Astron. Astrophys. 93, 269–281 (1981)

ASTRONOMY
AND
ASTROPHYSICS

The Maximum Entropy Method of Image Restoration-properties and Limitations

M. M. Komesaroff*, R. Narayan, and R. Nityananda
Raman Research Institute, Bangalore, India-560006

Received February 19, accepted August 5, 1980

Summary. The maximum entropy method (MEM) of spectral analysis is examined from a somewhat novel viewpoint.

A spectral function being necessarily positive, its (discrete) Fourier transform is an autocorrelation sequence. We show that associated with an autocorrelation sequence there is a sequence of unit vectors whose scalar products give the autocorrelations. Specifying the autocorrelation up to some finite lag imposes constraints on the vector sequence and thus on unspecified autocorrelation terms. Each unspecified term lies within a convex connected area in the complex plane. The first lies within a circle; we give expressions for its centre and radius. Any choice within this "circle of constraint" is consistent with positivity, but choosing the centre maximizes the area available to the next unspecified term, which is a circle of the same radius. Sequentially setting each term to the centre of its circle maximizes the area available for each subsequent term; this procedure also yields the MEM spectral estimate.

The MEM is shown to be a type of model fitting; the model consists of a number of asymmetric "peaks" equal to the number of autocorrelation terms originally specified. Since each peak requires two complex numbers to specify it, the individual characteristics of the peaks are not independently adjustable. This can lead to undesirable results, for example spurious splitting of peaks.

If the circles of constraint contract to points the peaks reduce to delta functions. Measuring error drives the autocorrelations towards this fully constrained limit, and thus tends to enhance the "peakiness" of the spectrum.

These and other features of the MEM are substantiated by numerical examples.

Key words: maximum entropy – spectrum – autocorrelation – deconvolution

I. Introduction

The most obvious basic problem in radio astronomy is to measure the distribution of the radio "brightness" as a function of angular position in the sky. To achieve angular resolution finer than that provided by the largest single reflector, one has recourse to the technique of aperture synthesis. In principle, aperture synthesis may be thought of as a procedure involving only two aerials, each aerial having its own receiver, and each aerial receiver combination accepting the same band of radio frequencies. A single observation involves measuring the correlation between the radio-frequency signals from the two aerials for one particular spacing between them. A set of such measurements made for all possible spacings yields the spatial autocorrelation function of the electromagnetic field. The brightness distribution is then the Fourier transform of this autocorrelation function. It is not necessary to measure the autocorrelation as a continuous function of the spacing between the aerials. It need only be measured at integral multiples of a certain basic spacing calculated from the sampling theorem. Thus the measurements yield a discrete set of numbers which constitute an autocorrelation sequence. An immediate and obvious difficulty appears; the brightness distribution is derived from the measured autocorrelation sequence by a Fourier Transformation. But this requires that we know the autocorrelation sequence out to infinite spacing. The question then arises: since our measurements can necessarily only yield a finite set of samples of the autocorrelation sequence, what is the best estimate we can make of the brightness distribution? The same problem arises in the estimation of radio frequency power spectra by the autocorrelation technique and also in various other branches of experimental physics.

Traditionally two procedures have been adopted to cope with this difficulty. In the first, all unmeasured values of the autocorrelation sequence are set to zero and a Fourier inversion carried out on the truncated sequence. The disadvantages of the procedure are obvious. Setting unknown values to zero is completely arbitrary, and leads to ripple, or sidelobes, in the resulting brightness distribution. Furthermore, it leads to areas of negative brightness which are clearly non-physical since brightness by its nature must be positive.

Another procedure is to apply a set of weighting factors to the measured autocorrelation values before carrying out the Fourier inversion. The weighting factors decrease monotonically with increasing spacing. This procedure can largely eliminate side lobes, but at the cost of a degradation in angular resolution. An objection which is perhaps more basic follows from the fact that the autocorrelation sequence and the brightness distribution are related by a Fourier transformation. A Fourier inversion of the estimated brightness distribution should therefore yield the measured autocorrelation sequence, but, of course, it now will not do so. In this respect the result does not represent the measurements.

Another approach was adopted by Biraud (1969). Starting from a set of interferometric measurements on the source, he

Send offprint requests to: R. Narayan
* Permanent address: CSIRO, Division of Radiophysics, P.O. Box 76, Epping, NSW 2121, Australia

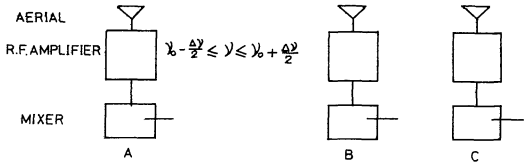


Fig. 1. Schematic representation of three identical aerial receiver combinations

used a recursive technique to calculate a function, nowhere negative, the Fourier transform of which fitted the data best in a least squares sense. He was able to show that the brightness distribution so derived was a better representation of the actual source than the result of the more conventional technique.

The feasibility of a procedure such as that used by Biraud depends on the fact that an autocorrelation sequence, which is the Fourier transform of a non-negative function, is not the most general type of Fourier transform, and therefore the terms of such a sequence are not mutually independent. This means that if we are given the values, ϱ_0, ϱ_1 ... ϱ_n, of the first $(n+1)$ terms of such a sequence, we can set limits on the values which terms of higher order can take. This fact is basic to the Maximum Entropy approach to spectral analysis although several of the papers on this subject do not state this explicitly.

In Sect. II of the present paper the possibility of extrapolating an autocorrelation sequence is made plausible using a simple example from aperture synthesis. We show that if we know the values of two terms of an autocorrelation sequence, we can set limits on a third term.

Sections III and IV represent an extension of an argument due to Komesaroff and Lerche (1978). In Sect. III it is shown that if we have a set of numbers $\varrho_0, \ldots \varrho_n$ normalized so that $\varrho_0 = 1$, a necessary and sufficient condition that these should form part of an autocorrelation sequence is that a set of unit vectors V_j should exist such that each autocorrelation term can be represented as a scalar product of the form

$$\varrho_m = V_j V^*_{j-m}; \quad 0 \leqq m \leqq n$$

for all j. This requirement can be restated in terms of the Toeplitz matrix $T(n)$, defined by

$$t_{jk} = \varrho_{k-j} = \varrho^*_{j-k}$$

in which j and k run from zero to n. If $\varrho_0 \ldots \varrho_n$ are members of an autocorrelation sequence, then $T(n)$ must be non-negative definite. A further consequence is that if we know the values of $\varrho_0, \ldots \varrho_n$, each term of higher order can be localized within a single connected convex region in the complex plane. This result is true in any number of dimensions (Appendix II).

Returning to the one-dimensional case, it is shown in Sect. IV that if the first term of higher order is set to the centre of its permitted region and this procedure is repeated indefinitely, the result is equivalent to the maximum entropy spectral estimate. This gives a new derivation, by the vector approach, of a result which Burg (1975) has derived from the time series point of view. This choice is shown to maximize the Toeplitz determinant $\det(T(n+m))$ for all $m > 0$. As discussed in Sect. IV, van den Bos (1971) has given a related but weaker result.

The form of the maximum entropy estimate is discussed in Sect. V. It is first shown that if $\varrho_0, \ldots \varrho_n$ are members of an autocorrelation sequence, then provided the Toeplitz matrix $T(n)$ is non-singular, the maximum entropy extrapolation procedure yields an autocorrelation sequence $\varrho(n+m)$ which decreases exponentially to zero as $m \to \infty$. The corresponding spectrum can be resolved into a set of n "peaks", each of finite width, and we can set a lower limit on the width of any such peak. It is shown that in general the individual peaks are not Lorentzian nor even symmetric, as seems to have been assumed earlier.

For the special case in which the Toeplitz matrix is singular the areas in the complex plane corresponding to terms of higher order contract to points. The solution is then fully constrained, and it is shown that the corresponding spectrum is a set of n delta functions. Again we note that Burg (1975) has already derived this result. The alternative derivation given here from the vector point of view is particularly simple and appealing.

In Sect. VI, the special nature of the model which the maximum entropy method (MEM) fits to the measured data is discussed. Numerical examples are given demonstrating the very succesful results which the MEM can yield in some cases, but also demonstrating some of the artefacts and distortions which it can introduce. The effect of measuring error on the MEM spectrum is then discussed, both analytically and numerically.

II. A Simple Example

In this section a simple example from aperture synthesis is used to illustrate the method adopted in this paper, and also to show that if even the values of only two terms in an autocorrelation sequence are specified, this allows us to set limits on the values which unspecified terms may take.

Consider three identical receiving systems, as shown in Fig. 1. Each consists of an aerial, a radio frequency amplifier and a mixer. Each aerial-receiver combination has a uniform response within the frequency range $\nu_0 - \Delta\nu/2 \leqq \nu_0 \leqq \nu_0 + \Delta\nu/2$ and zero response outside that range.

In order to measure the (real part of the) spatial autocorrelation term associated with the spacing AB (and thus the corresponding Fourier component of the brightness distribution) we correlate the i.f. voltage from mixer A with that from mixer B, and similarly for the other two correlations. Since each system is band limited, we need only sample each i.f. voltage at integral multiples of the time interval

$$\Delta t = 1/2\ \Delta\nu.$$

Denoting the normalized autocorrelation term corresponding to the spacing AB by ϱ_{AB} etc., we can then write

$$\varrho_{AB} = \sum_{k=1}^{M} V_{Ak} V_{Bk}, \quad \varrho_{BC} = \sum_{k=1}^{M} V_{Bk} V_{Ck}, \quad \varrho_{AC} = \sum_{k=1}^{M} V_{Ak} V_{Ck} \tag{1}$$

where V_{Ak}, V_{Bk}, and V_{Ck} are the kth voltage samples from receiving systems A, B, and C respectively, normalized so that

$$\sum_{k=1}^{M} V^2_{Ak} = \sum_{k=1}^{M} V^2_{Bk} = \sum_{k=1}^{M} V^2_{Ck} = 1 \tag{2}$$

and $M\Delta t$ is the total duration of the observations.

