J. Astrophys. Astr. (1982) **3**,. 419–450

Maximum Entropy Image Reconstruction–A Practical Non-Information-Theoretic Approach

Rajaram Nityananda and Ramesh Narayan
Raman Research Institute, Bangalore 560080

Received 1982 April 28; accepted 1982 September 27

Abstract. The maximum entropy method (MEM) of image reconstruction is discussed in the context of incomplete Fourier information (as in aperture synthesis). Several current viewpoints on the conceptual foundation of the method are analysed and found to be unsatisfactory. It is concluded that the MEM is a form of model-fitting, the model being a non-linear transform of a band-limited function. A whole family of 'entropies' can be constructed to give reconstructions which (a) are individually unique, (b) have sharpened peaks and (c) have flattened baselines. The widely discussed $1nB$ and $-B1nB$ forms of the entropy are particular cases and lead to Lorentzian and Gaussian shaped peaks respectively. However, they hardly exhaust the possibilities–for example, $B^{1/2}$ is equally good.

The two essential features of peak sharpening and baseline flattening are shown to depend on a parameter which can be controlled by adding a suitable constant to the zero spacing correlation ρ_{00}. This process, called FLOATing, effectively tames much of the unphysical behaviour noted in earlier studies of the MEM. A numerical scheme for obtaining the MEM reconstruction is described. This incorporates the FLOAT feature and uses the fast Fourier transform (FFT), requiring about a hundred FFTs for convergence. Using a model brightness distribution, the MEM reconstructions obtained for different entropies and different values of the resolution parameter are compared. The results substantiate the theoretically deduced properties of the MEM.

To allow for noise in the data, the least-squares approach has been widely used. It is shown that this method is biased since it leads to deterministic residuals which do not have a Gaussian distribution. It is suggested that fitting the noisy data exactly has the advantage of being unbiased even though the noise appears in the final map. A comparison of the strengths and weaknesses of the MEM and CLEAN suggests that the MEM already has a useful role to play in image reconstruction.

Key words: image reconstruction—maximum entropy method

1. Introduction

In a pioneering paper, Ables (1974) drew the attention of astronomers to the maximum entropy method (MEM) of restoring a spectrum from partial knowledge of its Fourier components (autocorrelations) and exhibited examples showing that (i) artifacts such as sidelobes could be suppressed and (ii) resolution could be increased. Since many fields *e.g.* aperture synthesis, very-long baseline interferometry (VLBI), speckle interferometry, share the common problem of incomplete Fourier information, it is natural that the MEM should have received a great deal of attention (see for example the conference proceedings edited by van Schooneveld 1979). In X-ray astronomy, some use on real data has begun (Willingale 1981; Pye *et al.* 1981) but it seems fair to say that the method has not lived up to its initial promise. We attribute this to two factors: (i) There has been a continuing and inconclusive discussion of the merits of different 'entropies' such as $\ln B$ and $-B\ln B$ (B being the surface brightness), each backed by statistical/thermodynamic/information-theoretic/combinatorial reasoning; (ii) A clear picture of the properties of MEM restorations is lacking and a method of controlling the resolution and sensitivity to noise is not available. Numerical implementation has also been a problem, although the scheme of Gull and Daniell (1978) has been found satisfactory by later workers (Willingale 1981). In this paper we attempt to give the MEM an alternative motivation and discuss its practical implementation. Section 2 discusses the need for nonlinear restoration methods in general, thus arguing in favour of 'non-classical' techniques such as the MEM. In Section 3 we summarise earlier work on the MEM and introduce our basic approach. In Section 4 we discuss the whole family of restoration methods based on maximising the integral of some function $f(B)$ of the brightness. The popular $\ln B$ and $-B\ln B$ entropies are members of this family. Criteria for the choice of f and the properties of different choices are given. A parameter R, which is a measure of the resolution in the restored map, is identified and a scheme for controlling R by adding a constant to the zero-spacing correlation ('FLOATing') is introduced. Section 5 discusses numerical schemes for implementing the MEM and Section 6 compares and contrasts the restorations obtained with various choices of f as well as with the CLEAN algorithm. Section 7 is devoted to the effect of noise in the data and shows that the Standard least-squares approach leads to a bias in the restoration. It is argued that our methods, though fitting the data exactly, have desirable properties in the presence of noise. In Section 8 we summarise our conclusions. The MEM is compared with CLEAN, bringing out the strengths and weaknesses of both the methods. We conclude that the MEM is now sufficiently well understood to be applied in certain practical situations. Some problematic areas calling for further work are identified. Appendix A discusses the uniqueness of the MEM solution, Appendix B considers the appearance of delta functions in MEM restorations and Appendix C discusses a multidimensional restoration scheme proposed by Newman (1977, 1978) which is shown to be incorrect.

2. The need for nonlinearity in restoration methods

Let us, for simplicity, consider a function in the range $0 \leqslant (x, y) \leqslant 1$. It is specified completely by its Fourier coefficients $\rho_{m,n}$ where

$$B(x,y) = \sum_{m,n=-\infty}^{+\infty} \rho_{m,n} \exp\,[2\pi i(mx+ny)]\,;$$

$$\rho_{m,n} = \int\int B(x,y) \exp\,[-2\pi i\,(mx+ny)]dx\,dy. \tag{1}$$

If we are given measurements of $\rho_{m,n}$ for some specified values M, N belonging to the set K(for known.), we can form the so-called principal solution $B_p(x, y)$ by restricting the summation in Equations (1) to this range

$$B_p(x,y) = \sum_{M,N\in K} \rho_{M,N} \exp\,[2\pi i\,(Mx+Ny)]. \tag{2}$$

$B_p(x, y)$, as is well known, is the convolution of $B(x, y)$ with the so called 'dirty beam' $d(x, y)$ which is. a function with Fourier coefficients 1 for $(M, N) \in K$ and zero otherwise.

Any attempt to improve on $B_p(x, y)$ without modifying the measurements would need to make some nonzero choice for the unmeasured autocorrelations *viz.* $\rho_{m,n}$ for $(m, n) \in U$ (U for unknown). One might at first think that the standard methods of interpolation and extrapolation would suffice for this. These methods are linear. Let us therefore restrict ourselves to restored maps $B_r(x, y)$ resulting from *linear* operations on $B_p(x, y)$. Let us also require that the operation should be *translation invariant*, *i.e.* that a given source should be restored in the same way regardless of its position. These two requirements restrict our linear operation to a convolution

$$B_r(x, y) = \int\int dx'\,dy'\,B_p\,(x', y')\,K(x-x', y-y'). \tag{3}$$

This means that the non-vanishing Fourier coefficients $\rho_{M,N}$ are multiplied by $K_{M,N}$ [the transform of $K(x, y)$] and $\rho_{m,n}$ for $(m, n) \in U$ continue to remain zero. Of course, this is what the classical weighting schemes achieve at the cost of modifying the measurements. We therefore conclude that (a) to leave the measured $\rho_{M,N}$ unaltered, (b) to retain translation invariance and (c) to give nonzero values to the unmeasured $\rho_{m,n}$, we need to make $B_r(x, y)$ a *nonlinear* function of $B_p(x, y)$. This generates new spatial frequencies in a manner that is well known in the time domain. Another way of appreciating the need for nonlinearity is to consider model-fitting procedures. For example, if we have to fit a given data set to the sum of three gaussian peaks, adding the fits obtained with two data sets would give six peaks whereas fitting the sum of the data would give us three. Thus we again have a non-linear operation being performed on the data.

The MEM is a non-linear technique as discussed below. The above arguments show that this is inevitable. However, the nonlinearity leads to some peculiar properties which one should thoroughly understand before applying the technique. This is further discussed in the rest of the paper.

3. Prevailing views on the MEM

The ME methods of restoration are all based on choosing the unmeasured Fourier coefficients to maximise a quantity which is the integral of a function $f(B)$ of the brightness; *i.e.*, maximise

$$E\{B(x,\ y)\} = \int\int f(B)\ dx\ dy. \tag{4}$$

Since $B(x,\ y)$ should satisfy the measurements,

$$\int\int B(x,\ y)\ \exp\ [\ -2\pi i(Mx + Ny)]\ dx\ dy = \rho_{MN}\ ,\ (M,\ N) \in K. \tag{5}$$

Equations (5) represent constraints to be obeyed in maximising Equation (4). It is clear from the form of Equation (4) that the restoration is translation invariant.

Before discussing our interpretation of Equations (4) and (5), we summarise some of the views prevalent in the literature on the nature of the MEM. If one is given only the integral of the brightness *viz.* ρ_{00}, it is clear that a uniform brightness distribution maximises E in Equation (4) for the usual choices $f = \ln B$, $-B\ln B$. This has perhaps led to the view that the MEM would like to produce as featureless a map as is consistent with the data (Ponsonby 1973), and that it is maximally noncommittal with regard to the unmeasured data (Ables 1974). A related point of view (Gull and Daniell 1978, 1979) is to define the 'most probable' map consistent with the data. The difficult step in this approach is assigning an *a priori* probability distribution for brightness maps in general. Gull and Daniell (1978, 1979) have adopted a combinatorial approach, based on building up a map by random distribution of 'quanta', to estimate the *a priori* distribution.* Others (*e.g.* Kikuchi and Soffer 1977; Ponsonby 1979) have invoked the Bose-Einstein distribution for photons over available modes. Another approach to the MEM has been through time series. Parzen (1968) and van den Bos (1971) have pointed out the relationship between the MEM and the autoregressive model for a time series. However, attempts to generalise this to two dimensions (Newman 1977, 1978) have not been successful (see Appendix C). Yet another very interesting approach is that of Komesaroff and Lerche (1979) in which the positivity constraint confines the first unmeasured Fourier coefficient (in one dimension) to a circle in the complex plane, and the MEM is shown to be equivalent to choosing sequentially the centre of each such, circle. In fact, this approach was used by Komesaroff and the present authors (Komesaroff, Narayan and Nityananda 1981) to study the properties of one-dimensional MEM restorations in some detail. However, our attempts to generalise this approach to two or more dimensions have not been successful. Meanwhile, Högbom (1979) and Subrahmanya (1979, 1980) have suggested that the successes of the MEM are simply because of the 'penalty' which functions like $\ln B$ and $-B\ln B$ impose on the undesirable baseline ripples in the brightness distribution. In their view, the choice of an entropy function is just a means of incorporating *a priori* information into the reconstruction.

*The referee has drawn our attention to a recent preprint by Gull and Skilling (1982) giving a new information-theoretic derivation of the $-B\ln B$ form and a criticism of other forms. This does not affect our subsequent discussion, which is concerned with the influence of different 'entropies' on the reconstructed image.

In this paper, we adopt and extend the last-mentioned approach. We find it difficult to assign, in any reasonable way, an *a priori* probability distribution for the brightness levels in a map. The argument from statistical mechanics (Kikuchi and Soffer 1977; Ponsonby 1979) is surely inappropriate since the photons from different parts of the source, corresponding to different cells in the map, have hardly been in thermal contact with one another. The combinatorial argument (Gull and Daniell, 1978, 1979) again seems to have in mind a specific model for how the ensemble of sources which we observe was generated. It appears to us that seeking a universal *a priori* probability distribution for map brightness levels is too rigid an approach. We have therefore explored the implications of a whole family of restorations, not in terms of underlying statistical principles (which may not exist), but in terms of their impact on specific features of the restored map, *viz.* peak width, peak shape, baseline and sensitivity to noise. It turns out that this family of ME restorations has two parameters which influence the degree of peak sharpening and baseline flattening in the restoration process, as well as the sensitivity to noise. In fact, Subrahmanya (1979, 1980) has emphasized that useful reconstruction schemes contain one or more parameters to be set by the user.

4. The family of maximum entropy restorations

We return to the problem of maximising the 'entropy' (4) subject to the constraints (5). Differentiating Equation (4) with respect to an unknown Fourier coefficient ρ_{mn}, we get

$$\frac{\partial E}{\partial \rho_{mn}} = 0 = \int\int f'(B) \exp\left[2\pi i\,(mx + ny)\right]\, dx\, dy;\ \ (m, n) \in U. \tag{6}$$

This implies that the Fourier coefficients of $f'(B)$, which we denote by σ_{mn}, vanish outside the set $(m, n) \in K$. We call such a function band limited [the principal solution $B_{\mathrm{p}}(x, y)$ is an example]. We thus have

$$\sigma(x, y) \equiv f'(B(x, y)) = \sum_{(M, N) \in K} \sigma_{MN} \exp\left[2\pi i\,(Mx + Ny)\right]. \tag{7}$$

Denoting the inverse of the function f' by g, we have from Equation (7)

$$B(x, y) = g\left[\sigma(x, y)\right] = g\left[\sum_{(M, N) \in K} \sigma_{MN} \exp\left[2\pi i\,(Mx + Ny)\right]\right]. \tag{8}$$

Equations (7) and (8) also follow from maximising Equation (4) subject to the constraints (6) by the method of Lagrange multipliers, whose role is played by the σ_{MN}.

