J. Phys. Earth, **26**, 123-128, 1978

POSSIBLE INSTABILITY IN THE BURG MAXIMUM ENTROPY METHOD

Masanori SAITO

Geophysical Institute, University of Tokyo, Tokyo, Japan

(Received October 21, 1977)

The maximum entropy method (MEM) suggested by Burg is claimed to give power spectral estimates with high resolution, especially for short records. However, we can produce examples for which Burg's algorithm gives unrealistic spectra. It is shown that the algorithm fails when applied to minimum phase wavelets. A new algorithm, one-sided Burg algorithm, is proposed to circumvent this difficulty. Both the new and original algorithms give reasonable spectral estimates when they are applied to stationary data.

The maximum entropy method (MEM) suggested by Burg is claimed to give power spectral estimates with high resolution, especially for short records. Various applications have been reported and results are now accumulating (ULRYCH, 1972a, b; SMYLIE *et al.*, 1973; ULRYCH *et al.*, 1973). The MEM is a natural extension of the autoregressive analysis and is based on the Wold decomposition theorem which says that a discrete stationary stochastic process can always be written as an autoregressive process. It consists of estimating an optimum prediction error filter and the variance of prediction errors. Optimum prediction error filter will be solved most conveniently by Levinson's algorithm (see, WIGGINS and ROBINSON, 1965), provided that the autocorrelation functions are known. For a short record, however, an estimate of the autocorrelation function is, if not impossible, a difficult problem.

Burg proposed a slightly different approach by which not only the prediction error filter but also the autocorrelation function is obtained (see, for example, ULRYCH, 1972b; SMYLIE *et al.*, 1973; ANDERSEN, 1974). Let $a_{m,t}$ ($t=0, 1, \ldots, m$, $a_{m,0}=1$) be the $(m+1)$-term optimum prediction error filter. By a well-known recursive solution (WIGGINS and ROBINSON, 1965), we have

$$\beta_m + k_m \alpha_m = 0 , \tag{1}$$

$$a_{m+1,t} = a_{m,t} + k_m a^*_{m,m+1-t} \qquad t=0, 1, \ldots, m+1 , \tag{2}$$

$$\alpha_{m+1} = \alpha_m + k_m \beta^*_m = \alpha_m(1-|k_m|^2) , \qquad \alpha_0 = r_0 , \tag{3}$$

where * denotes complex conjugate and β_m is defined in terms of the autocorrelation function of input data x_t by

$$\beta_m = \sum_{t=0}^{m} a_{m,t} r_{m+1-t} \,. \tag{4}$$

Instead of giving the autocorrelation function r_t and finding k_m by the use of Eqs. (4) and (1), Burg proposed that k_m be determined so that the square sum of the actual forward *and* backward prediction errors might be minimized. The backward prediction error (or hindsight error) is defined as the output of $a_{m,t}^*$ when applied to the same data but in the time reversed direction. Thus, the forward prediction error $e_{m,t}$ and the backward prediction error $f_{m,t}$ are defined by

$$\begin{aligned} e_{m,t} &= \sum_{\tau=0}^{m} a_{m,\tau} x_{t-\tau} \qquad t=m,\, m+1,\, \ldots,\, N\,, \\ f_{m,t} &= \sum_{\tau=0}^{m} a_{m,\tau}^* x_{t+\tau} \qquad t=0,\, 1,\, 2,\, \ldots,\, N-m\,, \end{aligned} \tag{5}$$

where it is assumed that x_t is given at $t=0, 1, \ldots, N$. Using Eq. (2), we have

$$\begin{aligned} e_{m+1,t} &= e_{m,t} + k_m f_{m,t-m-1}\,, \\ f_{m+1,t} &= f_{m,t} + k_m^* e_{m,t+m+1}\,. \end{aligned} \tag{6}$$

Thus, Burg's requirement that

$$S_{2,m+1} = \sum_{t=m+1}^{N} |e_{m+1,t}|^2 + \sum_{t=0}^{N-m-1} |f_{m+1,t}|^2 = \min, \tag{7}$$

will be met by

$$k_m = -\frac{2\sum_{t=0}^{N-m-1} e_{m,t+m+1} f_{m,t}^*}{\sum_{t=0}^{N-m-1} \left[|e_{m,t+m+1}|^2 + |f_{m,t}|^2 \right]}\,. \tag{8}$$

However, condition (7) is not a unique choice. On the contrary it might be irrelevant in some cases. For example, suppose x_t is a minimum phase wavelet. It is well known that a minimum phase wavelet is completely predictable in the forward direction with a minimum phase prediction filter, except at the first point. Since a prediction error filter has to be the minimum phase (otherwise the process x_t becomes unrealizable), it turns out that a minimum phase wavelet cannot be predicted in the backward direction. Minimization of the sum of the forward and backward prediction errors might therefore result in a non-optimum prediction error filter in a sense that its output is non-white.