Each normalized autocorrelation term has the form of the scalar product of two unit vectors. We may write

$$\varrho_{AB} = V_A V_B \tag{3}$$

where the kth components of the unit vectors V_A and V_B are respectively V_{Ak} and V_{Bk}. These vectors as so defined exist in a space of M dimensions, and usually M is very large. In principle, to measure the true autocorrelation ϱ_{AB} requires $M \to \infty$. However, as we are interested in the relations between the vectors and

not in their absolute orientations, we may choose a system of coordinates such that V_A lies along the x axis and V_B lies in the plane defined by the x and y axes. Thus the three unit vectors can be represented in a space of three dimensions, and each correlation term is equal to the cosine of the angle between the corresponding vectors.

Therefore, as illustrated in Fig. 2, if we know the autocorrelation terms ϱ_{AB} and ϱ_{BC}, we know that the angle between V_A and V_B is θ_{AB} and that V_C must lie on a cone, of which V_B is the axis, and of which the semi-apex angle is θ_{BC}, where

$$\cos\theta_{AB}=\varrho_{AB},\quad \cos\theta_{BC}=\varrho_{BC}. \tag{4}$$

It is than clear from Fig. 2 that the angle θ_{AC} must satisfy the inequality

$$|\theta_{AB}-\theta_{BC}|\leqq\theta_{AC}\leqq\text{the smaller of } \theta_{AB}+\theta_{BC} \text{ and } 2\pi-(\theta_{AB}+\theta_{BC})$$

from which it follows that

$$\cos(\theta_{AB}-\theta_{BC})\geqq\cos\theta_{AC}\geqq\cos(\theta_{AB}+\theta_{BC})$$

$$\varrho_{AB}\varrho_{BC}+(1-\varrho_{AB}^2)^{1/2}(1-\varrho_{BC}^2)^{1/2}\geqq\varrho_{AC}\geqq\varrho_{AB}\varrho_{BC} \tag{6}$$
$$-(1-\varrho_{AB}^2)^{1/2}(1-\varrho_{BC}^2)^{1/2}.$$

Thus given the values of ϱ_{AB} and ϱ_{BC}, we can set limits on the values which ϱ_{AC} can take. If the magnitudes of the known terms are close to unity, these limits will be very narrow.

For the particular case in which the three receiving systems are colinear and equally spaced we can write

$$\varrho_{AB}=\varrho_{BC}=\varrho_1$$

since the correlation depends on the element spacing but not on the location of the interferometer pair. Writing $\varrho_{AC}=\varrho_2$, it then follows that

$$1\geqq\varrho_2\geqq2\varrho_1^2-1. \tag{7}$$

If $1-\varrho_1^2$ is sufficiently small we can use this method to set lower limits to higher order terms corresponding to successively higher integral multiples of the basic spacing.

III. The Fourier Transform of a Non-negative Function

a) A Finite Autocorrelation Sequence

The general problem we are investigating may be stated as follows. If we know that the function

$$F(x)\geqq0 \tag{8}$$

for all x within the range

$$-1/2\leqq x\leqq+1/2 \tag{9}$$

and zero elsewhere, what constraints does this impose on the Fourier transform of $F(x)$?

In view of (8), $F(x)$ may be expressed by

$$F(x)=f(x)f^*(x)=|f(x)|^2. \tag{10}$$

Since we are interested in the values of $F(x)$ and $f(x)$ only within the range of x given by (9) it is sufficient to specify the Fourier transform of each of these functions at discrete intervals and thus we may write

$$F(x)=\sum_{m=-\infty}^{+\infty}\varrho_m\exp(2\pi imx) \tag{11}$$

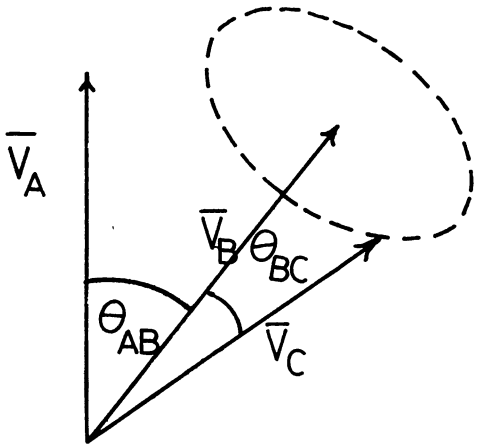


Fig. 2. Representation of the three unit vectors defined in Sect. II

$$f(x)=\sum_{m=-\infty}^{+\infty}g_m\exp(2\pi imx). \tag{12}$$

Taking account of Eqs. (10), (11), and (12) and the convolution theorem, it then follows that

$$\varrho_m=\sum_{k=-\infty}^{+\infty}g_k g_{k-m}^*. \tag{13}$$

A set of numbers ϱ_m, in general complex, given by a relation of the form (13) is said to be an autocorrelation sequence. The right hand side of (13) may be regarded as the scalar product of two (complex) vectors. Let us normalize the ϱ_k so that $\varrho_0=1$. Then, if the set of $n+1$ complex numbers $\varrho_0,\ldots\varrho_n$ form part of an autocorrelation sequence, a set of $n+1$ unit vectors must exist such that

$$\varrho_m=V_j V_{j-m}^* \tag{14}$$

for all j and m. Thus the existence of a set of vectors satisfying the condition (14) is a necessary condition for the ϱ_m to form part of an autocorrelation sequence.

Let us now construct an arbitrary vector

$$V=\sum_{k=0}^{n}a_k V_k \tag{15}$$

where the a_k are arbitrary complex numbers. Taking the squared modulus of both sides of Eq. (15)

$$VV^*=\sum_{l=0}^{n}\sum_{k=0}^{n}a_l a_k^* V_l V_k^* \tag{16}$$

$$=\sum_{l=0}^{n}\sum_{k=0}^{n}a_l a_k^*\varrho_{l-k} \tag{17}$$

from (14).

Equations (17) may be written in the equivalent form

$$A^\dagger T(n)A\geqq0 \tag{18}$$

where A is the column vector of which a_k is the kth component, and the Toeplitz matrix $T(n)$ is defined by

$$T_n=\begin{pmatrix}\varrho_0 & \varrho_1 & \varrho_2 & \cdots & \varrho_n\\ \varrho_1^* & \varrho_0 & \varrho_1 & \cdots & \varrho_{n-1}\\ \vdots & \vdots & \vdots & & \vdots\\ \varrho_n^* & \varrho_{n-1}^* & \varrho_{n-2}^* & \cdots & \varrho_0\end{pmatrix} \tag{19}$$

Since the elements of the vector A are completely arbitrary, Eq. (18) implies that $T(n)$ must be non-negative definite.

But according to Bochner's Theorem (see for example Papoulis, 1965) inequality (18) is not only a necessary but also a sufficient condition for $\varrho_0, \varrho_1, \ldots \varrho_n$ to be members of an autocorrelation sequence and thus part of the Fourier transform of a non-negative function. It follows that the existence of a set of unit vectors satisfying Eq. (14) for all j and m is a sufficient as well as a necessary condition.

b) Extending the Autocorrelation Sequence

The question now arises – if we know the values $\varrho_0 \equiv 1, \varrho_1, \ldots \varrho_n$ of the first $n+1$ terms of a uniformly sampled autocorrelation sequence, what constraints does this allow us to place on the terms of higher order?

The starting point is Eq. (14). In geometric terms, knowing $\varrho_0, \varrho_1, \ldots \varrho_n$ is equivalent to knowing the (complex) scalar products among any $n+1$ consecutive members of the infinite sequence of vectors V_j. This completely determines their "internal geometry". Therefore one can express a given unit vector V_j uniquely as a linear combination of the n vectors V_{j-1}, V_{j-2}, $\ldots$, V_{j-n} and a vector orthogonal to these. Thus

$$V_j = \sum_{k=1}^{n} \alpha_{k,n} V_{j-k} + \beta_n \varepsilon_j \tag{20}$$

where β_n (clearly of modulus <1) and the $\alpha_{k,n}$ are constants independent of j and the unit vector ε_j is such that

$$\varepsilon_j V^*_{j-p} = 0; \quad 0 < p \leqq n. \tag{21}$$

Taking the scalar product of (20) with V^*_{j-p}

$$V_j V^*_{j-p} = \sum_{k=1}^{n} \alpha_{k,n} V_{j-k} V^*_{j-p} + \beta_n \varepsilon_j V^*_{j-p} \tag{22}$$

and, in view of (14) and (21), this may be written as

$$\varrho_p = \sum_{k=1}^{n} \alpha_{k,n} \varrho_{p-k}; \quad 0 < p \leqq n. \tag{23}$$

From (20) it also follows that

$$\left(V_j - \sum_{k=1}^{n} \alpha_{k,n} V_{j-k}\right)\left(V^*_j - \sum_{l=1}^{n} \alpha^*_{l,n} V^*_{j-l}\right) = \beta_n^2 \tag{24}$$

and taking account of (14) and (23), we have

$$\varrho_0 - \sum_{k=1}^{n} \alpha_{k,n} \varrho^*_k = \varrho_0 - \sum_{k=1}^{n} \alpha^*_{k,n} \varrho^*_k = \beta_n^2. \tag{25}$$

Equations (23) and (25) may be written

$$\begin{pmatrix} \varrho_0 & \varrho_1^* & \varrho_2^* & \cdots & \varrho_n^* \\ \varrho_1 & \varrho_0 & \varrho_1^* & \cdots & \varrho_{n-1}^* \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \varrho_n & \varrho_{n-1} & \varrho_{n-2} & \cdots & \varrho_0 \end{pmatrix} \begin{pmatrix} 1 \\ -\alpha_{1,n} \\ \cdot \\ \cdot \\ -\alpha_{n,n} \end{pmatrix} = \begin{pmatrix} \beta_n^2 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}. \tag{26}$$