It is important to ensure that there is only one function maximising the entropy expression (4). For a set of nonlinear equations such as (5) and (6), there is no general way of determining whether the solution is unique. However, it is shown in Appendix A that $f''(B) < 0$ is a sufficient condition for unique solutions and we therefore consider only such functions in this paper. Another natural question is whether a solution of the form (8) can be found at all. This again places restrictions

on the choice of the function f depending on the dimension d of the space over which the map B is defined. This problem is discussed in Appendix B which shows that with the familiar ln$B$ entropy, for instance, one may sometimes find no solution of the form (8) in three and higher dimensions.

Equations (7) and (8) state that the restored brightness distribution $B_r(x, y)$, after undergoing the nonlinear transformation $f'(B_r)$, becomes a band-limited function. We can use this fact to identify the particular features of f' that are important for obtaining 'good' restorations. To do this, we introduce *a priori* knowledge about $B_r(x, y)$, namely that it has a *flat baseline* and *sharp peaks*.* In contrast, the band-limited function $f'(B_r)$ has a rippled baseline and rounded peaks. The schematic argument in Fig. 1 shows that the two required characteristics (of flat baseline and sharp peaks) can be achieved if the 'amplification factor' $f''(B)$ is large for small B and small for large B. We thus see that in addition to $f''(B) < 0$ (required for uniqueness), we also need $f'''(B) > 0$. When these requirements are met, the transformation g defined in Equation (8) constructs $B(x, y)$ by sharpening the peaks and suppressing the baseline ripple in the band-limited function $\sigma(x, y)$ (Fig. 1). From the above arguments, it is also clear that there are no sign restrictions on either $f(B)$ or $f'(B)$.

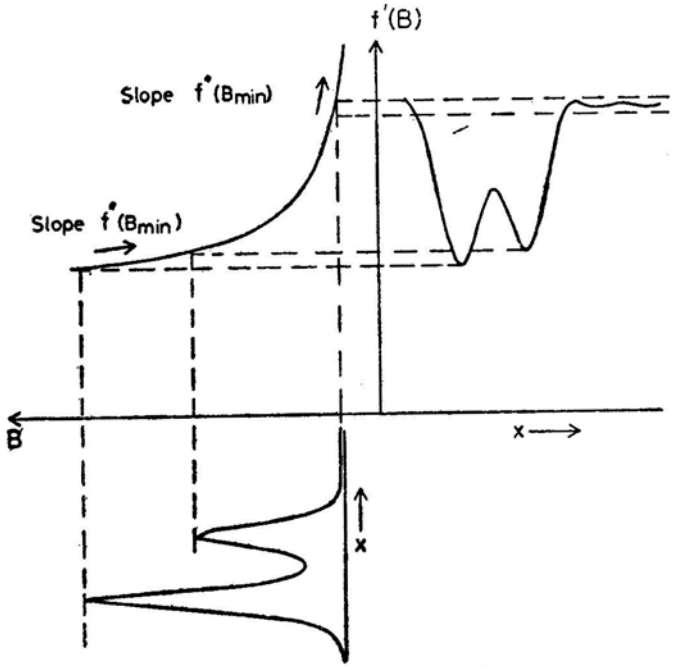

Figure 1. Illustration of the peak-sharpening and baseline-flattening properties of the MEM. The case shown is $f(B) = \ln B$. According to Equation (7) $f'(B) = 1/B$ is a band-limited function, shown to the right of the vertical axis. The function shown below is B. It is clear that the large slope [$f''(B)$] of the curve of $f'(B)$ *vs.* B at small B ensures a flat baseline, and the small slope at large B ensures sharp peaks.

*It is noteworthy that this is similar to the assumptions made in the CLEAN algorithm (Högbom 1974) though the information is used in a different way.

To lowest order, one expects the restoration to be determined by the two extreme values $f''(B_{\min})$ and $f''(B_{\max})$. Since the absolute scale of f is immaterial, we define the following ratio.

$$R = f''(B_{\min}) / f''(B_{\max}) \tag{9}$$

which decides how nonlinear the transformation in Equation (8) is. R thus measures the extent of peak sharpening and baseline flattening compared to a band-limited function such as the principal solution. Alternatively, we can think of R as a measure of how far we are extrapolating the measured Fourier coefficients. Surely, such an important characteristic of the restoration process should be chosen by the user, whereas in most present realisations of the MEM, the ratio R would be decided by the value of ρ_{00} (the average of the map) and therefore predetermined by the data. Bhandari (1978) had suggested that ρ_{00}, should be varied to achieve useful results and had illustrated this in one dimension. There is a very simple technique by which the value of R can be set by the user. In our implementation of the MEM, a suitable constant C is added to the map so that $f''(B_{\min}) / f''(B_{\max})$ equals the preassigned number R. We refer to this process as 'FLOATing' the map. Those who regard ρ_{00} as a sacred part of the data not to be modified could instead imagine that we are maximising $f(B+C)$ instead of $f(B)$. As pointed out by R. D. Ekers (personal communication), the true value of ρ_{00} in many aperture-synthesis observations at high frequencies would be dominated by the 3 K background which most practitioners of the MEM would regard as irrelevant! We note in this context that CLEAN makes no use of ρ_{00}.

Having chosen a suitable value for R (examples are given in Section 6), we can next classify the different functions f as 'soft' or 'hard' depending on whether the transition from the high value of f'' at small B to the small value at large B is rapid or slow. We expect that the soft functions will only affect the low-level features of the map leaving the peaks relatively unsharpened, while the hard functions will sharpen the peaks significantly with correspondingly lower efficiency in flattening the baseline. Let us consider a family of functions f with f'' being negative and varying as an inverse power of B (hence $f''>0$). In this family, the cases $f'' \propto -1/B^2$ and $-1/B$ correspond to the $\ln B$ and $-B\ln B$ entropies. For brightness distributions B varying from 0 to 1 and a given value of R, the normalised form of f'' is

$$f_n''(B) = -\frac{1}{[1 + B(R^{1/n} - 1)]^n}. \tag{10}$$

Fig. 2 shows the shape of $f_n''(B)$ at R=10 for typical values of n. For large n, Equation (10) asymptotically tends to the form $-\exp(-B\ln R)$, which is the hardest entropy in the above family. At the other end, we notice that as $n \to 0$, $f_n''(B)$ is a constant over most of the range, rising (in absolute value) from $1/R$ to 1 only for B very close to zero. Since the amplification factor f'' tends to a constant, the restoration tends to a band-limited function. More directly, n=0 corresponds to $f(B)=\frac{1}{2}B^2$ which implies that $f'(B) = B$ is a band-limited function. A band-limited restoration which fits the data is of course nothing but the principal solution. Another limit in which we obtain the principal solution is as $R \to 1$, since $f_n''(B)$ then

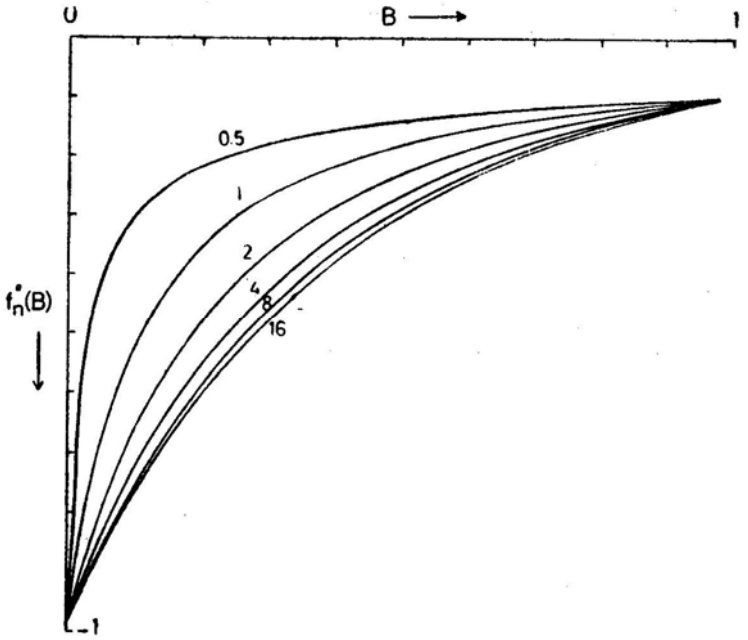

Figure 2. Variation of $f''(B)$ *vs. B* for various entropies characterised by $f'' = 1/(B + C)^n$. f'' varies from –10 to – 1 as B varies from 0 to 1. The values of n go from 0·5 to 16 and are indicated against the corresponding curves. Note that all the entropies with large n tend to a single curve (corresponding to the exponential entropy $f = -e^{-\alpha\beta}$). The 'softer' entropies with smaller values of n produce a flatter baseline and blunter peaks than the 'hard' entropies with large n.

tends to a constant as shown in Equation (10). This case is realised when we add a large constant C to ρ_{00}. Since $B_{\min}/B_{\max}$ tends to 1, so does R. Bhandari (1978) noted that for large ρ_{00} the ME restoration tends to the principal solution and proved it analytically in the one dimensional, $f = 1nB$ case. We now see that the result is completely general.

From the above arguments we see that there is a two-dimensional family of ME reconstructions described by two parameters R and n. R gives a measure of the compression of the baseline relative to the peaks when compared to a band-limited function. The parameter n in the expression $f''(B) \propto -1/B^n$ is a measure of how far this flattening effect extends upwards from the baseline. As $n \to 0$, the 'soft' limit, only the very lowest features in the map undergo the nonlinear transformation and we tend to the principal solution. The common form $-B\ln B$ is near this end. Restorations with this entropy would have flat baselines but relatively wide peaks (though the peaks could be sharpened by increasing R). As n increases, we obtain 'harder' entropies which are less efficient at baseline flattening but are better at sharpening peaks (*i.e.* increasing the resolution). The $\ln B$ entropy is 'harder' than $-B\ln B$ while $-\exp(-B\ln R)$ is the hardest in the whole family. However, we should point out that R and n can only be semiquantitative measures of the above tendencies, since even at a fixed value of R, the properties of the restoration obviously depend on the data as well. We have chosen $B_{\min}$ and B_{max} as representing the baseline and peak regions but it could happen for example that B_{min} is only an isolated fluctuation and much of the baseline lies above it. Similarly a single large peak may get

sharpened relative to lower-level peaks. This latter tendency was noted by Komesaroff, Narayan and Nityananda (1981).

We now show that the parameter n can also be interpreted as introducing *a priori* information on the *shapes* of peaks. Let us expand the band-limited function $\sigma(x, y)$ in Equation (7) about any minimum:

$$\sigma\,(x, y) = \sigma_0 + ax^2 + by^2. \tag{11}$$

For convenience, the minimum has been chosen as the origin of x and y and the axes have been chosen to make the second-order term a sum of squares. For the two entropies $f_1 = \ln B$ and $f_2 = -B\ln B$, $f_1' = 1/B$ and $f_2' = -1 - \ln B$. The shape of the peak thus becomes (from Equations 7 and 8),

$$\begin{aligned} B\,(x, y) &\simeq [\sigma_0 + ax_2 + by_2]_{-1} \qquad \text{for } f_1, \\ &\simeq \exp\,(-1 - \sigma_0 - ax^2 - by^2) \qquad \text{for } f_2 \end{aligned} \tag{12}$$

We thus have a generalised Lorentzian peak for f_1 and a gaussian peak for f_2. Each choice of $f(B)$ similarly leads to a corresponding peak shape. These results are valid when the width of the peak in B is much less than the width of the minimum in $\sigma(x, y)$, so that the Taylor expansion in Equation (11) is accurate. This will be true for high values of R, as discussed earlier. Apart from the earlier arguments regarding baseline flattening and peak sharpening, the required peak shapes in the restored map could also be used as a criterion to select the form of entropy $f(B)$. The question of peak shapes is discussed further in Appendix B.

