirical rule adopted for optimal M generally holds and there is no hope of improving it.

As a further step in studying the statistical properties of peak location, we have calculated the empirical distribution function of peak location and compared it with that of the Fourier method. Denoting the frequency shift by Δf, the signal amplitude by A, and the standard deviation of noise by σ, the probality $H(x)$ that the value of $\sqrt{N}\, N\Delta t\, |\Delta f|\, A/\sigma$ is greater than x is plotted in Figure 3 for the Fourier method (continuous line) for the optimal filter order in the one sinusoid case (dots) and for two different filter orders in the two sinusoids case (crosses). (In the Fourier method the distribution functions were not sensitive to the number of sinusoidal components, therefore only the curve related to the one sinusoid case is plotted.) The test signals were the same as used for the calculation of the curves in Figure 2. Because the probability distribution of peak location in the Fourier method depends only on the signal parameters in the combination mentioned above (Kovács, 1981), the same combination was used here too, though the validity of this statement is not proved in the case of MEM. Nevertheless, if we adopt the combination of parameters mentioned above, it makes it easier to compare the probability distribution functions of peak location of signals with various parameters.

For a given set of parameters one hundred realizations of a random series were used to calculate the empirical distribution function of peak location. The calculation was repeated ten times in the one sinusoid and five times in the two sinusoids case with different realizations and the distribution functions were averaged and plotted. For the one sinusoid case the following parameters were used: $N = 20, 30, 40$; $\sigma = 0.005, 0.1, 0.2, 0.4, 0.8$ and $M =$ optimal M, $A = 1$, $\Delta t = 1$ s. It can be seen that the stability of the MEM against noise is weaker than that of the Fourier method. The situation becomes even more unfavourable for the MEM when a test signal with two sinusoidal components is used (parameters: $N = 30$, $\sigma = 0.02$, $\Delta t = 1$ s, $A = 1$, $M = 12, 16$). However, it is important to remark that whereas MEM was able to give the correct average frequency tested (i.e. 0.07009 Hz for $M = 12$ and 0.06999 Hz for $M = 16$), the Fourier method

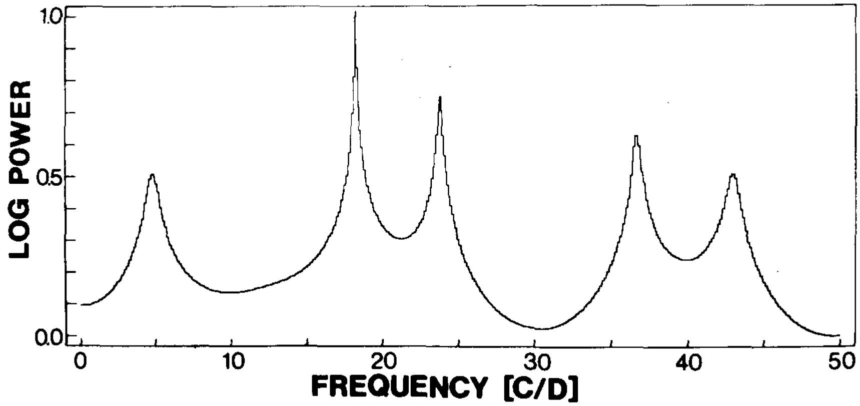

Fig. 4. LS AR spectrum of *SX Phe* (data from Stock *et al.*, 1972).

suffered a serious frequency shift (i.e. the average frequency was 0.075 00 Hz). As a conclusion we may say that numerical tests only are unable to give confidence intervals in a general case for the frequency stability of the two-way LS spectrum.

3. Tests by Real Astronomical Data

Finally, we show the efficiency of the two-way LS frequency spectrum estimator by using real astronomical data.

The LS frequency spectrum of *SX Phe* is shown in Figure 4. The data published by Stock *et al.* (1972) were used with $M = 24$, $L = 4$ (the data were made equidistant by quadratic interpolation and the data number was left unchanged). The power was normalized to unity at the highest peak. The primary (0.054 91 day) and the secondary (0.042 02 day) periods show very good agreement with their previously known values (0.054 964 8 and 0.042 773 day, respectively). For comparison, Percy (1977) obtained the values 0.0557 and 0.0423 day respectively using Burg's technique. The nonlinear interaction frequencies are also clearly seen and are very close to their predicted values.