Knowing the values of $\varrho_0, \varrho_1, \ldots \varrho_n$, we may solve these $(n+1)$ equations for the $n+1$ unknowns β_n^2 and $\alpha_{k,n}$. In fact, the full matrix inversion is not required. An economic procedure for evaluating β_n^2 and the $\alpha_{k,n}$ is described in Appendix I. The values of these constants are independent of j and thus we may write equations analogous to (20) and (21) for the set of vectors $V_{j+1}, V_j, \ldots, V_{j+1-n}$

$$V_{j+1} = \sum_{k=1}^{n} \alpha_{k,n} V_{j+1-k} + \beta_n \varepsilon'_{j+1} \tag{27}$$

$$\varepsilon'_{j+1} V^*_{j+1-p} = 0; \quad 0 < p \leqq n. \tag{28}$$

We note that the scalar product of V_{j+1} with V_{j-n} is ϱ_{n+1} which is not known. This implies from (27) that the scalar product of the unit vector ε'_{j+1} with V_{j-n} is not fixed by the available information. We will now determine the freedom which ε'_{j+1} has and thus obtain bounds for ϱ_{n+1}. We decompose ε'_{j+1} into two mutually orthogonal parts

$$\varepsilon'_{j+1} = \lambda_1 \eta + \Lambda_1 \varepsilon_{j+1} \tag{29}$$

$$|\lambda_1|^2 + |\Lambda_1|^2 = 1 \tag{30}$$

where the unit vector η is a linear combination of the set V_j, $V_{j-1}, \ldots, V_{j-n}$ and ε_{j+1} is a unit vector orthogonal to this set.

$$\varepsilon_{j+1} V^*_{j+1-p} = 0; \quad 0 < p \leqq n+1. \tag{31}$$

Substituting (29) into (28) and using (31) we see that η has to be orthogonal to $V_j, V_{j-1}, \ldots, V_{j+1-n}$.

$$\eta V^*_{j+1-q} = 0; \quad 0 < q \leqq n. \tag{32}$$

Equation (32) and the requirement that η be a linear combination of the set $V_j, V_{j-1}, \ldots, V_{j-n}$ show us that the role of η is similar to that of ε_j in (20) and (21). In fact we can write

$$V_{j-n} = \sum_{k=1}^{n} \alpha^*_{k,n} V_{j-n+k} + \beta_n \eta. \tag{33}$$

The only difference between (33) and (20) is that the complex conjugates of $\alpha_{k,n}$ appear. This is easily verified by constructing the equation corresponding to (26) for the coefficients in (33).

Having determined η from (33) we can substitute Eq. (29) for ε'_{j+1} into Eq. (27) for V_{j+1} to obtain

$$V_{j+1} = \sum_{k=1}^{n} \alpha_{k,n} V_{j+1-k} + \lambda_1 \left(V_{j-n} - \sum_{k=1}^{n} \alpha^*_{k,n} V_{j-n+k}\right) + \Lambda_1 \beta_n \varepsilon_{j+1} \tag{34}$$

Taking the scalar product with V^*_{j-n}, we obtain, from (14), (25), (31), and (34),

$$\varrho_{n+1} = \sum_{k=1}^{n} \alpha_{k,n} \varrho_{n+1-k} + \lambda_1 \beta_n^2 \tag{35}$$

where λ_1 is an unknown constant, the only condition which it must satisfy being [from (30)]

$$|\lambda_1| \leqq 1. \tag{36}$$

Thus, given the values of $\varrho_0 \equiv 1, \varrho_1, \ldots, \varrho_n$, we can localize ϱ_{n+1} to a circle in the complex plane. The radius of the circle is β_n^2 and its centre is at the point represented by the complex number

$$R_n = \sum_{k=1}^{n} \alpha_{k,n} \varrho_{n+1-k}. \tag{37}$$

All "acceptable" values of ϱ_{n+1} lie within this circle. If we assign to ϱ_{n+1} a value lying outside this circle, there is no non-negative function whose Fourier transform passes through the points $\varrho_0, \varrho_1, \ldots, \varrho_n, \varrho_{n+1}$.

An argument essentially identical with that given above yields

$$\varrho_{n+2} = \sum_{k=1}^{n} \alpha_{k,n} \varrho_{n+2-k} + \lambda_1 \left(\varrho_1 - \sum_{k=1}^{n} \alpha^*_{k,n} \varrho_{k+1}\right) + \lambda_2 (1 - |\lambda_1|^2) \beta_n^2 \tag{38}$$

where $|\lambda_2| \leqq 1$. Thus, acceptable values of ϱ_{n+2} depend on the value of ϱ_{n+1} which in turn depends on λ_1. Any combination of λ_1 and λ_2 such that $|\lambda_1| \leqq 1$ and $|\lambda_2| \leqq 1$ yields an acceptable value of ϱ_{n+2}. All such acceptable values of ϱ_{n+2} lie within an area in the complex plane which in general is not circular and which is always larger than the circle within which ϱ_{n+1} lies. Any particular

choice of λ_1, and thus of ϱ_{n+1} imposes a further constraint on ϱ_{n+2}, that is, further restricts the area within which ϱ_{n+2} must lie. This area is a circle of radius $(1-|\lambda_1|^2)\beta_n^2$. The area of the circle is a maximum when $\lambda_1=0$, in which case the radius becomes β_n^2.

In general, the values of $\varrho_0, \varrho_1, \ldots, \varrho_n$ constrain all ϱ_{n+m} (where $m>0$) to successively larger areas in the complex plane. If we assign "acceptable" values to $\varrho_{n+1}, \varrho_{n+2}, \ldots, \varrho_{n+m-1}$, these impose additional constraints on ϱ_{n+m} which must then lie within a circle, the radius of which cannot exceed β_n^2.

Appendix II shows that the region in which higher members of the autocorrelation sequence are allowed to lie is convex.

IV. Transition to the Maximum Entropy Spectral Estimate

On what basis can we select a combination of the higher order ϱ's from all the acceptable combinations? A priori we cannot exclude any combination; but the following argument suggests a procedure for making a selection. As we shall see, this procedure leads to the maximum entropy spectral estimate.

Let us first consider the effect of setting ϱ_{n+1} to one of its extreme values [see Eq. (35)]. That is, we set $|\lambda_1|=1$ and also fix its phase, so that $\lambda_1=e^{i\varphi}$ where φ is some definite real number. In that case, as Eq. (38) shows, we no longer have any freedom in our choice of ϱ_{n+2}. It is constrained to one particular value. It is also easily shown that each ϱ of higher order is constrained to a particular value. If we had some independent method of determining the value of ϱ_{n+10} (say) we have no reason to suppose that this value would agree with the one derived from the fully constrained extrapolation procedure. Furthermore, as shown in Sect. V below, the function for which we have now determined the Fourier transform is a set of $(n+1)$ delta functions. Thus in all respects, this procedure is arbitrary, and it also implies infinite resolution on the basis of a finite number of measurements.

Consider on the other hand the effect of setting $\lambda_1=0$, so that ϱ_{n+1} lies at the centre of its permitted range (Eq. 35). This minimizes the greatest possible difference between the unknown true value of ϱ_{n+1} and our estimate of it. Further, the radius of the circle to which ϱ_{n+2} becomes confined by our choice of ϱ_{n+1} attains its largest value, viz., β_n^2 (Eq. 38). The procedure of setting each successive term in the sequence to the centre of its permitted range minimally constrains all higher order terms in the sequence. As shown in Sect. V, this procedure leads to an autocorrelation sequence ϱ_{n+m} which decreases exponentially to zero as $m\to\infty$, and thus does not imply infinite resolution. It is equivalent to the iterative algorithm

$$\varrho_m-\sum_{k=1}^{n}\alpha_{k,n}\varrho_{m-k}=0;\ m>0$$
$$=\beta_n^2;\ m=0. \tag{39}$$

It is also equivalent to setting

$$\varepsilon_{j+m}\varepsilon_j^*=\delta_{m,0} \tag{40}$$

where ε_{j+m} is given by

$$V_{j+m}=\sum_{k=1}^{n}\alpha_{k,n}V_{j+m-k}+\beta_n\varepsilon_{j+m} \tag{41}$$

or

$$\sum_{k=0}^{n}\gamma_k V_{j+m-k}=\beta_n\varepsilon_{j+m} \tag{42}$$

where

$$\gamma_0=1,\quad \gamma_k=-\alpha_{k,n};\quad 0<k\leqq n. \tag{43}$$

Thus from Eqs. (40) and (42)

$$\left(\sum_{k=0}^{n}\gamma_k V_{j+m-k}\right)\left(\sum_{l=0}^{n}\gamma_l^* V_{j-l}^*\right)=\beta_n^2\delta_{m,0}$$

$$\sum_{k=0}^{n}\gamma_k\sum_{l=0}^{n}\gamma_l^*\varrho_{m+l-k}=\beta_n^2\delta_{m,0} \tag{44}$$

from Eq. (14).

Taking Fourier transforms we obtain

$$\sum_{m=-\infty}^{+\infty}\exp(2\pi imx)\sum_{k=0}^{n}\gamma_k\sum_{l=0}^{n}\gamma_l^*\varrho_{m+l-k}=\beta_n^2\sum_{m=-\infty}^{+\infty}\delta_{m,0}\exp(2\pi imx)$$

Rearranging the right hand side and setting $M=m+l-k$

$$\sum_{M=-\infty}^{+\infty}\varrho_M\exp(2\pi iMx)\sum_{k=0}^{n}\gamma_k\exp(2\pi ikx)\sum_{l=0}^{n}\gamma_l^*\exp(-2\pi ilx)$$
$$=\beta_n^2.$$

Thus, if the autocorrelation sequence ϱ_M has its first $n+1$ members equal to the known autocorrelation terms $\varrho_0\ldots\varrho_n$ and if the values of ϱ_M for $M>n$ are derived by the extrapolation procedure described previously, then the Fourier transform, $S(x)$ of ϱ_M is given by

$$S(x)=\sum_{M=-\infty}^{+\infty}\varrho_M\exp(2\pi iMx)$$
$$=\frac{\beta_n^2}{\sum_{k=0}^{n}\gamma_k\exp(2\pi ikx)\sum_{l=0}^{n}\gamma_l^*\exp(-2\pi ilx)}. \tag{45}$$

The function $S(x)$ is identical to the "maximum entropy" spectral estimate, where the definition of the entropy H of a function $F(x)$ is

$$H=\int\ln F(x)dx. \tag{46}$$

(See Burg, 1967; Ables, 1974; Newman, 1977).