5. Numerical procedures for obtaining the MEM reconstruction

We now describe the numerical procedures that we have used to implement the MEM. We are given a set of Fourier coefficients ρ_{MN}, $(M, N) \in K$ and are required to find values for ρ_{mn}, $(m, n) \in U$ satisfying Equation (6), which is the condition that the entropy be maximised. A Standard approach to maximisation of any function E is to start from a trial solution and evaluate the gradient of E there. In this case the components of the gradient are given by $\sigma_{\_m,\ \_n}$, the Fourier coefficients of $f'(B)$. We have

$$\frac{\partial E}{\partial \rho_{mn}} = \int\int f'(B) \exp\,[2\pi i\,(mx + ny)]\; dx\, dy \equiv \sigma_{-m,\,-n}. \tag{13}$$

In practice, the integrals in Equations (1) and (4) and hence also in Equation (13) are replaced by discrete sums with x and y varying on an $N\times N$ grid, where N is usually a power of 2 in order to implement the Fast Fourier Transform (FFT) conveniently. Correspondingly, m and n also vary on an $N\times N$ grid. The number of unknown Fourier coefficients to be determined is therefore finite. A more convenient form of Equation (13) is obtained by introducing the real and imaginary parts (denoted by single and double primes) of ρ, $\delta\sigma$, and σ. We have

$$\delta E = 2\,\sigma''_{mn}\,\delta\rho''_{mn} + 2\,\sigma'_{mn}\,\delta\rho_{mn}'. \tag{14}$$

From Equation (14), we identify the derivatives of E with respect to ρ'_{mn} and ρ''_{mn} to be $2\ \sigma'_{mn}$ and $2\ \sigma''_{mn}$. In the gradient method of implementing the MEM, one increments the unknowns by a multiple μ of the gradient

$$\rho'^{(k+1)}_{mn} = \rho'^{(k)}_{mn} + \mu\ \sigma'^{(k)}_{mn}, \quad \rho''^{(k+1)}_{mn} = \rho''^{(k)}_{mn} + \mu\ \sigma''^{(k)}_{mn}. \tag{15}$$

The coefficient μ is chosen so as to reach the highest value of E on the line defined by Equation (15). In practice, this is achieved by making a small shift in the ρ_{mn} proportional to $\delta\mu$ and computing the gradient at the new point. By linear extrapolation, one can find the value of μ at which the component of the gradient vector parallel to the search direction is zero. Starting with the new values of ρ_{mn}, one again computes the gradient and repeats the process iteratively.

An essential part of the gradient scheme as we implement it is the choice of ρ_{00} In each Iteration the ratio of the highest and lowest values of if is computed. It is then possible to add a suitable constant C to all values in the map such that $f''(B_{\min} + C)/f''(B_{\max} + C)$ is equal to the preassigned value of the ratio R. Such a 'FLOAT' step may be introduced at every cycle, or every few cycles, but is mandatory whenever $B_{\min}$ is negative at any stage.

A well-known improvement to the gradient method is the conjugate gradient method (Fletcher and Reeves 1964). In this method, each search direction is a weighted sum of the previous search direction and the present gradient, with the weights involving squared gradients. Without going into details, we point out that all the quantities required for this scheme are readily computable via the FFT, as shown by Equation (13). We have implemented the conjugate gradient method, with each cycle requiring four FFT's. At a value of R=100, for a 32×32 map, we need typically about 40 iterations for convergence. Much of the improvement in the map really takes place within the first 10 iterations. Execution time on a minicomputer system (PDP 11/34) is about 15 min for the above example. The time required for convergence increases with R, which is to be expected since R measures the degree of extrapolation of the data.

We have also explored the so-called fixed-point iteration scheme which exploits the fact that $f'(B)$ is a band-limited function. For concreteness, we describe the case f=ln$B$, f' =1/B. The stages are: (i) One starts from the principal solution, (ii) The reciprocal of the map is taken after FLOATing. (iii) The Fourier transform is then computed. In general, it will not vanish outside the known set *i.e.* $\sigma_{mn} \neq 0$, $(m, n) \in U$. (iv) These values are set equal to zero and $\sigma(x, y)$ is computed from $\sigma_{M,N}$ using Equation (7). (v) After a FLOAT step, the reciprocal is taken to obtain a map which is then transformed to give a new set ρ_{mn}. (vi) This will not agree with the measured ρ_{MN} which are therefore reset to their true values. However, we have now extrapolated ρ_{MN} outside the measured region. The new map is computed and the entire cycle from (ii) to (vi) is repeated until convergence is obtained.

Our experience is that the straightforward scheme described above converges only for small R. For the large values of R which are of interest, it is necessary to improve upon step (iv). Setting the σ_{mn} for $m, n \in U$ equal to zero is not the best strategy since, in any case, they become nonzero after one iteration. The following alternative approaches have proved to be quite successful. In the first iteration, the σ_{mn} are set

equal to zero. At the next iteration, one now finds new nonzero values $\sigma_{mn}^{(1)}$ We can express these in the form

$$\sigma_{mn}^{(1)} = \alpha\, \sigma_{mn}^{(0)} + \tau_{mn}, \tag{16}$$

where the vector τ_{mn} is orthogonal to $\sigma_{mn}^{(0)}$. The motive behind Equation (16) is to find a pattern in the as which, after one iteration, repeats with a constant multiple α. The τ_{mn} represent the error or noise in this simplified picture. Setting $\sigma_{mn}^{(0)}$ to zero gives $\sigma_{mn}^{(1)}$ in the next iteration. Instead we set $\sigma_{mn}^{(0)}$ equal to $\beta\sigma_{mn}^{(0)}$ where β is given by

$$\beta\, \sigma_{mn}^{(0)} = -\left(\frac{\alpha}{1-\alpha}\right) \sigma_{mn}^{(0)}. \tag{17}$$

One can check by linear interpolation that this choice will give $\sigma_{mn}= 0$ in the next iteration. The argument is illustrated in Fig. 3. In practice, the presence of the τ_{mn} term leads to a non zero residual and the process has to be repeated.

The following refinement of the above technique has proved quite effective. Using a simplified linear picture, we may consider $(1 - \alpha)$ to be the 'eigenvalue' of the approximate eigenvector $\sigma_{mn}^{(N)}$ for the transformation corresponding to one cycle of the fixed point algorithm. At each iteration, one can obtain the 'eigenvalue' of the previous set of σ_{mn}. By keeping track of the largest and smallest eigenvalues, one can optimize β in Equation (17) for maximum convergence in the residuals. We have had good results with this approach. This scheme is related to the one described by Willingale (1981) but has a more flexible procedure of averaging successive iterations to achieve fast convergence.

6. Illustrative examples of MEM reconstructions

In this section we present some examples of two-dimensional reconstructions with different choices of the function $f(B)$ to illustrate the general ideas put forward in

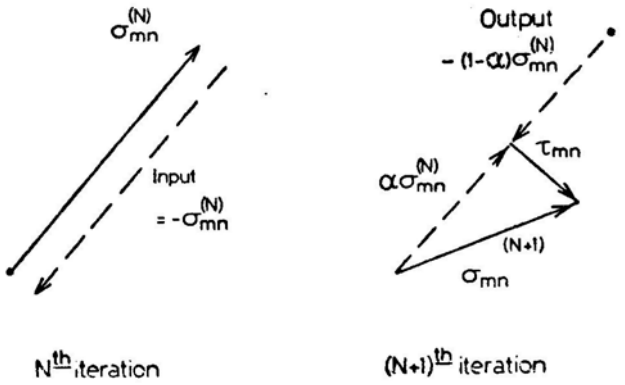

Figure 3. Illustrating the fixed-point iterative scheme for numerical computation of the MEM solution. The space of the Fourier coefficients of the reciprocal (for the ln$B$ entropy) of the brightness, is shown. It is desired to reach the origin, σ_{mn} = 0(ρ_{mn} not measured). In the simplest scheme, $\sigma_{mn}^{(N)}$, the values at the Nth iteration, are set equal to zero and after one cycle they take the values $\sigma_{mn}^{(N+1)}$, which are decomposed into $\alpha\sigma_{mn}^{(N)}$ plus 'noise' τ_{mn}. The dashed vectors labeled 'input' and 'output' show the shifts from the original position before and after one iteration. It is possible to improve the scheme to make the 'output' vector terminate near the origin (Equation 17)

Section 4. These results were all computed by the conjugate gradient method described in the previous section. We also compare our results with those obtained with the widely used CLEAN algorithm (Högbom 1974).

The model source is represented in Fig. 4 and consists of two strong unequal anisotropic gaussian peaks near each other and a broader low-level feature further off. The values of brightness were sampled on a 32×32 grid and Fourier transformed to produce the visibility function ρ_{mn}. The values of ρ_{mn} for $|m|$, $|n|$ greater than 3 were then set equal to zero. The response to a point source is now the 'dirty beam' shown in Fig. 5. The 'dirty map' obtained by transforming the truncated data back to the map plane is shown in Fig. 6. The two strong features have merged and their sidelobes drown the weak feature. The contours in the figures have been chosen to highlight the low-level features in the map since the improvement obtained by using the MEM or CLEAN is most apparent there.

Figure 4. A model source consisting of two strong anisotropic overlapping gaussian peaks with a weaker circular peak some distance away. The three peaks are at (13, 9), (11, 17) and (27, 21) on a 32 × 32 grid with peak heights 1, 1 and 0·1. Their major axes are located at position angles 135°, 100° and 90° and the pairs of rms values in the two principal directions are (3, 1·5), (3, 2) and (2, 2) respectively. In all the following figures the x axis runs downwards and the y axis from left to right. The contour levels are 0·01, 0·03, 0·05, 0·1, 0·2, 0·3, 0·5, 0·7, 1·0, 1·5 in all the 'maps shown in this paper except those in Fig. 12.

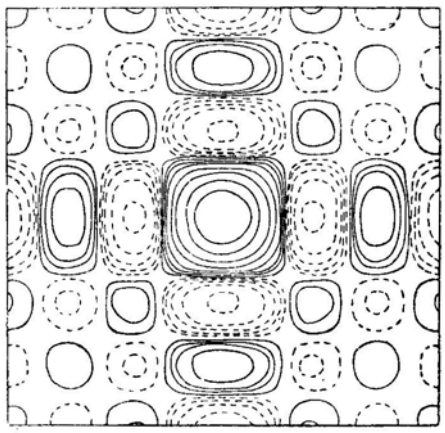

Figure 5. The reconstruction of a point source using only Fourier coefficients ρ_{mn} with $|m|$, $|n| \leqslant 3$. This is the so-called dirty beam. Negative contours are shown by broken lines.

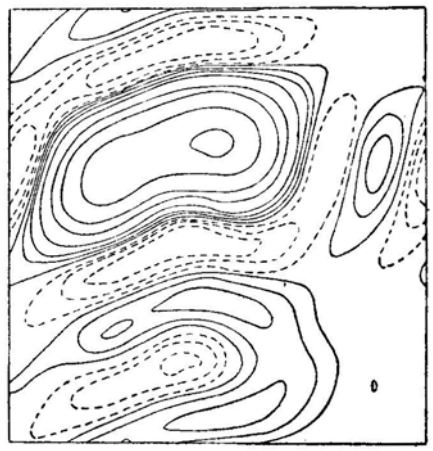

Figure 6. The dirty map made with the Fourier coefficients ρ_{mn} of the model in Fig. 4 ($|m|, |n| \leqslant 3$). This is the convolution of Fig. 4 with Fig. 5.

For comparison with the MEM solutions given later, Fig. 7 shows the result of applying the CLEAN algorithm with a gain factor of 0·25 to the dirty map of Fig. 6. Figs 7a, b and c are the results for CLEAN restoring beams of decreasing half width. Clearly the resolution in Fig. 7c is too large and false detail appears while Fig. 7a has degraded the resolution of the original map (Fig. 3). The suppression of the ripple, revealing the weak component, is excellent in the CLEAN restorations.

Now we look at the restorations performed with the ln$B$ MEM with R values of 25 and 100 respectively. These are given in Figs 8a and b. The enhanced peak sharpening and baseline suppression for R=100 relative to R=25 are readily apparent. One can also see that in this case, the restoration with R=100 is probably more faithful to the original than any of the CLEAN restorations in Fig. 6. Of course, we only take this to mean that the *a priori* information introduced by the MEM was more appropriate *for this map.* Figs 9a and b give the restorations with $f = -B\ln B$ and R=25 and 100. Note the approximate correspondence with Figs 8a and b, showing that R is a good measure of resolution, approximately independent of the form of entropy. We find that Fig. 9b is very close to the original. This can be understood since the starting model has gaussian peaks, which are naturally produced by the $-B\ln B$ entropy (Section 4).

Figs 10a and b show reconstructions from the dirty map of Fig. 5 using an un conventional form of 'entropy', $B^{1/2}$. It is notable that this form, with nothing in it to suggest the name 'entropy', is as good as either of the other two shown in Figs 8 and 9. In terms of our one-parameter classification of entropies, $B^{1/2}$ is midway between $-B\ln B$ and $\ln B$. Table 1 clearly illustrates the gradation in properties from $-B\ln B$ to $B^{1/2}$ to $\ln B$. At a constant R, peak height as well as ripple (maximum negative value) both increase down the sequence. For a given entropy, increasing R leads to an increase in peak height and decrease in ripple. Using the two parameters R and n one should be able to achieve independently the desired degree of peak sharpening and ripple suppression.