The original Burg algorithm, which will be called a 'two-sided' algorithm, because the prediction errors in the two directions are minimized, will be modified so that it might work on minimum phase data. We simply replace condition (7) with

$$S_{1,m+1} = \sum_{t=m+1}^{N} |e_{m+1,t}|^2 = \min. \tag{9}$$

Proceeding as before we find that the optimum k_m in this case is given by

$$k_m = -\frac{\sum_{t=0}^{N-m-1} e_{m,t+m+1} f^*_{m,t}}{\sum_{t=0}^{N-m-1} |f_{m,t}|^2} . \tag{10}$$

Hereafter we call this algorithm 'one-sided'.

The one-sided algorithm has a definite statistical meaning that it produces the maximum likelihood estimate of $a_{m,t}$ provided that the prediction error is normally distributed (JENKINS and WATTS, 1968, p. 189), while the two-sided algorithm has no such meaning. There is no point in minimizing the backward prediction error. It should be mentioned, however, that for stationary data both algorithms give almost the same results. Thus, even if the two may lead to the same estimates for sufficiently long stationary data, it is possible that they result in different estimates for finite non-stationary data.

To illustrate this point, let us consider a minimum phase wavelet

$$x_t = \sin\,[2\pi f_0(t+1)]e^{-\alpha(t+1)} \qquad t \geq 0 . \tag{11}$$

The exact prediction error filter coefficients for this wavelet are

$$a_1 = -2e^{-\alpha} \cos(2\pi f_0) \qquad a_2 = e^{-2\alpha} ,$$

and the energy spectral density of (11) is

$$E(f) \propto |1 + a_1 e^{-2\pi i f} + a_2 e^{-4\pi i f}|^{-2} ,$$

at frequency f. Figure 1 shows the energy spectrum of (11) computed by the two-sided Burg algorithm using 100 data points ($0 \leq t < 100$) for $f_0 = 0.1$ and $\alpha = 0.05$. As seen in this figure the two-sided algorithm produced a very sharp peak near $f = 0.1$ at $m = 2$. Without the knowledge of the true spectrum, one would be tempted to draw a conclusion that the two-sided algorithm gave a sharp resolution. On the contrary, the estimate is quite unreliable and the instability seems to develop with increasing m, although the square sum $S_{2,m}$ is steadily decreasing with m. The instability is clearly seen in the plot of $a_{m,t}$, and the systematic errors in the autocorrelation function computed by the algorithm are also evident (Fig. 2).

The one-sided Burg algorithm and the conventional method, in which the autocorrelation function is first estimated by a zero extension of the given data and the normal equation is solved by Levinson's algorithm, were also applied to the same data, and it was found that they both reproduced the true spectrum after two interactions. The estimated and the true spectra could not be distinguished in Fig. 1. Thus it is concluded that the one-sided algorithm and the conventional algorithm are preferred over the original two-sided Burg algorithm, at least for this specific example.

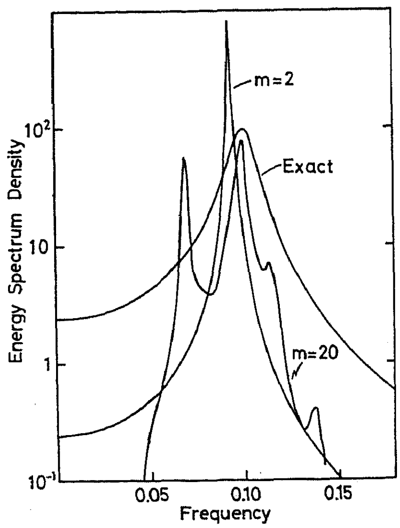

Fig. 1. Energy spectral density estimates computed by the two-sided Burg algorithm. Estimates by the one-sided and the conventional algorithms could not be distinguished from the exact curve in this figure.

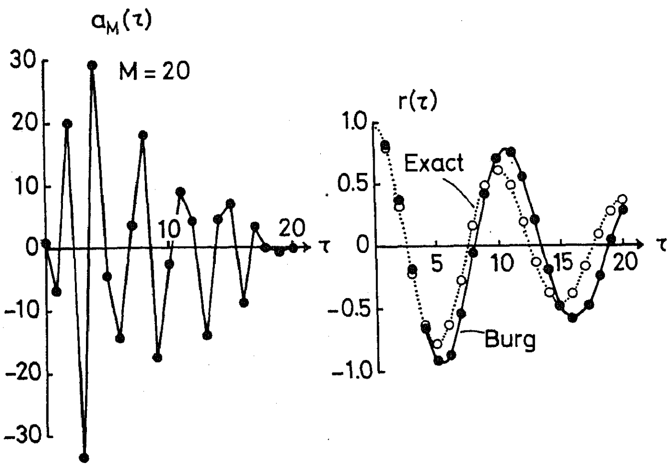

Fig. 2. Prediction error filter coefficients (left) and autocorrelation function (right) computed by the two-sided Burg algorithm.