From Eq. (26) we can easily solve for β_n^2 to obtain

$$\beta_n^2=\det(T(n+1))/\det(T(n)). \tag{47}$$

From this equation we obtain

$$\det T(n+m)=\beta_{n+m}^2\beta_{n+m-1}^2\ldots\beta_{n+1}^2\det(T(n)). \tag{48}$$

Now Eq. (38) shows that the largest possible value of β^2 at each stage is β_n^2 and is only attained if we choose $\lambda=0$ at every stage, as in (39) and (40). Thus, our procedure of giving β^2 its largest value at each stage is equivalent to maximizing $\det(T(n+m))$ with respect to all the unknown correlations which occur in it.

Van den Bos (1971), has noted that each new choice of an unknown autocorrelation according to the maximum entropy scheme maximizes the determinant at that stage, but asserts that this is not equivalent to giving $\det(T(n+m))$ its absolute maximum value with respect to all unknown autocorrelations in it. The above argument shows that the procedure does indeed yield the absolute maximum value of $\det(T(n+m))$.

We see directly that the above procedure maximizes the entropy when we note that as $m\to\infty$, $\ln\beta_m^2$ tends to a value proportional to the entropy as defined in (46). An elegant proof of this result can be found in the review by Smylie et al. (1973, Eq. 17).

V. Properties of the Maximum Entropy Autocorrelation and Spectrum

a) Decay of the Autocorrelation Coefficients

Here we examine the form of the autocorrelation and the corresponding spectrum resulting from the extrapolation procedure described in Sect. IV. We show that, provided the matrix $T(n)$ is non-singular and, therefore, from (48) $\beta_n^2 > 0$, the autocorrelation ϱ_m decreases exponentially for very large values of m, and thus the spectrum contains no δ-functions. In fact the spectrum consists of n asymmetric "peaks" each of finite width. It is shown that an approximate lower limit can be set on the width of any peak, this lower limit being proportional to β_n^2. For $\beta_n^2 = 0$, the spectrum consists of n δ-functions.

We first define a linear operation S (for shift) which takes any vector in the sequence into the next one, i.e.

$$SV_j = V_{j+1}. \tag{49}$$

Also

$$(SV_j)\,(SV_k)^* = V_{j+1}V_{k+1}^* = \varrho_{j-k} = V_jV_k^*. \tag{50}$$

Equation (50) tells us that the scalar product of any two vectors in the sequence is preserved if we apply the shift operation to both of them. This is true for linear combinations of the vectors V_j also. Thus S is a unitary operation.

Now we define a space N, consisting of the n vectors V_{j-1} to V_{j-n} and their linear combinations. The basic recurrence relation between the vectors, Eq. (20), reads

$$V_j = \sum_{k=1}^{n} \alpha_{k,n} V_{j-k} + \beta_n \varepsilon_j \tag{20}$$

with ε_j orthogonal to all vectors in N by Eq. (21). The first term on the right hand side of (20) lies in the space N. Applying the shift S to both sides of (20) we find $S\varepsilon_j = \varepsilon_{j+1}$, and in general,

$$S^m\left(V_j - \sum_{k=1}^{n} \alpha_{k,n} V_{j-k}\right) = \beta_n S^m \varepsilon_j = \beta_n \varepsilon_{j+m}$$

or

$$V_{j+m} - \sum_{k=1}^{n} \alpha_{k,n} V_{j+m-k} = \beta_n \varepsilon_{j+m}. \tag{51}$$

The argument in Sect. IV has shown that this recurrence relation by itself does not uniquely determine the ε_{j+m} or the V_{j+m}. We therefore add to our definition of S the requirement that the ε_{j+m} as given by Eq. (51) be such that

$$\varepsilon_{j+m} V_{j-p}^* = 0; \quad 1 \leqq p \leqq n, \quad m \geqq 0. \tag{52}$$

Thus the shift operator so defined keeps a vector orthogonal to the space N if it is so to start with.

We consider two projection operators, P and Q, which acting on any vector, pick out the component of it in the space N and the component orthogonal to N respectively. Clearly, these two components together make up the original vector, as expressed by the following equation

$$PV + QV = V \quad \text{or} \quad P + Q = I$$

where I denotes the identity operation. Further

$$P^m = P \quad \text{and} \quad Q^m = Q \tag{53}$$

for all $m > 0$, since projection does not alter a vector which has already been projected. In what follows we can ignore the scalar product of any vector in N with any vector of the form QV since the two are orthogonal.

We are interested in the behaviour of ϱ_l expressed as a scalar product (Eq. 14)

$$\varrho_l = V_{j-n+l}V_{j-n}^* = (P+Q)\,V_{j-n+l}V_{j-n}^*$$
$$= PV_{j-n+l}V_{j-n}^*.$$

But

$$PV_{j-n+l} = PS^lV_{j-n}$$
$$= (PS)^lV_{j-n} \tag{54}$$

in view of (53), and thus

$$\varrho_l = (PS)^lV_{j-n}V_{j-n}^*. \tag{55}$$

The advantage of Eq. (55) is that it refers only to vectors in the finite dimensional space N. It can be analysed by introducing the n eigenvectors e_k and corresponding eigenvalues λ_k of the operator PS. Expanding the vector V_{j-n} in terms of these eigenvectors we find

$$\varrho_l = \left[(PS)^l \sum_{k=1}^{n} a_k e_k\right]\left[\sum_{m=1}^{n} a_m^* e_m^*\right] = \sum_k C_k \lambda_k^l \tag{56}$$

where

$$C_k = \sum_{m=1}^{n} a_k a_m^* (e_k e_m^*). \tag{57}$$

The required expression for the autocorrelation is given by (56).

The Eqs. (20) and (54) indicate that the operation $(PS)^n$ strictly decreases the length of any vector in the space N. This is because any vector acquires a component along V_j in at most n operations of shifting, and will then decrease under the last projection. This implies that

$$|\lambda_k| < 1$$

where λ_k is any of the n eigenvalues of (PS). Thus from (56) the autocorrelation ϱ_l decreases exponentially as $l \to \infty$. Burg (1975) has used a different approach to show that ϱ_l corresponding to the maximum entropy spectral estimate, decays to zero as $l \to \infty$.

b) Nature of the Peaks in the MEM *Spectrum*

Equation (56) gives the autocorrelation ϱ_l for $l \geqq 0$. For $l < 0$ we write

$$\varrho_{-l} = \sum_{k=1}^{n} C_k^* \lambda_k^{*l}$$

and

$$\varrho_0 = \sum_{k=1}^{n} C_k = \sum_{k=1}^{n} C_k^* = \sum_{k=1}^{n} (C_k + C_k^*)/2. \tag{58}$$

Letting $z = \exp(2\pi i x)$, the spectrum $S(x)$ corresponding to the autocorrelation ϱ_l may be written

$$S(x) = \sum_{l=-\infty}^{+\infty} \varrho_l z^l$$
$$= \sum_{k=1}^{n} \left\{C_k/2 + \sum_{l=1}^{\infty} C_k(\lambda_k z)^l + C_k^*/2 + \sum_{l=1}^{\infty} C_k^*(\lambda_k^*/z)^l\right\}$$
$$= \sum_{k=1}^{n} \left\{C_k/2 + \frac{C_k\lambda_k z}{1-\lambda_k z} + C_k^*/2 + \frac{C_k^*\lambda_k^*/z}{1-\lambda_k^*/z}\right\}. \tag{59}$$

Thus the spectrum $S(x)$ may be represented as the sum of n terms $S_k(x)$ given by

$$S_k(x) = C_k/2 + \frac{C_k\lambda_k z}{1-\lambda_k z} + C_k^*/2 + \frac{C_k^*\lambda_k^*/z}{1-\lambda_k^*/z}.$$

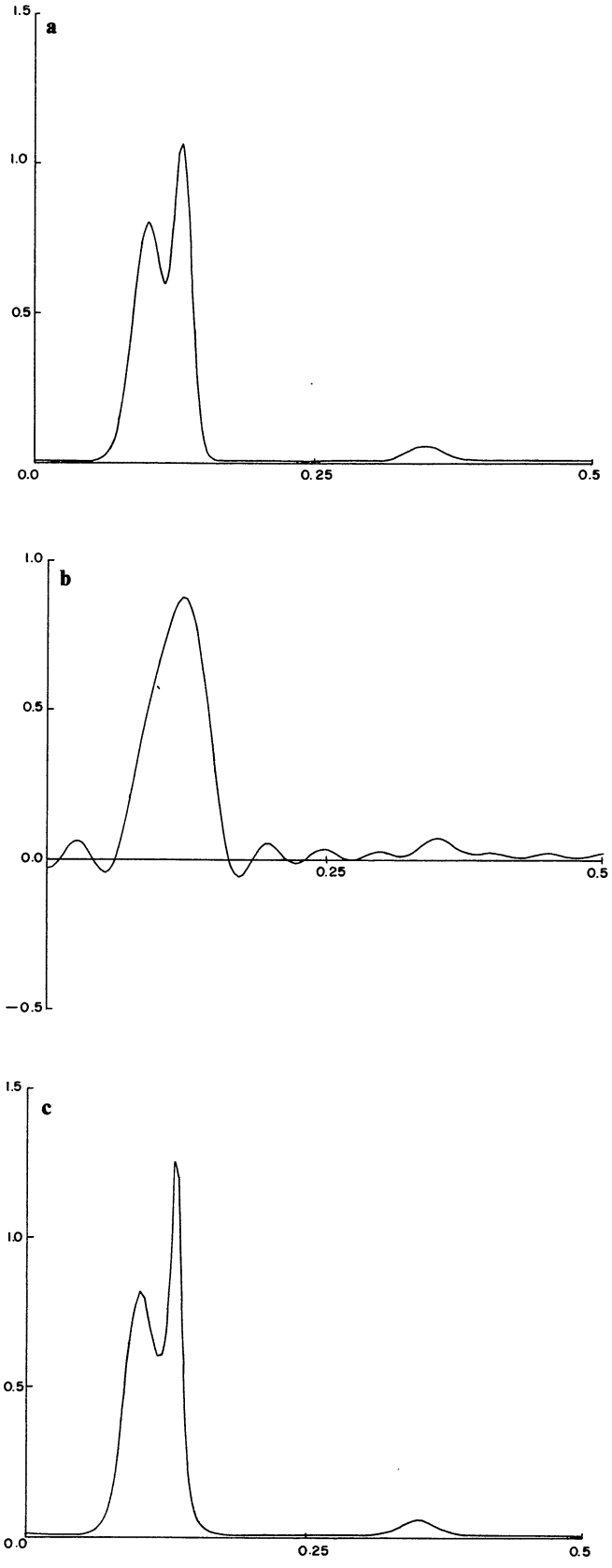


Fig. 3. a The model spectrum $S(x)=0.02+0.8\exp(-(0.102-x)^2/25)+\exp(-(0.133-x)^2/9)+0.05\exp(-(0.352-x)^2/36)$ **b** Reconstructed spectrum using only the first twenty Fourier coefficients viz. ϱ_0–ϱ_{19} **c** MEM spectrum calculated using ϱ_0–ϱ_{19}