Figs 11b and 11a illustrate reconstructions with the same data as before with two choices $-e^{-aB}$, R=25 and $B^{3/2}$, R=100 which represent very 'hard' and very 'soft' functions respectively, as explained in Section 4. As expected, the peak sharpening in

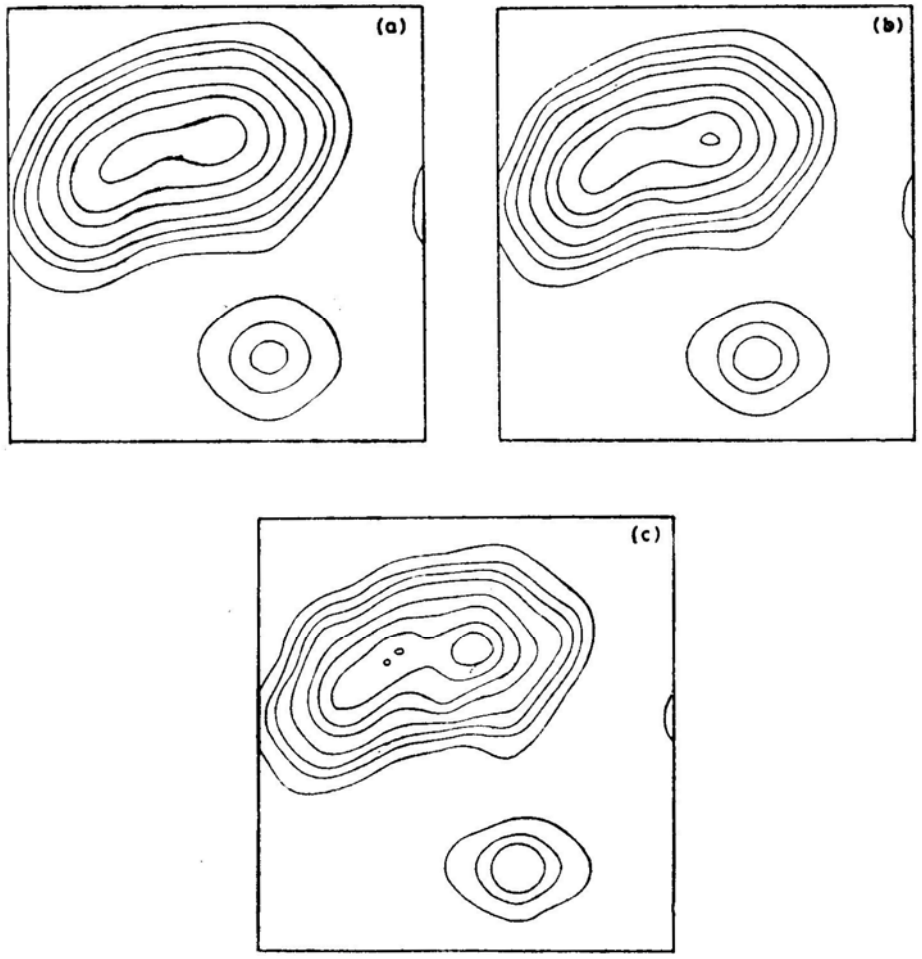

Figure 7. Restored brightness distribution obtained by applying to Fig. 6 Högbom's CLEAN algorithm with gain factor 0·25. (a), (b) and (c) correspond to gaussian restoring beams of σ = 2, 1·75 and 1·5. Note the false detail in (c) and the lowering of the peak height in (a).

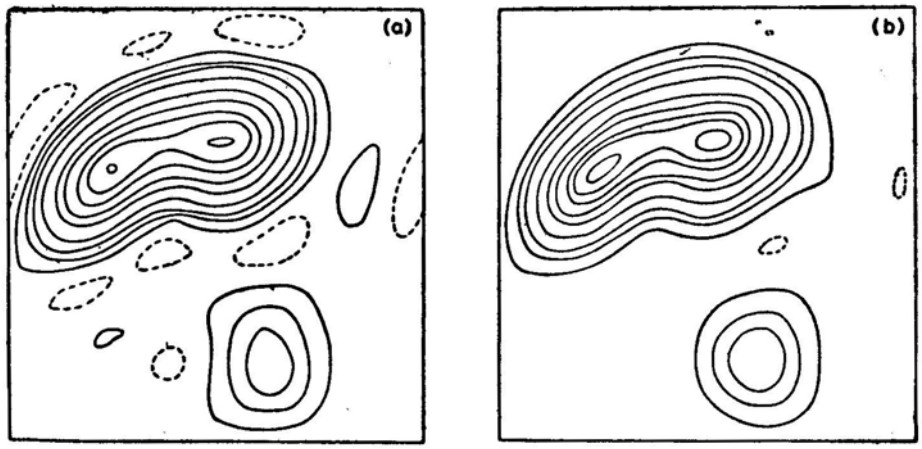

Figure 8. MEM restoration starting from the data of Fig. 6 with the ln$B$ entropy; resolution darameter R = 25 (a), and R= 100 (b).

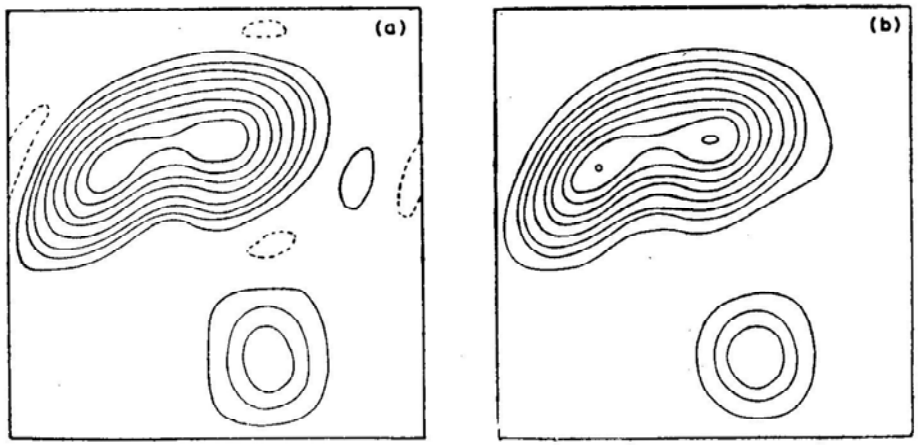

Figure 9. MEM restoration starting from the data of Fig. 6 with the $-B\ln B$ form of entropy; (a) $R = 25$, (b) $R = 100$.

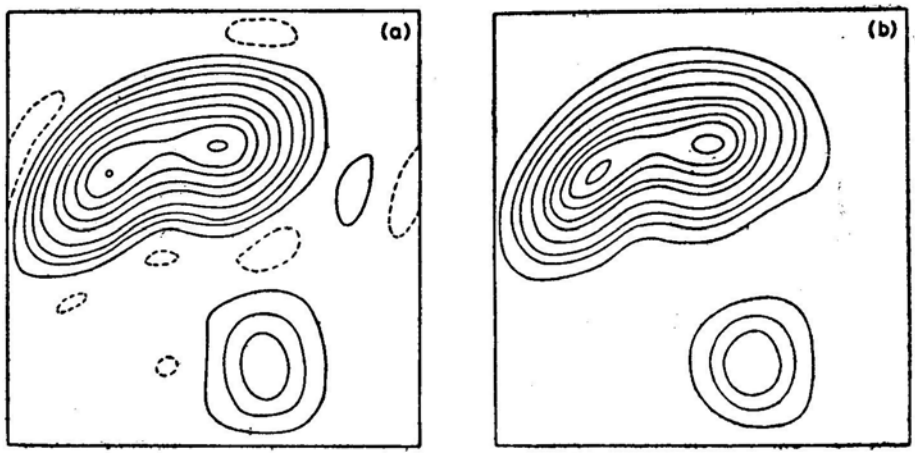

Figure 10. MEM restoration starting from the data of Fig. 6 with the $B_{1/2}$ form of entropy; (a)$R = 25$, (b)$R = 100$.

Table 1. Gaussian map of Fig. 4 restored with different forms of the MEM and with CLEAN.

Entropy	Maximum peak height		Maximum negative value	
	$R = 25$	$R = 100$	$R = 25$	$R = 100$
$B^{3/2}$		0·8800		−0·0174
$-B \ln B$	0·9808	1·0180	−0·0145	−0·0059
$B^{1/2}$	1·0420	1·1469	−0·0167	−0·0080
$\ln B$	1·0711	1·2484	−0·0189	−0·0111
$-1/B^2$	1·1072	1·3571	−0·0242	−0·0185
$-e^{-B \ln R}$	1·1890		−0·0280	

True map	1·0	0·0
Dirty map	0·7483	−0·0783
CLEAN (High resolution)	0·8619	−0·0036
CLEAN (Medium resolution)	0·7314	−0·0027
CLEAN (Low resolution)	0·6393	−0·0017

Fig. 1 la is tremendous with very little ripple suppression. In fact at R = 100, this form of entropy is so active that it begins to split peaks. This form is obviously not recommended for radio astronomy applications though it could have its uses elsewhere. The reconstruction with $B^{3/2}$ has flatter peaks than any of the earlier figures as expected but surprisingly has more ripple than some of the harder forms. The reason is that the series of entropies going from $B^{1+\epsilon}$ to $B^{2-\epsilon}$ (i.e. n going from $-1+\epsilon$ to $-\epsilon$) are in a separate class. These forms do not have an infinite barrier at B=0 and positivity can be maintained only by FLOATing. By this argument, $-B\ln B$ is the last of the positivity-enforcing entropies. The form $B^{3/2}$ is interesting for another reason. Since reconstructions with this entropy have $B^{1/2}$ band limited, this means Fourier extrapolation does not exceed twice the measured range (this, in fact, explains why the baseline cannot be flattened beyond a point).

Figs 12ad show an interesting example of a case where the MEM as well as CLEAN fail. The true map (Fig. 12a) is an elliptic plateau. When the 'measured' range is restricted to $|m|, |n| \leqslant 3$, one obtains the dirty map in Fig. 12b. The MEM reconstruction with the $-B\ln B$ entropy at R=100 is shown in Fig. 12d. The ripple at the base has been largely removed but the ripple at the top of the plateau has actually increased. This is a two-dimensional analogue of an effect first discussed by Komesaroff, Narayan and Nityananda (1981) for one-dimensional plateaus. As a matter of interest we show in Fig. 12c the CLEAN reconstruction with gain=0·5. CLEAN, being peak seeking (as is the MEM), has fitted peaks on the plateau. Moreover, being a sequential technique (which the MEM is not), it has broken the symmetry of the map. Fig. 12e is discussed in Section 8.

The $-B\ln B$ MEM (for example) has introduced sharp ripple whose wavelength is related to the cutoff in the data. Since this method has been interpreted as maximising a certain *α priori* probability for brightness distributions (Gull and Daniell 1978, 1979), we conclude that this *α priori* distribution does not agree with usual ideas on which features in a map are reasonable and which ones are artefacts. In our way of looking at the MEM, it is clear that the assumption of sharp peaks at high B and a flat baseline at low B (Section 4) has been violated. The ripples on top of the plateau

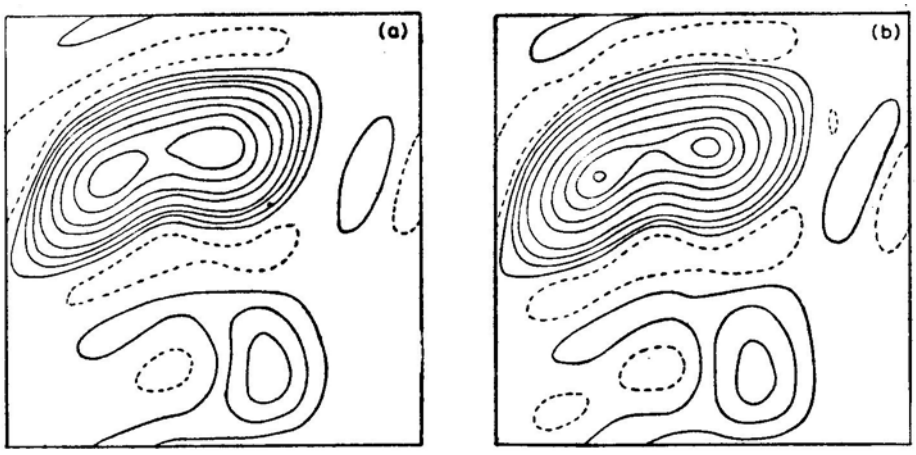

Figure 11. Reconstruction from the data of fig. 6 using (a) a very 'soft' entropy $-B^{3/2}$ with R = 100 (note the ripple in the baseline and rounded peak), (b) a very hard entropy $-\exp(-\alpha B)$ with R = 25; note the very sharp peaks and rippled baseline.