As a second example a part of the photoelectric data of Θ *Tuc* published by Stobie and Shobbrook (1976) was analysed. The data were made equidistant in the same way as for *SX Phe*. The frequency spectra of the individual nights of observation are shown in Figure 5, where each spectrum has been normalized to unity at its highest peak. The letters *a*, *b*, *c*, *d* stand for JD 2441000 + 597, 611, 612, 634, respectively. $M = 14$, $L = 3$ in each case except in case *d*, where $M = 14$, $L = 5$. Except for *d* (where it is indicated only), the spectra show (similarly to Percy's (1977) result) a more or less stable double peak structure in the vicinity of 17 and 20.4 cycle/day (however, by using $M = 30$, $L = 2$ in case *d*, the peak at 19.8 cycle/day splits into peaks at 17.5 and 19.8 cycle/day, respectively). This result is consistent with the Fourier method based frequency spectra obtained by Kurtz (1980) and Pelt (1980) and the small instability in our spectra may partly be accounted for by the frequencies grouped around 20, 18, and 16 cycle/day as was claimed by those studies.

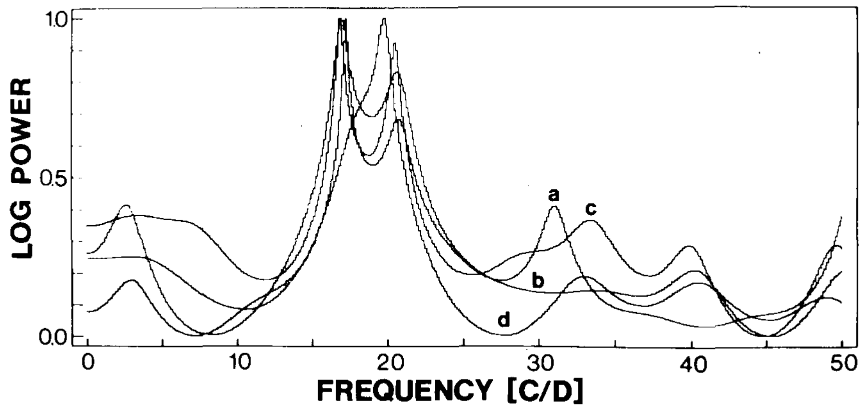

Fig. 5. LS AR spectra of the individual nights of observation of Θ *Tuc* (data from Stobie and Shobbrook, 1976). The letters a, b, c, d stand for JD 2441000 + 597, 611, 612, 634, respectively.

Acknowledgements

It is a pleasant duty to acknowledge that a considerable part of the computation for this note was performed with the PDP 11/45 computer donated to the N. Copernicus Astronomical Center by the U.S. Academy of Sciences following the initiative and through the action of Dr C.R. O'Dell and Mr A. M. Baer.

The author sincerely thanks the Director of the Center, Prof. W. Dziembowski, for hospitality during his stay in Warsaw; Director of the Konkoly Observatory, Dr B. Szeidl, for supporting the author's initiatives to work there; and Dr M. Kozłowski for aspects related to computer handling.

The financial support of the Hungarian and of the Polish Academy of Sciences is also acknowledged.

References

Berk, K. N.: 1974, *Ann. Statistics* **2**, 489.
Burg, J. P.: 1975, 'Maximum Entropy Spectral Analysis', Ph.D. Thesis, Stanford University, Palo Alto, Calif.
Chen, W. Y. and Stegen, G. R.: 1974, *J. Geophys. Res.* **79**, 3019.
Fougere, P. F.: 1977, *J. Geophys. Res.* **82**, 1051.
Kovács, G.: 1981, *Astrophys. Space Sci.* **78**, 175.
Kromer, R. E.: 1970, 'Asymptotic Properties of the Autoregressive Spectral Estimator'; Ph.D. Thesis, Dept. of Statistics Tech. Rpt. No. 13, Stanford University, Stanford, Calif.
Kurtz, D. W.: 1980, *Monthly Notices Roy. Astron. Soc.* **193**, 61.
Pelt, J.: 1980, 'Fequency Analysis of Astronomical Time Sequences', Publication of Tartu Astrophysical Observatory.
Percy, J. R.: 1977, *Monthly Notices Roy. Astron. Soc.* **181**, 647.
Reid, J. S.: 1979, *J. Geophys. Res.* **84**, 5289.
Stock, J., Kunkel, W. E., Hesser, J. E., and Lasker, B. M.: 1972, *Astron. Astrophys.* **21**, 249.
Stobie, R. S. and Shobbrook, R. R.: 1976, *Monthly Notices Roy. Astron. Soc.* **174**, 401.
Swingler, D. N.: 1979, *J. Geophys. Res.* **84**, 679.
Ulrych, T. J. and Clayton, R. W.: 1976, *Phys. Earth Planet. Inter.* **12**, 188.
Ulrych, T. J. and Ooe, M.: 1979, in S. Haykin (ed.), *Nonlinear Methods of Spectral Analysis*, Topics in Applied Physics, Springer-Verlag, Berlin, Heidelberg, New York, 1979, Vol. 34, p. 81.