Denoting λ_k by $r_k\exp(-2\pi i x_k)$, C_k by $C_k'+iC_k''$, and remembering that $z=\exp(2\pi i x)$, we find

$$S_k(x)=\frac{C_k'(1-r_k^2)-2C_k''r_k\sin 2\pi(x-x_k)}{1+r_k^2-2r_k\cos 2\pi(x-x_k)}. \tag{60}$$

As expected this function is periodic in x with period 1. We can visualize it better by assuming that r_k is close to 1 and x is close to x_k. We recover the approximate form

$$S_k(x)=\frac{a+b(x-x_k)}{c+d(x-x_k)^2}.$$

With $b=0$, we have the familiar Lorentzian peak. However, the b term makes the peak asymmetric since it changes sign when $(x-x_k)$ changes sign. While many authors, guided no doubt by the exponential form of the autocorrelation function generated by MEM, have stated that it corresponds to a sum of Lorentzian peaks, the point about the asymmetry seems to be new and gives us insight into the MEM as a model fitting procedure as discussed in the next section. It must be remembered that while Eq. (60) is valid in general, it is useful to interpret the spectrum as the sum of many peaks only when they do not overlap severely. This will certainly be true for small β_n^2, since as shown in Sect. Vc, we start from δ-functions at $\beta_n^2=0$.

In Appendix III, the "weighting factor", C_k of the kth peak is evaluated; it is given by

$$C_k=\beta_n^2/(1-\lambda_k^*\lambda_k)\prod_{i\neq k}(1-\lambda_i/\lambda_k)(1-\lambda_i^*\lambda_k). \tag{61}$$

It is also shown there that an approximate lower bound can be set on the width at half maximum a of the kth peak. It is

$$a\geqq\beta_n^2/2^{2n-1}\pi. \tag{62}$$

c) The Special Case

A zero value for β_k^2 implies that the Toeplitz matrix (Eq. 48) is singular. For that case, in place of Eq. (20), we write

$$V_{j+m}-\Sigma\alpha_{k,n}V_{j+m-k}=0$$

for all $m>0$. Any unit vector in the sequence can now be expressed as a linear sum of vectors in the space N. Furthermore any vector in the sequence can be transformed into the next member of the sequence by a unitary operator S. Since the operator is unitary it preserves the lengths of all vectors and its eigenvalues are of modulus unity. The equation analogous to (56) is now

$$\varrho_l=\sum_m e_m\exp(il\varphi_m) \tag{63}$$

where φ_m is real. Equation (63) implies that we have n terms in ϱ_l the *mth* of which is the Fourier transform of a δ-function located at $\varphi_m/2\pi$. The strength of this term is obtained from Eq. (57) using the orthogonality of the eigenvector of a unitary matrix

$$C_m=a_m a_m^*.$$

As expected, the C_m are positive in this case. When we put $l=0$ we have the relation, true in general

$$\sum_{m=1}^{n}C_m=\varrho_0=1$$

Burg (1975), using a different method, has shown that when the Toeplitz matrix is singular, the spectrum is a set of δ-functions. Bhandari (1978) has noted the appearance of sharp spikes in the MEM spectrum when the data are close to violating positivity (i.e. β^2 close to zero).

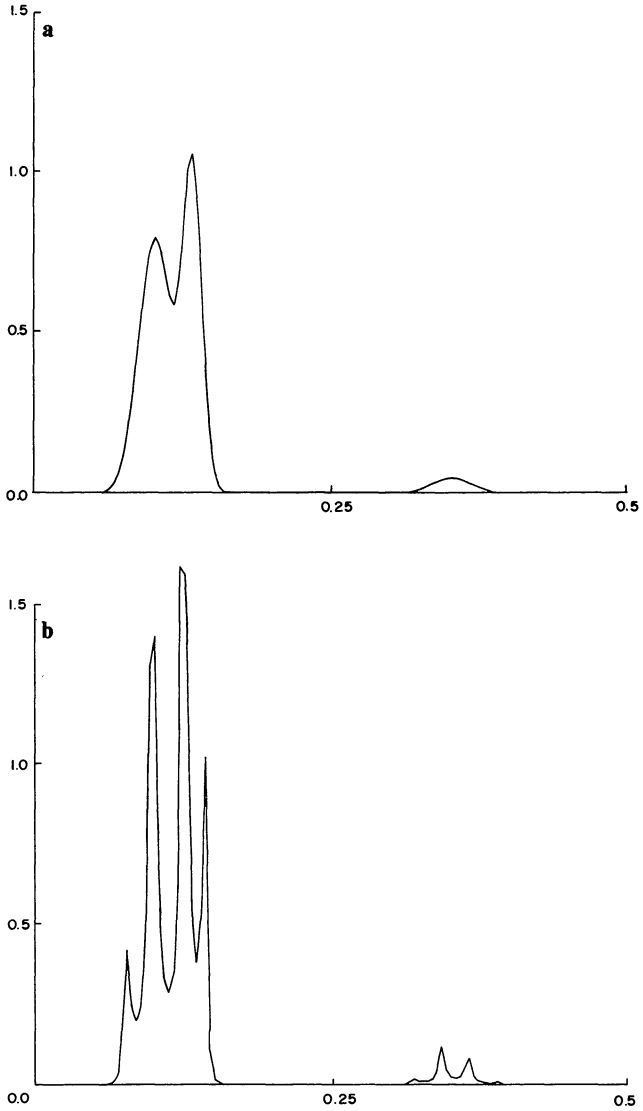


Fig. 4a and b. same as in Figs. 3a and c except that the baseline of 0.02 is omitted

VI. Discussion of the Maximum Entropy Approach

a) Typical Good Results with the MEM

Figure 3 presents a model calculation which illustrates the tremendous improvement that is possible with MEM in favourable cases. The model spectrum[1] (Fig. 3a) consists of two strong closely spaced Gaussian peaks and a much weaker Gaussian peak some distance away, all on a baseline. We assume that only the first twenty Fourier coefficients (ϱ_0–ϱ_{19}) of this spectrum are available. If the rest of the coefficients are put equal to zero, one reconstructs the spectrum shown in Fig. 3b where we notice that the two strong peaks are not resolved at all while the weak peak is almost unidentifiable among the termination ripples. Figure 3c shows the spectrum obtained with the MEM. We notice that the MEM introduces two improvements: (i) it greatly improves the resolution; (ii) it suppresses the termination ripples almost completely.

Other examples highlighting the marked improvement obtainable with the MEM have been presented in the literature (e.g., Ables, 1974; Komesaroff and Lerche, 1979).

In view of results such as Fig. 3c, it is important to understand the strengths and weaknesses of the MEM.

b) MEM *as a Form of Model Fitting*

The MEM, as its name suggests, was originally introduced (Burg, 1967, 1975) as a scheme to maximize the entropy of the spectrum (as defined in 46) subject to the constraints imposed by the measured data. Ables (1974) later reinterpreted the approach as being maximally noncommital with respect to the unmeasured autocorrelations. The MEM has also been described as an autoregressive method by Parzen (1968) and as an all-pole model by van den Bos (1971). According to the approach of Komesaroff and Lerche (1978) described in Sect. IV, each successive autocorrelation term is chosen so as to maximize the area in the complex plane available to the next term in the sequence.

The many distinct interpretations mentioned above do not directly answer the basic question as to what sort of spectrum the MEM generates. Following van den Bos, we regard the MEM as a kind of model fitting. We have shown in Sect. V that given $(n+1)$ autocorrelations viz., $\varrho_0, \ldots \varrho_n$, the spectrum consists of n asymmetric quasi-Lorentzian peaks. Now each peak is described by two complex numbers viz., the eigenvalue λ_i and its weight C_i. These two complex numbers (or four real numbers) describe four characteristics of the peak, viz., (i) Position, (ii) Width, (iii) Strength, and (iv) Asymmetry.

The important point to be noted is that the data consists of only n complex autocorrelations ($\varrho_0 = 1$) while the spectrum is described by $2n$ complex numbers (the λ_k and C_k). It is therefore clear that not all the characteristics of the spectrum are independently adjustable. For instance, suppose the $n\,\lambda_k's$ are given which means we specify the positions and widths of the n peaks. Then the $C_k's$, which describe the strengths and asymmetries of the peaks, are automatically determined (see Eq. 61). This is a rather remarkable feature in the MEM. Equation (61) relates the weights of the peaks to the eigenvalues λ_k. In general, we can state the following:

a) The term $(1-\lambda_k\lambda_k^*)$ in the denominator suggests that sharper peaks are stronger than wider ones.

b) The remaining terms in the denominator of (61) indicate that individual peaks are stronger if they occur in clusters than if they are isolated.

c) Smaller values of β_n^2 imply sharper peaks, a point that is further elaborated in the next section and in Appendix III.

The preceding discussion shows that one should be careful in interpreting the details shown by a MEM spectrum. The authors' experience is that the positions of peaks are usually correct. The improvement in resolution is also often remarkable (though, on occasion, the spectrum can be "over-resolved" to the point of being useless as shown in the next section). However, the weights and widths of the peaks are less reliable. When there is a large amount of data and a number of quasi-Lorentzians go into building up each peak in the spectrum, than the weights and widths too can be reasonably well reproduced. But, in less favourable cases, the results cannot be relied on. In brief, the MEM is very good at resolving and identifying peaks but less efficient at finding the *shapes* of the peaks.