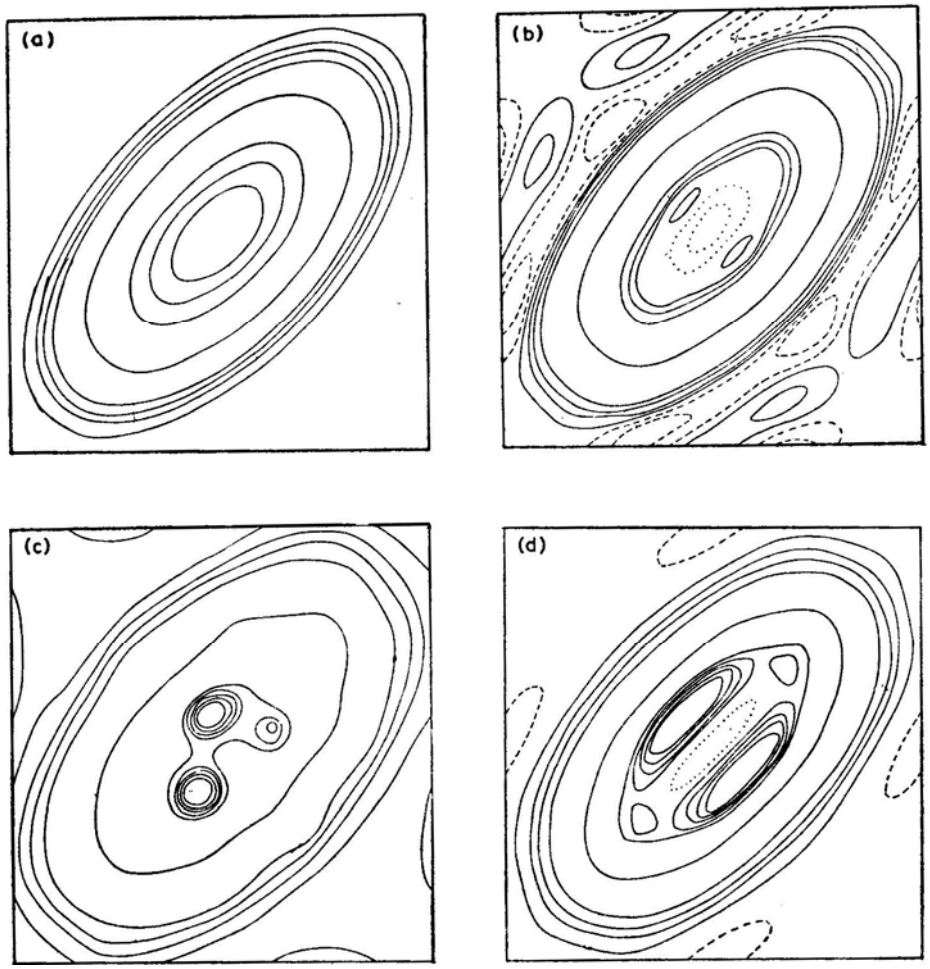

Figure 12. (a) Model source consisting of an elliptic plateau. The contour levels are 0.01, 0.03, 0.05, 0.1, 0.5, 0.9, 0.95, 0.97, 0.99, 1.01, 1.03, 1.05, 1.1. the contour levels have been chosen so as to bring out the low-level ripple as well as the ripple on the plateau, whose parameters are give in. Table 3. (b) Dirty map obtained from (a) by retaining only Fourier components ρ_{mn} with $|m|$, $|n| \leqslant 3$. Dashed contours represent negative values while dotted contours indicate depressions in which the height decreases inwards. (c) Restoration by CLEAN starting from (b). The gain factor used is 0·5 and the rms of the restoring gaussian beam is $\sigma = 2$. Note the effective suppression of low-level ripple accompanied by spurious, symmetry-breaking peaks on top of the plateau. (d) Restoration by $-B\ln B$ MEM with $R = 100$. Note that low-level ripple is suppressed. However, the spurious features on the plateau are similar to and even stronger than those in the dirty map.

have been treated as peaks to be sharpened. Clearly, CLEAN also fails because its underlying *α priori* assumptions about the map are not true in this case.

7. Treatment of noise in the measurements

In any real problem the measured Fourier coefficients $\rho_{mn}^{(m)}$ will have errors which we take, for simplicity, to be gaussian with rms σ (please note that σ_{mn} and σ are different objects!). The superscript (m) stands for 'measured' and we continue to use ρ_{mn} for

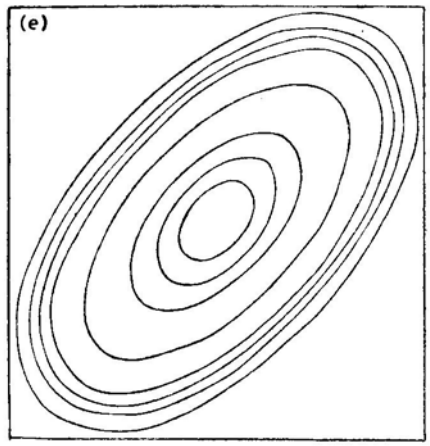

Figure 12. (e) Restoration with a modified entropy in $(B + C_1) + \ln (C_2 - B)$ with C_1 and C_2 FLOATing constants determined by $(B_{\max} + C_1)^2/ (B_{\min} + C_1)^2 = 100$, $B_{\min} + C_1 = C_2 - B_{\max}$. with this entropy the ripple has disappeared both from the low-lying region and the plateau.

the Fourier coefficients of the reconstructed brightness distribution (which need not agree with the measured values). Ables (1974) suggested that in the presence of noise one should satisfy the condition

$$\sum_{(m,\, n)\in K} | (\rho_{mn} - \rho_{mn}^{(\mathrm{m})}) |^2 = 2N\sigma^2 \tag{18}$$

where N equals the number of (complex) correlations measured. We refer to this approach as least-squares MEM. Ables' suggestion was implemented by Gull and Daniell (1978). Willingale (1981) has given the appropriate generalisation for the deconvolution problem in X-ray astronomy where the errors (σ) vary over the map. The condition (18) would appear to be reasonable since it gives deviations of the right overall size between the model and the measurements. However, we show below that these deviations ($\rho_{MN} - \rho_{MN}^{(\mathrm{m})}$) are far from randomly distributed and indeed introduce a systematic bias into the reconstruction.

We introduce a Lagrange multiplier λ for the single constraint (18) (which replaces the set of constraints (4) in the noise-free case). We thus have

$$\frac{\partial}{\partial \rho_{mn}} \left\{ \int\int f(B)\, dx\, dy - \lambda \sum_{(M,\, N)\in K} \left[(\rho'_{MN} - \rho'^{(\mathrm{m})}_{MN})^2 + (\rho''_{MN} - \rho''^{(\mathrm{m})}_{MN})^2 - 2\sigma^2 \right] \right\} = 0 \tag{19}$$

where we again use the real and imaginary parts ρ' and ρ'' of the Fourier coefficients. For clarity, we discuss the $f = \ln B$ case though the argument is general. Differentiating with respect to an unmeasured Fourier coefficient gives the same result as before *viz.* B is the reciprocal of a band-limited function $\sigma(x, y)$. Differentiating with respect to a measured coefficient ρ_{MN} in Equation (19) gives

$$\rho_{MN} = \rho_{MN}^{(m)} + \frac{1}{\lambda}\int\int f'(B)\,\exp\,[-2\pi i\,(Mx+Ny)]\,dx\,dy$$

i.e.

$$\rho_{MN} = \rho_{MN}^{(m)} + \frac{1}{\lambda}\sigma_{MN}. \tag{20}$$

Equation (20) immediately suggests an iterative scheme for implementing Equation (19). We can find the MEM solution with zero noise ($\lambda=\infty$), which has $\rho_{MN}=\rho_{MN}^{(m)}$, by any of the methods already described in Section 5. We then have available a good first approximation to the coefficients σ_{MN} in Equation (20). It is now easy to choose λ to satisfy Equation (18). Equation (20) would then give modified values of ρ_{MN} in the measured range. Since–in practical cases–the shifts are expected to be small, the process can be iterated and convergence is quite rapid.

The important lesson of the iterative scheme described by Equation (20) is that the residuals ($\rho_{MN} - \rho_{MN}^{(m)}$) are just proportional to the Fourier coefficients of the reciprocal of the MEM brightness distribution(!) and have no relation to the true errors in the measurements. Fig. 13 illustrates the nature of the bias which least-squares MEM can introduce. Schematically, we have a 2*N*dimensional space of data in which the coordinates of a point are the 2*N* measured Fourier coefficients. P_0 represents the true values of ρ_{MN} and a sphere of radius $\sigma\sqrt{2N}$ around P_0 represents the data sets obtained with various realisations of the noise. Given any point P_1 on this sphere, the least-squares MEM method modifies the measurements by the same residuals

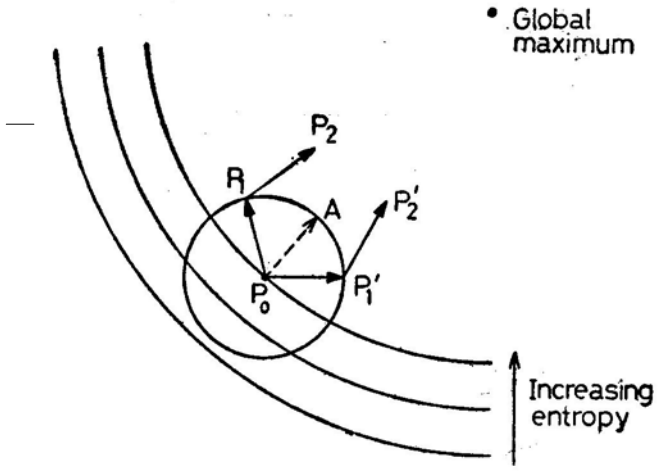

Figure 13. The space of measured Fourier coefficients. P_0 represents the true values, while P_1, P_1' etc. represent the values that would be obtained in measurements with different realisations of the noise. P_1, P_1' *etc.* lie on a sphere of radius $\sigma\sqrt{2N}$ where 2*N* is the number of (real) measurements and σ the rms of each. Starting from the noisy data, P_1, P_1' *etc;* the least-squares approach moves a distance $\sigma\sqrt{2N}$ to P_2, P_2' *etc.* The average over all realisations of the MEM solutions P_2, P'_2 *etc.* is the point A which has a systematic and predictable bias P_0A from the true data. The vectors P_1P_2, P_1' P'_2 represent the residuals and, for small σ , are independent of position on the sphere, pointing in the direction of increasing entropy, perpendicular to the constant entropy contours. At the centre of these contours is the unconstrained maximum entropy solution *viz.* a flat map (ρ_{00} is assumed fixed).

since, by Equation (20), these are dominated by the zero-order values of σ_{MN} (*i.e.* those corresponding to the noise-free case). The average of the solutions obtained with different realisations of the noise is *not* the same as the noise-free solution but differs from it to *first order* in σ(Fig.13). We can treat the two vectors P_0P_1 and P_1P_2 in Fig. 13 as orthogonal since there is no correlation between the noise, represented by P_0P_l, and the residuals introduced by the MEM, represented by P_1P_2. Therefore $P_0P_2 \simeq \sqrt{2}\ P_0P_1$, showing that the MEM has really taken us further from the true data.