The failure of the two-sided algorithm is not surprising because a minimum phase wavelet is predictable with a realizable minimum phase filter. But as mentioned before, both algorithms will work for stationary time series. Numerical experiments have been performed in which computer generated random numbers were fed into a two-pole recursive filter and the resulting time series were processed by the three methods. It was found that they all gave almost identical spectral estimates within the accuracy of computation.

Although the two-sided algorithm might give unreliable spectra, it is numerically stable and it produces minimum phase prediction error filters almost always. To show this we note $|k_m| \leq 1$ because of Eq. (8) and use a theorem due to Rouché:

If two polynomials $P(z)$ and $Q(z)$ satisfy

$$|P(z)| > |Q(z)| \quad \text{on } |z| = 1 ,$$

then $P(z)$ and $P(z)+Q(z)$ have the same number of zeros inside the unit circle, $|z| < 1$.

Let $A_m(z)$ be the z-transform of $a_{m,t}$, i.e., $A_m(z) = \sum_t a_{m,t} z^t$, then Eq. (2) will be written as

$$A_{m+1}(z) = A_m(z) + k_m z^{m+1} A_m^*(1/z) .$$

We take $P(z) = A_m(z)$ and $Q(z) = k_m z^{m+1} A_m^*(1/z)$ and the theorem holds true provided that $|k_m| < 1$. Hence $P(z) = A_m(z)$ and $P(z)+Q(z) = A_{m+1}(z)$ have the same number of zeros inside the unit circle on the complex z-plane. Repeating the same argument we find that $A_m(z)$ and $A_0(z) = 1$, which has no zero, have the same number of zeros inside the unit circle. Thus it is proved that $A_m(z)$ has no zeros in the unit circle provided that $|k_m| < 1$. This completes the proof that $a_{m,t}$ obtained by the two-sided Burg algorithm is always the minimum phase (the proof by SMYLIE *et al.* (1973) seems valid only when m is infinite).

On the other hand $|k_m| \leq 1$ is not always true in the one-sided algorithm and it is probable that it diverges. It certainly fails when applied to maximum phase wavelets, but this could be an advantage rather than a drawback, because one can easily find that the algorithm has been applied in a wrong direction.

Finally, the time requirement and storage requirement for each algorithm are estimated (Table 1). The large differences are due to the fact that $e_{m,t}$ and

Table 1. Approximate estimates of computation time and storage requirements.

Algorithm	Time	Storage
Two-sided Burg algorithm	$M(5N-2M)$	$2N+M$
One-sided Burg algorithm	$M\left(4N-\frac{3}{2}M\right)$	$2N+M$
Conventional algorithm	$M\left(N+\frac{1}{2}M\right)$	$2M$

N denotes the number of data and M the length of the prediction error filter. The computational time is represented by the number of multiply-and-add operations.

$f_{m,t}$ have to be computed in the Burg algorithms. The conventional algorithm is best from a computational point of view, provided that the length of data is sufficiently long.

REFERENCES

ANDERSEN, N., On the calculation of filter coefficients for maximum entropy spectral analysis, *Geophysics*, **39**, 69–72, 1974.

JENKINS, G. M. and D. G. WATTS, *Spectral Analysis and Its Applications*, Holden-Day, 1968.

SMYLIE, D. E., G. K. C. CLARKE, and T. J. ULRYCH, Analysis of irregularities in the earth's rotation, in "*Methods in Computational Physics*," Vol. 13, pp. 391–430, Academic Press, 1973.

ULRYCH, T. J., Maximum entropy power spectrum of long period geomagnetic reversals, *Nature*, **235**, 218–219, 1972a.

ULRYCH, T. J., Maximum entropy power spectrum of truncated sinusoids, *J. Geophys. Res.*, **77**, 1396–1400, 1972b.

ULRYCH, T. J., D. E. SMYLIE, O. G. JENSEN, and G. K. C. CLARKE, Predictive filtering and smoothing of short records by using maximum entropy, *J. Geophys. Res.*, **78**, 4959–4964, 1973.

WIGGINS, R A. and E. A. ROBINSON, Recursive solution to the multi-channel filtering problem, *J. Geophys. Res.*, **70**, 1885–1891, 1965.