1 In all the simulations presented in this paper, the spectra are taken to be symmetric with only the positive half displayed. Consequently, the autocorrelations are all real

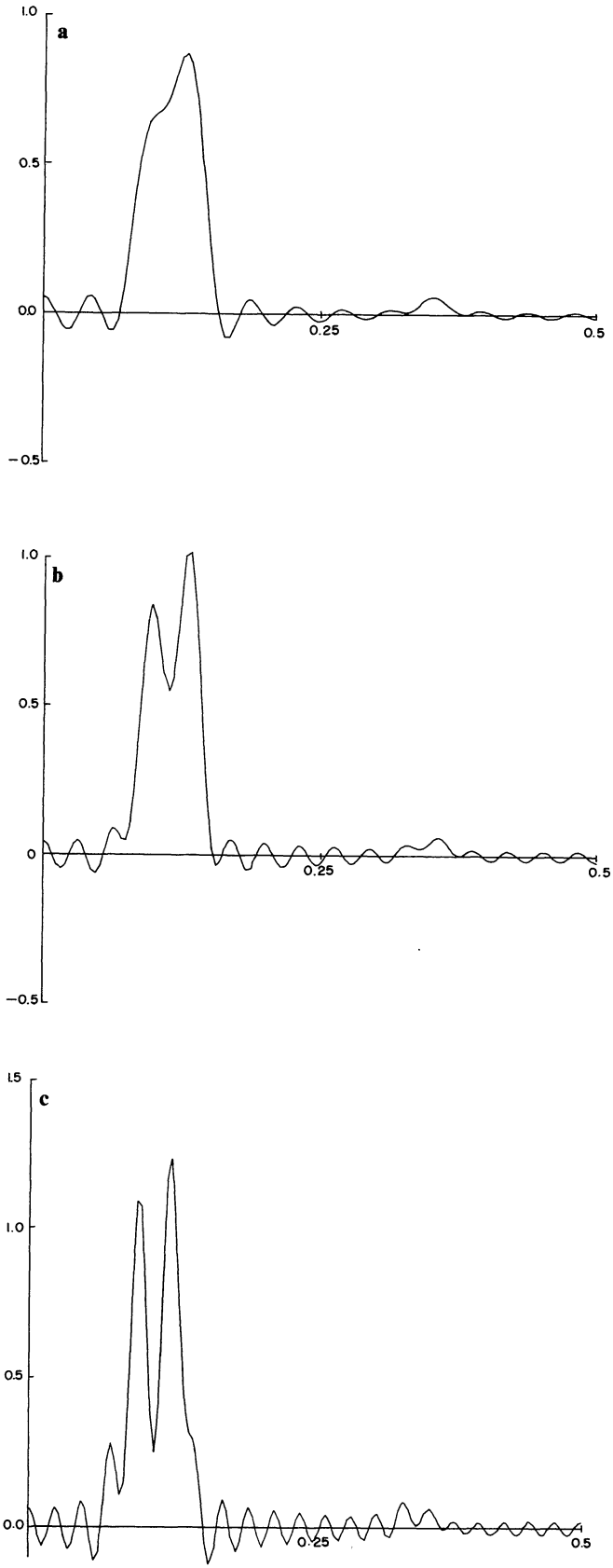


Fig. 5. Spectra obtained by extrapolating the autocorrelations of Fig. 3 to various finite ranges using the MEM formula **a** Extrapolation up to ϱ_{23} **b** up to ϱ_{31} **c** up to ϱ_{43}

c) Some Dangers of the MEM

We discuss here a few unfavourable cases where the MEM gives startling results.

The spectrum in Fig. 4 is identical to that in Fig. 3 except for the absence of the raised baseline. The MEM spectrum however is an apparently meaningless collection of sharp spikes. To understand the reason for this, one notes that the value of β_n^2 in this case is 0.0024 whereas it is 0.22 for the case shown in Fig. 3. We have shown in Sect. V that $\beta_n^2=0$ is the fully constrained case resulting in a spectrum of n δ-functions. It is therefore not surprising that when β_n^2 is close to zero, the spectrum has many sharp peaks. To understand the effect in detail, we note that a small value of β_n^2 implies that the unknown autocorrelations are restricted to small areas in the complex plane. This implies that the true spectrum is very close to violating the positivity constraint i.e., over a large range it has values close to zero. Consequently, the expression for the "entropy" H in (46) is dominated by the large negative values of $\ln F(x)$. In order to maximize H, the MEM algorithm will push up the baseline. This, to be consistent with the measured autocorrelations, appears to result in peak sharpening and splitting.

Thus, when β_n^2 is very small, the MEM spectrum is unreliable since it involves a very large extrapolation of the autocorrelation sequence beyond the measured values. In such a case, it might appear to be far more sensible to limit the resolution by some means. A possible suggestion is to use the MEM extrapolation formula (39) only upto a finite lag instead of up to infinity. The results of such a procedure on the spectrum of Fig. 4 are shown in Fig. 5. We note that for extrapolations up to say $\varrho_{3n/2}$, there is improved resolution without the introduction of any spurious peaks, whereas for extrapolation up to ϱ_{2n} or beyond, the spectrum is beginning to resemble the MEM spectrum of Fig. 4. We also note that all finite extrapolation spectra have fairly strong termination ripples.

Figure 6 shows another example where the non-linear nature of the logarithm in the MEM leads to poor results. The true spectrum has a step falling smoothly to a raised baseline. The normal spectrum obtained by the truncated transform is compared with the MEM spectrum. The notable feature is that the ripples in the baseline of the MEM spectrum have greatly reduced while the ripples in the step have actually *increased.* The reason for this is that the logarithm in H is sensitive only to ripples in the regions where $S(x)$ is small. Thus, the MEM effectively suppresses termination ripples only in peaky spectra with large regions near zero while, if there are extensive steps or plateaus, it is quite insensitive to ripples in these regions and may, in some cases, actually enhance them.

Figure 7 shows another danger in the MEM, where a single Gaussian peak is reconstructed as *two* peaks. The value of β_n^2 here is 0.30 and so this is different from the earlier discussion for small β_n^2. We believe that the effect here is a result of the form of model fitting employed by the MEM. Since the first few autocorrelations of a Gaussian peak are sought to be fitted by those of quasi-Lorentzians, it appears that the best fit is often obtained with *two* peaks separated from each other. Simulations show that the splitting of Gaussian peaks occurs over a wide range of n and β_n^2. It is a rather disturbing feature in the MEM since Gaussians do occur in nature.

We note that the defects of MEM discussed here so far do not result from random errors in either the data or the computations. This is contrary to the suggestion of Gull and Daniell (1979). However, we show below that noise does tend to aggravate the undesirable features in MEM.

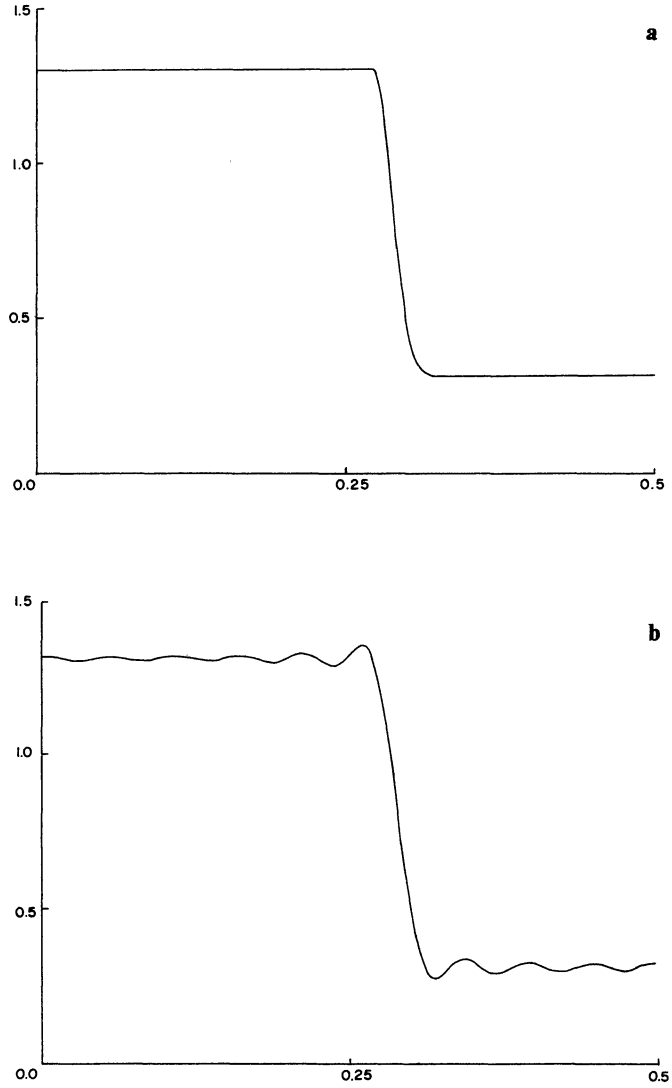


Fig. 6. a Model spectrum $S(x)=S(-x)=1.3$, $0<x\leqq 0.273$; $=0.3+\exp(-(0.273-x)^2/100)$, $x>0.273$ **b** Reconstructed spectrum using ϱ_0–ϱ_{19} only **c** MEM reconstruction using ϱ_0–ϱ_{19}