An alternative approach is to fit the noisy data exactly, using the methods described in Section 5. The noise will propagate to the unmeasured correlations as well. However, there is the advantage that the measurements are not further degraded or biased. Fig. 14 illustrates the two approaches to noise. Fig. 14a shows the result of

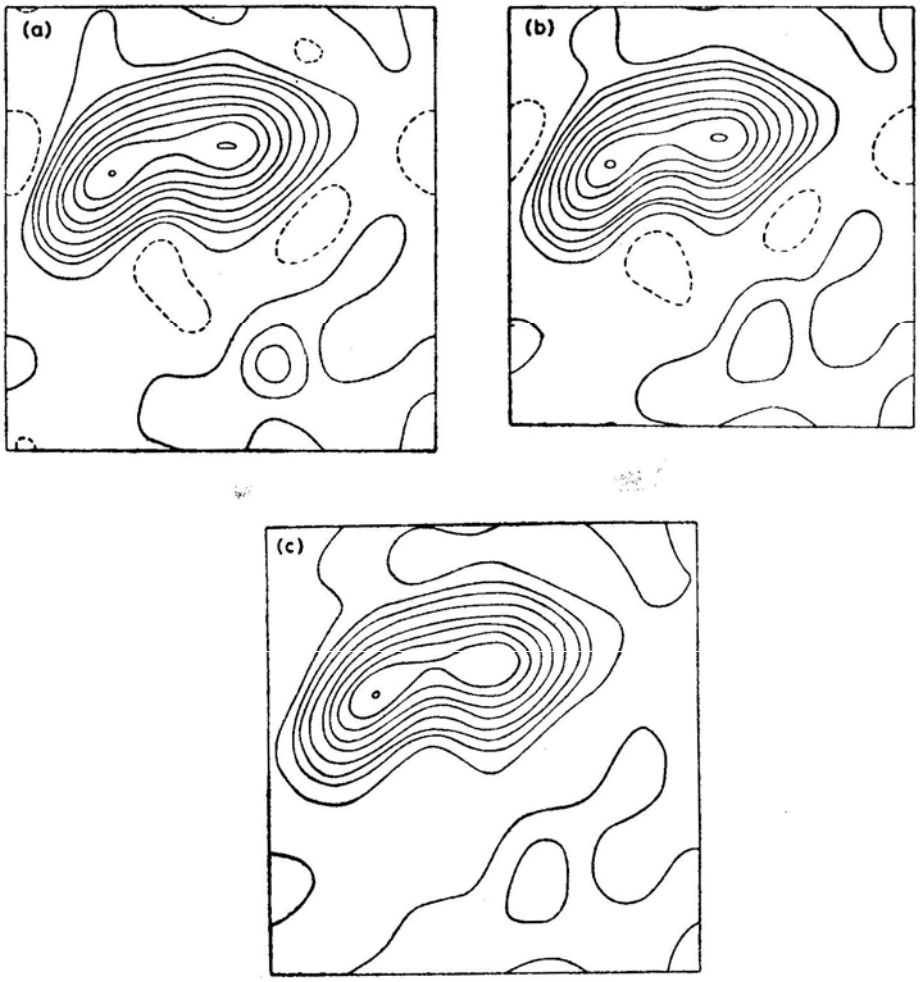

Figure 14. Illustration of the different ways of treating nosy data with the MEN. (a) R = 100, $-B\ln B$ MEM map of Fig. 9b with gaussian noise of rms = 1 added to the Fourier coefficients ρ_{mn} $|m\ |, |\ n\ | \leqslant 3$; (b) $-B\ln B$ MEM restoration from the noisy data in (a) with R = 100 fitting the data exactly; (c) $-B\ln B$ MEM restoration fitting the data in (a) by the least-squares method with σ = 1. Note that the result is not less noisy than (b) but has rounded off the peaks and distorted the ripple so that it no longer goes negative.

adding random noise with, a gaussian distribution of standard deviation 1 to the known coefficients $\rho_{mn}(|m|, |n| \leqslant 3)$ of the MEM map of Fig. 9b. Fig. 14b shows a reconstruction ($-B\ln B$ entropy, R=100) which fits the noisy data exactly. It is seen that the spurious features are not identical to those in Fig. 14a, since the noise has also affected the values of extrapolated coefficients. Fig. 14c shows the result of using the least-squares MEM scheme (Equations 1820). It can be seen that the spurious features have been modified more than in Fig. 14b. The peak and baseline values given in Table 2 confirm this. The reason is that now not only unknown but even measured coefficients have been changed. It is clear that this change has not reduced the effects of noise. In fact, the rms deviation of the $\rho_{mn}(|m|, |n| \leqslant 3)$ from the true values is 1 for the reconstruction in Fig. 14b but 1·32 for Fig. 14c, showing that the least-squares approach is taking us further from the noise-free MEM solution by approximately the $\sqrt{2}$ factor expected.

The unphysical nature of the residuals in conventional least-squares MEM has been pointed out by Bryan and Skilling (1980). They have proposed a new algorithm which forces the residuals to have a gaussian distribution. This indeed represents the use of additional *a priori* information, though with added computational complexity. The bias discussed above is not basically surprising. When the measurements are inaccurate, the *a priori* distribution of brightness levels can make its presence felt and pull the solution towards the global maximum–a flat distribution. This is illustrated by the constant entropy contours in Fig. 13, which enforce a specific choice of the error vector normal to them (Bryan and Skilling 1980). The net effect on the map is to flatten and raise the baseline and lower the peaks.

We illustrate this in Fig. 15. Fig. 15a shows the result of adding uncorrelated gaussian noise of rms 0·25 to *all* the correlations of Fig. 9b. The result of the least-squares MEM procedure treating all the correlations as measured is shown in Fig. 15b. Here, there is no question of extrapolation and we see quite clearly that the least-squares approach has not suppressed the noise but only modified it by removing the negative parts. Table 2 gives the peak and baseline values for the two maps in Fig, 15.

We conclude that using *a priori* information to solve for the noise is likely to lead to bias. This bias should always be kept in mind in interpreting maps made by the least-squares method. In many situations the alternative approach of fitting the noisy data may be preferable.

Table 2. Peak and minimum values of restorations from noisy data with the $-B\ln B$ MEM.

Figure	Description	Peak	Minimum
14a	Map of Fig. 9b with noise added to ρ_{mn} ($\|m\|, \|n\| < 3$) (R = 100, rms = 1)	1·0187	−0·0229
14b	Reconstruction fitting the noisy data exactly	1·0334	−0·0178
14c	Reconstruction fitting the noisy data by least squares with σ = 1	1·0020	−0·0080
15a	Fig. 9b with noise of rms = 0·25	1·0223	−0·0355
15b	Fig. 15a fitted by least squares with σ = 0·25	0·9832	−0·0088

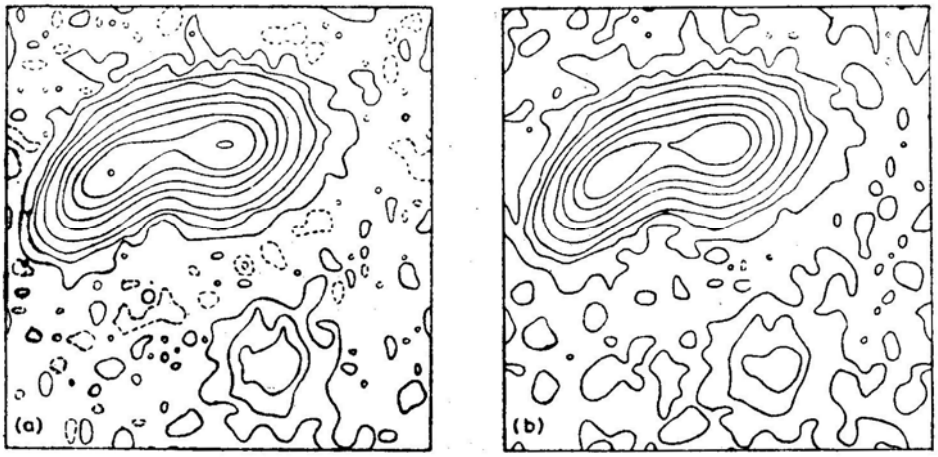

Figure 15. (a) A map obtained by adding gaussian noise with rms = 0.25 to all the Fourier coefficents of Fig. 9b. (b) Result of applying the least-squares method with σ =0.25 to the map in (a). There is no extrapolation of the data involved here. Comparing (b) to (a) show off the last-squares method distorts (and increases) the noise already present in (a) so as to round off the peaks and remove negative regions.

8. Summary and discussion

The major conclusions of our study are the following:

(a) The MEM is basically a model-fitting procedure which works well if the reconstruction is required to be peaky with a flat baseline. We do not agree that it is 'maximally noncommittal' with respect to the missing data. In fact, the unknown correlation values are 'committed' to developing peaks and a baseline in conformity with the model. Nor is it right to say that the reconstruction is as featureless as possible. Given the wrong problem, such as the plateau in Fig. 12, MEM produces spurious features stronger than even those in the dirty map! Regarding peak strengths and widths, it should be noted that higher peaks are likely to be sharper and stronger than lower ones. This result was pointed out by Komesaroff, Narayan and Nityananda (1981) for the one-dimensional ln$B$ entropy. We have been able to generalise it to all entropies and all dimensions in this paper. Moreover, the reconstructed peaks could sometimes have very curious properties, as discussed in Appendix B. All these peculiarities should be kept in mind in practical applications of the MEM.

(b) We have shown that the common—Bln$B$ and ln$B$ forms of entropy are members of a continuous family of functions characterised by a parameter n (see Equation 10). All these functions satisfy the important restrictions $f''(B) < 0$, $f'''(B) > 0$, which are necessary for acceptable reconstructions. The character of the reconstructions changes gradually from one end of this family to the other. For any particular application, depending on the requirements, a restricted range of n would be suitable [see (c) below]. What we would like to emphasize, however, is that no single form of 'entropy' can be considered fundamental and better than all others. This has been a source of considerable controversy, with two schools backing the — Bln$B$ and ln$B$ entropies respectively. Arguments and analogies from statistics, thermodynamics and information theory have all been applied in order to identify the best entropy. However, the results of Fig. 12 (which can be obtained with any

form of entropy) show that any 'ideal' entropy would fail badly if the conditions are not right.

To settle the controversy we recommend widespread use of the $B^{1/2}$ entropy. This is midway between $-B\ln B$ and $B\ln B$ and gives equally good reconstructions. As a major benefit, there is no temptation to associate any information-theoretic or thermodynamic concept with it. We also believe the misleading name of 'Maximum Entropy Method' should be changed to something more appropriate such as 'Variational Method'.

(c) If fundamental arguments are not appropriate, how should one select the form of f ? We suggest the following criteria, discussed in Section 4.

(i) Soft functions (small but positive values of n in Equation 10) generate fairly broad peaks but extremely flat baselines. On the other hand, hard functions (large n) reconstruct sharp peaks and leave a lot of residual ripple. For radio astronomy applications, soft functions are probably more appropriate.

(ii) The shapes of peaks reconstructed by any $f(B)$ can be predicted approximately. For gaussian peaks one should use $-B\ln B$ entropy, for Lorentzian peaks $\ln B$, *etc.* It should, of course, be kept in mind that all the functions considered are geared towards sharp peaks and flat baselines. If this is not an appropriate model in a particular case, then the results can be poor. Thus, in the case of the plateau (Fig. 12), if one knew in advance about the flat top, then a more appropriate function to maximise would be $\ln (C_1 + B) + \ln (C_2 - B)$ where C_1 and C_2 are FLOATing constants. Fig. 12e shows the remarkable improvement that this can make.

(d) There are two basic symmetries in all forms of the MEM.

(i) The reconstructions are translationally invariant in the sense that if the data are modified to correspond to a shift in the origin of (x, y), the reconstructed map also displays the same shift, but does not otherwise change.

(ii) If the data are scaled by a constant, so is the reconstruction. However, an important symmetry is missing in MEM reconstructions.

(iii) If one trivially modifies the dirty map by adding a constant to $B(x, y)$ at all x, y, the MEM reconstruction shows non-trivial changes.

One way of understanding this is to note that all the forms of the MEM with $n \geqslant 0$ in Equation (10) impose positivity. If one adds a large constant to the map, positivity becomes a weak or irrelevant consideration, while if one subtracts a constant, one may even end up with no positive solution at all. Thus, the form of the reconstruction is apparently at the mercy of the experimental value of ρ_{00} (the zero Fourier coefficient). If one had a large value of ρ_{00}, the MEM would give back the principal solution, a lower value would give a 'good' reconstruction, a still lower value would give a highly spiky function, *etc.* (Bhandari 1978; Komesaroff, Narayan and Nityananda 1981).

The modification of the MEM which we call FLOAT avoids this undesirable feature and has the added advantage that the user, can control the resolution of the restored map. In our scheme, we maximise the integral of $f(B + C)$ where C is a number which is set as follows. We have identified the ratio $R = f''(B_{\min} + C)/$

f'' (B_{max} + C) (see Equation 9) as a measure of the resolution in the restoration. C is chosen so that R attains a preset value. As R increases, the resolution also increases. We have found that R = 100 gives good maps. The results are however not significantly changed for a fairly large range of R.

When the FLOAT feature is incorporated, the results are insensitive to ρ_{00} and MEM becomes identical to the classical methods of restoration as far as symmetries are concerned. Another point to be noted is that one is no longer strictly enforcing positivity. It is $B + C$ which is positive. In fact most of the restorations displayed in this paper go negative at some point.

(e) Komesaroff, Narayan and Nityananda (1981) showed that, for fixed ρ_0, as the noise in the measured data increases, MEM restorations in one dimension with the ln$B$ entropy become progressively more peaky until, at a given noise level, there are no positive functions which fit the data. This is no doubt a feature of all entropies and all dimensions. However, with the FLOAT modification, this is not a problem. Since R is specified, a larger constant will be automatically added when the data are noisy, and the resolution of the restoration is not affected. If necessary one could even empirically reduce the value of R with increasing noise.

An alternative approach to noise is the least-squares approach (Ables 1974; Gull and Daniell 1978, 1979) which allows the map to deviate from the measured data, thus ensuring positivity. In this scheme, the resolution is apparently automatically reduced with increasing noise level. The main criticism we have of this apprroach is that the residuals between the fitted and measured data are *biased* (see Section 7). We believe it is betterto fit the data exactly.

(f) Although the method CLEAN (Högbom 1974) was proposed at about the same time as MEM, it has found far wider acceptance in radio astronomy. Most observations are CLEANed while MEM maps are a rarity. Based on the results of this and earlier papers, we make a comparison of the two methods.

(i) Both CLEAN and MEM are ideally suited to a map with well-separated point sources and an extensive fiat background. As the sources become extended, both methods become relatively less effective, but it appears that MEM may be marginally superior (see Figs 7–10 which correspond to a model with fairly extended sources).

(ii) CLEAN is excellent at ripple suppression even when the dirty beam has very large sidelobes. In the examples discussed so far in this paper, the sidelobe level is not very high because the uv coverage is compact ($|u|$, $|v| \leqslant 3$ in all cases). Fig. 16 shows a more diflicult case where the data are missing in two sectors (as could happen in a practical case when the source 'sets').* The CLEAN restoration is excellent. However the MEM restoration with $-B\ln B$ (the results are the same for other forms) is rather bad, with ripple suppression being relatively poor. The reason is not difficult to see. Equation (8) shows that the logarithm of the reconstructed brightness is band-limited with nonzero Fourier coefficients corresponding only to measured points. The large sidelobes of this function cannot be adequately removed by taking the exponential unless very high values of R are used. An alternative explanation is to note that the MEM generates the unmeasured ρ_{mn} by non-linear operations on the

*Table 3 gives details of the missing data in the u-v plane.

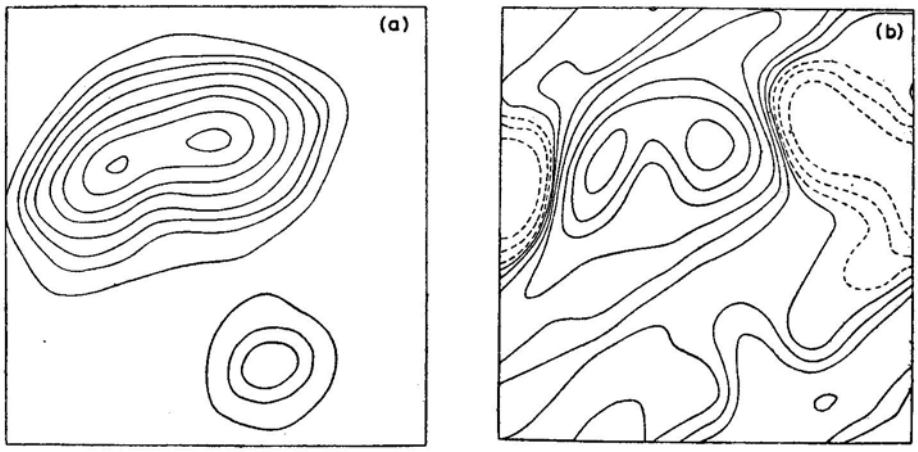

Figure 16. (a) Result of a CLEAN restoration with the model of Fig. 4 When a sector of data in the u –v plane is missing. Details of the missing sector are given in Table 3. (b) Result of the $-B\ln B$ MEN restoration with R = 100 on the same 'sectored' data as in (a). The method is now unable to suppress the sidelobes.

Table 3. Particulars of Figs. 12 and 16.

The plateau in Fig. 12 has the form

$$B(x, y) = 1/[1 + (\text{gaussian function})^{-1}]$$

When the parameters of the gaussian are

strength = e^4 = 54.598
position angle of major axis = 135°
rms along major axis = 4.5
rms along minor axis = 2.25

The data in the uv plane used for Fig. 16: for each value of u the range of positive v s is given (of course u, v is accompanied by – u, –v). The points (approximately) form a sector in the u–v plane.

u	$v_{\min}$	$v_{\max}$	u	$v_{\min}$	$v_{\max}$
−4	1	3	0	1	7
−3	1	5	1	1	7
−2	1	6	2	2	6
−1	1	7	3	3	5

Band-limited function $\sigma(x, y)$. In all earlier examples, the extrapolated values of ρ_{mn} were relatively small. However, in Fig. 16, some of the unmeasured ρ_{mn} have large values. To produce these by interpolation needs large non-linearity *i.e.* large R. At very large R, the numerical schemes converge very slowly.

(iii) One of the problems that prevents a deeper understanding of CLEAN is the sequential nature of the method. Schwarz (1979) has shown that the position and strength of each point source identified by CLEAN (with unit gain) can be understood as a least-squares fit to the current map. However, there is no least-squares interpretation of the entire collection of point sources. Because of this we believe one can at most expect to have only an empirical under standing of CLEAN. On the other hand, the MEM is a well-posed mathematical formulation and therefore in principle one should be able to appreciate its properties better. We hope the present paper has made a beginning in this direction.

Acknowledgements

It is a pleasure to thank M. M. Komesaroff for introducing us to the field and R. D. Ekers for discussions about CLEAN and related matters. We are grateful to R. K. Lakshmi and Nambu Ramam for coping cheerfully with the long and much revised manuscript and the numerous drawings.

Appendix A

Uniqueness of the MEM Reconstruction

Burg (1975) has shown that the problem of maximising the entropy $\int f(B(x))\, dx$ (where x is a position vector in a d-dimensional space) keeping certain Fourier co efficients of B fixed has a unique solution if $f''(B)$ is negative. We give here a simplified rederivation of this result.

Let $B_1(x)$ and $B_2(x)$ be two brightness distributions which fit the same set of measurements. We can interpolate between them using a parameter p which goes from 0 to 1, and calculate the entropy $E(p)$ corresponding to the interpolated function

$$B(p) = (1 - P)\, B_1(x) + P\, B_2(x), \tag{A1}$$

$$E(P) = \int f[(1 - P)\, B_1 + P\, B_2]\, dx. \tag{A2}$$

Differentiating Equation (A2) twice with respect to p

$$\frac{d^2}{dp^2} E(p) = \int f''(B(p))\, (B_1 - B_2)^2\, dx \leqslant 0 \text{ if } f''(B) \leqslant 0. \tag{A3}$$

The brightness distribution for each value of p satisfies the measured data since these are linear in B (Equation A1). As we interpolate between two brightness distributions, Equation (A3) shows that the graph of entropy versus p is convex upwards. If we now suppose that there are two maxima of E, we could interpolate between them and there would clearly have to be a minimum in between. Since this contradicts the convexity property just proved, the brightness distribution maximising the entropy is unique.

Appendix B

Limiting Behaviour of the MEM Restoration at High Resolution

Komesaroff, Narayan and Nityananda (1981) showed that one-dimensional MEM reconstructions with $f = \ln B$ approach a sum of δ functions as ρ_0 is lowered (which is equivalent to increasing the resolution parameter R defined in Equation 9). In fact, the limiting map (as $R \to \infty$) is uniquely fixed by the data and positivity, and is independent of the entropy form used. The behaviour in two and higher dimensions can be more complicated as shown by the following example.

Take $\rho_{01} = 1$, $\rho_{11} = 0{\cdot}8$, $\rho_{10} = 0{\cdot}8$. Clearly, the lowest possible value for ρ_{00} is 1. However, even after putting $\rho_{00} = 1$, we can fit the above data with a map of the form $\delta(y)\ b(x)$ where the function $b(x)$ only has to fit one number (*viz.* 0·8). Thus even in the $R \to \infty$ limit, there are many solutions consistent with the data, out of which each entropy selects one.

In this appendix, we draw attention to another unexpected property of many of the forms of MEM in certain dimensions *viz.* the appearance of δ functions in the restorations, over a *range* of values of ρ_0. Although the argument is general, it is convenient to treat the $f = \ln B$ form of MEM in three dimensions and consider a specific example *viz.* $\rho_{100} = \rho_{010} = \rho_{001} = 1$. Clearly, the lowest possible value of ρ_{000} is 1 and for any higher values, it is easy to find nonsingular maps–for example a gaussian like $A \exp [-\alpha(x^2 + y^2 + z^2)]$–which fit the data. The ln B MEM solution for this data set is the reciprocal of a band-limited function (Equation 8).

$$B(x, y, z) = \left(\sigma_0 - \frac{\sigma_1}{3}(\cos 2\pi x + \cos 2\pi y + \cos 2\pi z)\right)^{-1} \tag{B1}$$

where σ_0 and σ_1 are determined by the constraint equations

$$\rho_{000} = \int_0^1\int\int \sigma_0 \frac{dx\ dy\ dz}{\left[1 - \frac{\sigma_1}{3\sigma_0}(\cos 2\pi x + \cos 2\pi y + \cos 2\pi z)\right]}$$

$$\equiv \frac{1}{\sigma_0} I_0\left(\frac{\sigma_1}{\sigma_0}\right), \tag{B2}$$

$$\rho_{100} = 1 = \int_0^1\int\int \frac{\cos 2\pi x\ dx\ dy\ dz}{\sigma_0\left[1 - \frac{\sigma_1}{3\sigma_0}(\cos 2\pi x + \cos 2\pi y + \cos 2\pi z)\right]}$$

$$\equiv \frac{1}{\sigma_0} I_1\left(\frac{\sigma_1}{\sigma_0}\right). \tag{B3}$$

The crucial new feature in 3 dimensions is that the integrals in Equations (B2) and (B3) converge even when $\sigma_1 = \sigma_0$. This is because the denominator behaves as r^2, r being the distance from the point where it vanishes. In 'd' dimensions, we thus have $\int r^{d-1}\ dr/r^2$ which converges for $d \geqslant 3$. At the same time, it is impossible (because of positivity) to make σ_0 less than σ_1. Thus the ratio of ρ_{000} to ρ_{100} is restricted to the range over which the ratio of the two definite integrals defined in Equations (B2) and (B3) can vary. The ratio $(I_1/I_0) = 0$ at $\sigma_1/\sigma_0 = 0$ and increases monotonically to $(0{\cdot}516/1{\cdot}516) = 0{\cdot}34$ at $\sigma_1/\sigma_0 = 1$. Thus, when ρ_{000}/ρ_{00} becomes less than 2·94, there is no solution to the pair of equations (B2) and (B3). Put differently, this example shows that in three (or higher) dimensions, the reciprocal of a band-limited function is not able to fit a set of data with sufficiently low ρ_0, even though these data are perfectly consistent with smooth positive maps. It is easy to

see that the entropy $1/B^n$ will encounter similar problems in d and higher dimensions if $2/(n+1) < d$. Thus, the $(1/B)$ entropy is in trouble in two and higher dimensions, while this can happen for harder entropies even in one dimension.