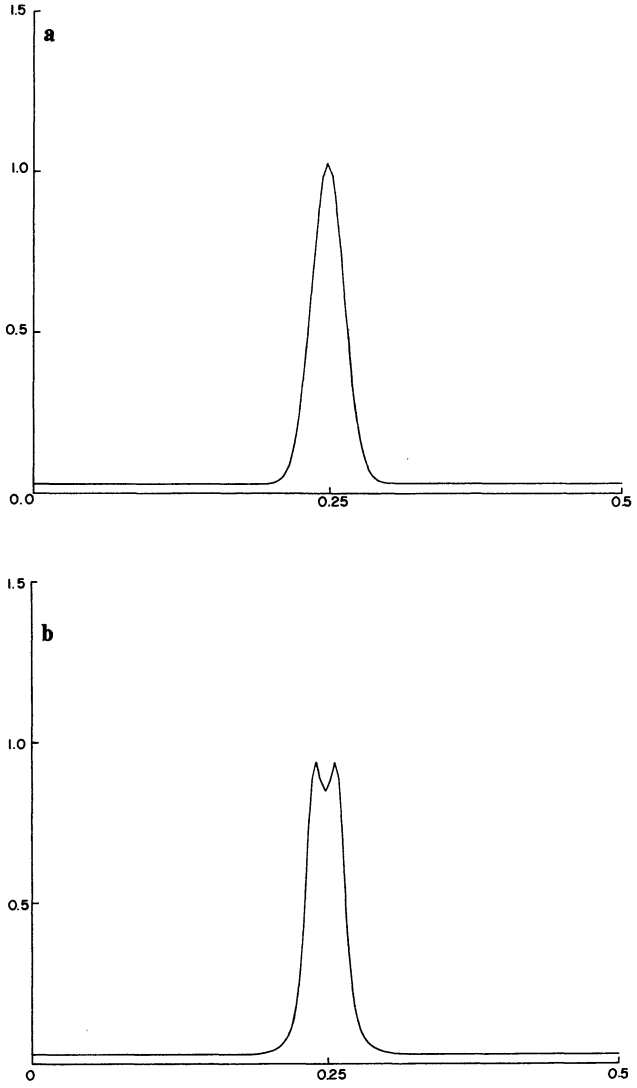


Fig. 7. a Model spectrum $S(x)=S(-x)=0.03+\exp(-(0.25-x)^2/25)$, $x\geqq 0$ **b** MEM reconstruction using ϱ_0–ϱ_{19}

d) MEM *with Noisy Data*

In this section we show that the presence of random noise in the data tends to reduce the value of β_n^2 i.e., drives the spectrum closer to the fully constrained limit. We use the following recurrence relations discussed in Appendix I.

$$\lambda_{m+1}=(\varrho_{m+1}-R_m)/\beta_m^2, \quad R_m=\sum_{k=1}^{m}\alpha_{k,m}\varrho_{m+1-k}$$

$$\beta_{m+1}^2=\beta_m^2(1-|\lambda_{m+1}|^2); \quad m=1,2,\ldots,n. \tag{64}$$

These equations can be pictorially represented in the complex plane as in Fig. 8. The autocorrelations $\varrho_0, \varrho_1 \ldots \varrho_m$ restrict ϱ_{m+1} to a circle with centre at R_m and radius β_m^2. Then, $|\lambda_{m+1}|$ is the ratio of the distance of ϱ_{m+1} from R_m to β_m^2, and β_{m+1}^2 is proportional to both β_m^2 and $(1-|\lambda_{m+1}|^2)$.

Let us assume that the effect of noise is to randomly and isotropically shift the ϱ's from their precise positions in the complex plane. It is clear from Fig. 8 that the most probable effect on ϱ_{m+1}

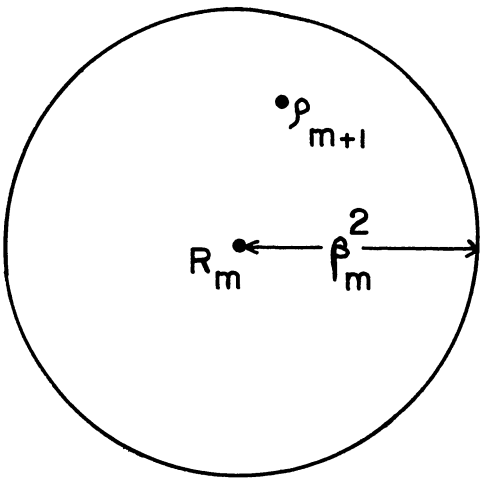


Fig. 8. Positivity requires ϱ_{m+1} to be within a circle in the complex plane with centre at R_m and radius β_m^2

is to increase its distance from R_m. This effect is more pronounced since R_m will itself undergo a random displacement because of the errors in $\varrho_0, \ldots \varrho_m$. Thus, $|\lambda_{m+1}|$ tends to increase and β_{m+1}^2 to decrease. Clearly, the effect increases rapidly with both the noise amplitude and the number of autocorrelations.

Figure 9 shows some numerical results on the model of Fig. 3. Random, normally distributed errors of various amplitudes were added to the Fourier transform of this model and the MEM spectrum was calculated in each case. By the time the errors were of amplitude 1.65% of ϱ_0, β_n^2 was practically zero. The variation of β_n^2 in this case as a function of the *rms* magnitude of the errors is plotted in Fig. 9d. The catastrophic effect of error on β_n^2 is well illustrated.

Figure 9 brings out the fact that when there are random errors in the measured correlations, there is a real danger of the data being *not* an autocorrelation sequence. The MEM cannot be directly used in such a case because there is evidently no positive spectrum fitting the data.

In the present paper, we have mainly dealt with the properties of the MEM and its strengths and weaknesses. We have pointed out a number of problem areas where the MEM gives wrong or even absurd results. It is not the purpose of this paper to discuss schemes for improving the MEM. However, we are currently considering the following possibilities to increase the reliability of the method:

a) Extrapolating the autocorrelation data to only a finite range determined either as a function of the measured range or in terms of the noise in the data.

b) Artificially increasing ϱ_0 with respect to the other autocorrelations by an amount sufficient to avoid the bad effects of the MEM (Bhandari, 1978, has discussed this in general terms.)

c) Extrapolating the $\varrho's$ by choosing the higher $\lambda's$ to be a smooth continuation of the first $n\lambda's$ instead of taking them to be zero as the MEM does (Note that any choice of $\lambda's$ such that all $|\lambda| \leqq 1$ will lead to a positive spectrum).

Appendix I

We briefly present an economical way of calculating $\alpha_{k,n}$ and β_n^2, based on the algorithm of Trench (1974) for the inversion of Toeplitz matrices. Replacing n by any $m(\leqq n)$, we write (27) and (28) as

$$V_{j+1} = \sum_{k=1}^{m} \alpha_{k,m} V_{j+1-k} + \beta_m \varepsilon'_{j+1,m} \tag{65}$$

$$\varepsilon'_{j+1,m} V^*_{j+1-p} = 0; \quad 0 < p \leqq m. \tag{66}$$

Similarly, we rewrite (34) and (32) as

$$V_{j+1} - \sum_{k=1}^{m} \alpha_{k,m} V_{j+1-k} = \lambda_m \left(V_{j-m} - \sum_{k=1}^{m} \alpha^*_{k,m} V_{j-m+k} \right) + (1-|\lambda_m|^2)^{1/2} \beta_n \varepsilon_{j+1,m} \tag{67}$$

$$\varepsilon_{j+1,m} V^*_{j+1-q} = 0; \quad 0 < q \leqq m+1. \tag{68}$$

Taking the scalar product of (67) with V^*_{j-m}, we obtain a result similar to (35) which can be re-arranged to read

$$\lambda_m = \left(\varrho_{m+1} - \sum_{k=1}^{m} \alpha_{m,k} \varrho_{m+1-k} \right) / \beta_m^2. \tag{69}$$

Also, rearranging (67) into the form (65) for $m \to m+1$, one can make the following identifications.

$$\alpha_{k,m+1} = \alpha_{k,m} - \lambda_m \alpha^*_{m+1-k,m}; \quad 0 < k \leqq m \tag{70}$$

$$\alpha_{m+1,m+1} = \lambda_m \tag{71}$$

$$\beta_{m+1}^2 = (1-|\lambda_m|^2)\beta_m^2. \tag{72}$$

Thus, starting from the initial values

$$\beta_0^2 = 1, \quad \alpha_{1,1} = \varrho_1/\varrho_0 \tag{73}$$

Eqs. (69)–(72) can be recursively used to calculate $\alpha_{k,n}$ and β_n^2.

Incidentally, when the given $\varrho's$ do not form an autocorrelation sequence, at least one of the intermediate λ_m will have modulus greater than unity.

Appendix II

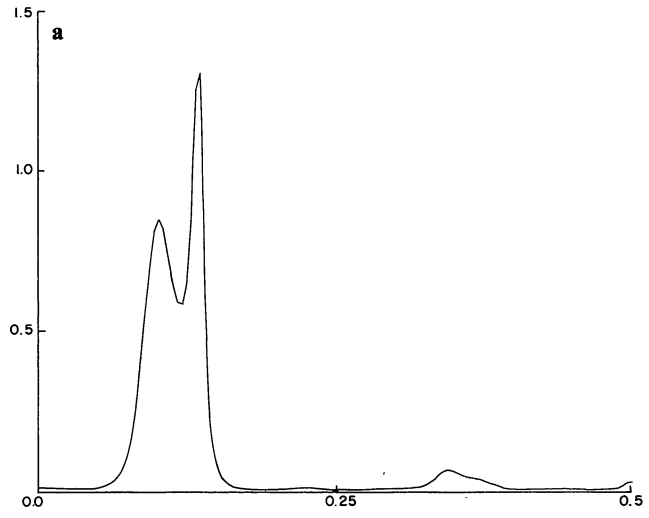
Given a positive function of many variables $S(x,y,\ldots)$, let some values of its Fourier series coefficients $\varrho(n_1, n_2, \ldots)$ be measured. For brevity, denote a known autocorrelation value by $\varrho(n)$ and an unknown one by $\varrho(m)$. Let ϱ_1 and ϱ_2 be two possible extrapolations of the measured data, both consistent with positivity. Agreement with the measured data means $\varrho_1(n) = \varrho_2(n)$. However, the two solutions will in general make different predictions for an unknown autocorrelation value, $\varrho(m)$. Consider the sum $w\varrho_1 + (1-w)\varrho_2$, $0 \leqq w \leqq 1$. Evaluated at a known point n where $\varrho_1 = \varrho_2$, it agrees with the known value. Its transform, being the sum with positive weights of two positive functions, is itself positive and so $w\varrho_1 + (1-w)\varrho_2$ is an admissible solution. At some point m, the complex number $w\varrho_1(m) + (1-w)\varrho_2(m)$ lies on the straight line segment joining $\varrho_1(m)$ and $\varrho_2(m)$. Thus every point on this line is an allowed value in an extrapolation consistent with positivity and the measured data. The region of allowed values for $\varrho(m)$ thus has to be convex.