The question now arises–what is the nature of the solution when ρ_{000} becomes less than the critical value? A deeper study shows that we can still find generalized solutions which consist (for the $f = \ln B$ case) of the reciprocal of a band-limited function plus δ-functions located at the places where the band-limited function vanishes. In the example considered earlier, when $\rho_{000} < 2{\cdot}94$ we seek a solution of the form

$$B(x, y, z) = \frac{1}{\left[\sigma_0 - \frac{\sigma_1}{3}(\cos 2\pi x + \cos 2\pi y + \cos 2\pi z)\right]} + a\,\delta(x)\,\delta(y)\,\delta(z). \tag{B4}$$

We give a non-rigorous justification of Equation (B4). The δ-function makes no contribution to the reciprocal* of (B4), which is therefore still a band-limited function, thus fulfilling the condition for a maximum (Equation 6 of the main text) with respect to variation of the individual Fourier coefficients. We further have to verify that we have a maximum of the entropy with respect to variations in the strength and position of the δ-function. Consider the equations

$$E = \int d\mathbf{x} \ln\left[\frac{1}{\sigma(\mathbf{x})} + \sum a_i\,\delta(\mathbf{x} - \mathbf{x}_i)\right] = \text{maximum}, \tag{B5}$$

$$\rho_{\mathbf{N}} = \int \left[\frac{1}{\sigma(\mathbf{x})} + \sum a_i\,\delta(\mathbf{x} - \mathbf{x}_i)\right] \exp(-2\pi i\,\mathbf{N}\cdot\mathbf{x})\,d\mathbf{x}, \tag{B6}$$

where N is a vector with integer components labelling the Fourier coefficients. Equations (B5) and (B6) represent the quantity E being maximized and the constraints respectively. Again, using one of the standard representations of the δ-function, we can check that the value of E in Equation (B5) is independent of the a_i and x_i. Introducing a Lagrange multiplier for each constraint, we have to investigate the variation of $E + \Sigma\,\lambda_{\mathrm{N}}\mathrm{x}_{\mathrm{N}}$ as we change the function σ (x) by $\delta\sigma$ (x) and the positions and strengths of the δ-functions by $\delta\mathrm{x}_i$ and δa_i respectively. The result is

$$\begin{aligned}\delta\left(E + \mathcal{W}\,\lambda_{\mathbf{N}}\,\mathbf{x}_{\mathbf{N}}\right) = &-\int d\mathbf{x}\,[\delta\sigma(\mathbf{x})/\sigma(\mathbf{x})] \\ &+ \sum_{\mathbf{N}} \lambda_{\mathbf{N}} \left\{\int d\mathbf{x} \exp(-2\pi i\,\mathbf{N}\cdot\mathbf{x})\left(-\frac{\delta\sigma}{\sigma^2} + \sum_i [\delta a_i\,\delta(x - \mathbf{x}_i)\right.\right. \\ &\left.\left. - a_i\,\delta\mathbf{x}_i\cdot\nabla_{\mathbf{x}}\,\delta(\mathbf{x} - \mathbf{x}_i)]\right)\right\} = 0\end{aligned} \tag{B7}$$

*A general nonlinear function of the delta function is not defined, but in the case of the reciprocal we can justify putting it equal to zero by using any one of the sequences (*e.g.* gaussians) which tend to the δ-function.

where the differentiation with respect to x_i has been transferred to x in the last term. Note that Equation (B5) is insensitive to a_i and x_i which appear only in the variation of Equation (B6). The variation with respect to $\delta\sigma$ gives

$$\sigma(\mathbf{x}) = -\sum \lambda_{\mathbf{N}} \exp(-2\pi i\, \mathbf{N}\cdot\mathbf{x}) \tag{B8}$$

implying that σ(x) is a band-limited function as before. Varying with respect to x_i in Equation (B7) gives

$$\nabla\left(\sum_{\mathbf{N}} \lambda_{\mathbf{N}} \exp(-2\pi i\, \mathbf{N}\cdot\mathbf{x})\right)_{\mathbf{x}=\mathbf{x}_i} = \nabla\sigma(\mathbf{x}_i) = 0. \tag{B9}$$

According to Equation (B9), the δfunctions can only be located at stationary points of σ(x). The variation of Equation (B7) with respect to a_i gives

$$\sum_{\mathbf{N}} \lambda_{\mathbf{N}} \exp(-2\pi i\, \mathbf{N}\cdot\mathbf{x}_i) = \sigma(\mathbf{x}_i) = 0. \tag{B10}$$

Equation (B10) restricts the locations still further to the zeros of σ(x) [at these places Equation (B9) is automatically fulfilled since σ(x) is a nonnegative function].

As an example, we consider the problem $\rho_{100} = \rho_{010} = \rho_{001} = 1$, $\rho_{000} = 2$. The required solution is of the form (B4). To fit the data we require that

$$\frac{1}{\sigma_0} I_0(1) + a = 2, \tag{B11}$$

$$\frac{1}{\sigma_0} I_1(1) + a = 1, \tag{B12}$$

where a is the strength of the δ-function, located at the origin by the symmetry of the data. I_0 and I_1 defined in Equations (B2) and (B3), have the values 1·516 and 0·516. Equations (B11) and (B12) then give $\sigma_0 = \sigma_1 = 1$, a =0·484.

The appearance of δ-functions in the solution of a variational problem as a parameter is varied, has a well-known physical counterpart in Bose condensation (see, for example, the text by Landau and Lifshitz 1969), where the number of particles with momentum p, is given by

$$n(P) = \{\exp[(P^2/2m - \mu)/k_B T] - 1\}^{-1}$$

For temperature T above a value T_0, the normalisation of $n(p)$ is maintained by adjusting μ. At T_0, μ becomes zero, but the integral of $n(p)$ still converges in three dimensions. For T less than T_0 we can no longer vary μ to obtain the correct number of particles, since it remains fixed at zero. The excess particles are accommodated in a δ-function at p=0. The quantity being maximised in this problem is the (thermodynamic) entropy, with the total energy and number of particles as constraints.

The arguments discussed above are true for the problem in which x, y,... are continuous variables. In any computational scheme, we are dealing with a discrete

set of points and a summation. Unlike an integral, a sum diverges if the value at any given point approaches infinity, however weakly. Thus, such a discrete computational scheme would be equivalent, in the example considered, to solving the equations (B13) and (B14) below instead of (B2) and (B3).

$$\rho_{000} = \frac{1}{N^3} \sum_{p,q,r=0}^{N-1} \frac{1}{\left[\sigma_0 - \frac{\sigma_1}{3}\left(\cos\frac{2\pi p}{N} + \cos\frac{2\pi q}{N} + \cos\frac{2\pi r}{N}\right)\right]}, \tag{B13}$$

$$\rho_{100} = \frac{1}{N^3} \sum_{p,q,r=0}^{N-1} \frac{\cos(2\pi p/N)}{\left[\sigma_0 - \frac{\sigma_1}{3}\left(\cos\frac{2\pi p}{N} + \cos\frac{2\pi q}{N} + \cos\frac{2\pi r}{N}\right)\right]}. \tag{B14}$$

We now see that the ratio ρ_{000}/ρ_{100}can vary from ∞ to 1 as (σ_1/σ_0) varies from 0 to 1, and there is no lower limit for ρ_{000} other than that set by positivity. However, one expects, even in the *discrete* case, that when r_{000} becomes less than the critical value below which there is no solution in the *continuous* case, there should be some qualitative differences in the reconstructed distribution. We conjecture that there will be an anomalous sensitivity to the grid size since the discrete analogue of the δ-function has a height varying inversely with the grid size.

Appendix C

An apparently straightforward generalization of the ln$B$ form of the MEM to two or more dimensions was given by Newman (1977, 1978). His starting point was the elegant one-dimensional form due to Burg (1967) in which B is expressed as the reciprocal of a band-limited function (as in Equation 8). This band-limited, positive, function is in turn factorised into two complex conjugate terms, one with positive and the other with negative frequencies. That is

$$B(x) = \frac{1}{\sum_{n=-N}^{N} \sigma_n \exp(2\pi i n x)} = \frac{1}{\left|\sum_{n=0}^{N} \alpha_n \exp(2\pi i n x)\right|^2}. \tag{C1}$$

In the one-dimensional case, given $\rho_0, \rho_1 \ldots \rho_N$, the nonlinear equations (5 and 6) for σ_n ($2N + 1$ real unknowns) reduce to a system *of linear* equations for the quantities α_n in the factorised form (C1) (N complex plus one real unknown).

In two dimensions, Newman (1977) writes by analogy

$$B(x,y) = \frac{1}{\sum_{m,n=-N}^{+N} \sigma_{mn} \exp[2\pi i(mx + ny)]} = \frac{1}{\left|\sum_{m,n=0}^{N} \alpha_{mn} \exp[2\pi i(mx + ny)]\right|^2}. \tag{C2}$$

He presents a system of linear equations to solve for α_{mn} in terms of ρ_{mn} ($-N \leqslant m,n \leqslant N$). There are now $(2N+1)^2$ real data in the ρ_{mn}. However, since the α_{mn} are restricted to positive frequencies, there are only $2[(N+1)^2 - 1] + 1$ real free parameters in the factorised form (C2). Therefore, a general band-limited function of x and y cannot be factorised and the solution (C2) cannot fit a general data set (of ρ_{mn}'s). The failure of factorisation in two dimensions has been noted by Burg (quoted by Woods 1972) but apparently ignored in later discussions (Newman 1978, van Schooneveld 1979).

The difficulty with Newman's solution to the multidimensional MEM problem can also be realised by considering the following alternative distinct factorised form

$$B(x,y) = \frac{1}{\left| \sum_{n=0}^{-N} \sum_{m=0}^{N} \beta_{mn} \exp\left[2\pi i(mx+ny)\right] \right|^2}. \qquad \text{(C3)}$$

Here, the allowed frequencies mn occupy the fourth quadrant of the m,n plane instead of the first (as in Equation C2). By symmetry, we must be able to find a general solution of the type (C3) if there is one of the type (C2). However, the true MEM solution is unique and is thus not of the form (C2) or (C3).

We present below a simple example illustrating the failure of Newman's solution to fit the data. Table A1 gives a two dimensional autocorrelation measured at the points (0,0), (0,1) (1,0), (1,1) and (1, –1). The α and β coefficients corresponding to these ρ's were computed using the method described by Newman (1977) and the corresponding brightness distributions were transformed to obtain the autocorrelations shown in Table A1. As expected, these do not agree with the original data. Newman (1978) has suggested that except for 'pathological' cases, his method will give the MEM solution. The simple arguments and example given above show that just the opposite is true.

Table Al. A test of the two-dimensional MEM scheme proposed by Newman (1977). The α and β Solutions represent two variants of his method described by Equations (C2) and (C3).

True	α solution	β solution
1·0	1·079	0·995
0·3	0·417	0·291
0·2	0·338	0·186
0·1	0·181	0·055
−0·2	0·131	−0·205

References

Ables, J. G. 1974, *Astr. Astrophys. Suppl. Ser.*, **15,** 383.

Bhandari, R. 1978, *Astr. Astrophys.*, **70,** 331.

Bryan, R. K., Skilling, J. 1980, *Mon. Not. R. astr. Soc*, **191,** 69.

Burg, J. P. 1967, in *Annual Meeting of the International Society of Exploration Geophysicists*, Reprinted in *Modern Spectrum Analysis*, Ed. D. G. Childers, IEEE Press, New York (1978) p. 34.

Burg, J. P. 1975, *Ph D thesis,* Stanford University.

Fletcher, R., Reeves, C. M. 1964, *Comput. J.*, **7,** 149.

Gull, S. F., Daniell, G. J. 1978, *Nature*, **272,** 686.

Gull, S. F., Daniell, G. J. 1979, in *Image Formation from Coherence Functions in Astronomy*, Ed. C. van Schooneveld, D. Reidel, Dordrecht, p. 219.

Gull, S. F., Skilling, J.1982, *Preprint.*
Högbom, J. A. 1974, *Astr. Astrophys. Suppl. Ser.*, **15,** 417.
Högbom, J. A. 1979, in *Image Formation from Coherence Functions in Astronomy*, Ed. C. van Schooneveld, D. Reidel, Dordrecht, p. 237.
Kikuchi, R., Soffer, B. H. 1977, *J. Opt. Soc. Am.*, **67,** 1656.
Komesaroff, M. M., Lerche, I. 1979, in *Image Formation from Coherence Functions in Astronomy.* Ed. C. van Schooneveld, D. Reidel, Dordrecht, p. 241.
Komesaroff, M. M., Narayan, R., Nityananda, R.1981, *Astr. Astrophys.*, **93,** 269.
Landau, L. D., Lifshitz, E. M. 1969, *Statistical Physics*, 2 edn Pergamon Press, Oxford, p. 159.
Newman, W. I. 1977, *Astr. Astrophys.*, **54,** 369.
Newman, W. I. 1978, *Astr. Astrophys.*, **70,** 409.
Parzen, E. 1968, in *Multivariate Analysis*, Vol. 2, Academic Press, New York.
Ponsonby, J. E. B. 1973, *Mon. Not. R. astr. Soc.*, **163,** 369.
Ponsonby, J. E, B. 1979, in *Image Formation from Coherence Functions in Astronomy*, Ed. C. van Schooneveld, D. Reidel, Dordrecht, p. 235.
Pye, J. P., Pounds, K. A., Rolf, D. P., Seward, F. D., Smith, A., Willingale, R. 1981, *Mon. Not. R, astr. Soc.*, **194,** 569.
Schwarz, U. J. 1979, in *Image Formation from Coherence Functions in Astronomy*, Ed. C. van Schooneveld, D. Reidel, Dordrecht, p. 261.
Subrahmanya, C. R. 1979, in *Image Formation from Coherence Functions in Astronomy*, Ed. C. van Schooneveld, D. Reidel, Dordrecht, p. 287.
Subrahmanya, C. R.1980, *Bull. astr. Soc. India*, **8,** 5.
van den Bos, A. 1971, *IEEE Trans. Inf. Theory*, **17,** 493.
van Schooneveld, C. 1979, in *Image Formation from Coherence Functions in Astronomy,* Ed. C. van Schooneveld, D. Reidel, Dordrecht, p. 197.
Willingale, R. 1981, *Mon. Not. R. astr. Soc.*, **194,** 359.
Woods, J. W. 1972, *IEEE Trans. Inf. Theory*, **18,** 232.