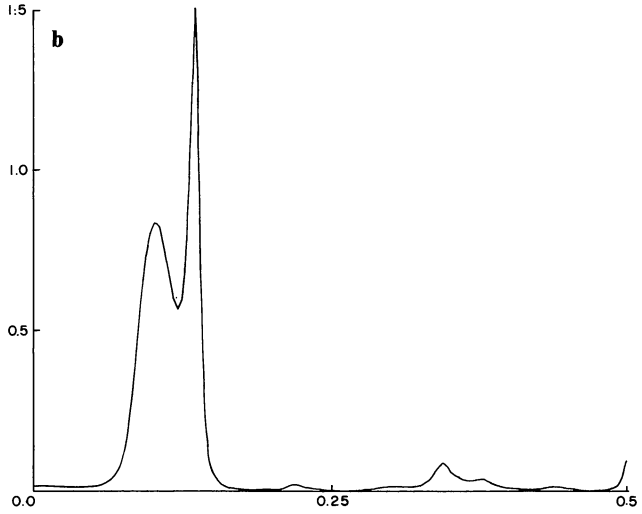
Appendix III

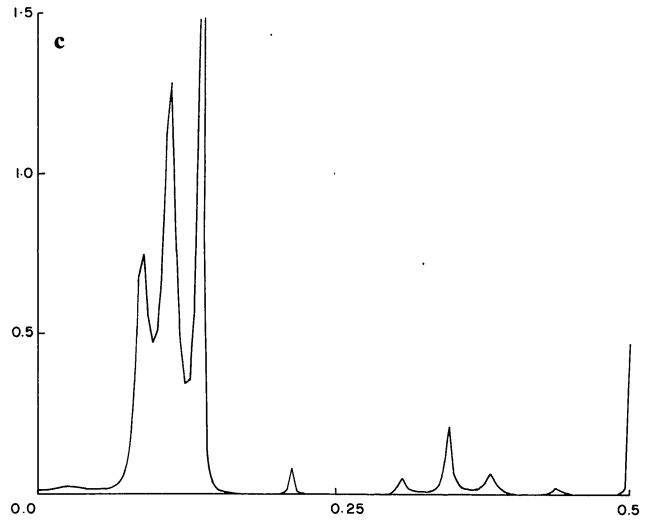
a) The Weighting Factors C_k

The maximum entropy spectrum, $S(x)$, is given by Eq. (45) which may be written

$$S(x) = \beta_n^2 / \sum_{k=0}^{n} \gamma_k z^k \sum_{l=0}^{n} \gamma_l^* z^{-l} \tag{74}$$







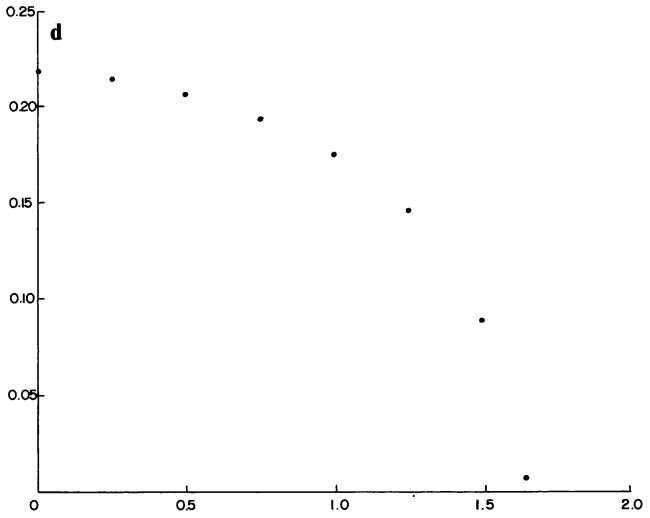


Fig. 9. MEM reconstructions of the spectrum of Fig. 3 with Gaussian noise of increasing amplitude added to ϱ_0–ϱ_{19} (a) the rms value of the noise is 0.5% of ϱ_0 (b) 1% (c) 1.5% (d) plot of β^2 as a function of the rms noise. For more than 1.65% noise the data cease to be an autocorrelation sequence

---

where $z=\exp(2\pi i x)$.

From Eqs. (23), (43), and (54) we may write

$$\left[(PS)^n-\sum_{k=1}^{n}\alpha_{k,n}(PS)^{n-k}\right]V_{j-n}V^*_{j-n}=\sum_{k=0}^{n}\gamma_k(PS)^{n-k}V_{j-n}V^*_{j-n}=0$$

and thus the characteristic equation of the matrix (PS) is

$$\sum_{k=0}^{n}\gamma_k z^{n-k}=0$$

the roots of this equation being the eigenvalues $\lambda_1 \ldots \lambda_k$ of (PS). Hence (since $\gamma_0=1$),

$$\sum_{k=0}^{n}\gamma_k z^{n-k}=\prod_{l=1}^{n}(z-\lambda_k)$$

and provided $z\neq 0$,

$$\sum_{k=0}^{n}\gamma_k z^{k}=\prod_{k=1}^{n}(1-\lambda_k z)$$

$$\sum_{l=0}^{n}\gamma^*_l z^{-k}=\prod_{j=1}^{n}(1-\lambda^*_j/z).$$

Substituting in (74)

$$S(x)=z^n\beta_n^2\Big/\prod_{i=1}^{n}(1-\lambda_i z)\prod_{j=1}^{n}(z-\lambda^*_j) \tag{75}$$

This expression for $S(x)$ may be expanded in partial fractions

$$\frac{z^n\beta_n^2}{\prod_{i=1}^{n}(1-\lambda_i z)(z-\lambda^*_i)}=\sum_{k=1}^{n}\left[\frac{a_k}{1-\lambda_k z}+\frac{b_k}{z-\lambda^*_k}\right]. \tag{76}$$

We evaluate a_k by the customary procedure of multiplying by $(1-\lambda_k z)$ and setting $z=1/\lambda_k$, yielding

$$a_k=\beta_n^2/(1-\lambda_k\lambda^*_k)\prod_{i\neq k}(1-\lambda_i/\lambda_k)(1-\lambda^*_i\lambda_k) \tag{77}$$

An analogous procedure yields

$$b_k=\lambda_k^* a_k^*. \tag{78}$$

Now, taking account of the relation,

$$\Sigma C_k=\Sigma C_k^*=\Sigma(C_k+C_k^*)/2$$

We may write Eq. (59) in the form

$$S(x)=\sum_{k=1}^{n}\left[\frac{C_k}{1-\lambda_k z}+\frac{\lambda_k^* C_k^*}{z-\lambda_k^*}.\right] \tag{79}$$

Comparing all the Eqs. (75)–(79) we find

$$C_k=a_k=\beta_n^2/(1-\lambda_k\lambda_k^*)\prod_{i\neq k}(1-\lambda_i/\lambda_k)\,(1-\lambda_i^*\lambda_k). \tag{61}$$

b) The Minimum Width of any Peak:

The argument in Sects. V and VI indicates that for small β_n^2 we expect the MEM spectrum to consist of n distinct "peaks". From Eq. (60), the width at half maximum of the kth peak is given approximately by

$$a_k=\frac{2(1-|\lambda_k|)}{2\pi}=\frac{2(1-r_k)}{2\pi}. \tag{80}$$

From the discussion in Sect. V we know that

$$\varrho_0=1=\sum_{k=1}^{n}C_k=\sum_{k=1}^{n}\beta_n^2/(1-\lambda_k\lambda_k^*)\prod_{i\neq k}(1-\lambda_i/\lambda_k)\,(1-\lambda_i\lambda_k^*) \tag{81}$$

Now, let the i^{th} peak be very much narrower than all the others. Then because $(1-r_k^2)\leqq 2(1-r_k)$ occurs in the denominator, the i^{th} term in Eq. (81) dominates all the rest and so

$$1\gtrsim\beta_n^2/2(1-r_i)\prod_{j\neq i}(1-\lambda_j/\lambda_i)\,(1-\lambda_j^*\lambda_i)$$

The value of $|\lambda_i|\leqq 1$, and thus the largest value that $|1-\lambda_j/\lambda_i|$ or $|1-\lambda_j^*\lambda_i|$ can take is 2, since all the eigenvalues λ_j lie within the unit circle. Putting in this extreme condition and referring to (80) we find

$$a\geqq\beta_n^2/2^{2n-1}\pi \tag{62}$$

which gives a strict lower bound on the width of any peak.

Acknowledgements. The authors thank C. R. Subramanya for helpful discussions and R. Bhandari for the computer programme used to carry out the simulations.

References

Ables, J.G.: 1974, *Astron. Astrophys. Suppl.* **15**, 383
Bhandari, R.: 1978, *Astron. Astrophys.* **70**, 331
Biraud, Y.: 1969, *Astron. Astrophys.* **1**, 124
Burg, J.P.: 1967, Paper presented at the Annual International Society of Exploration Geophysicists Meeting, Oklahoma City, Oklahoma
Burg, J.P.: 1975, Thesis "Maximum Entropy Spectral Analysis", Dept. Geophys., Stanford Univ.
Gull, S.F., Daniell, G.J.: 1979, in *Image Formation from Coherence Functions in Astronomy*, ed. C. van Schooneveld, D. Reidel Publ. Comp., Dordrecht, Holland, p. 219
Komesaroff, M.M., Lerche, I.: 1979, in *Image Formation from Coherence Functions in Astronomy*, ed. C. Van Schooneveld, D. Reidel Publ. Comp., Dordrecht, Holland, p. 241
Papoulis, A.: 1965, Probability, Random Variables and Stochastic Processes, McGraw Hill, p. 349
Parzen, E.: 1968, Multiple Time Series Modelling in Multivariate Analysis II, Academic Press, New York
Smylie, D.E., Clarke, G.K.C., Ulrych, T.J.: 1973, Methods in Computational Physics Vol. 13, Academic Press, New York
Trench, W.F.: 1974, *Mathematics of Computation* **28**, 1089
van den Bos, A.: 1971, IEEE *Trans. Inform. Theory*, IT-17, 493